

# **Computational Vision and Medical Image Processing**

## **VIPIMAGE 2015**

**João Manuel R.S.Tavares  
A.M. Natal Jorge**

**EDITORS**

 **CRC Press**  
Taylor & Francis Group  
A BALKEMA BOOK

COMPUTATIONAL VISION AND MEDICAL IMAGE PROCESSING V

This page intentionally left blank

PROCEEDINGS OF VIPIMAGE 2015 – V ECCOMAS THEMATIC CONFERENCE ON COMPUTATIONAL VISION AND MEDICAL IMAGE PROCESSING, TENERIFE, SPAIN, 19–21 OCTOBER 2015

# Computational Vision and Medical Image Processing V

*Editors*

João Manuel R.S. Tavares & R.M. Natal Jorge

*Faculdade de Engenharia, Universidade do Porto, Porto, Portugal*



CRC Press

Taylor & Francis Group

Boca Raton London New York Leiden

---

CRC Press is an imprint of the  
Taylor & Francis Group, an **informa** business

A BALKEMA BOOK

*CRC Press/Balkema is an imprint of the Taylor & Francis Group, an informa business*

© 2016 Taylor & Francis Group, London, UK

Typeset by V Publishing Solutions Pvt Ltd., Chennai, India  
Printed and bound in Great Britain by CPI Group (UK) Ltd, Croydon, CR0 4YY

All rights reserved. No part of this publication or the information contained herein may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, by photocopying, recording or otherwise, without written prior permission from the publisher.

Although all care is taken to ensure integrity and the quality of this publication and the information herein, no responsibility is assumed by the publishers nor the author for any damage to the property or persons as a result of operation or use of this publication and/or the information contained herein.

Published by: CRC Press/Balkema  
P.O. Box 11320, 2301 EH Leiden, The Netherlands  
e-mail: [Pub.NL@taylorandfrancis.com](mailto:Pub.NL@taylorandfrancis.com)  
[www.crcpress.com](http://www.crcpress.com) – [www.taylorandfrancis.com](http://www.taylorandfrancis.com)

ISBN: 978-1-138-02926-2 (Hbk)  
ISBN: 978-1-315-64279-6 (eBook PDF)

## Table of contents

Acknowledgements	ix
Preface	xi
Invited lectures	xiii
Scientific committee	xv
<i>Invited lectures</i>	
An adaptive non-rigid image registration technique using hierarchical B-splines <i>A. Pawar, Y. Zhang, X. Wei, Y. Jia, T. Rabczuk, C.L. Chan &amp; C. Anitescu</i>	3
Medical image segmentation using Object Shape Models: A critical review on recent trends, and alternative directions <i>A.X. Falcão, T.V. Spina, S.B. Martins &amp; R. Phellan</i>	9
Primal-dual method for continuous max-flow approaches <i>K. Wei, X.-C. Tai, T.F. Chan &amp; S. Leung</i>	17
Image restoration using variational approaches: Some recent advances <i>A. Lanza, S. Morigi &amp; F. Sgallari</i>	25
Virtual and augmented medical imaging environments: Applications to simulation, training, surgical planning and interventional guidance <i>C.A. Linete</i>	33
<i>Contributed papers</i>	
The fast method of creating High Dynamic Range image for fluorescent microscopy applications <i>A. Bal</i>	41
Automatic cheek detection in digital images <i>M. Frackiewicz, H. Palus &amp; K. Radlak</i>	49
A variational model for image fusion with simultaneous cartoon and texture decomposition <i>M. Dodangeh, I.N. Figueiredo &amp; G. Gonçalves</i>	57
Image contrast enhancement using split Bregman method <i>S.Gh. Bardeji, I.N. Figueiredo &amp; E. Sousa</i>	63
Extraction of spectral drowsy component from the resting electroencephalographic signal for quick, objective and direct testing of sleepiness in absolute terms <i>A.A. Putilov, O.G. Donskaya &amp; E.G. Verevkin</i>	69
The illusion of a blackbody at the human ear and the human temperature measurement <i>A. Cardoso</i>	75
Video-based Turkish Sign Language recognition systems <i>M. Aktaş &amp; E.B. Sonmez</i>	81

A nonsmooth nonconvex sparsity-promoting variational approach for deblurring images corrupted by impulse noise <i>A. Lanza, S. Morigi &amp; F. Sgallari</i>	87
Classification-based blood vessel segmentation in retinal images <i>J. Odstrcilik, R. Kolar, V. Harabis &amp; R.P. Tornow</i>	95
Line extraction via phase congruency with a novel adaptive scale selection for Poisson noisy images <i>V.A. Krylov &amp; J.D.B. Nelson</i>	101
Diagnosis of human intestinal parasites by deep learning <i>A.Z. Peixinho, S.B. Martins, J.E. Vargas, A.X. Falcão, J.F. Gomes &amp; C.T.N. Suzuki</i>	107
Texture-energy features for microaneurysms detection <i>D. Veiga, N. Martins, C. Pereira, M. Ferreira &amp; J. Monteiro</i>	113
An iterative algorithm for Total Variation minimization in DBT imaging <i>A.M. Mota, N. Matela, N. Oliveira &amp; P. Almeida</i>	119
No-reference wavelet based Retinal Image Quality Assessment <i>L.S. Abdel Hamid, A. El-Rafei, S. El-Ramly, G. Michelson &amp; J. Hornegger</i>	123
Vessel segmentation of retinal images with fuzzy morphology <i>P. Bibiloni, M. González-Hidalgo &amp; S. Massanet</i>	131
Identification of subendocardial infarction—a feasibility study using synthetic ultrasonic image data of a left ventricular model <i>J. Źmigrodzki, S. Cygan, B. Leśniak-Plewińska &amp; K. Kalużyński</i>	137
Microcalcification segmentation in full field digital mammography <i>N. Martins, D. Veiga, C. Pereira, M. Ferreira, N. Alves &amp; M. Delgado</i>	143
Unsupervised delineation of the vessel tree in retinal fundus images <i>N. Strisciuglio, M. Vento, G. Azzopardi &amp; N. Petkov</i>	149
A registration algorithm for microscopy images of the capillary bed <i>H. Rieiro, J.L. Alba Castro, S. Martinez-Conde &amp; S.L. Macknik</i>	157
Bilateral filtering based biomedical image colorization <i>A. Popowicz &amp; B. Smolka</i>	163
Comparison of the internal structures of bones by microtomography <i>J.C.L. Stutz, J.S. Domínguez &amp; J.T. de Assis</i>	171
Model adaptation for mesh generation of biological structures <i>A. Ramos-de-Miguel, R. Montenegro &amp; J.M. Escobar</i>	175
Medical volume rendering based on gradient information <i>T.F. de Moraes, P.H.J. Amorim, J.V.L. da Silva, H. Pedrini, M.I. Meurer</i>	181
Chaos theory-based quantification of ROIs for mammogram classification <i>J. Kurek, B. Świderski, S. Dhahbi, M. Kruk, W. Barhoumi, G. Wieczorek &amp; E. Zagrouba</i>	187
M5L: A web-based Computer Aided Detection system for automated search of lung nodules in thoracic Computed Tomography scans <i>A. Traverso, M. Agnello, P. Cerello, M. Saletta, S. Bagnasco, C. Peroni, E. Fiorina, M.E. Fantacci, A. Retico &amp; E. Lopez Torres</i>	193
Non-intrusive and calibration free visual exploration analysis in children with Autism Spectrum Disorder <i>D. Cazzato, F. Adamo, G. Palestro, G. Crifaci, P. Pennisi, G. Pioggia, L. Ruta, M. Leo &amp; C. Distante</i>	201
Semi-automatic tumor contouring method using PET and MRI medical images <i>S. Urbán, L. Ruskó &amp; A. Nagy</i>	209

Automatic detection of bones based on the confidence map for Rheumatoid Arthritis analysis <i>K. Radlak, N. Radlak &amp; B. Smolka</i>	215
Improved computer recognition of Fuhrman grading system in analysis of Clear-Cell Renal Carcinoma <i>M. Kruk, J. Kurek, S. Osowski &amp; R. Koktysz</i>	221
A proof of concept of an augmented reality system for Nuss surgery <i>A. Ferreira, P. Morais, S. Queirós, F. Veloso, N.F. Rodrigues, J. Correira-Pinto &amp; J.L. Vilaça</i>	227
Usage of mobile devices in a bone fracture reduction process <i>J.R. Jiménez, F. Paulano, J.M. Noguera &amp; J.J. Jiménez</i>	233
Automated peritumoral edema segmentation in preoperative brain MRI scans <i>E. Binaghi, P. Melchiorre, L. Romitielli, S. Balbi &amp; D. Lattanzi</i>	239
Evaluation of segmentation techniques for wound area identification <i>G. Zhang, P. Xiberta, A. Bardera, I. Boada &amp; A. Romero</i>	245
Automated image segmentation based on multiobjective optimization and machine learning <i>S.M. Shontz, J.S. Tahara, D.O. McLaurin, D.J.L. Colbry &amp; B. Parikh</i>	251
Fractal descriptor on holographic images of cervical cells <i>M. Mihailescu, E.I. Scarlat, I.A. Paun, I. Grigorescu, R. Radu &amp; O.T. Nedelcu</i>	255
An integrated two Time-of-Flight camera system to determine knee flexion movement: Comparison with standard motion analysis system <i>E. Veron-Tocquet, V. Burdin, J. Savéan, J. Leboucher &amp; O. Rémy-Néris</i>	261
Mechanics-based analysis of the left atrium via echocardiographic imaging <i>S. Gabriele, L. Teresi, V. Varano, P. Nardinocchi, P. Piras, G. Esposito, P.E. Puddu, C. Torromeo &amp; A. Evangelista</i>	267
A new thick-walled conical model of the Left Ventricle <i>B. Leśniak-Plewińska, S. Cygan, J. Źmigrodzki &amp; K. Kalużyński</i>	273
Augmented Reality in radiofrequency ablation of the liver tumors <i>L.T. De Paolis &amp; F. Ricciardi</i>	279
Anthropomorphic robot forefinger virtually simulated <i>H.J. Rabiela, B.V. González &amp; G.D.A. Miranda</i>	283
CT based identification problem for the multicompartment model of blood perfusion <i>E. Rohan, V. Lukeš &amp; J. Brašnová</i>	289
A study on discrimination of SIFT feature applied to binary images <i>I. Setitra &amp; S. Larabi</i>	295
Time-constrained detection of colored objects on raw Bayer data <i>A.J.R. Neves, A. Trifan &amp; J.L. Azevedo</i>	301
Image scanning techniques for speeded-up color object detection <i>A. Trifan, A.J.R. Neves &amp; B. Cunha</i>	307
Improving deep neural networks classification by preprocessing images <i>H. Erdmann, F.T. Ito, D. Takabayashi &amp; D.N. dos Santos</i>	313
Surface reconstruction of bone fragments: A comparative study <i>F. Paulano, J.J. Jiménez &amp; J.R. Jiménez</i>	321
Towards a robust patch based multi-view stereo technique for textureless and occluded 3D reconstruction <i>B. Haines &amp; L. Bai</i>	327

Forming tool inspection using fiber-optic sensor heads <i>S. Matthias, M. Kästner, E. Reithmeier, P. Sieczkarek &amp; A.E. Tekkaya</i>	335
Navigation of robotics platform using advanced image processing navigation methods <i>L. Beran, P. Chmelar &amp; L. Rejsek</i>	341
SNIP: Smile—neutral facial display intensity predictor <i>K. Nurzynska &amp; B. Smolka</i>	347

## Acknowledgements

The editors and the Conference co-chairs acknowledge the support towards the publication of the Book of Proceedings and the organization of the V ECCOMAS Thematic Conference VipIMAGE to the following organizations:

- Universidade do Porto (UP)
- Faculdade de Engenharia da Universidade do Porto (FEUP)
- Instituto de Ciência e Inovação em Engenharia Mecânica e Engenharia Industrial (INEGI)
- European Community on Computational Methods in Applied Sciences (ECCOMAS)
- International Association for Computational Mechanics (IACM)
- Fundação para a Ciência e a Tecnologia (FCT)
- Associação Portuguesa de Mecânica Teórica Aplicada e Computacional (APMTAC)

This page intentionally left blank

## Preface

This book contains invited lectures and full papers presented at VipIMAGE 2015—V ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing, which was held in Tenerife, Canary Islands, Spain, during the period 19–21 October 2015. The event had 6 invited lectures, and 50 contributed presentations originated from 19 countries: Algeria, Brazil, China, Czech Republic, Egypt, France, Germany, Hungary, Italy, Mexico, Norway, Poland, Portugal, Romania, Russia, Spain, The United Kingdom, The United States of America and Turkey.

Computational methodologies of signal processing and analyses have been commonly used in our society. For instances, full automatic or semi-automatic Computational Vision systems have been increasing used in surveillance tasks, traffic analysis, recognition process, inspection purposes, human-machine interfaces, 3D vision, deformation analysis and aided medical procedures.

One of the notable aspects of the Computational Vision domain is the inter- and multi-disciplinarily. Actually, principles and methodologies of other sciences, such as Informatics, Mathematics, Statistics, Psychology, Mechanics and Physics, are regularly embraced into this domain. One of the key motives that contributes for the continually effort done in this field of the human knowledge is the high number of applications that can be easily found in Medicine. For instance, computational algorithms can be applied on medical images for shape reconstruction, motion and deformation analysis, tissue characterization or computer-assisted diagnosis and therapy.

The main objective of these ECCOMAS Thematic Conferences on Computational Vision and Medical Image Processing, initiated in 2007, is to promote a comprehensive forum for discussion on the recent advances in the related fields in order to identify potential collaboration between researchers of different sciences. Henceforth, VipIMAGE 2015 brought together researchers representing fields related to Biomechanics, Biomedical Engineering, Computational Vision, Computer Graphics, Computer Sciences, Computational Mechanics, Electrical Engineering, Mathematics, Statistics, Medical Imaging, Medicine and Rehabilitation.

The expertises spanned a broad range of techniques for Signal Processing and Analysis, Image Acquisition, Image Processing and Analysis, Data Interpolation, Registration, Acquisition and Compression, Image Segmentation, Tracking and Analysis of Motion and Deformation, 3D Vision, Computer Simulation, Medical Imaging, Computer Aided Diagnosis, Surgery, Therapy and Treatment, Computational Bio-imaging and Visualization, Telemedicine, Virtual and Enhanced Reality, Satellite imagery, Software Development and Applications.

The conference co-chairs would like to take this opportunity to express gratitude for the support given by The International European Community on Computational Methods in Applied Sciences and The Portuguese Association of Theoretical, Applied and Computational Mechanics, and thank to all sponsors, to all members of the Scientific Committee, to all Invited Lecturers, to all Session-Chairs and to all Authors for submitting and sharing their knowledge.

João Manuel R.S. Tavares  
Renato M. Natal Jorge  
*Conference Co-Chairs*

This page intentionally left blank

## Invited lectures

During VipIMAGE 2015, Invited lectures were delivered by 6 Expertises from 4 countries:

- Alexandre Xavier Falcão, *Universidade de Campinas, Brazil*
- Cristian A. Linte, *Mayo Clinic, USA*
- Fiorella Sgallari, *University of Bologna, Italy*
- Yongjie Zhang, *Carnegie Mellon University, USA*
- Xiaochuan Pan, *The University of Chicago, USA*
- Xue-cheng Tai, *University of Bergen, Norway*

This page intentionally left blank

## Scientific committee

All works submitted to VipIMAGE 2015 were evaluated by an International Scientific Committee composed by 116 expert researchers from recognized institutions of 26 countries:

Ahmed El-Rafei, *Ain Shams University, Egypt*  
Alberto Gambaruto, *Barcelona Supercomputing Centre, Spain*  
Alexandre Cunha, *California Institute of Technology, USA*  
Amr Abdel-Dayem, *Laurentian university, Canada*  
Ana Mafalda Reis, *Universidade do Porto, Portugal*  
Ana Maria Mendonça, *Universidade do Porto, Portugal*  
André Correia, *Universidade do Porto, Portugal*  
André R.S. Marçal, *Universidade do Porto, Portugal*  
Angel Alberich-Bayarri, *Quiron Hospitals Group, Spain*  
António Luís Pereira do Amaral, *Instituto Superior de Engenharia de Coimbra, Portugal*  
Antonios Gasteratos, *Democritus University of Thrace, Greece*  
Armando Jorge Sousa, *Universidade do Porto, Portugal*  
Audenaert Emmanuel, *Ghent University, Belgium*  
Begoña Calvo, *Universidad de Zaragoza, Spain*  
Bernard Gosselin, *University of Mons, Belgium*  
Christos Grecos, *University of the West of Scotland, UK*  
Constantine Kotropoulos, *Aristotle University of Thessaloniki, Greece*  
Cristian A. Linte, *Mayo Clinic, USA*  
Da-Chuan Cheng, *China Medical University, Taiwan*  
Daniel Cremers, *Technische Universität München, Germany*  
Daniela Iacoviello, *Sapienza Università di Roma, Italy*  
Djemel Ziou, *Université de Sherbrooke, Canada*  
Eduardo Borges Pires, *Universidade de Lisboa, Portugal*  
Eduardo Soudah, *International Center for Numerical Methods in Engineering, Spain*  
Elisabetta Binaghi, *Università degli Studi dell'Insubria, Italy*  
Emmanuel Audenaert, *Ghent University, Belgium*  
Emre Celebi, *Louisiana State University in Shreveport, USA*  
Enrique Alegre, *University of Leon, Spain*  
Fátima Nunes, *Universidade de São Paulo, Brazil*  
Fernao Vistulo de Abreu, *Universidade de Aveiro, Portugal*  
Filipa Sousa, *Universidade do Porto, Portugal*  
Francesco Iacoviello, *University of Cassino and South Lazio, Italy*  
Francisco P.M. Oliveira, *Universidade de Coimbra, Portugal*  
Gerald Schaefer, *Loughborough University, UK*  
Gerhard Holzapfel, *Graz University of Technology, Austria*  
Giuseppe Placidi, *University of L'Aquila, Italy*  
Helcio R.B. Orlande, *Universidade Federal do Rio de Janeiro, Brazil*  
Hemerson Pistori, *Universidade Católica Dom Bosco, Brazil*  
Henryk Palus, *Silesian University of Technology, Poland*  
Hongen Liao, *Tsinghua University, China*  
Isabel N. Figueiredo, *Universidade de Coimbra, Portugal*  
J. Paulo Vilas-Boas, *Universidade do Porto, Portugal*  
Jaime S. Cardoso, *Universidade do Porto, Portugal*  
Jan D'hooge, *Catholic University of Leuven, Belgium*

Javier Melenchón, *Universitat Oberta de Catalunya, Spain*  
Jeff Weiss, *The University of Utah, USA*  
Jimmy T. Efird, *East Carolina Heart Institute, USA*  
João L. Vilaça, *Instituto Politécnico do Cávado e do Ave, Portugal*  
João Paulo Papa, *Universidade Estadual de São Paulo, Brazil*  
Jordi González, *Universitat Autònoma de Barcelona, Spain*  
Jorge A. Silva, *Universidade do Porto, Portugal*  
Jorge Barbosa, *Universidade do Porto, Portugal*  
Jorge Manuel Miranda Dias, *Universidade de Coimbra, Portugal*  
Jorge S. Marques, *Universidade de Lisboa, Portugal*  
Jos Vander Sloten, *Katholieke Universiteit Leuven, Belgium*  
José Augusto Ferreira, *Universidade de Coimbra, Portugal*  
José C. Reis Campos, *Universidade do Porto, Portugal*  
Josef Slapal, *Brno University of Technology, Czech Republic*  
Kristian Sandberg, *Computational Solutions, Inc., USA*  
Krzysztof Rechowicz, *Old Dominion University, USA*  
Lasse Riis Østergaard, *Aalborg University, Denmark*  
Laurent Cohen, *University Paris Dauphine, France*  
Luis Amaral, *Universidade do Minho, Portugal*  
Luis Paulo Reis, *Universidade do Minho, Portugal*  
Luiz Neves, *Universidade Federal do Paraná, Brazil*  
Lyuba Alboul, *Sheffield Hallam University, UK*  
Mads Nielsen, *University of Copenhagen, Denmark*  
Mahmoud El-Sakka, *The University of Western Ontario, Canada*  
Manuel Filipe Costa, *Universidade do Minho, Portugal*  
Manuel González-Hidalgo, *University of the Balearic Islands, Spain*  
Manuel Ujaldón, *University of Málaga, Spain*  
Manuele Bicego, *Università degli Studi di Verona, Italy*  
Masud Rahman, *Department of Infrastructure and Regional Development, Australia*  
Marc Thiriet, *Université Pierre et Marie Curie, France*  
Mário Forjaz Secca, *Universidade Nova de Lisboa, Portugal*  
Mengxing Tang, *Imperial College London, UK*  
Michela Cigola, *University of Cassino and South Latium, Italy*  
Miguel A. González Ballester, *Universitat Pompeu Fabra, Spain*  
Miguel Tavares da Silva, *Universidade de Lisboa, Portugal*  
Miguel Velhote Correia, *Universidade do Porto, Portugal*  
Nguyen Dang Binh, *Hue University, Vietnam*  
Nikos Chrisochoides, *Old Dominion University, USA*  
Nuno Machado, *Instituto Politécnico de Lisboa, Portugal*  
Paola Lecca, *University of Trento, Italy*  
Paola Nardinocchi, *Sapienza Università di Roma, Italy*  
Paolo Di Giambardino, *Sapienza Università di Roma, Italy*  
Paolo Emilio Puddu, *Sapienza Università di Roma, Italy*  
Paulo Flores, *Universidade do Minho, Portugal*  
Pedro Almeida, *Universidade de Lisboa, Portugal*  
Pedro Martins, *Universidade do Porto, Portugal*  
Petia Radeva, *University of Barcelona, Spain*  
Philippe Young, *University of Exeter, UK*  
Ragnar Nortvedt, *Haukeland University Hospital, Norway*  
Reneta Barneva, *SUNY Fredonia, USA*  
Rui Lima, *Universidade do Minho, Portugal*  
Sabina Tangaro, *Istituto Nazionale di Fisica Nucleare, Italy*  
Sanderson L. Gonzaga de Oliveira, *Universidade Federal de Lavras, Brazil*  
Sandra Rua Ventura, *Instituto Politécnico do Porto, Portugal*  
Sebastian Kurtek, *The Ohio State University, USA*  
Sgallari Fiorella, *University of Bologna, Italy*  
Slimane Larabi, *USTHB University, Algeria*  
Susana Oliveira Branco, *Instituto Politécnico de Lisboa, Portugal*

Suzanne M. Shontz, *Mississippi State University, USA*  
Thierry Brouard, *Université François Rabelais, France*  
Vaclav Skala, *University of West Bohemia, Czech Republic*  
Valentin E. Brimkov, *SUNY Buffalo State College, USA*  
Victor Hugo C. de Albuquerque, *Universidade de Fortaleza, Brazil*  
Waldir Leite Roque, *Universidade Federal da Paraíba, Brazil*  
Xiongbiao Luo, *Nagoya University, Japan*  
Xue-Cheng Tai, *University of Bergen, Norway*  
Yongjie (Jessica) Zhang, *Carnegie Mellon University, USA*  
Young Lae Moon, *Chosun University Hospital, South Korea*  
Yubo Fan, *Beihang University, China*  
Zbislaw Tabor, *Cracow University of Technology, Poland*  
Zeyun Yu, *University of Wisconsin at Milwaukee, USA*  
Zhen Ma, *Universidade do Porto, Portugal*

This page intentionally left blank

*Invited lectures*

This page intentionally left blank

# An adaptive non-rigid image registration technique using hierarchical B-splines

Aishwarya Pawar, Yongjie Zhang & Xiaodong Wei

*Department of Mechanical Engineering, Carnegie Mellon University, Pennsylvania, USA*

Yue Jia, Timon Rabczuk, Chiu Ling Chan & Cosmin Anitescu

*Institute of Structural Mechanics, Bauhaus-Universität Weimar, Weimar, Germany*

**ABSTRACT:** An efficient non-rigid image registration framework aims at accurately constructing a spatial transformation, which enables the smooth deformation of an image in order to correctly align it with the target image. We present a new approach to implement non-rigid image registration, in which the spatial transformation is constructed by using hierarchical B-spline basis functions. Based on the concept of minimization of an energy functional, which drives the registration process by minimizing the difference between the images, the control points are dynamically updated resulting in a new spatial transformation function at each stage of the registration process. Moreover the registration is carried out on a set of successively refined grids, which makes the method more efficient and faster. Instead of uniformly refining the entire control grid, this dynamic implementation is carried out by an adaptive grid refinement procedure, wherein we detect large deformation areas and locally refine these regions. This makes the computation of the spatial transformation more efficient and accurate with fewer number of control points. A comparison between the adaptive grid refinement and uniform grid refinement methods is explained using synthetic and medical images as examples.

## 1 INTRODUCTION

Image registration is an image processing technique, which has a broad range of applications, but is most prominently applied in the field of computer vision and medical image analysis (Goshtasby 2005), (Zhang et al. 2012). It involves the alignment of two or more images so that there is proper integration of useful data from separate images. The important features from each of these images can be combined together to achieve a better understanding of the complete picture of the image data that is being studied. There have been many methods, by which image registration is carried out. A comprehensive survey of several such methods with their respective advantages and limitations has been covered in detail in (Maintz and Viergever 1998) and (Zitova and Flusser 2003).

The process of image registration involves deforming one image, usually called the source image, by computing a spatial transformation function, which aligns it with the other image, called the target image. The spatial transformation is defined in a parametric form, and the parameters are computed through minimization of an energy functional that is based on using a similarity metric to drive the registration process. Such an

implementation has been proposed in (Rueckert et al. 1999) and (Tustison et al. 2007). According to the type of transformation, image registration methods are classified into rigid and non-rigid registration models. Three types of transformations carried out under rigid registration are translation, rotation and reflection. Rigid registration is generally used to register images that undergo small deformations. In non-rigid registration, we can deal with large deformation problems. However, certain regularization constraints need to be imposed on the deformation to ensure the smoothness of the solution and proper convergence to get the desired result.

In spline based registration (Jia et al. 2015), (Leng et al. 2013) and (Szeliski and Coughlan 1994), a spline-based control grid overlays the image, and a spatial transformation is constructed to deform this grid. An efficient and stable algorithm was suggested by (Jia et al. 2015), in which the spatial transformation was based on B-spline basis functions, and the control points were computed dynamically to enable non-rigid registration. Moreover the method was made more efficient by computing the spatial transformation on a set of successively refined grids. In our work, instead of uniformly refining the grid, we employ an adaptive mesh refinement scheme using

hierarchical B-splines. Hierarchical B-spline based image registration methods have been previously used in (Hansen et al. 2008) and (Xie and Farin 2004). In the current work, we formulate an adaptive grid refinement strategy by using hierarchical B-splines and interfacing it with the minimization of the energy functional as given in (Jia et al. 2015). Thus we can solve large deformation problems with adaptive grid refinement strategy through the dynamic computation of fewer control points. This method is more efficient and faster as compared to the uniform grid refinement method.

The paper is organized as follows. In section 2, we first present the framework of image registration and introduce our adaptive grid refinement strategy using hierarchical B-splines. In section 3, we apply our approach to solve two numerical examples, by comparing the results in terms of computational time and accuracy with the results obtained using uniform grid refinement strategy. The paper ends with a brief summary in section 4.

## 2 IMAGE REGISTRATION MODEL

### 2.1 Registration framework

The process of registration of two images,  $I_1(\mathbf{x})$  (the source image) and  $I_2(\mathbf{x})$  (the target image), is based on aligning  $I_1(\mathbf{x})$  with  $I_2(\mathbf{x})$  through a computed spatial transformation function. We use bi-cubic B-splines to obtain a  $C^2$  mapping from  $R^2 \rightarrow R^2$ . In particular, the spatial transformation function  $f(\mathbf{x})$  is defined as:

$$f(\mathbf{x}) = \sum_{k=1}^{N_b} \mathbf{P}_k \phi_k(\mathbf{x}), \quad (1)$$

where  $\mathbf{P}_k$  are a set of control points and  $\phi_k(\mathbf{x})$  are the bivariate basis functions.  $N_b$  represents the total number of basis functions used.  $\phi_k(\mathbf{x})$  is the tensor product of univariate B-spline basis functions  $N_{i,p}(\xi)$  and  $N_{j,q}(\eta)$ , defined on the knot vectors,  $\Xi = \{\xi_1, \dots, \xi_{n_1+p+1}\}$  and  $H = \{\eta_1, \dots, \eta_{n_2+q+1}\}$  in  $\xi$  and  $\eta$  directions, respectively.  $n_1$  and  $n_2$  are the number of univariate basis functions.  $p$  and  $q$  are the degree of polynomials in  $\xi$  and  $\eta$  directions, respectively. This is written as

$$\phi_{i,j}(\mathbf{x}) = N_{i,p}(\xi)N_{j,q}(\eta), \quad (2)$$

where  $i = 1, 2, \dots, n_1$  and  $j = 1, 2, \dots, n_2$ . The details on B-spline functions can be found in (Piegl and Tiller 2012) and (Cottrell et al. 2009). The initial control points are defined to create an identity map,  $f(\mathbf{x}) = \mathbf{x}$ , at the beginning of the registration. This represents the initial source image. After representing  $I_1(\mathbf{x})$ , we proceed to find the optimal transformation

$f(\mathbf{x})$ , such that,  $I_1(f(\mathbf{x})) \approx I_2(\mathbf{x})$ . The transformation of the source image is driven by minimizing an energy functional, which uses the Sum of Squared Differences (SSD) of the two images as the driving force to align the images. The detailed description of the formulation of the energy functional can be found in (Jia et al. 2015) and (Leng et al. 2013). The energy functional is given as:

$$\begin{aligned} E(\mathbf{x}) = & \int_{\Omega} (I_2(\mathbf{x}) - I_1(f(\mathbf{x})))^2 d\Omega \\ & + \lambda_1 \int_{\Omega} (\|f_{,\xi}(\mathbf{x})\|_2^2 + \|f_{,\eta}(\mathbf{x})\|_2^2) d\Omega \\ & + \lambda_2 \int_{\Omega} (\|f_{,\xi}(\mathbf{x})\|_2^2 \|f_{,\eta}(\mathbf{x})\|_2^2 \\ & - \langle f_{,\xi}(\mathbf{x}), f_{,\eta}(\mathbf{x}) \rangle^2), \end{aligned} \quad (3)$$

where  $\lambda_1$  and  $\lambda_2$  are regularization parameters that are set according to the characterization of the problem.  $\lambda_1$  and  $\lambda_2$  are set as 0.0001 in our method.  $f_{,\xi}(\mathbf{x})$  and  $f_{,\eta}(\mathbf{x})$  are the derivatives of  $f(\mathbf{x})$  in  $\xi$  and  $\eta$  directions respectively and  $\langle f_{,\xi}(\mathbf{x}), f_{,\eta}(\mathbf{x}) \rangle$  is the inner product operator which is equal to  $f_{,\xi}(\mathbf{x})^T f_{,\eta}(\mathbf{x})$ . The first term in Equation (3) drives the registration process by minimizing the differences between the images. Unlike rigid deformations, non-rigid deformations can achieve image deformations of larger magnitude. But this sometimes results in unrealistic changes within the image. Thus the second and third terms in Equation (3), called the regularization constraints, are used to prevent these unrealistic changes and also make the transformation as smooth as possible.

Since the energy functional is a function of  $f(\mathbf{x})$ , we differentiate  $E(\mathbf{x})$  with respect to the control points  $\mathbf{P}$ , and update their values according to the equation:

$$\mathbf{P}^{t+1} = \mathbf{P}^t - \epsilon \delta E(\mathbf{x}), \quad (4)$$

where  $\mathbf{P}^{t+1}$  and  $\mathbf{P}^t$  are control points evaluated at the current and previous iterations respectively.  $\epsilon$  denotes the time step which is chosen based on the problem.  $\delta E(\mathbf{x})$  is defined as given in (Jia et al. 2015).

Similarity metrics such as Sum of Squared Differences (SSD) and the Ratio of Similarity (RS) (Jia et al. 2015), provide a way to quantitatively evaluate the results of the registration process. RS is defined as

$$RS(\%) = \left( 1 - \frac{\|I_2(\mathbf{x}) - I_1^k(f(\mathbf{x}))\|_{L^2}}{\|I_2(\mathbf{x}) - I_1(\mathbf{x})\|_{L^2}} \right) \times 100, \quad (5)$$

where  $I_1^k(f(\mathbf{x}))$  is the image obtained at the  $k$ th time step. When  $RS = 100\%$ , this corresponds to the completely registered image.

## 2.2 Hierarchical B-splines

This section begins with a brief review of hierarchical B-spline basis functions. Let us consider univariate basis functions defined on the open knot vectors  $\Xi = \{\xi_1, \dots, \xi_{n_l+p+1}\}$ , where  $n_l$  and  $p$  are the number of univariate basis functions and the degree of the polynomials in the  $\xi$  direction, respectively. The local support of the basis function  $N_{i,p}(\xi)$  is in the knot interval  $[\xi_i, \xi_{i+p+1}]$ , denoted by  $\text{supp}(N_{i,p}(\xi))$ . We define the maximum level of refinement as  $l_{\max}$ . The knot vectors of the current level are obtained by bisecting those of the previous level. A basis function  $N_{i,p}^l(\xi)$  at level  $l$  can be represented as a linear combination of a subset of the basis functions from the next refinement level. These basis functions are called as the children basis functions of  $N_{i,p}^l(\xi)$ . We have

$$N_{i,p}^l(\xi) = \sum_{k=0}^{N_c-1} S_{k,p} N_{k,p}^{l+1}(\xi), \quad (6)$$

where  $N_{k,p}^{l+1}(\xi)$  are the children basis functions and  $N_c$  is the number of children basis functions. The refinement coefficients  $S_{k,p}$  are calculated using the Oslo Algorithm (Cohen et al. 1980). In the similar manner the children of the bivariate basis functions  $\phi_{i,j}^l(\mathbf{x})$  (Equation (2)), are the basis functions found in the next refinement level, which are completely contained in the region of local support,  $[\xi_i, \xi_{i+p+1}] \times [\eta_j, \eta_{j+q+1}]$ .

The hierarchical refinement procedure involves replacing a selected set of basis functions at a particular refinement level with their children basis functions. This enables local refinement of the selected basis functions at different refinement levels, ( $l = 0, 1, \dots, l_{\max}$ ). Further details on the implementation of local refinement by using hierarchical B-splines can be found in (Bornemann and Cirak 2013). We explain the implementation of local refinement of B-spline basis functions on two consecutive levels,  $l$  and  $l+1$ , in the following steps:

1. At a particular refinement level  $l$ , a subset of B-spline basis functions satisfying the refinement criterion are identified, which form the set  $\phi_a^l$ . The remaining B-spline basis functions are set as active ( $\phi_a^l$ ).
2. The children B-spline basis functions of all the basis functions  $\phi_a^l$  are set as active ( $\phi_a^{l+1}$ ).
3. The process ends by collecting all the active B-spline basis functions at levels  $l$  and  $l+1$ , in order to get the hierarchical B-splines basis functions,  $\phi_{hb}^{l+1}$ .

Thus we can obtain,

$$\phi_{hb}^{l+1} = \phi_a^l \cup \phi_a^{l+1}. \quad (7)$$

The refinement procedure is implemented in a recursive manner till we reach the refinement level  $l_{\max} - 1$ .

## 2.3 Adaptive local refinement

According to the multilevel technique described in (Jia et al. 2015), the errors in the registration will blow up after certain number of iterations when we use only one mesh to compute the spatial transformation  $f(\mathbf{x})$ . Hence a new technique is developed, wherein the registration process is carried out on a set of grids which are refined at each level. The final image obtained at a particular level  $l$  is used as the source image to compute the solution on the refined grid at level  $l+1$ . This is done to prevent large distortion of the coarse level elements when the image undergoes large deformation. This process continues till we reach the grid at the maximum refinement level. Thus the solution is computed on a set of grids that are refined uniformly in the entire image domain. It can be seen however, that in the image domain, there are some regions where there will be large deformation, such as areas where the differences in the intensities of the source and target image are large, whereas the remaining regions are either well-matched or contain low information, such as image background. Thus we can use this information and locally refine the regions where there is large deformation keeping the rest of the grid elements coarse since they will not distort too much to give unrealistic results. We endeavor to use the dynamical implementation on a hierarchy of locally refined grids so that we can avoid uniformly increasing the number of control points in the whole domain.

The detailed outline of the algorithm is described in the following: Given a pair of images,  $I_1(\mathbf{x})$  and  $I_2(\mathbf{x})$ . The process starts at level-0, which is a uniform grid. The initial set of control points  $\mathbf{P}_{hb}^0$  are defined such that  $f(\mathbf{x}) = \sum_{i=1}^N \mathbf{P}_{hb,i}^0 \phi_{hb,i}^0 = \mathbf{x}$ .  $N$  is the total number of basis functions at level-0. For  $l = 0, 1, \dots, l_{\max} - 1$ ,

1. Compute the absolute difference in the intensities,  $I_d$ , between  $I_1(f(\mathbf{x}))$  and  $I_2(\mathbf{x})$ , at the cell coordinates of the hierarchical B-spline control grid at level- $l$ . Here  $I_1(f(\mathbf{x}))$  is the evolved image obtained from the previous level- $(l-1)$ . It is the initial source image if  $l = 0$ . Let  $D_{\text{mean}}$  be the average difference in the intensities between the two images.
2. Loop over the active B-spline basis functions  $\phi_j^l$ ,  $j = 1, 2, \dots, N_b$ , where  $N_b$  is the total number of active B-spline basis functions at level- $l$ . Loop over  $\text{supp}(\phi_j^l)$ , and find the average of the  $I_d$  values in  $\text{supp}(\phi_j^l)$ , denoted as  $D_j$ . If  $D_j > \rho D_{\text{mean}}$ , then we refine the particular B-spline according

to the refinement procedure explained in section 2.2.  $\rho$  is the threshold parameter set by the user. A higher value of  $\rho$  results in less control grid refinement.

3. After carrying out the refinement, we collect all the active B-spline basis functions, control points and cells at level- $l$ , to get the hierarchical basis functions, control points and cells,  $\phi_{hb}^{l+1}$ ,  $\mathbf{P}_{hb}^{l+1}$  and  $\Omega^{l+1}$  respectively.
4. Begin the loop for iterations,  $k = 1, 2, \dots, i_{max}$ , where  $i_{max}$  is the maximum number of iterations for a particular level.
  - a. Compute the energy functional value,  $\delta E^{k,l+1}(\mathbf{x})$ .
  - b. We update the control points, according to the relation  $\mathbf{P}_{hb}^{k+1,l+1} = \mathbf{P}_{hb}^{k,l+1} - \epsilon \delta E^{k,l+1}(\mathbf{x})$ .
  - c. Using the updated control points, the spatial transformation is computed and we have  $f(\mathbf{x}) = \sum_{j=1}^{N_{hb}} \mathbf{P}_{hb,j}^{k+1,l+1} \phi_{hb,j}^{l+1}$ , where  $N_{hb}$  is the total number of hierarchical B-splines.
  - d. Using the spatial transformation  $f(\mathbf{x})$ , we obtain image  $I_1(f(\mathbf{x}))$ .
5. The final image obtained at the end of the iterations loop is used as the source image for the next refinement level.
6. The process terminates upon reaching the desired accuracy, that is, the best possible match between the source and target images.

### 3 IMPLEMENTATION AND DISCUSSIONS

The efficiency of the proposed registration framework is evaluated on a pair of synthetic images and a pair of medical images. The resulting images obtained at the end of each refinement level are recorded. The difference between the resulting image and the target image at each refinement level is illustrated. The results were generated on a computer with 2.5 GHz quad-core Intel Core i7 processor and 16 GB RAM. Finally, a comparison of the convergence of the solution using the similarity metrics and the time taken for each level are shown in a tabulated manner along with the results obtained with uniform grid refinement technique.

#### 3.1 Synthetic image example (ring to sun)

The registration model is tested on a pair of synthetic images (see Figs. 1 (a–b)) to evaluate the efficiency of the algorithm. At the first level, we use a uniform control grid consisting of  $30 \times 30$  elements and  $33 \times 33$  control points. The threshold parameter  $\rho$ , is set as 1.5. The number of control points in the second and third refinement levels are 2,328 and 6,254 respectively. The time step for the first 20 iterations is set as 0.002. For the

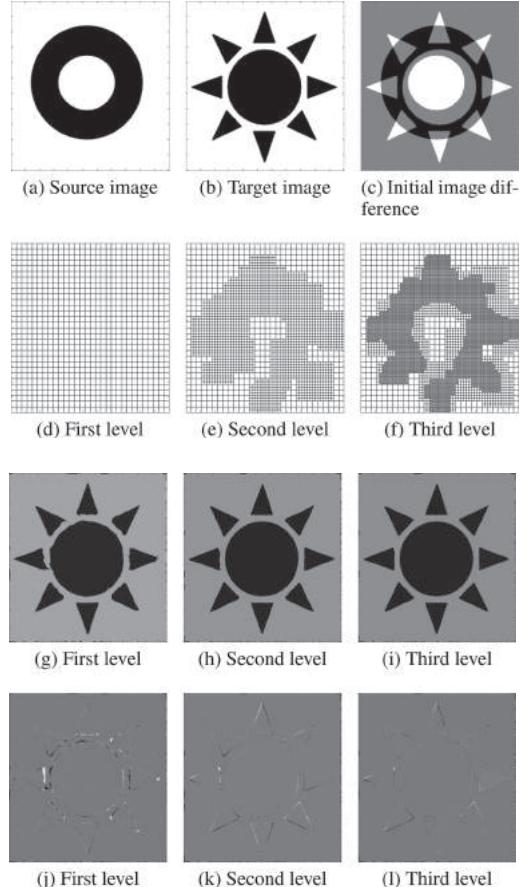


Figure 1. Synthetic image results. The initial source image, the target image and initial differences are shown in (a–c). Meshes generated at different levels based on hierarchical B-spline basis refinement are shown in (d–f). The resulting images obtained after each stage of registration are shown in (g–i). The differences between the target and resulting images obtained at each level are shown in (j–l).

remaining iterations, the time step was reduced to 0.001 in order to achieve a smooth convergence. The solution reached the desired accuracy at the end of the third refinement level. For the first two levels, the total number of iterations for each level is 40, whereas for the third level, 20 iterations are used. The values of  $RS$  at the end of each level are 91.36%, 97.02% and 97.96%. The final images obtained after each stage of refinement are shown in Figures 1 (g–i). The difference between the resulting image and target image at the end of each level are shown in Figures 1 (j–l). It can be seen that the maximum deformation in the image is observed on the coarse level and the spatial

Table 1. Synthetic image (Ring to Sun example, image size =  $220 \times 220$ ): Comparison of the convergence of the solution and the time taken using our method (HBS) with uniform B-spline grid refinement (UBS).

Levels (Iterations)	RS (%)		CPU Time (in seconds)		SSD (initial = $2.2943 \times 10^4$ )		Control points used	
	HBS	UBS	HBS	UBS	HBS	UBS	HBS	UBS
Level 1 (40)	91.36	91.00	74.06	70.84	1,982.28	2,064.87	1,089	1,089
Level 2 (40)	97.02	96.65	155.25	303.60	682.76	768.80	2,328	3,969
Level 3 (20)	97.96	97.80	333.67	393.27 (at 6th iteration)	468.62	505.55	6,254	15,129
Total time (seconds)			562.98	767.71				

Table 2. Medical image (Brain MRI example, image size =  $320 \times 320$ ): Comparison of the convergence of the solution and the time taken using our method (HBS) with uniform B-spline grid refinement (UBS).

Levels (Iterations)	RS (%)		CPU time (in seconds)		SSD (initial = $2.55 \times 10^3$ )		Control points used	
	HBS	UBS	HBS	UBS	HBS	UBS	HBS	UBS
Level 1 (10)	54.05	54.05	81.81	37.40	1,170.30	1,170.30	1,849	1,849
Level 2 (20)	75.41	73.73	208.60	326.61	626.25	669.12	3,997	6,889
Level 3 (20)	79.07	79.51	770.10	2,247.60	532.95	521.93	11,089	26,569
Total time (seconds)			1,060.51	2,611.61				

transformation is then evaluated on the finer levels to further improve the accuracy. Table 1 shows the comparison of the time taken, the value of *RS* and the number of control points used for each refinement level of the grid. From Table 1, it can be seen that the total time taken to achieve the desired accuracy using the adaptive refinement scheme is less than that of the uniform refinement strategy. Although the initialization time taken to set up the hierarchical data structure in adaptive grid refinement is more than that taken for the uniform grid refinement, during the computation of the deformed image, the time taken by the adaptive refinement method is much less. This is because the number of control points used to compute the spatial transformation of the deforming image is much lower than the uniformly refined grid. This makes the adaptive registration process much more efficient.

### 3.2 Medical image example (Brain MRI)

The registration model is tested on a pair of medical images as shown in Figures 2 (a–b). Unlike synthetic images, medical images have more complex feature information. Thus we use more control points at the first refinement level so as

to capture the features of the images correctly. Also, the time step is reduced as compared to synthetic images in order to gradually and smoothly deform the image. In this example, we use a uniform control grid consisting of  $40 \times 40$  elements and  $43 \times 43$  control points at the first level. The threshold parameter  $\rho$ , is set as 1.5. The number of control points in the second and third refinement levels are 3,997 and 11,089 respectively. The time step for the entire registration process is set to be 0.00025. The solution reached the desired accuracy at the end of the third refinement level. For the first level, the total number of iterations equals to 10, whereas for the second and third level they were increased to 20. The values of *RS* at the end of each level are 54.05%, 75.41% and 79.07%. The final images obtained at each refinement level are shown in Figures 2 (g–i). The difference between the resulting and target images at the end of each level are shown in Figures 2 (j–l). Table 2 shows the comparison of the time taken, the value of *RS* and the number of control points used for each refinement level of the grid. From Table 2, it can be seen that the total time taken to achieve the desired accuracy using adaptive refinement scheme was less than that of the uniform refinement strategy.

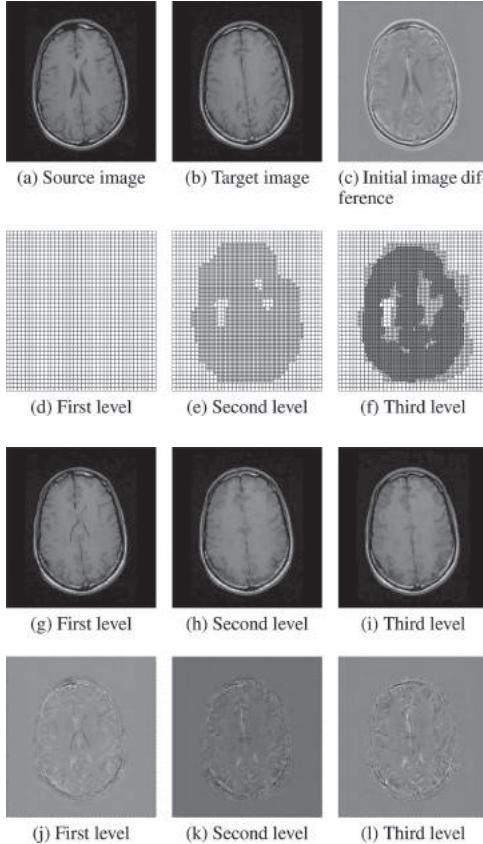


Figure 2. Medical image results. The initial source image, the target image and initial differences are shown in (a–c). Meshes generated at different levels based on hierarchical B-spline basis refinement are shown in (d–f). The resulting images obtained after each stage of registration are shown in (g–i). The differences between the target and resulting images obtained at each level are shown in (j–l).

#### 4 CONCLUSION

A non-rigid image registration method using adaptive grid refinement based on hierarchical B-spline basis functions has been presented. By detecting the large deformation areas in the image, the control grid is locally refined in these regions. Hence we use fewer number of control points resulting in an efficient and accurate computation of the spatial transformation function. A comparison between our method and uniform grid refinement shows that our method performs better by achieving the same accuracy in a shorter amount of time. We demonstrate the method on artificial and medical images. The algorithm has worked successfully on medical images and hence provides a promising direction for future work in several applications

such as medical image interpolation and further extension to 3-D registration problems.

#### ACKNOWLEDGEMENTS

The medical images were provided from (<http://overcode.yak.net/15>). The research at Carnegie Mellon University was supported in part by NSF CAREER Award OCI-1149591. The research at Bauhaus University Weimar was supported in part by the ITN-INSIST funded by the EU-FP7.

#### REFERENCES

- Bornemann, P.B. & F. Cirak (2013). A subdivision-based implementation of the hierarchical B-spline finite element method. *Computer Methods in Applied Mechanics and Engineering* 253, 584–598.
- Cohen, E., T. Lyche, & R. Riesenfeld (1980). Discrete B-splines and subdivision techniques in computer-aided geometric design and computer graphics. *Computer graphics and image processing* 14(2), 87–111.
- Cottrell, J.A., T.J. Hughes, & Y. Bazilevs (2009). *Isogeometric analysis: toward integration of CAD and FEA*. John Wiley & Sons.
- Goshtasby, A.A. (2005). *2-D and 3-D image registration: for medical, remote sensing, and industrial applications*. John Wiley & Sons.
- Hansen, M.S., R. Larsen, B. Glocker, & N. Navab (2008). Adaptive parametrization of multivariate B-splines for image registration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8.
- Jia, Y., Y. Zhang, & T. Rabczuk (2015). A novel dynamic multilevel technique for image registration. *Computers & Mathematics with Applications* 69(9), 909–925.
- Leng, J., G. Xu, & Y. Zhang (2013). Medical image interpolation based on multi-resolution registration. *Computers & Mathematics with Applications* 66(1), 1–18.
- Maintz, J.A. & M.A. Viergever (1998). A survey of medical image registration. *Medical image analysis* 2(1), 1–36.
- Piegls, L. & W. Tiller (2012). *The NURBS book*. Springer Science & Business Media.
- Rueckert, D., L.I. Sonoda, C. Hayes, D.L. Hill, M.O. Leach, & D.J. Hawkes (1999). Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging* 18(8), 712–721.
- Szeliski, R. & J. Coughlan (1994). Hierarchical spline-based image registration. *IEEE Conference on Computer Vision and Pattern Recognition*, 194–201.
- Tustison, N.J., B. Avants, G.J.C., et al. (2007). Improved FFD B-spline image registration. In *IEEE 11th International Conference on Computer Vision*, pp. 1–8.
- Xie, Z. & G. Farin (2004). Image registration using hierarchical B-splines. *IEEE Transactions on Visualization and Computer Graphics* 10(1), 85–94.
- Zhang, Y., Y. Jing, X. Liang, G. Xu, & L. Dong (2012). Dynamic lung modeling and tumor tracking using deformable image registration and geometric smoothing. *Molecular & Cellular Biomechanics* 9(3), 213–226.
- Zitova, B. & J. Flusser (2003). Image registration methods: a survey. *Image and vision computing* 21(11), 977–1000.

# Medical image segmentation using Object Shape Models: A critical review on recent trends, and alternative directions

Alexandre Xavier Falcão, Thiago Vallin Spina & Samuel Botter Martins\*

*Institute of Computing, University of Campinas, Campinas, Brazil*

Renzo Phellan

*Eldorado Research Institute, Campinas, Brazil*

**ABSTRACT:** Segmentation is important to define the spatial extension of body anatomic structures (objects) in medical images for quantitative analysis. In this context, it is desirable to eliminate (at least minimize) user interaction. This aim is feasible by combining object delineation algorithms with Object Shape Models (OSMs). While the former can better capture the actual shape of the object in the image, the latter provides shape constraints to assist its location and delineation. We review two important classes of OSMs for medical image segmentation: Statistical (SOSMs) and Fuzzy (FOSMs). SOSMs rely on the image mapping onto a reference coordinate system, which indicates the probability of each voxel to be in the object (a probabilistic atlas built from a set of training images and their segmentation masks). Imperfect mappings due to shape and texture variations asks for object delineation algorithms, but the methods usually assume that the atlas is at the best position for delineation. Multiple atlases per object can mitigate the problem and a recent trend is to use each training mask as an individual atlas. By mapping them onto the coordinate system of a new image, object delineation can be accomplished by label fusion. However, the processing time for deformable registration is critical to make SOSMs suitable for large scale studies. FOSMs appear as a recent alternative to avoid reference systems (deformable registration) by translating the training masks to a common reference point for model construction. This relaxes the shape constraints, but asks for a more effective object delineation algorithm and some efficient approach for object's location. One of the solutions, named *optimum object search*, translates the model inside an estimated search region in the image while a criterion function guides translation and determines the best delineated object among candidates. This makes segmentation with FOSMs considerably faster than with SOSMs, but SOSMs that adopt the optimum object search can be more effective and with less atlases per object. We then discuss the pros and cons of the recent FOSM and SOSM approaches by providing alternative directions, which also include the user to correct segmentation errors and improve the models.

## 1 INTRODUCTION

Medical image segmentation is a crucial step in the quantitative analysis of body anatomic structures (objects) for the non-invasive diagnosis and treatment of diseases. We may divide image segmentation into two tightly coupled tasks: object recognition and delineation (Falcão et al. 1998). Recognition involves locating the whereabouts of the object in an image and verifying errors in segmentation, while delineation defines the object's spatial extent.

Humans usually outperform computers in object recognition, while computers are more precise for

object delineation. This has motivated the use of interactive image segmentation methods that exploit a synergism between user and computer, by combining user input for approximately locating the object with some image-based delineation method to extract it from the background (Falcão et al. 1998, Falcão et al. 2000, Falcão et al. 2004, Rother et al. 2004, Boykov and Funka-Lea 2006, Grady 2006, Couplie et al. 2011, Spina et al. 2014). The main drawback in interactive segmentation is its limited capacity for large scale studies in medical image analysis. Moreover, image segmentation is a tedious task prone to human errors, specially when dealing with 3D MR and CT images, since the results must be visually assessed in multiple planes (sagittal, axial, and coronal).

In order to eliminate or at least minimize user intervention in segmentation, the shape and texture properties of the object are an important

---

\*The authors would like to thank CNPq (grants 302970/2014-2 and 479070/2013-0), FAPESP (grant 2011/01434-9), and CAPES for the financial support.

prior information. *Active Shape Models* (ASMs) (Cootes et al. 1995), for instance, were proposed to learn the possible manifestations of the object's shape from a set of training images with their segmentation masks, and encode them into a linearly deformable shape model. The shape is represented as a set of boundary points (landmarks) whose correspondence must be specified in each mask. The landmarks are mapped onto a common coordinate system in order to estimate their mean position and possible variations, limiting the deformations of the model. Similarly, texture variations can be incorporated into an *Active Appearance Model* (AAM) (Cootes et al. 1998).

ASM/AAM approaches may be interesting solutions for object delineation, but they do not solve recognition. Usually, the user has to place the model near the object for subsequent model deformation (with possible translation) to match the landmarks with the strongest edges for object delineation (Rueda 2010, Ghose et al. 2010, Inamdar and Ramdasi 2013). The specification of corresponding landmarks in different masks is a challenging task (Rueda 2010), whose the interactive solution makes model construction a tedious and error-prone step. Such a problem can be avoided with deformable image registration (Pluim et al. 2003), but this essentially suggests the use of another type of Object Shape Model (OSM)—a probabilistic atlas model (Mazziotta et al. 1995).

In this paper, we review two important classes of OSMs for medical image segmentation: Statistical (SOSMs) and Fuzzy (FOSMs). SOSMs are popular approaches, being traditionally called statistical atlas-based models (Cabezas et al. 2011, Harrigan et al. 2015, Rusu et al. 2011, Lötjönen et al 2011). Similarly to ASMs, the model is built from a set of training images with their corresponding segmentation masks. The simplest approach for model construction requires the selection of one reference image from the training set and the deformable mapping of the remaining ones onto the reference coordinate system. The resulting deformation field is applied to the corresponding masks and the object label frequency at each coordinate defines a *prior probability map* (a probabilistic atlas). By mapping a new image onto the atlas coordinate system, the object can be delineated by thresholding the prior probability map (Mazziotta et al. 1995), by estimating and thresholding a posterior probability map (Vos et al. 2013), or by applying some image-based delineation algorithm (Grau et al. 2004).

The main drawbacks of classical SOSMs are related to their dependency on the deformable image registration algorithm. They assume that object recognition is solved by registration, which grants a near-accurate mapping from any test

image to the model's reference image. This is not valid in practice, given that a single reference image cannot represent all object instances available in nature. This problem has motivated the use of multiple SOSMs per object, converging to the point of considering each training mask as an independent atlas in recent works (Wang et al. 2013, Akhondi-Asl and Warfield 2013). That is, for a new test image, each training image must be registered with the test image, the deformation field is applied to the corresponding mask, and the registered segmentation masks go through a process referred to as *label fusion* to output the final delineation of the target image. The argument is that the strategy of Multiple Atlas Label Fusion (MALF) makes better use of the training dataset, by increasing the chance of correctly segmenting the target image.

Although MALF strategies seem to improve accuracy, the time consumed by the registration step increases prohibitively, thereby making them unfeasible for large scale studies. Moreover, they assume that the dataset used for training is fixed and will never increase during testing. Updating OSMs is necessary in clinical settings to ensure adaptability to new challenging images. Lastly, as pointed out in (Phellan Falcão & Udupa 2014), registration does not solve object recognition, since the object's boundary might not even be confined in the *uncertainty region* of the model afterwards (Phellan, Falcão, & Udupa 2014) (i.e., the set of voxels with prior probability between (0, 1)).

FOSMs have been recently proposed as an alternative to SOSMs, with the aforementioned issues in mind (Miranda et al 2008, Miranda et al. 2009). Instead of registering the training images to obtain a prior probability map, the segmentation masks are aligned by translation to a common reference point (e.g., the masks' geometric centers) and combined to output a *fuzzy image*. As a consequence, the construction of a FOSM is much faster. The values in the fuzzy image are proportional to the membership that each voxel has to the object of interest, although they do not reflect prior probabilities since the voxel values do not correspond to a same location in the actual anatomy. The lack of registration makes the uncertainty region of a fuzzy image thicker than that of a classic SOSM's prior probability map (Fig. 1). As a consequence, the uncertainty region of a FOSM requires more effective delineation, although it better accommodates the boundary of the real object when the model is placed at the object's center in the test image.

After construction, the automatic search for the object of interest and segmentation using FOSM involves finding the optimal position for the model in the test image. This is typically achieved by a local search for the object inside a given region of the test image. In one of the possible solutions,

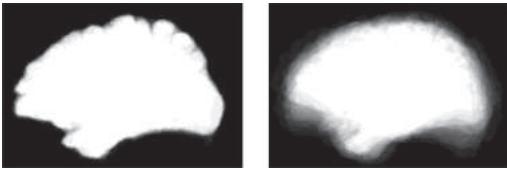


Figure 1. Sagittal view of an SOSM's prior probability map for the brain's left hemisphere on the right, and of a FOSM's fuzzy image on the left.

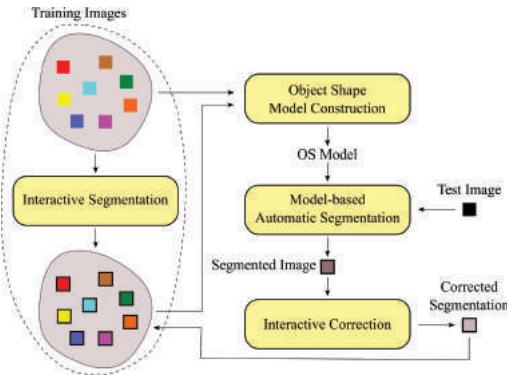


Figure 2. Overall pipeline for effective medical image segmentation using OSMs.

named *optimum object search*, the model translates inside an estimated search region in the test image while a criterion function guides translation and determines the best delineated object among candidates. Delineation often considers the external and internal portions of the fuzzy image as seeds (i.e., voxels with values {0, 1}, respectively), but other delineation methods can be used as well.

The key idea is that FOSMs assume that the model is initially mispositioned and try to correct it. This assumption is in stark contrast to SOSMs' reliance on the registration algorithm. Hence, Phellon, Falcão, & Udupa (2014) have proposed to use the FOSM's optimum object search procedure to improve object delineation in SOSMs, after initially positioning the model through deformable registration. Their promising results suggest a new interesting direction of research, which would combine classic SOSMs and FOSMs instead of using MALF.

Both statistical and fuzzy OSMs have been actively improved in recent years to address their shortcomings. Notwithstanding, we envision that effective image segmentation systems must allow user intervention when automatic segmentation fails. Figure 2 depicts such scheme. Giving the user the opportunity to correct mistakes is key, since we

cannot assume that the dataset used to compute the models captures all possible nuances observed in nature, nor that the models are capable of handling abnormalities. The corrected segmentation is also paramount to update the model in order to improve the delineation of future test images.

In the remainder of the paper, we highlight recent improvements developed for SOSMs (Section 2) and FOSMs (Section 3), with their pros and cons. We then provide details on how the automatic segmentation result may be interactively corrected by using image-based delineation methods (Section 4), before stating conclusions in Section 5.

## 2 STATISTICAL OBJECT SHAPE MODELS

MALF approaches typically compute a confidence weight for each atlas (template), which dictates the importance that each mask should have in the fusion process. In (Wang et al. 2013), the local intensity similarity between the test image and each pair of template images is considered for this purpose. To mitigate errors in registration of the atlases, the test image is further divided into patches and, for each test patch, the method searches locally for the most similar patch in each atlas image (finding the corresponding patch in each mask). The labels in the test patch are defined by weighted label fusion from the labels of the corresponding mask patches. The object is then defined by local shape imposition from the models. In contrast, Akhondi-Asl and Warfield (2013) argue that local intensity similarity is a poor measure for relating the true target image segmentation with the dataset templates. They learn instead Gaussian Mixture Model classifiers using the intensity and label maps of the registered atlases, which output probabilistic segmentations of the target image. These probabilistic segmentations are then fused together to determine the final binary segmentation mask.

To reduce the time consumed by the registration of multiple atlases, Sanroma et al. (2014) argue that not all atlases are necessarily useful for segmentation of a given target image. In fact, the inclusion of misleading atlases may undermine labeling performance, thereby making atlas selection a necessity. They propose a selection technique based on supervised learning that is applied before deformable registration, which ranks the atlases according to the expected segmentation accuracy that each one will have on the target image.

The work of Sanroma et al. (2014) supports our earlier claims when they indicate that smaller and more selective may perform better in practice. We believe that improving the classic framework is more fruitful since it is supported by a solid statistical background. During the optimum object

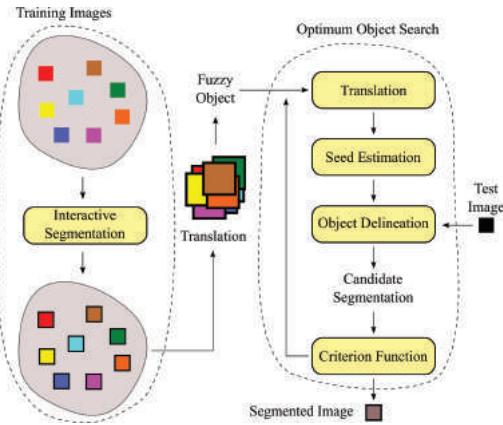


Figure 3. Example of segmentation using FOSM.

search with SOSM in (Phellan, Falcão, & Udupa 2014), the assessment of whether the model is correctly positioned is done synergistically by using the probability map to constrain the location where a delineation algorithm is applied (i.e., only on the voxels of the test image superimposed by the uncertainty region of the model). The delineation generates a candidate segmentation mask that is evaluated according to a criterion function, whose score is expected to be maximum when the mask corresponds to the real object. Such an approach object search (Fig. 3) comes from the inception of FOSMs (Miranda et al. 2008).

### 3 FUZZY OBJECT SHAPE MODELS

We may define a FOSM as a triple composed of a fuzzy image, a delineation algorithm, and a criterion function (Miranda et al. 2008). Different approaches devised for each of these elements may lead to improvements of the model. For instance, an important issue to be considered when using FOSMs is that a thicker uncertainty region may degenerate the model's fuzzy image, thereby decreasing performance. In (Miranda et al. 2009), we proposed a solution to circumvent this problem by grouping the training images according to their shape similarities, before generating multiple FOSMs to represent a single object of interest. The proposed algorithm starts by computing the similarity between all training shapes using the Dice coefficient. Then, maximal cliques are selected to form groups whose similarity is above a given threshold. Each group produces a FOSM, previously denoted as *Object Cloud Model*, that is used to search for the object of interest using a multi-scale version of the procedure shown in Figure 3.

The FOSM whose best candidate segmentation mask achieves the highest score of the criterion function, among all of the ones in the *Cloud Bank*, is selected to output the final delineation.

The Cloud Bank Model was later extended to search and segment multiple objects in the Cloud System Model (Miranda et al. 2010). The interesting advancement in this case was that the objects may be searched together, taking their relative positions into account to improve object recognition. This inspired recent works on body-wide hierarchical Automatic Anatomy Recognition (AAR) of multiple organs, which separately explored the construction, recognition, and delineation aspects of FOSMs (Udupa et al. 2011, Udupa et al. 2012, Udupa et al. 2014). investigated the modeling stage to create hierarchical FOSMs. They included the scaling factor of each object of interest and alignment based on the orientation of each principal axis in the fuzzy model's construction. These extra data were used in (Udupa et al. 2012) to help in the object recognition step. In this case, some criterion functions were evaluated, including supervised learning by Fisher Linear Discriminant, the minimization of a ball-scale energy, and a one-shot approach that positions every object directly, from the parent positioning relationship observed during training. In (Udupa et al. 2014), the work on hierarchical AAR was completed by analyzing the delineation stage using the Iterative Relative Fuzzy Connectedness framework (Ciesielski et al. 2007). Lastly, (Rittner et al. 2014) included multiple FOSMs per object for hierarchical AAR, similarly to the Cloud Bank Model (Miranda et al. 2009) and the Cloud System Model (Miranda et al. 2010), although only for object recognition.

A few aspects distinguish the works in FOSMs. In (Udupa et al. 2011, Udupa et al. 2012, Udupa et al. 2014), the hierarchy of the objects guides segmentation, and the object search involves limited translation and rotation of the models, according to parameters learned by training. Object recognition is done first and then delineation is applied as a separated step. In contrast, the FOSMs in (Miranda et al. 2008, Miranda et al. 2009, Miranda et al. 2010, Phellan 2014, Phellan et al. 2015) exploit the synergism between model and delineation algorithm for recognition and delineation as a single step in the optimum object search guided by a criterion function. This process requires only translations (since there is no object alignment for model construction) in multiple scales (Miranda et al. 2008, Miranda et al. 2009, Miranda et al. 2010) or, as in (Phellan 2014, Phellan et al. 2015), translations constrained within an estimated search region and guided by an optimization algorithm, named Multiscale Parameter Search (Ruppert et al. 2015). The search region is estimated based on the

observed displacements between training masks and the center of a reference object (e.g., the body region), which can be automatically segmented by image-based techniques.

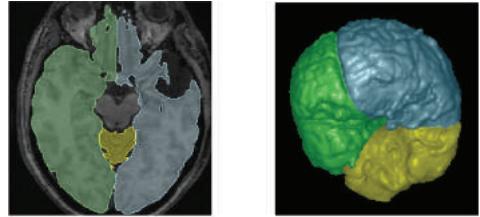
The optimized object search of FOSM has indeed improved SOSM (Phellan 2014), making it more effective than FOSM for single object delineation. Its extension to multiple objects is straightforward, being the criterion function applied to all delineated boundaries and taking into account the relative position of the objects with respect to the center of the multiobject cloud system (Miranda et al. 2010). On the other hand, the speed gains of FOSMs over SOSMs make the former more attractive for large scale studies and interactive applications.

#### 4 INTERACTIVE CORRECTION OF OSM SEGMENTATION

Regardless of the object model selected for automatic segmentation, errors may still occur after the model is applied to a new image. It is important to have a generic image-based delineation method that allows the user to make corrective actions. To the best of our knowledge, there is only one method capable of taking as input *any* automatic segmentation result and allowing the user to correct errors without starting over the process. We have devised such a method in (Miranda et al. 2010, Miranda et al. 2011). The principle behind it is that our method automatically transforms the input segmentation into an *Optimum Path Forest* (OPF) with minimum number of roots—i.e., a data structure that captures the optimum connectivity between voxels and seeds (root voxels), as used by seed competition methods for object delineation (IFT-SC) based on the Image Foresting Transform (Falcão et al. 2004). This allows very efficient corrections by adding or removing trees of the forest (Falcão and Bergo 2004).

IFT-based delineation treats an image as a graph by taking all of the voxels as the nodes and an adjacency relation to define the arcs (e.g., 6-neighbors in 3D images). In interactive segmentation, a *seed* set composed of user-drawn scribbles on the foreground and background is fed to IFT for delineation. These seeds compete among themselves to conquer every voxel in the image, partitioning the graph into an OPF where each seed is the root of an optimum path tree. The object is defined as the set of trees rooted at internal seeds.

Given any input segmentation mask, such as the one in Figure 4a the resuming method does the reverse process of IFT-based segmentation, by computing a set of seeds (forest roots) with



(a) FOSM segmentation with small errors and 3D rendition.



(b) Zoom on a couple of automatically computed resuming seeds for the background (white dots indicated with arrows).



(c) Segmentation correction using IFT-SC with one user-selected background scribble (in white), and its 3D rendition.

Figure 4. Interactive correction of a pre-existing segmentation computed using our FOSM on a brain MR image of a patient who underwent surgery to treat epilepsy, in which part of his left brain hemisphere was removed.

minimum cardinality (Fig. 4b). The user may then draw new scribbles (add trees) or delete scribbles (remove trees) and the forest is updated differentially (Falcão and Bergo 2004), taking time proportional to the modified regions (Fig. 4c). Such a procedure provides interactive response time to the user and the corrected segmentation may be used to improve the OSM.

#### 5 CONCLUSIONS

We have presented a critical review on trends in SOSMs and provided alternative directions that include the more efficient FOSMs and SOSMs with a reduced number of models and optimum object search. In the latter, it is also interesting to create a multi-object system model in order to exploit their relative position and simultaneous delineation for more accurate segmentation.

OSM-based segmentation methods have been developed by assuming an initial training set with images and their corresponding segmentation masks. The training set is created by interactive image segmentation methods. However, for a large number of images, the user does not know the minimum training set size and which images are more representative for training the OSMs, such that the segmentation of the remaining images will be more accurate. These challenges actually require an active learning approach for model construction. In this context, the efficiency gain of is a considerable advantage over SOSMs.

On the other hand, SOSMs with optimum object search can be more effective than FOSMs (Phellan 2014). This suggests hybrid approaches that take advantage of the tighter shape imposition of to delineate object parts with lack of boundary contrast, while can better capture the object's boundary as it appears in other parts of the image.

Additionally, SOSMs- and FOSMs-based have not considered that automatic segmentation may fail and solutions to correct them without starting over the process are required. We have presented an efficient solution for this problem based on the Image Foresting Transform.

## REFERENCES

- Akhondi-Asl, A. & S. Warfield (2013, Oct). Simultaneous truth and performance level estimation through fusion of probabilistic segmentations. *IEEE Trans. Med. Imaging* 32(10), 1840–1852.
- Boykov, Y. & G. Funka-Lea (2006, Nov). Graph cuts and efficient N-D image segmentation. *Int. J. Comput. Vis.* 70(2), 109–131.
- Cabezas, M., A. Oliver, X. Llad, J. Freixenet, & M. B. Cuadra (2011). A review of atlas-based segmentation for magnetic resonance brain images. *Comput. Methods Programs Biomed.* 104(3), e158–e177.
- Ciesielski, K.C., J.K. Udupa, P.K. Saha, & Y. Zhuge (2007). Iterative relative fuzzy connectedness for multiple objects with multiple seeds. *Comput. Vis. Image Underst.* 107(3), 160–182.
- Cootes, T.F., G.J. Edwards, & C.J. Taylor (1998, Jun). Active Appearance Models. In *ECCV*, Volume 2, Freiburg, Germany, pp. 484–498.
- Cootes, T.F., C.J. Taylor, D.H. Cooper, & J. Graham (1995). Active shape models – their training and application. *Comput. Vis. Image Underst.* 61(1), 38–59.
- Couprise, C., L. Grady, L. Najman, & H. Talbot (2011). Power watershed: A unifying graph-based optimization framework. *IEEE Trans. Pattern. Anal. Mach. Intell.* 33(7), 1384–1399.
- Falcão, A.X. & F.P.G. Bergo (2004, sept.). Interactive volume segmentation with differential image foresting transforms. *IEEE Trans. Med. Imag.* 23(9), 1100–1108.
- Falcão, A.X., J. Stolfi, & R.A. Lotufo (2004). The image foresting transform: theory, algorithms, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* 26(1), 19–29.
- Falcão, A.X., J.K. Udupa, & F.K. Miyazawa (2000). An ultrafast user-steered image segmentation paradigm: Live-wire-on-the-fly. *IEEE Trans. Med. Imag.* 19(1), 55–62.
- Falcão, A.X., J.K. Udupa, S. Samarasekera, S. Sharma, B.E. Hirsch, & R.A. Lotufo (1998). User-steered image segmentation paradigms: Live-wire and live-lane. *Graph. Model. Im. Proc.* 60(4), 233–260.
- Ghose, S., A. Oliver, R. Mart, X. Llad, J. Freixenet, J.C. Vilanova, & F. Meriaudeau (2010). Prostate Segmentation with Texture Enhanced Active Appearance Model. In *SITIS*, Volume 1, Kuala Lumpur, Malaysia, pp. 18–22.
- Grady, L. (2006). Random walks for image segmentation. *IEEE Trans. Pattern. Anal. Mach. Intell.* 28(11), 1768–1783.
- Grau, V., A. U.J. Mewes, M. Alcaiz, R. Kikinis, & S.W. Warfield (2004). Improved Watershed Transform for Medical Image Segmentation Using Prior Information. *IEEE Trans. Med. Imaging* 23(4), 447–458.
- Harrigan, R.L., A.J. Plassard, L.A. Mawn, R.L. Gallopoway, S.A. Smith, & B.A. Landman 2015). Constructing a statistical atlas of the radii of the optic nerve and cerebrospinal fluid sheath in young healthy adults. In *SPIE*, Volume 9413, Orlando, USA, pp. 941303–941303–7.
- Inamdar, R.S. & D.S. Ramdasi (2013). Active Appearance Models for Segmentation of Cardiac MRI Data. In *ICCS*, Volume 1, pp. 96–100.
- Lötjönen, J., J. Koikkalainen, L. Thurfiel, R. Lundqvist, G. Waldemar, H. Soininen, & D. Rueckert (2011, March). Improved Generation of Probabilistic Atlases for the Expectation Maximization Classification. In *ISBI*, Chicago, USA, pp. 1839–1842.
- Mazziotta, J.C., A.W. Toga, A. Evans, P. Fox, & J. Lancaster (1995). A Probabilistic Atlas of the Human Brain: Theory and Rationale for Its Development: The International Consortium for Brain Mapping (ICBM). *NeuroImage* 2(2), 89–101.
- Miranda, P., A. Falcão, & J. Udupa (2008). Clouds: A model for synergistic image segmentation. In *ISBI*, pp. 209–212.
- Miranda, P., A. Falcão, G. Ruppert, & F. Cappabianco (2011, March). How to fix any 3d segmentation interactively via image foresting transform and its use in mri brain segmentation. In *ISBI*, Chicago, USA, pp. 2031–2035.
- Miranda, P.A.V., A.X. Falcão, & J.K. Udupa (2009). Cloud bank: a multiple clouds model and its use in MR brain image segmentation. In *ISBI*, Boston, USA.
- Miranda, P.A.V., A.X. Falcão, & G.C.S. Ruppert (2010). How to complete any segmentation process interactively via image foresting transform. In *SIBGRAPI*, Gramado, Brazil, pp. 309–316.
- Miranda, P.A.V., A.X. Falcão, & J.K. Udupa (2010, March). Cloud models: Their construction and employment in automatic MRI segmentation of the brain. Technical Report IC-10-08, IC, University of Campinas.
- Phellan, R. (2014). Medical image segmentation using statistical and fuzzy object shape models. Master's thesis, Institute of Computing—University of Campinas (UNICAMP).

- Phellan, R., A.X. Falcão, & J.K. Udupa (2014). Improving atlas-based medical image segmentation with a relaxed object search. In Y. Zhang and J. Tavares (Eds.), *CompIMAGE*, Volume 8641, pp. 152–163. Pittsburgh, USA.
- Phellan, R., A.X. Falcão, & J.K. Udupa (2015). Medical image segmentation using object atlas versus object cloud models. In *SPIE*, Volume 9415, Orlando, USA, pp. 94151M–94151M–11.
- Pluim, J. P. W., J. Maintz, & M. Viergever (2003, Aug). Mutualinformation-based registration of medical images: a survey. *IEEE Trans. Med. Imaging* 22(8), 986–1004.
- Rittner, L., J.K. Udupa, & D.A. Torigian (2014). Multiple fuzzy object modeling improves sensitivity in automatic anatomy recognition. In *SPIE*, Volume 9034, San Diego, USA, pp. 90342U–90342U–7.
- Rother, C., V. Kolmogorov, & A. Blake (2004). “Grab-Cut”: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* 23(3), 309–314.
- Rueda, S. (2010, june). *2D and 3D Digital Shape Modeling Strategies*. Ph. D. thesis, University of Nottingham, Nottingham.
- Ruppert, G.C.S., G. Chiachia, F.P.G. Bergo, F.O. Favretto, C.L. Yasuda, A. Rocha, & A. X. Falcão (2015). Medical image registration based on watershed transform from greyscale marker and multi-scale parameter search. *Comput. Methods Biomed. Eng.: Imaging & Visual.*, 1–19.
- Rusu, M., N. Bloch, C.C. Jaffe, N.M. Rofsky, E.M. Genega, E. Feleppa, R.E. Lenkinski, & A. Madabhushi (2013). Statistical 3D Prostate Imaging Atlas Construction via Anatomically Constrained Registration. In *SPIE*, Volume 8669, Lake Buena Vista, USA, pp. 1–15.
- Sanroma, G., G. Wu, Y. Gao, & D. Shen (2014, Oct). Learning to rank atlases for multiple-atlas segmentation. *IEEE Trans. Med. Imaging* 33(10), 1939–1953.
- Spina, T.V., P.A.V. Miranda, & A.X. Falcão (2014, Dec). Hybrid approaches for interactive image segmentation using the Live Markers paradigm. *IEEE Trans. Image Process.* 23(12), 5756–5769.
- Udupa, J.K., D. Odhner, A.X. Falcão, K.C. Ciesielski, P.A.V. Miranda, M. Matsumoto, G.J. Grevera, B. Saboury, & D.A. Torigian (2012). Automatic anatomy recognition via fuzzy object models. In *SPIE*, San Diego, USA.
- Udupa, J.K., D. Odhner, A.X. Falcão, K.C. Ciesielski, P.A.V. Miranda, P. Vaideeswaran, S. Mishra, G.J. Grevera, B. Saboury, & D.A. Torigian (2011). Fuzzy object modeling. In *SPIE*, Lake Buena Vista, USA.
- Udupa, J.K., D. Odhner, L. Zhao, Y. Tong, M.M. Matsumoto, K.C. Ciesielski, A.X. Falcão, P. Vaideeswaran, V. Ciesielski, B. Saboury, S. Mohammadianrasanani, S. Sin, R. Arens, D.A. Torigian (2014). Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images. *Med. Image Anal.* 18(5), 752–771.
- Vos, P.C., I. Işgum, J.M. Biesbroek, B.K. Velthuis, & M.A. Viergever (2013). Combined Pixel Classification and Atlasbased Segmentation of the Ventricular System in Brain CT images. In *SPIE*, Volume 8669, pp. 1–6.
- Wang, H., J.W. Suh, S.R. Das, J.B. Pluta, C. Craige, & P.A. Yushkevich (2013). Multi-atlas segmentation with joint label fusion. *IEEE Trans. Pattern Anal. Mach. Intell.* 35(3), 611–623.

This page intentionally left blank

# Primal-dual method for continuous max-flow approaches

Ke Wei

*Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong*

Xue-Cheng Tai

*Department of Mathematics, University of Bergen, Bergen, Norway*

Tony F. Chan

*Office of the President, Hong Kong University of Science and Technology, Hong Kong*

Shingyu Leung

*Department of Mathematics, Hong Kong University of Science and Technology, Hong Kong*

**ABSTRACT:** We review the continuous max-flow approaches for the variational image segmentation models with piecewise constant representations. The review is conducted by exploring the primal-dual relationships between the continuous min-cut and max-flow problems. In addition, we introduce the parameter free primal-dual method for solving those max-flow problems. Empirical results show that the primal-dual method is competitive to the augmented Lagrangian method.

## 1 INTRODUCTION

Multi-phase image segmentation (or multi-labelling) is an important problem in image processing and has a wide range of applications in related areas such as computer vision (Paragios, Chen, & Faugeras 2005), stereo (Kolmogorov & Zabih 2002, Kolmogorov & Zabih 2004a) and 3D reconstruction (Vogiatzis, Esteban, Torr, & Cipolla 2007). The basic task in image segmentation is to optimally assign different labels to different pixels of an image with respect to some energy functional.

In the spatially discrete setting, the image is usually modelled as a graph, and the solution of the multi-phase image labelling problem can be found by computing the min-cut or max-flow solutions of the graph, see (Kolmogorov & Boykov 2005, Kolmogorov & Zabih 2004b, Boykov, Veksler, & Zabih 2001, Boykov & Kolmogorov 2003) and references therein. In the spatially continuous setting, the variational approach has been widely studied, where the problem is formulated as the minimization problem of a continuous energy functional. Compared with the graph-based method in the discrete setting, the variational method has several advantages: 1) it can avoid the metrical errors thanks to the crucial rotation invariance property; 2) a wide range of reliable numerical schemes are available, and these schemes can be easily implemented and accelerated; 3) it requires less memory in computation; 4) it is easy to use GPU and parallel processors.

In this paper, we focus on the variational approach for the image segmentation problem. Let  $\Omega$  denote the domain of an input image. Then the task is to find a partition  $\{\Omega_i\}_{i=1}^n$  of  $\Omega$  which minimises the following energy functional

$$\min_{\{\Omega_i\}_{i=1}^n} \sum_{i=1}^n \int_{\Omega_i} f_i(x) dx + \alpha R(\{\Omega_i\}_{i=1}^n) \\ \text{s.t. } \cup_{i=1}^n \Omega_i = \Omega \text{ and } \Omega_k \cap \Omega_l = \emptyset \text{ for } k \neq l, \quad (1)$$

where  $R(\cdot)$  is a regularisation term. Concrete representations for  $\{\Omega_i\}_{i=1}^n$  are needed so that we can design numerical algorithms to solve (1). Over the last 30 years, many representations have been proposed, for example the level set method (Osher & Sethian 1988, Chan & Vese 2001). In this work, we are interested in the piecewise constant representations where each region is represented by a unique binary or integer value. With the piecewise constant representations, convex relaxations can often be constructed for (1) and so efficient and robust numerical algorithms can be developed based on those relaxations. So far there are three typical piecewise constant representations for  $\{\Omega_i\}_{i=1}^n$  and each representation will result in a particular form of (1).

*Binary value representation:* For each partition  $\Omega_i$ , we can define an indicator function

$$u_i(x) = \begin{cases} 1 & x \in \Omega_i \\ 0 & x \notin \Omega_i, \end{cases} \quad i = 1, \dots, n.$$

With these indicator functions, model (1) with the Potts regulariser  $R(\{\Omega_i\}_{i=1}^n) = \sum_{i=1}^n |\partial\Omega_i|$  can be rewritten as<sup>1</sup>

$$\min_{u_i \in \{0,1\}, u \in S} \sum_{i=1}^n \int_{\Omega} u_i f_i dx + \alpha \sum_{i=1}^n \int_{\Omega} |\nabla u_i| dx \quad (2)$$

where

$$S = \left\{ (u_1, \dots, u_n) : \sum_{i=1}^n u_i = 1, u_i \geq 0 \right\}$$

Such a representation was used in (Lellmann, Kappes, Yuan, Becker, & Schnörr 2009, Zach, Gallup, Frahm, & Niethammer 2008). In (Bae, Yuan, & Tai 2011), global minimization property for this model was proven. In (Yuan, Bae, Tai, & Boykov 2010), this model was interpreted as a min-cut problem with a corresponding max-flow model.

*Integer value representation:* Let  $u : \Omega \rightarrow \{1, \dots, n\}$  be a labelling function such that  $u(x) = i$  if  $x \in \Omega_i$ ,  $i = 1, \dots, n$ . Let  $\psi_i(u)$  be the corresponding indicator functions for  $\Omega_i$  which is related to  $u$ . Then model (1) with the regulariser  $R(\{\Omega_i\}_{i=1}^n) = \sum_{i=1}^n |\partial\Omega_i|$  reduces to

$$\min_{u \in \{1,2,\dots,n\}} \sum_{i=1}^n \int_{\Omega} f_i \psi_i(u) dx + \alpha \sum_{i=1}^n \int_{\Omega} |\nabla \psi_i(u)| dx, \quad (3)$$

This model was first introduced in (Lie, Lysaker, & Tai 2005, Lie, Lysaker, & Tai 2006b). Augmented Lagrangian was used in (Lie, Lysaker, & Tai 2005, Lie, Lysaker, & Tai 2006b) to solve it. In (Bae & Tai 2009), discrete graph cut was used to solve this model which is very much related to the graph based method in (Ishikawa 2003, Darbon & Sigelle 2006). The interpretation of this model as a min-cut problem and its corresponding max-flow model was given in (Bae, Yuan, Tai, & Boykov 2010, Bae, Yuan, Tai, & Boykov 2014a). This problem can also be transformed into a binary minimization problem. If we replace the regulariser by  $R(\{\Omega_i\}_{i=1}^n) = \int_{\Omega} |\nabla u| dx$  and introduce

$$\lambda_i(x) = \begin{cases} 1 & \text{if } u(x) > i \\ 0 & \text{if } u(x) \leq i \end{cases}, \quad i = 1, \dots, n-1$$

and  $\lambda_0(x) = 1$ ,  $\lambda_n(x) = 0$ . Then model (1) can also be written as a binary minimization problem (Liu, Tai, Leung, & Huang 2014)

<sup>1</sup>In the sequel, we will omit the notation  $x$  when there is no confusion from the context.

$$\min_{\lambda_i \in \{0,1\}, \lambda \in B} \sum_{i=1}^n \int_{\Omega} (\lambda_{i-1} - \lambda_i) f_i dx + \alpha \sum_{i=1}^{n-1} \int_{\Omega} |\nabla \lambda_i| dx, \quad (4)$$

where

$$B = \{(\lambda_0, \dots, \lambda_n) : 1 = \lambda_0 \geq \dots \geq \lambda_n = 0\}.$$

After solving (4), the labelling function  $u$  can be recovered by  $u = \sum_{i=1}^n (\lambda_{i-1} - \lambda_i) \cdot i$ .

*Product of binary values representation:* This representation was first given in (Lie, Lysaker, & Tai 2005, Lie, Lysaker, & Tai 2006a) extending the multiphase level set framework of Vese-Chan (Vese & Chan 2002). In the case of two-phase labelling, its convex relaxation and global minimization was proposed in (Chan, Esedoglu, & Nikolova 2006). For simplicity, assume  $n = 2^m$ . Define  $w_0(s) = 1 - s$  and  $w_1(s) = s$  and introduce  $m$  binary function  $\phi_1, \dots, \phi_m : \Omega \rightarrow \{0,1\}$  such that  $x \in \Omega_i$  if and only if  $\phi_1 \dots \phi_m$  is a binary representation of the integer  $i$ . With these notations, model (1) with the regulariser  $R(\{\Omega_i\}_{i=1}^n) = \sum_{i=1}^n |\nabla \phi_i| dx$  can be formulated as

$$\min_{\phi_i \in \{0,1\}} \int_{\Omega} \sum_{i=1}^n \prod_{k=1}^m w_{a_k^i}(\phi_k) f_i dx + \alpha \sum_{k=1}^m \int_{\Omega} |\nabla \phi_k| dx, \quad (5)$$

where  $a_1^i \dots a_m^i$  is a binary representation of  $i$ , denoted by  $i = [a_1^i \dots a_m^i]$ . The interpretation of this model as a min-cut problem and its corresponding max-flow model was given in (Bae, Lellmann, & Tai 2013, Bae & Tai 2015, Liu, Tai, Leung, & Huang 2014).

## 2 CONTINUOUS MAX-FLOW APPROACHES

Three piecewise constant representations and the corresponding formulations for labelling problems were introduced in the last section. All the three formulations are non-convex because either the constraints are non-convex or both the objective functional and the constraints are non-convex. A number of approaches have appeared in the literature recently showing that these models can be interpreted as continuous min-cut problems. There exists a corresponding continuous max-flow model for each of these three min-cut problems. Exploring the connection between these continuous min-cut and max-flow problems, convex global minimization method could be derived for these non-convex problems.

Since the pioneer work of Chan et al. (Chan, Esedoglu, & Nikolova 2006), one of the important research topics in image segmentation is to find good convex relaxations for the non-convex minimization problems arising from the variational methods and then design numerical algorithms for the corresponding convex relaxations. In a series of

research papers (Yuan, Bae, Tai, & Boykov 2014, Bae & Tai 2015, Bae, Yuan, Tai, & Boykov 2014b, Bae, Lellmann, & Tai 2013, Yuan, Bae, Tai, & Boykov 2010), the authors have developed a set of computationally efficient algorithms for the models in the last section. The essential idea is to interpret these models as continuous min-cut problems and each of them has a corresponding continuous max-flow problem. This section reviews the continuous max-flow approaches from the primal-dual perspective. However, we use a different derivation to get these continuous max-flow models.

## 2.1 Two-phase image segmentation

We first consider the Chan-Vese model (Chan, Esedoglu, & Nikolova 2006) for the two phase segmentation problem

$$\min_{u \in [0,1]} \int_{\Omega} (1-u)C_s dx + \int_{\Omega} uC_t dx + \alpha \int_{\Omega} |\nabla u| dx. \quad (6)$$

Model (6) can be viewed as a special case of either the formulation (2) or (5). Since (6) is a non-convex optimization problem, it is not tractable to find its global solution. Thus the authors propose to relax the binary valued function  $u$  to  $0 \leq u \leq 1$  and solve the following convex problem

$$\min_{u \in [0,1]} \int_{\Omega} (1-u)C_s dx + \int_{\Omega} uC_t dx + \alpha \int_{\Omega} |\nabla u| dx. \quad (7)$$

Moreover, it is proved in (Chan, Esedoglu, & Nikolova 2006) that the solution to (6) can be obtained by thresholding the solution to (7). However, the numerical algorithms for (7) usually suffer from the non-smoothness of the TV term. We can form a dual problem of (7) as follows:

$$\begin{aligned} & \min_{u \in [0,1]} \int_{\Omega} (1-u)C_s dx + \int_{\Omega} uC_t dx + \alpha \int_{\Omega} |\nabla u| dx \\ &= \min_{u \in [0,1]} \max_{\substack{p_s \leq C_s \\ p_t \leq C_t \\ |q| \leq \alpha}} \int_{\Omega} (1-u)p_s dx + \int_{\Omega} up_t dx + \int_{\Omega} u \operatorname{div} q dx \\ &= \min_{u \in \mathbb{R}} \max_{\substack{p_s \leq C_s \\ p_t \leq C_t \\ |q| \leq \alpha}} \int_{\Omega} (1-u)p_s dx + \int_{\Omega} up_t dx + \int_{\Omega} u \operatorname{div} q dx \\ &= \max_{\substack{p_s \leq C_s \\ p_t \leq C_t \\ |q| \leq \alpha}} \min_{u \in \mathbb{R}} \int_{\Omega} (1-u)p_s dx + \int_{\Omega} up_t dx + \int_{\Omega} u \operatorname{div} q dx \\ &= \max_{\substack{p_s \leq C_s \\ p_t \leq C_t \\ |q| \leq \alpha}} \min_{u \in \mathbb{R}} \int_{\Omega} p_s dx + \int_{\Omega} u(\operatorname{div} q - p_s + p_t) dx \\ &= \max_{\substack{p_s \leq C_s \\ p_t \leq C_t \\ |q| \leq \alpha}} \int_{\Omega} p_s dx \text{ s.t. } \begin{cases} p_s \leq C_s, p_t \leq C_t, |q| \leq \alpha \\ \operatorname{div} q - p_s + p_t = 0, \end{cases} \end{aligned} \quad (8)$$

where in the first equality, we use the fact (Giusti 1977)

$$\int_{\Omega} \alpha |\nabla u| dx = \max_{|q| \leq \alpha} \int_{\Omega} u \operatorname{div} q dx$$

in the distributional sense; and the third equality follows from the minmax theorem (Ekeland & Teman 1999, Chapter 6, Proposition 2.4). Instead of regarding (7) as a convex relaxation, our view point is that (6) is a *min-cut problem* over a continuous domain and the last maximization problem in (8) is a *max-flow problem*. Problem (7) is the *dual problem* of (8). The pair of min-max problems (7) and (8) are continuous analogue of the min-cut and max-flow problems in the graph based methods (Yuan, Bae, & Tai 2010, Yuan, Bae, Tai, & Boykov 2014). In the discrete case, it is well-known that the min-cut problem is equivalent to the max-flow problem. In the continuous setting, the following three problem are equivalent, (Yuan, Bae, Tai, & Boykov 2014, Prop. 3.1), and thus gives another way to explain the convex relaxation model (7):

$$\begin{aligned} \text{Min-cut problem (6)} &\Leftrightarrow \text{Max-flow problem (8)} \\ &\Leftrightarrow \text{Dual of max-flow problem (7)} \end{aligned}$$

From the above derivation, we can see that the so called cut for (7) is the Lagrangian multiplier for the flow conservation constraint  $\operatorname{div} q - p_s + p_t = 0$  of the max-flow problem.

## 2.2 Multi-phase image segmentation

We will derive the max-flow models for the multi-phase image formulations (2), (4) and (5) one by one and show that the equivalence between the min-cut and max-flow problems can also be extended to multi-phase cases.

### 2.2.1 Binary value representation

Following the convex relaxation technique for the two-phase Chan-Vese model, a natural convex relaxation for (2) is

$$\min_{u \in S} \sum_{i=1}^n \int_{\Omega} u_i f_i dx + \alpha \sum_{i=1}^n \int_{\Omega} |\nabla u_i| dx. \quad (9)$$

We can form a dual problem of (9) by introducing another  $n$  variables  $p_i, i = 1, \dots, n$  which corresponds to the data fidelity term and an auxiliary variable  $p_s$ :

$$\begin{aligned} & \min_{(u_1, \dots, u_n) \in S} \sum_{i=1}^n \int_{\Omega} u_i f_i dx + \alpha \sum_{i=1}^n \int_{\Omega} |\nabla u_i| dx \\ &= \min_{u_i \geq 0} \max_{p_s} \int_{\Omega} \left( 1 - \sum_{i=1}^n u_i \right) p_s dx + \sum_{i=1}^n \int_{\Omega} u_i f_i dx \\ & \quad + \alpha \sum_{i=1}^n \int_{\Omega} |\nabla u_i| dx \end{aligned}$$

$$\begin{aligned}
&= \min_{u_i \in \mathbb{R}} \max_{\substack{p_s \\ p_i \leq f_i \\ |q_i| \leq \alpha}} \int_{\Omega} \left( 1 - \sum_{i=1}^n u_i \right) p_s dx + \sum_{i=1}^n \int_{\Omega} u_i f_i dx \\
&\quad + \sum_{i=1}^n \int_{\Omega} u_i \operatorname{div} q_i dx \\
&= \max_{\substack{p_s \\ p_i \leq f_i \\ |q_i| \leq \alpha}} \min_{u_i \in \mathbb{R}} \int_{\Omega} \left( 1 - \sum_{i=1}^n u_i \right) p_s dx + \sum_{i=1}^n \int_{\Omega} u_i f_i dx \\
&\quad + \sum_{i=1}^n \int_{\Omega} u_i \operatorname{div} q_i dx \\
&= \max_{\substack{p_s \\ p_i \leq f_i \\ |q_i| \leq \alpha}} \min_{u_i \in \mathbb{R}} \int_{\Omega} p_s dx + \sum_{i=1}^n \int_{\Omega} u_i (\operatorname{div} q_i - p_s + p_i) dx \\
&= \max \int_{\Omega} p_s dx \text{ s.t. } \begin{cases} p_i \leq f_i, |q_i| \leq \alpha \\ \operatorname{div} q_i - p_s + p_i = 0, \end{cases} \quad (10)
\end{aligned}$$

where the third equality follows from the minmax theorem. The maximization problem (10) corresponds to a type of max-flow model in the spatially continuous domain (Yuan, Bae, Tai, & Boykov 2010).

### 2.2.2 Integer value representation

Relaxing  $\lambda_i \in \{0,1\}$  to  $\lambda_i \in [0,1]$  gives the convex relaxation of (11)

$$\min_{\lambda \in B} \sum_{i=1}^n \int_{\Omega} (\lambda_{i-1} - \lambda_i) f_i dx + \alpha \sum_{i=1}^{n-1} \int_{\Omega} |\nabla \lambda_i| dx. \quad (11)$$

By fixing  $\lambda_0 = 1$  and  $\lambda_n = 0$ , the dual problem of (11) can be computed as follows:

$$\begin{aligned}
&\min_{\lambda \in B} \sum_{i=1}^n \int_{\Omega} (\lambda_{i-1} - \lambda_i) f_i dx + \alpha \sum_{i=1}^{n-1} \int_{\Omega} |\nabla \lambda_i| dx \\
&= \min_{\substack{\lambda_i \in \mathbb{R} \\ p_i \leq f_i \\ |q_i| \leq \alpha}} \max_{\phi_i} \sum_{i=1}^n \int_{\Omega} (\lambda_{i-1} - \lambda_i) p_i dx + \sum_{i=1}^{n-1} \int_{\Omega} \lambda_i \operatorname{div} q_i dx \\
&= \max_{\substack{p_i \leq f_i \\ |q_i| \leq \alpha}} \min_{\lambda_i \in \mathbb{R}} \sum_{i=1}^n \int_{\Omega} (\lambda_{i-1} - \lambda_i) p_i dx + \sum_{i=1}^{n-1} \int_{\Omega} \lambda_i \operatorname{div} q_i dx \\
&= \max_{\substack{p_i \leq f_i \\ |q_i| \leq \alpha}} \min_{\lambda_i \in \mathbb{R}} \int_{\Omega} p_i dx + \sum_{i=1}^{n-1} \lambda_i (\operatorname{div} q_i - p_i + p_{i+1}) dx \\
&= \max \int_{\Omega} p_i dx \text{ s.t. } \begin{cases} p_i \leq f_i, |q_i| \leq \alpha \\ \operatorname{div} q_i - p_i + p_{i+1} = 0, \end{cases} \quad (12)
\end{aligned}$$

where in the first equality  $1 \leq i \leq n-1$  for  $\lambda_i$  and  $q_i$ , whereas  $1 \leq i \leq n$  for  $p_i$ ; again the second equality

follows from the minmax theorem. The maximisation problem (12) also corresponds to a continuous maxflow model (Bae, Yuan, Tai, & Boykov 2014b).

### 2.2.3 Product of binary values representation

The energy functional in (5) is also non-convex in addition to the non-convexity of the constraints. In order to find a convex relaxation of (5), we can compute the pixel-wise convex envelop of the data term, which can be obtained by computing the bi-conjugate of the functional twice (Bae, Lellmann, & Tai 2013). The bi-conjugate of  $g(u) := \sum_{i=1}^n \prod_{k=1}^m w_{a_k^i}(u_k) f_i$ ,  $u_k \in \{0,1\}$  ( $k = 1, \dots, m$ ) is

$$g^*(p) = \max_{u_k \in \{0,1\}} \left\{ \sum_{k=1}^m p_k u_k - \sum_{i=1}^n \prod_{k=1}^m w_{a_k^i}(u_k) f_i \right\}.$$

Consequently the bi-conjugate of  $g^*$  is

$$\begin{aligned}
g^{**}(\phi) &= \max_{p \in \mathbb{R}^m} \left\{ \sum_{k=1}^m \phi_k p_k - g^*(p) \right\} \\
&= \max_{p \in \mathbb{R}^m} \left\{ \sum_{k=1}^m \phi_k p_k + \min_{u_k \in \{0,1\}} \left\{ \sum_{i=1}^n \prod_{k=1}^m w_{a_k^i}(u_k) f_i - \sum_{k=1}^m p_k u_k \right\} \right\} \\
&= \max_{p \in \mathbb{R}^m} \sum_{k=1}^m \{\phi_k p_k + p_0\} \\
&\text{s.t. } p_0 \leq - \sum_{k=1}^m u_k^i p_k + f_i, \text{ for any } i = [u_1^i \dots u_m^i],
\end{aligned}$$

where the constraint in the last equality follows from the fact

$$\sum_{i=1}^n \prod_{k=1}^m w_{a_k^i}(u_k^i) f_i = f_i \text{ if } i = [u_1^i \dots u_m^i].$$

Thus the problem (5) with the regularisation term can be relaxed to

$$\begin{aligned}
&\min_{\phi} \max_p \int_{\Omega} p_0 dx + \int_{\Omega} \sum_{k=1}^m \phi_k p_k dx + \alpha \int_{\Omega} \sum_{k=1}^m |\nabla \phi_k| dx \\
&= \min_{\phi} \max_{\substack{p \\ p \neq q_i \leq \alpha}} \int_{\Omega} p_0 dx + \int_{\Omega} \sum_{k=1}^m \phi_k p_k dx + \int_{\Omega} \sum_{k=1}^m \phi_k \operatorname{div} q_k dx \\
&= \max_{\substack{p \\ p \neq q_i \leq \alpha}} \min_{\phi} \int_{\Omega} p_0 dx + \int_{\Omega} \sum_{k=1}^m \phi_k (\operatorname{div} q_k + p_k) dx \\
&= \max \int_{\Omega} p_0 dx \text{ s.t. } \begin{cases} p_0 \leq - \sum_{k=1}^m u_k^i p_k + f_i \\ |q_k| \leq \alpha \\ \operatorname{div} q_k + p_k = 0. \end{cases} \quad (13)
\end{aligned}$$

Moreover, if the data term of (5) is sub-modular, (13) is a tight convex relaxation for (5). In particular, the data term is sub-modular when  $n = 4$  (Bae & Tai 2015).

### 3 GENERAL ALGORITHMIC FRAMEWORK

#### 3.1 Augmented lagrangian method

The maximisation problems (8), (10), (12) and (13) share the following general form

$$\max_p F(p) \text{ s. t. } \begin{cases} p \in C_p \\ q \in C_q, \\ Lq + Dp = 0, \end{cases} \quad (14)$$

where

- $F(p)$  is a (linear) functional of  $p$ ,
- $L$  and  $D$  are linear mappings,
- $C_p$  and  $C_q$  are convex sets.

The augmented Lagrangian formulation of (14) is

$$\mathcal{L}(p, q, \gamma) = F(p) + \langle \gamma, Lq + Dp \rangle - \frac{c}{2} \|Lq + Dp\|^2. \quad (15)$$

Then with a triple of intial points  $(p^0, q^0, \gamma^0)$ , we can solve (14) by the augmented Lagrangian method as follows (Alg. 1):

- $p^{l+1} = \arg \max_{p \in C_p} \mathcal{L}(p, q^l, \gamma^l)$ ,
- $q^{l+1} = \arg \max_{q \in C_q} \mathcal{L}(p^{l+1}, q, \gamma^l)$ ,
- $\gamma^{l+1} = \gamma^l - c(Lq + Dp)$ .

When there are no explicit solutions to the first two subproblems, they can be solved approximately by several projected ascent iterations (Yuan, Bae, Tai, & Boykov 2014, Bae & Tai 2015, Bae, Yuan, Tai, & Boykov 2014b, Bae, Lellmann, & Tai 2013, Yuan, Bae, Tai, & Boykov 2010). The penalization parameter  $c$  is chosen by trial-and-error in numerical tests.

Table 1. Stepsize selections for the primal-dual method (Alg. 2).

	$\tau_l^p$	$\tau_l^q$	$\beta_l$
Cameraman	$0.1 * l$	$0.04 * l - 0.02$	$\frac{1}{0.08 * l + 0.2}$
3 color	$0.005 * l + 0.03$	$0.005 * l$	$\frac{1}{0.02 * l}$

### 3.2 Primal-dual method

In this section, we will provide another method for solving the continuous max-flow problems in Sec. 2. As it can be observed from the derivation process, each continuous max-flow approach is associated with a min-max problem of the form

$$\min_{\gamma \in C_\gamma} \max_{p \in C_p} F(p, q, \gamma), \quad (16)$$

where  $F(p, q, \gamma)$  is (linear) functional of  $p$ ,  $q$  and  $\gamma$ . Then following the general set-up of the primal-dual method in (Chambolle & Pock 2011, Esser, Zhang, & Chan 2010), we can solve (16) as follows (Alg. 2):

- $p^{l+1} = p^l + \tau_l^p \frac{\partial F(p, q^l, \gamma^l)}{\partial p}$ ,
- $q^{l+1} = q^l + \tau_l^q \frac{\partial F(p^{l+1}, q, \gamma^l)}{\partial q}$ ,
- $\gamma^{l+1} = \gamma^l - \beta_l \frac{\partial F(p^{l+1}, q^{l+1}, \gamma)}{\partial \gamma}$ ,

where  $\tau_l^p, \tau_l^q$  and  $\beta_l$  are adaptive stepsizes. The primal-dual method is parameter free compared with the augmented Lagrangian method in Sec. 3.1 as there is no proximal regularizer in Alg. 2. Moreover, we can choose the stepsizes automatically to accelerate the algorithm, for example see Table 1. In general,  $\tau_l^i$  ( $i = p, q$ ) and  $\beta_l$  should be inverse to each other, which are proportional to  $l$  and  $1/l$  respectively.

### 4 NUMERICAL EXPERIMENTS

This section compares the aforementioned two algorithms on the image segmentation problems. The tests are conducted on a Mac Pro laptop with Intel Core i5 CPUs @ 2.4 GHz and executed from Matlab. For the discretized version of the gradient, divergence and TV operators, we refer the readers to (Zhu 2008, Ch. 1.3).

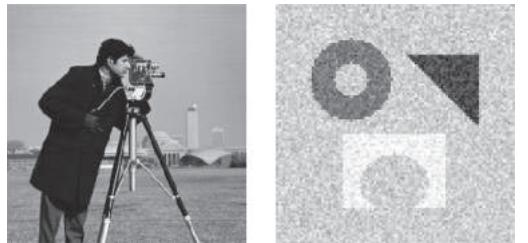


Figure 1. Input images for image segmentation. Left: cameraman for two-phase image segmentation; Right: 3 color for multi-phase image segmentation.

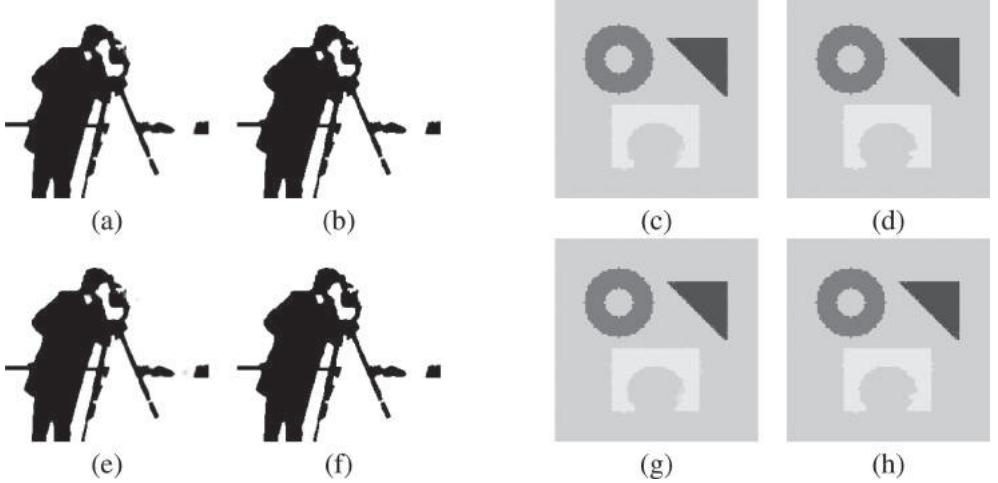


Figure 2. The first row: rescaled labelling function returned by the primal-dual method (Alg. 2) before ((a) and (c)) and after ((b) and (d)) thresholding; The second row: rescaled labelling function returned by the augmented Lagrangian method (Alg. 1) before ((e) and (g)) and after ((f) and (h)) thresholding. The thresholding value is 0.5.

Table 2. Computational results of the image segmentation for the two-phase (cameraman) and multi-phase (3 color) image segmentations by the primal-dual and the augmented Lagrangian methods. The algorithm terminates if  $e^l$  defined in (17) is less than  $\varepsilon$ .

	Cameraman				3 color			
	$\varepsilon = 1e - 3$		$\varepsilon = 1e - 5$		$\varepsilon = 1e - 3$		$\varepsilon = 1e - 5$	
	Iter	Time (s)						
Primal-dual (Alg. 2)	32	0.15	118	0.51	25	0.11	65	0.25
Augmented Lagrangian (Alg. 1)	32	0.16	149	0.67	40	0.18	215	0.89

In our tests, the data term has the form  $f_i = |u - c_i|^\beta$ , where  $u$  is the input image and  $c_i$  is the average intensity of the region  $i$  which is assumed to be fixed. Apparently,  $f_i$  is convex when  $\beta \geq 1$  and non-convex when  $\beta < 1$ . Here  $\beta$  is set to 1.

In order to compare the convergence rate of the two tested algorithms, we first (numerically) compute the optimal primal energy, denoted by  $E^*$ , by running either of the algorithms 10000 iterations. Then we evaluate the progress of the algorithm by computing the relative error of the primal energy defined as

$$e^l = \frac{|E^l - E^*|}{E^*}, \quad (17)$$

where  $E^l$  is the primal energy in the  $l$ th iteration.

We test the primal-dual and the augmented Lagrangian methods on the standard cameraman image (see the left of Fig. 1) for the two-phase image segmentation problem and on the standard noisy 3 color image (see the right of Fig. 1) for the multi-phase image segmentation problem. Due to the limited space, we only conduct the tests for the variational model with the integer value representation in Sec. 2.2.2. In Alg. 1,  $c$  is selected to be 0.3, and the adaptive stepsize selections in Alg. 2 are listed in Table 1.

The segmentation results are plotted in Figure 2. By comparing the plots before and after thresholding, we can see that both the primal-dual and the augmented Lagrangian methods can achieve very good performance since they all converge to a result which takes the values 0 and 1 almost everywhere. The computational results are presented

in Table 2 which includes the number of iterations and computational time needed for the algorithm to converge to certain accuracy. Table 2 clearly shows that the primal-dual method with the adaptive stepsize selections can be faster than the augmented Lagrangian method.

## REFERENCES

- Bae, E., J. Lellmann, & X.-C. Tai (2013). Convex relaxations for a generalized chan-vese model. In A. Heyden et al. (Eds.), *EMMCVPR 2013, LNCS 8081*, pp. 223–236. Springer-Verlag Berlin Heidelberg 2013.
- Bae, E. & X.-C. Tai (2009). Graph cut optimization for the piecewise constant level set method applied to multiphase image segmentation. In *Scale space and variational methods in computer vision*, pp. 1–13. Springer.
- Bae, E. & X.-C. Tai (2015). Efficient global minimization methods for image segmentation models with four regions. *J Math Imaging Vis* 51, 71–97.
- Bae, E., J. Yuan, & X.-C. Tai (2011). Global minimization for continuous multiphase partitioning problems using a dual approach. *International journal of computer vision* 92(1), 112–129.
- Bae, E., J. Yuan, X.-C. Tai, & Y. Boykov (2010). A fast continuous max-flow approach to non-convex multilabeling problems. UCLA, *Applied Mathematics, CAM-report-10-62*.
- Bae, E., J. Yuan, X.-C. Tai, & Y. Boykov (2014a). A fast continuous max-flow approach to non-convex multilabeling problems. In *Efficient Algorithms for Global Optimization Methods in Computer Vision*, pp. 134–154. Springer.
- Bae, E., J. Yuan, X.-C. Tai, & Y. Boykov (2014b). A fast continuous max-flow approach to non-convex multilabeling problems. In A. Bruh et al. (Eds.), *Global Optimization Methods, LNCS 8293*, pp. 134–154. Springer-Verlag Berlin Heidelberg 2014.
- Boykov, Y. & V. Kolmogorov (2003). Computing geodesics and minimal surfaces via graph cuts. In *ICCV 2003*, pp. 26–33.
- Boykov, Y., O. Veksler, & R. Zabih (2001). Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 1222–1239.
- Chambolle, A. & T. Pock (2011). A first-order primal-dual algorithm for convex problems with applications to image. *J Math Imaging Vis* 40, 120–145.
- Chan, T.F., S. Esedoglu, & M. Nikolova (2006). Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.* 66(5), 1632–1648.
- Chan, T.F. & L.A. Vese (2001). Active contours without edges. *IEEE Transactions on Image Processing* 10(2), 266–277.
- Darbon, J. & M. Sigelle (2006). Image restoration with discrete constrained total variation part i: Fast and exact optimization. *Journal of Mathematical Imaging and Vision* 26(3), 261–276.
- Ekeland, I. & R. Teman (1999). *Convex analysis and variational problems*. SIAM.
- Esser, E., X. Zhang, & T.F. Chan (2010). A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Science* 3(4), 1015–1046.
- Giusti, E. (1977). *Minimal surfaces and functions of bounded variation*. Australian National University.
- Ishikawa, H. (2003). Exact optimization for markov random fields with convex priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 1333–1336.
- Kolmogorov, V. & Y. Boykov (2005). What metrics can be approximated by geo-cuts, or global optimization of length/area and flux. In *ICCV 2005*, pp. 564–571.
- Kolmogorov, V. & R. Zabih (2002). Multi-camera scene reconstruction via graph cuts. In *European Conference on Computer Vision*, pp. 82–96.
- Kolmogorov, V. & R. Zabih (2004a). What energy functions can be minimized via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26, 65–81.
- Kolmogorov, V. & R. Zabih (2004b). What energy functions can be minimized via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26, 65–91.
- Lellmann, J., J. Kappes, J. Yuan, F. Becker, & C. Schnörr (2009). Convex multi-class image labeling by simplex-constrained total variation. In X.-C. Tai et al. (Eds.), *SSVM 2009, LNCS*, Volume 5567, pp. 150–162. Springer, Heidelberg 2009.
- Lie, J., M. Lysaker, & X.-C. Tai (2005). A piecewise constant level set framework. *Int. J. Numer. Anal. Model.* 2(4), 422–438.
- Lie, J., M. Lysaker, & X.-C. Tai (2006a). A binary level set model and some applications to mumford-shah image segmentation. *Image Processing, IEEE Transactions on* 15(5), 1171–1181.
- Lie, J., M. Lysaker, & X.-C. Tai (2006b). A variant of the level set method and applications to image segmentation. *Math. Comp.* 75(255), 1155–1174.
- Liu, J., X.-c. Tai, S. Leung, & H. Huang (2014). A new continuous max-flow algorithm for multiphase image segmentation using super-level set functions. *Journal of Visual Communication and Image Representation* 25(6), 1472–1488.
- Osher, S. & J. Sethian (1988). Fronts propagating with curvature dependent speed: algorithms based on hamilton-jacobi formulations. *J. Comput. Phys.* 79(1), 12–49.
- Paragios, N., Y. Chen, & O. Faugeras (2005). *Handbook of Mathematical Models in Computer Vision*. Springer New York.
- Vese, L. & T. Chan (2002). A new multiphase level set framework for image segmentation via the mumford and shahs model. *International Journal of Computer Vision* 50, 271–293.
- Vogiatzis, G., C.H. Esteban, P.H. Torr, & R. Cipolla (2007). Multi-view stereo via volumetric graph-cuts and occlusion robust photo-consistency. *PAMI* 29(12), 2241–2246.
- Yuan, J., E. Bae, & X. Tai (2010). A study on continuous maxflow and min-cut approaches. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 2217–2224. IEEE.

- Yuan, J., E. Bae, X.-C. Tai, & Y. Boykov (2010). A continuous max-flow approach to potts model. In K. Daniilidis, P. Maragos, and N. Paragios (Eds.), *ECCV 2010, Part VI, LNCS 6316*, pp. 379–392. Springer-Verlag Berlin Heidelberg 2010.
- Yuan, J., E. Bae, X.-C. Tai, & Y. Boykov (2014). A study on continuous max-flow and min-cut approaches. *Numerische Mathematik* 126(3), 559–587.
- Zach, C., D. Gallup, J.-M. Frahm, & M. Niethammer (2008). Fast global labeling for realtime stereo using multiple plane sweeps. In *Vision, Modeling and Visualization Workshop (VMV)*.
- Zhu, M. (2008). Fast numerical algorithms for total variation based image restoration. Thesis.

# Image restoration using variational approaches: Some recent advances

A. Lanza, S. Morigi & F. Sgallari

*Department of Mathematics, University of Bologna, Bologna, Italy*

**ABSTRACT:** In image formation, the observed images are usually blurred by optical instruments and/or transfer medium and contaminated by noise, which makes image restoration a classical inverse problem. A general principle for dealing with the intrinsic numerical instability of most inverse problems is that of regularization. Among the regularization approaches the most common is Tikhonov regularization. Another very popular choice in the literature for regularization is based on the Total Variation (TV) semi-norm. Since then, in order to provide more reliable and efficient models, several variational formulations have been investigated which involve two terms, the data-fidelity term and the regularization term, with the goal to capture characteristic and essential image features. In this review, we will focus on variational methods for the restoration of images corrupted by additive noise which we assume to be sampled from a known a-priori distribution. In particular, we will discuss recent advanced proposals for the regularization and fidelity terms in the optimization problem. Numerical results illustrate the efficacy of the different models.

## 1 INTRODUCTION

A real captured image is deteriorated during formation, transmission, and recording processes. Classically, this degradation is the result of two phenomena. The first one is deterministic and is related to the acquisition, to possible defects of the imaging system (blur created by an incorrect lens adjustment, by motion, for example) or to phenomena such as atmospheric turbulence. The second phenomena is a random one and corresponds to the noise coming from any signal transmission. There are basically three standard noise models in imaging systems: additive, multiplicative and impulse noise. Moreover, typical noise models are characterized by the shape of their probability density function, which in the discrete setting is represented by the noise histogram.

Image restoration refers to the recovery of a clean sharp image from a noisy, and potentially blurred, observation and it consists of both image deblurring and denoising processes. There are several advanced mathematical image restoration approaches that have been developed and are being used today: stochastic modeling, wavelets (Daubechies & Teschke 2005), and variational/Partial differential Equation (PDE) approaches. In general, the nonlinear-diffusion PDE-based methods and TV-based regularizations, succeeded in obtaining good quality, edge preserving restorations, especially for noise removal. However, they modify images towards piecewise constant

functions, in such a way that important information, encoded in image features like textures and details, is often compromised in the restoration process. In this review, we will focus on variational methods for the restoration of images corrupted by additive noise which we assume to be sampled from a known a-priori distribution.

Without loss of generality, we consider grayscale images with a square  $d \times d$  domain. Let  $\bar{u} \in \mathbb{R}^{d^2}$  be the unknown  $d \times d$  clean image concatenated into an  $d^2$ -vector,  $K \in \mathbb{R}^{d^2 \times d^2}$  be a known blurring operator and  $\bar{n} \in \mathbb{R}^{d^2}$  be an unknown realization of the random noise process, which we assume white with zero mean and standard deviation  $\sigma$ . The discrete imaging model of the degradation process which relates the observed degraded image  $g \in \mathbb{R}^{d^2}$  with  $\bar{u}$ , can be expressed as follows:

$$g = K\bar{u} + \bar{n}. \quad (1)$$

Given  $K$  and  $g$ , the goal of restoration is to solve the inverse problem of recovering an estimate  $u^*$  of  $\bar{u}$  from  $g$ , which is known as deconvolution or deblurring. When  $K$  is the identity operator, recovering  $\bar{u}$  from  $g$  is referred as denoising. It is well known that deblurring by direct inversion of  $K$ , i.e.  $u^* = K^{-1}g$ , can yield meaningless results due to noise and to the fact that the linear blur operator  $K$  is typically ill-conditioned or singular (Hansen 1998). To obtain a meaningful restoration, some prior information is commonly included in the formulation.

Variational approach to restoration typically consists in solving the optimization problem

$$u^* \leftarrow \arg \min_{u \in \mathbb{R}^{d^2}} \{ \mathcal{R}(u) + \mu \mathcal{F}(u, g) \}, \quad (2)$$

where the regularization term  $\mathcal{R}(u) = \|A(u)\|_q^q$  ( $A$  is a regularization operator, not necessarily linear) enforces certain prior constraints on the clean image  $u$ , the fidelity term  $\mathcal{F}(u, g) = \|Ku - g\|_p^p$  measures how  $u$  fits the observed image  $g$ , and  $\mu$  is a positive regularization parameter balancing the two terms. In particular, the fidelity term is typically related to the type of noise corrupting the image, while the regularization term commonly enforces smoothness of the solution.

Many different variational restoration models have been proposed, ranging from the classical Tikhonov ( $p=2$ ,  $q=2$ ) (Hansen 1998) (Calvetti & Reichel 2003), to Total Variation (Rudin, Osher, & Fatemi 1992), Mumford-Shah (Jung, Bresson, Chan, & Vese 2011), and many other variants of these.

Total variation regularization was originally introduced for noise reduction (Rudin, Osher, & Fatemi 1992), and is obtained from (2) by setting  $p=2$ ,  $q=1$ , and  $A(u)$  the gradient magnitude of  $u$ . If we let  $\nabla u_i := (G_{x,i}u, G_{y,i}u)^T$ , with  $G_{x,i}$ ,  $G_{y,i}$  representing the  $i$ th rows of the  $x$  and  $y$ -directional finite difference operators  $G_x$ ,  $G_y$ , respectively, then the regularization term  $\mathcal{R}(u)$  is defined by the TV-semi-norm

$$\|u\|_{TV} = \|A(u)\|_1 := \sum_{i=1}^{d^2} \sqrt{(G_{x,i}u)^2 + (G_{y,i}u)^2}.$$

The distinctive feature of TV regularization is that image edges can be preserved, but the restoration can present staircase effects.

For what concerns the fidelity term, the  $L_2$  norm is used for images corrupted by additive Gaussian noise. This standard choice corresponds to setting  $p=2$  in (2), which indirectly encodes the prior knowledge on the noise standard deviation  $\sigma$ . When noise is of impulsive type, it is well known that better performance are obtained by using the  $L_1$  norm.

In this work, we will first present two variants of the general variational restoration model in (2): the first relies on an adaptive regularization operator which is constructed by using fractional order derivatives (Sect. 2), the second proposes a fidelity term which imposes the resemblance of the residue image to a white noise realization by constraining its autocorrelation function (Sect. 3). These two approaches are particularly suitable for the restoration of partly-textured perturbed images.

One aspect generally missing from state-of-the-art image restoration methods is a full exploitation of all the available information about the noise.

Instead of the standard  $L_2$  fidelity term in (2), in order to explicitly exploit the assumption that the noise distribution is known, in the third approach we present (Sect. 4) we introduce a novel fidelity term that enforces the similarity between the residual distribution and a target noise distribution.

In Sect. 5 we will report some experimental results which allow to point out the potentials of the three considered variational approaches.

## 2 ADAPTIVE FRACTIONAL REGULARIZER

In this section, we briefly summarize the work in (Chan, Lanza, Morigi, & Sgallari 2013), which presents a variational model for restoration of textured images corrupted by blur and impulse/Gaussian additive noise. In particular, the model relies on an adaptive fractional order regularization term and on a  $L_1$  fidelity. The model is as follows:

$$u^* \leftarrow \arg \min_{u \in \mathbb{R}^{d^2}} \{ \|Ku - g\|_1 + \|\Lambda A_\alpha(u)\|_1 \}, \quad (3)$$

where  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_{d^2})$  is an  $d^2 \times d^2$  diagonal matrix with  $\lambda_i$  representing the regularization parameter for the  $i$ th pixel,  $A_\alpha(u_i) = \|\nabla^{\alpha_i} u_i\|_2$ , where  $\alpha_i$  represents the fractional order of differentiation for the  $i$ th pixel, and

$$\nabla^{\alpha_i} u_i := (G_{x,i}^{\alpha_i} u, G_{y,i}^{\alpha_i} u)^T \quad (4)$$

is the fractional-order discrete gradient operator, with components representing the  $x$  and  $y$ -directional fractional finite difference operators.

Let us motivate our model by analyzing the high-pass filtering character of the fractional order derivative operator. In Figure 1 we show the restored images of the blurred and noisy test image in Figure 3(b) by applying model (3) with different  $\alpha$  while keeping  $\lambda=1.0$ . With  $\alpha=1$  we get the  $\ell_1$ -TV model which preserves edges but

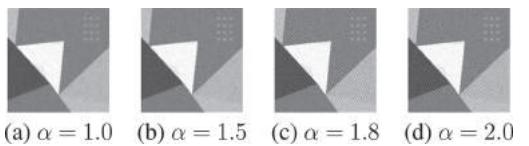


Figure 1. Restoration of test image shown in Figure 3(b) using model (3) with different values of  $\alpha$  while  $\lambda=1.0$  is fixed.

fails to preserve fine scale features such as textures in the images, see Figure 1(a). The high-pass capability becomes stronger with larger  $\alpha$ , so more and more texture regions are well preserved when  $\alpha$  increases to 1.5, 1.8 and then 2.0, see Figures 1(b)–(d), respectively. However, for  $\alpha=2.0$  the high pass capability of fractional order derivatives is too strong so that most components of the image remain and only little high frequencies noise is removed. Thus a best filtering is obtained by adaptively applying the fractional order regularization in texture regions and TV regularization in non-texture regions.

The method consists of two phases. In the first phase, we apply a texture detector (we refer to (Chan, Lanza, Morigi, & Sgallari 2013) for more details) to obtain a texture map. The texture map contains a texture measure (a real positive value) for each pixel and is thresholded to obtain the texture and non-texture classes. The texture class is further partitioned into  $C$  subclasses according to the texture measure. Following this classification a value  $\alpha_i$  is associated to each pixel  $i$ , with  $\alpha_i=1$  if the  $i$ th pixel belongs to the non-texture class, and  $\alpha_i \in \{\hat{\alpha}_1, \dots, \hat{\alpha}_C\}$  if the  $i$ th pixel belongs to one of the  $C$  texture subclasses. The regularization parameters  $\lambda_i$  in the diagonal matrix  $\Lambda$  in (3) are then chosen according to the  $\alpha_i$ 's. In particular, each texture class has an associated regularization value, while for the non-texture class we set  $\lambda=1.0$ . In the second phase, we apply the classical TV regularization ( $\ell_1$ -TV) to the non-texture regions while applying a fractional order TV regularization ( $\ell_1$ -TV $^\alpha$ ) in the texture classes. The corresponding minimization problem is solved by the half-quadratic method in (Nikolova & Ng 2005) (see (Chan, Lanza, Morigi, & Sgallari 2013) for details)

### 3 WHITENESS CONSTRAINING FIDELITY

When the standard deviation  $\sigma$  of noise is available, under the discrepancy principle, the *unconstrained* TV- $L_2$  model obtained by setting  $p=2$ ,  $q=1$  in (2), can be equivalently reformulated as the following *constrained* optimization problem:

$$u^* \leftarrow \arg \min_{u \in U} \|u\|_{TV}, \quad (5)$$

with the feasible set defined as

$$U = \left\{ u \in \mathbb{R}^{d^2} : \|Ku - g\|_2^2 \leq \tau^2 d^2 \sigma^2 \right\}, \quad (6)$$

where  $\tau$  is a scalar parameter controlling the variance of the residue image  $Ku - g$ .

In this section we illustrate the variational method proposed in (Lanza, Morigi, Sgallari, & Yezzi 2013) for image denoising and extended in (Lanza, Morigi, & Sgallari 2015) for image restoration, which relies on a TV regularization term (simply for its popularity, any other regularizer could be used as well) and on a fidelity that constrains the residual of the restoration to resemble a white noise process.

The model proposed in (Lanza, Morigi, & Sgallari 2015), referred to as TV-W, extends the classical constrained TV- $L_2$  restoration model (5), by incorporating constraints on the whiteness of the residue image. The model replaces the feasible set defined in (6) as follows:

$$u^* \leftarrow \arg \min_{u \in W} \|u\|_{TV}, \quad (7)$$

where the new feasible set  $W \subset \mathbb{R}^{d^2}$  contains solutions  $u$  such that the corresponding residue image  $Ku - g$  satisfies the discrepancy principle and is guaranteed to be white. We refer to (Lanza, Morigi, & Sgallari 2015) for a detailed description of the minimization algorithm based on the Alternating Directions Method of Multipliers (ADMM) (Chan, Tao, & Yuan 2013).

In order to define  $W$ , we consider the periodogram  $p_{\bar{n}} \in \mathbb{R}^{d^2}$  of the noise realization  $\bar{n}$  defined as  $p_{\bar{n}} = |T\bar{n}|$ , where  $|\cdot|$  denotes the component-wise modulus of the complex vector  $T\bar{n}$ ,  $T \in \mathbb{C}^{d^2 \times d^2}$  is obtained using the unitary matrix representing 1D Discrete Fourier Transform and the Kronecker product. The vector  $p_{\bar{n}}$  is suitably scaled by multiplication with a diagonal normalization matrix  $M$ , i.e.  $\hat{p}_{\bar{n}} = Mp_{\bar{n}}$ .

Then, the elements of the normalized periodogram are sorted in increasing order of spatial frequency by multiplication with a suitable permutation matrix  $\Pi$ . The *Normalized Cumulative Periodogram* (NCP) of  $\bar{n}$  is thus defined as the vector  $c_{\bar{n}} \in \mathbb{R}^{d^2}$  given by

$$c_{\bar{n}} := S\Pi\hat{p}_{\bar{n}}, \quad (8)$$

where  $S$  is the  $d^2 \times d^2$  partial sums matrix, i.e. a lower triangular matrix with all nonzero entries equal to one.

Hence, the whiteness set  $W$  used in model (7) is as follows:

$$W = \left\{ u \in \mathbb{R}^{d^2} : b^- \leq S\Pi M |T(Ku - g)| \leq b^+ \right\},$$

where the inequalities must be intended component-wise and  $b^+, b^- \in \mathbb{R}^{d^2}$  are vectors containing the upper and lower limits of the whiteness set, respectively.

## 4 HISTOGRAM CONSTRAINING FIDELITY

Similarly to the approach illustrated in the previous section, the work in (Lanza, Morigi, Sgallari, & Yezzi 2014) presents a constrained variational model for restoration based on a TV regularization and on an extension of the discrepancy-based constraint (6). In particular, instead of considering only the variance of the residue image, the approach in (Lanza, Morigi, Sgallari, & Yezzi 2014) constrains the overall distribution of the residual to resemble the distribution of noise, which is assumed to be known.

This concept is better illustrated in Figure 2. Three synthetic examples of additive image noise all with the same mean  $\mu = 127.5$  and standard deviation  $\sigma = 27.5$  are depicted in the first row, nevertheless the first noise example is completely different from the other two. In particular, the first is a simple step function, the second and the third are specific realizations of Gaussian and uniform noise distributions, respectively. In the second row the associated histograms are shown, which reflect the visual difference we can notice in the illustrations. Even if the second and the third noisy images in the top row appear visually similar, the associated histograms are significantly different from each other. Therefore, by exploiting the complete information on the noise distribution we expect to obtain a residual which better fits the probabilistic model of the noise process and, as a consequence, we expect an improvement in the restored image.

The variational model takes the form:

$$u^* \leftarrow \arg \min_{u \in \mathcal{H}} \|u\|_{TV}, \quad (9)$$

where the feasible set  $\mathcal{H}$  is defined as follows:

$$\mathcal{H} = \left\{ u \in \mathbb{R}^{d^2} : \bar{H}_L[i] \leq H_{g-u}[i] \leq \bar{H}_U[i] \forall i = 1, \dots, Q \right\},$$

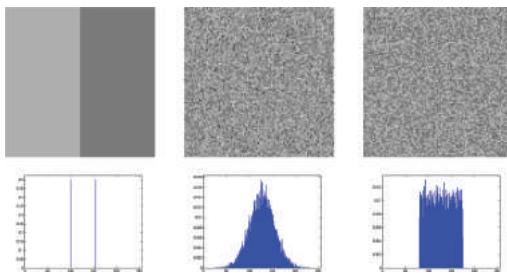


Figure 2. Top row: step image (left), realization of Gaussian (center) and uniform (right) noise distributions; bottom row: associated histograms.

where  $H_{g-u}$  denotes the normalized cumulative histogram of the residue image  $g - u$ ,  $i$  is the index of  $Q$  bins of the histogram, and  $\bar{H}_L$  and  $\bar{H}_U$  represents the lower and upper limits, respectively, of the band around the theoretical noise histograms. For the computation of the  $\bar{H}_L$  and  $\bar{H}_U$  limits, a Monte Carlo approach is used: given a target probability density function for the noise and a selected image dimension  $d$ , we generate a large number of different noise realizations by sampling from the target distribution. For each realization, the normalized and cumulative histograms are computed. Finally, the minimum and maximum values within all the realizations for all the histogram bins are obtained.

To compute the minimizer of the proposed model (i.e., the restored image), an efficient minimization algorithm based on the ADMM technique is presented (Chan, Tao, & Yuan 2013).

## 5 EXPERIMENTAL RESULTS

In this section we will give some examples related to the reviewed restoration algorithms when applied to images synthetically corrupted by blur and white Gaussian noise. We compare the proposed algorithms with the well-known Rudin-Osher-Fatemi (ROF) model (Rudin, Osher, & Fatemi 1992), based on the minimization of the  $TV-L_2$  or  $TV-L_1$  functionals.

The regularization parameter of the  $TV-L_2$  model is adjusted so that the solution is guaranteed to satisfy the discrepancy principle, that is the variance of the residual must be equal to the variance of the noise.

We stopped the algorithms as soon as the relative difference between two successive iterates satisfies  $\|u^k - u^{k-1}\|_2^2 / \|u^{k-1}\|_2^2 < 10^{-4}$  or after a maximum of 100 iterations.

The accuracy of the methods is evaluated by the Signal to Noise Ratio (SNR) defined by

$$SNR(u^*, \bar{u}) = 20 \log_{10} \frac{\|\bar{u}\|}{\|u^* - \bar{u}\|} dB, \quad (10)$$

where  $u^* \in \mathbb{R}^{d^2}$  is the computed estimate of the uncorrupted image  $\bar{u} \in \mathbb{R}^{d^2}$ . This quantity provides a quantitative measure of the quality of the restored image: a high SNR value indicates that  $u^*$  is an accurate approximation of  $\bar{u}$ .

The corrupted images taken as input by the algorithms have been obtained as follows. First, the original  $d \times d$  image  $\bar{u}$  is blurred by a Gaussian kernel characterized by two parameters band and sigma and generated through the MATLAB command `fspecial('Gaussian', band, sigma)`. The larger the sigma, the stronger the blur will be.

Then, the blurred image  $K\bar{u}$  is corrupted by additive zero-mean white Gaussian noise with standard deviation  $\sigma$ .

#### Example 1: Adaptive Fractional (AF) algorithm.

We consider the restoration of a blur- and noise-contaminated test image represented by  $255 \times 255$  pixels. The desired blur- and noise-free image is depicted in Figure 3(a). The resulting image is displayed in Figure 3(b).

In Table 1 the second column with header  $\text{SNR}_i$  reports SNR values for test image that have been corrupted by spatially-invariant Gaussian blur, characterized by  $\text{band} = 3$  and  $\text{sigma} = 1.5$ , and by noise with different levels  $\nu$ . In the case of Gaussian noise, let  $\tilde{g} \in \mathbb{R}^d$  be the associated vector with the unknown noise-free entries, i.e.,  $\tilde{g} = g + e$  where the vector  $e$  represents the noise. We define the noise-level  $\nu = \|e\|/\|\tilde{g}\|$ . The values of  $\nu$  are shown in the first column. Column three, labeled  $\text{SNR}_{\text{AF}}$ , shows

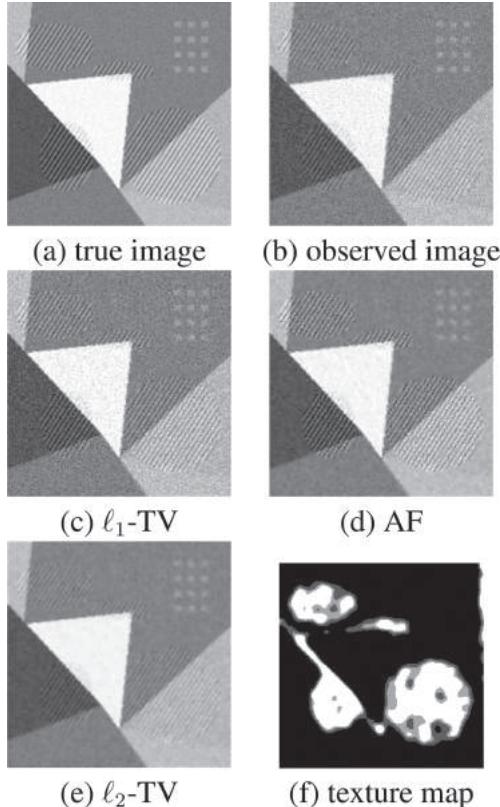


Figure 3. Example 1. test images: (a) blur- and noise-free image; (b) corrupted image by Gaussian blur,  $\text{band} = 3$  and  $\text{sigma} = 1.5$ , 10% of Gaussian noise; (c)  $L_1$ -TV restoration  $\lambda_1 = 0.1$ ; (d) restored image by AF algorithm; (e)  $L_2$ -TV restoration  $\lambda_1 = 0.1$ ; (f) texture classes computed on (b).

Table 1. Example 1. Results for restorations of test images that have been corrupted by spatially-invariant Gaussian blur, defined by  $\text{band} = 3$  and  $\text{sigma} = 1.5$ , and by noise corresponding to noise-level  $\nu$ .

$\nu$	$\text{SNR}_i$	$\text{SNR}_{\text{AF}}$	$\text{SNR}_{L_1\text{-TV}}$	$\text{SNR}_{L_2\text{-TV}}$
0.01	15.98	20.46	20.00	18.22
0.05	13.29	15.86	15.06	15.48
0.10	9.44	13.06	11.56	12.44

the SNR values for restorations by AF algorithm, and columns four and five, labeled  $\text{SNR}_{L_1\text{-TV}}$  and  $\text{SNR}_{L_2\text{-TV}}$ , show the SNR values for restorations by TV regularizer with  $L_1$  and  $L_2$  data fidelity terms, respectively.

The adaptive regularization process applies different levels of denoising in different image regions. This improves the result visually, in terms of texture preservation and smoothing of the homogeneous regions, and in terms of signal-to-noise-ratio. Figure 3(d) shows the restoration by AF algorithm and Figure 3(f) depicts the texture-map used for the restoration. The texture measures are obtained on the corrupted image in Figure 3(b) and the four texture classes are shown using gray-level values. In particular, in black regions the AF algorithm is applied with  $\alpha_4 = 1.0$ ,  $\alpha_2 = 1.7$ , and  $\alpha_3 = 1.8$  in the gray regions, and  $\alpha_1 = 1.9$  in white regions. The restoration by  $L_1$ -TV is illustrated in Figure 3(c) using  $\lambda = 0.1$  and it preserves textures but it does not remove the noise. However, when we apply  $L_2$ -TV restoration with  $\lambda = 1.0$  it does the opposite, that is, it preserves very well the edges but it destroys the textures. The restoration by  $L_2$ -TV, illustrated in Figure 3(e), shows less sharpness and smoothing of the texture regions.

#### Example 2. Whiteness constraining fidelity

We consider the restoration of skyscraper ( $d = 256$ ) image, which presents interesting mixtures of textures, flat regions and shaded areas; the noise-free version of the image is depicted in Figure (4)(a).

In this example we investigate the comparison between the restoration by TV-L2 and by TV-W algorithms for a fixed additive Gaussian noise with standard deviation  $\sigma = 5$  and variable Gaussian blur degradations, characterized by band and sigma, which significantly worsen the ill-conditioning of the image restoration problems. Table 2 reports in the first and second columns the parameters band and sigma of the Gaussian blur kernel, and the third column, labeled  $\text{SNR}_0$ , reports the SNR-values of the available corrupted image  $g$ , i.e.,  $\text{SNR}(\bar{u}, g)$ . The columns labeled TV-L<sub>2</sub> and TV-W display SNR-values for the restored images obtained by the two considered algorithms. For increasing values of the band and

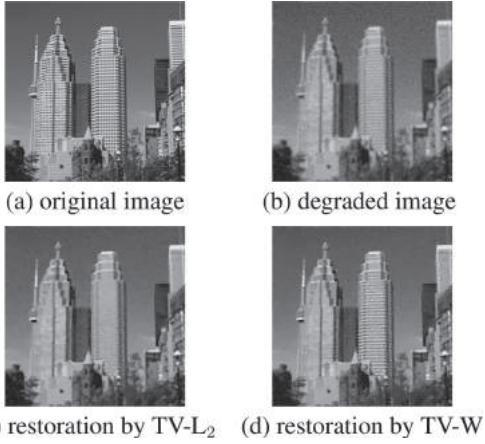


Figure 4. Example 2. Restoration obtained by TV-L<sub>2</sub> (c) and TV-W (d) models on the skyscraper image (a) degraded (b) by Gaussian blur with band = 5, sigma = 1.5 and additive zero-mean white Gaussian noise with standard deviation  $\sigma = 10$ .

Table 2. Example 2. SNR values from restoration of skyscraper image corrupted by additive zero-mean white Gaussian noise (fixed  $\sigma = 5$  and varying Gaussian blur).

Band	Sigma	SNR <sub>0</sub>	TV-L <sub>2</sub>	TV-W
7	2.5	4.89	6.54	<b>7.07</b>
7	1.2	7.18	9.06	<b>10.01</b>
5	1.0	8.39	10.82	<b>11.45</b>
5	0.7	11.35	14.15	<b>14.33</b>
3	1.0	9.88	12.24	<b>12.64</b>
3	0.5	15.85	17.66	<b>17.91</b>

sigma parameters, the textured parts of the images tend to disappear, consequently the benefits of the whiteness constraint decrease until the worst case with band = 7 and sigma = 2.5, where the SNR-values for the two algorithms are almost the same.

In Figure 4 we display the degraded images and the results by applying the two different image restoration methods.

A visual inspection of the figures of this example together with the results in Table 2 allow us to conclude that by constraining the whiteness of the residual, we can go beyond the use of just the residual norm or the variance of the noise model, thus providing better restorations.

### Example 3: Histogram constraining fidelity

We compare the proposed algorithm, referred to as TV-CDF, with the well-known Rudin-Osher-Fatemi (ROF) model (Rudin, Osher, & Fatemi 1992), based on the minimization of the TV-L<sub>2</sub> functional (5).

The regularization parameter of the TV-L<sub>2</sub> model is adjusted so that the solution is guaranteed to satisfy the discrepancy principle, that is the variance of the residual is equal to the variance of the noise.

The accuracy of the methods is evaluated by the Root Mean Squared Error (RMSE) defined as

$$\text{RMSE}(u, \bar{u}) := \sqrt{\|u - \bar{u}\|_2^2 / d^2}, \quad (11)$$

where  $u \in \mathbb{R}^d$  is the computed approximation of the desired noise-free image  $\bar{u} \in \mathbb{R}^d$ . This quantity provides a quantitative measure of the quality of the restored image  $u$ . A small RMSE value indicates that  $u$  is an accurate approximation of  $\bar{u}$ ; we recall, however, that the RMSE values are not always in agreement with visual perception.

We consider the restoration of test ( $d = 256$ ) image, which present interesting mixtures of textures, flat regions and shaded areas; the noise-free version of the image is depicted in Figure (5)(a).

The quantitative results in Table 3 demonstrate that the proposed TV-CDF model outperforms the TV-L<sub>2</sub> approach in terms of obtained RMSE values on the test image for all the considered noise levels.

The visual results for test image corrupted by zero-mean AWGN with standard deviation  $\sigma = 10$ ,

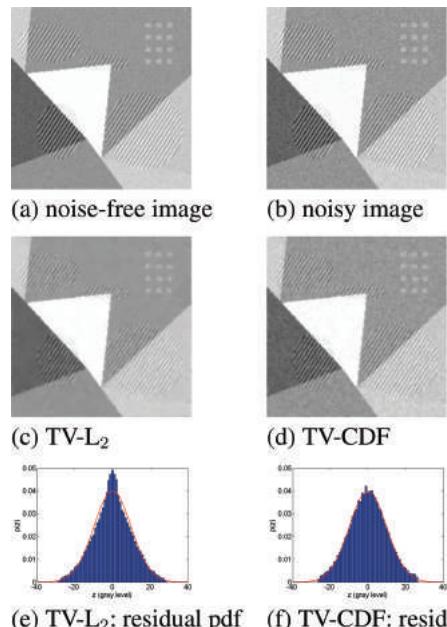


Figure 5. Example 3. Restoration results on test image corrupted by zero-mean AWGN with standard deviation  $\sigma = 10$ .

Table 3. RMSE values for test image corrupted by zero-mean AWGN with standard deviation  $\sigma$ .

$\sigma$	TV-L <sub>2</sub>	TV-CDF
10	7.41	<b>6.82</b>
20	11.00	<b>10.34</b>
30	13.55	<b>13.02</b>

are illustrated in Figure (5). In particular, the noise-free image is depicted in Figure (5)(a), the noise-corrupted image is shown in Figure (5)(b), the images restored by TV-L<sub>2</sub> and by TV-CDF are given in Figure (5)(c) and Figure (5)(d), respectively. In the last rows of Figure (5) we show the normalized histograms of the residuals associated with the restored images depicted in Figure (5)(c) and Figure (5)(d).

The effect of the proposed constraint is evident from the residual histograms. In fact, the residual histograms obtained by the proposed TV-CDF method and displayed in Figure (5)(f) adhere more faithfully to the target Gaussian probability density function (red curve) than those reported in Figure (5)(e) and computed by the TV-L<sub>2</sub> model.

## 6 CONCLUSIONS

We reviewed three recent proposals of variational approaches for image restoration. We have shown as the choice of suitable regularization and data fidelity terms is crucial for the success of the variational model. The aspect of computational efficiency in solving the optimization problem has not been considered in the present short review.

## REFERENCES

- Calvetti, D. & L. Reichel (2003). Tikhonov regularization of large linear problems. *BIT Numerical Mathematics* 43(2), 263–283.
- Chan, R.H., A. Lanza, S. Morigi, & F. Sgallari (2013, 2). An adaptive strategy for the restoration of textured images using fractional order regularization. *Numerical Mathematics: Theory, Methods and Applications* 6, 276–296.
- Chan, R.H., M. Tao, & X. Yuan (2013). Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers. *SIAM J. Imaging Sciences* 6(1), 680–697.
- Daubechies, I. & G. Teschke (2005). Variational image restoration by means of wavelets: Simultaneous decomposition, deblurring, and denoising. *Applied and Computational Harmonic Analysis* 19(1), 1–16.
- Hansen, P.C. (1998). *Rank-deficient and Discrete Ill-posed Problems: Numerical Aspects of Linear Inversion*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics.
- Jung, M., X. Bresson, T. Chan, & L. Vese (2011, June). Nonlocal mumford-shah regularizers for color image restoration. *Image Processing, IEEE Transactions on* 20(6), 1583–1598.
- Lanza, A., S. Morigi, & F. Sgallari (2015). Variational image restoration with constraints on noise whiteness. *Journal of Mathematical Imaging and Vision* 53(1), 61–77.
- Lanza, A., S. Morigi, F. Sgallari, & A.J. Yezzi (2013). Variational image denoising based on autocorrelation whiteness. *SIAM J. Imaging Sciences* 6(4), 1931–1955.
- Lanza, A., S. Morigi, F. Sgallari, & A.J. Yezzi (2014). Variational image denoising while constraining the distribution of the residual. *ETNA. Electronic Transactions on Numerical Analysis [electronic only]* 42, 64–84, electronic only.
- Nikolova, M. & M.K. Ng (2005). Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM Journal on Scientific Computing* 27(3), 937–966.
- Rudin, L.I., S. Osher, & E. Fatemi (1992, November). Nonlinear total variation based noise removal algorithms. *Phys. D* 60(1–4), 259–268.

This page intentionally left blank

# Virtual and augmented medical imaging environments: Applications to simulation, training, surgical planning and interventional guidance

Cristian A. Linde

*Department of Biomedical Engineering, Chester F. Carlson Center for Imaging Science,  
Rochester Institute of Technology, Rochester, NY, USA*

**ABSTRACT:** While effective therapy cannot be delivered without the appropriate equipment and instrumentation that enables access to the internal organs and tissues under minimally invasive conditions, real-time visualization and navigation information is critical to guide the instruments to the correct targets and manipulate the diseased tissues in absence of direct vision. Over the past decades, thanks to the advances in medical image acquisition, visualization and display, surgical tracking and image computing infrastructure, a wide variety of technology has emerged that facilitates diagnosis, procedure planning, intra-operative guidance and treatment monitoring while providing safer and less invasive approaches for therapy delivery. Nevertheless, successful exploitation, use and integration of the multi-modality images and signals available from both pre- and intra-operative sources require the development and validation of engineering tools that allow users to extract critical information for decision making during planning, visualization, guidance or training. This article serves as a quick snapshot of various applications of biomedical imaging and modeling, their integration into platforms for planning and guidance of interventions, and their potential to develop new paradigms for virtual anatomy visualization for simulation, teaching and training applications.

## 1 INTRODUCTION

Several decades ago many surgical interventions required the clinician to have full physical and visual access to the internal organ that required treatment: abdominal procedures involved large incisions and exposure of inner organs, neurosurgical interventions involved the opening of the skull, and cardiac interventions were performed on the open-chest, while the patient was placed on cardiac arrest and connected to a heart-lung machine. Despite the success of the therapy itself, often times patients would experience severe infections, delayed recovery and other complications mainly induced by the large incisions and tissue exposure needed to access the surgical targets rather than by surgically repairing the tissue. Hence, less invasive interventions have been motivated by improved patient outcome, fewer complications, reduced morbidity, shorter recovery times, and improved overall quality of life.

One crucial step toward reducing invasiveness was the substitution of direct vision by real-time camera views: the introduction of endoscopic imaging enabled the use of miniature cameras, dramatically reducing incision size (Litynski 1999). Another crucial step was the adoption of interventional radiology as a surgical sub-specialty, opening up the horizon for percutaneous, catheter-based interventions conducted under X-ray guidance

(Soares & Murphy 2005). Despite the limited soft tissue visualization provided by two-dimensional (2D) X-ray fluoroscopy imaging, this modality has become standard-of-care thanks to its real-time capabilities and wide clinical use.

The limited extent of direct visualization of the surgical scene and organs being treated, and often lack thereof, require significant reliance on medical imaging and computer-assisted navigation. While the less invasive treatment approaches have proved to be beneficial to patients, they have introduced new challenges for the physicians. One such challenge, for example, has been voiced by cardiac interventions being among the last disciplines to adopt the minimally invasive treatment techniques, mainly due to the challenges associated with access and visualization inside the beating heart.

In addition to its use in the planning and guidance of minimally invasive intervention, medical imaging and image-guided navigation present a significant potential for paradigm-shifting approaches for simulation, teaching and training. We live in the “digital” age surrounded by technology that enables acquisition of 3D, 4D and 5D medical images (Fig. 1) and stereoscopic and 3D visualization displays have become commodities. Yet, traditional pedagogical approaches still resort to using standard didactic equipment (i.e., textbooks, slides) to present human anatomy and describe the interconnection between



Figure 1. Visualizing the possibilities: examples of slice-, surface- and volume-rendered representations of different anatomical organ models generated from medical imaging datasets. *Image courtesy of the Biomedical Imaging Resource, Mayo Clinic, Rochester MN USA.*

different organs and systems. Teaching and training experiences would be much improved by leveraging the availability of open source, high-quality medical imaging datasets to generate 2D slices, 3D surfaces and 4D volume renderings of the anatomy. These methods enable the user to interact with the imaging data via commercially available tracking systems, visualize the interconnected data via widely available and many affordable virtual and/or augmented reality head-mounted displays or other routine visualization and display paradigms.

## 2 ENABLING TECHNOLOGIES FOR BIOMEDICAL IMAGE VISUALIZATION AND IMAGE-GUIDED NAVIGATION

Over the last ten years, all medical imaging modalities have evolved and their use has expanded much beyond their traditional diagnostic utilization. X-ray imaging, along with endoscopy, have been the first to provide real-time visualization during procedure guidance in the Operating Room (OR), and have departed significantly from their initial use. X-ray images are used to generate three-dimensional (3D) reconstructions of the anatomy and internal vasculature via Computer Tomography (CT), cone-beam CT and CT angiograms. Single-camera endoscopy has been upgraded to stereoscopic endoscopy, accompanied by several attempts toward augmented reality visualization.

Shortly afterwards Ultrasound (US) imaging found its way into the interventional suite, mainly thanks to its real-time capabilities, relatively low cost (compared to CT or Magnetic Resonance Imaging (MRI)), wide availability, portability and lack of ionizing radiation. Consequently,

trans-esophageal echocardiography has become standard-of-care for real-time patient monitoring during most cardiac interventions.

Lastly, MRI has enabled non-invasive, high-quality visualization inside the human body with exquisite soft tissue contrast, yet accompanied by its inherent high cost and demanding infrastructure.

While modalities such as CT and MRI have been predominately employed for diagnostic imaging and pre-procedural planning, recent efforts have been dedicated to adapting and developing the necessary infrastructure to use CT and MRI for intra-operative real-time guidance and visualization in the interventional suite. These interventional approaches require access to specialized technology compatible with specific high-end imaging equipment, imposing a highly invasive technological “footprint” in the interventional suite requiring significant modifications to procedure workflow and re-training. Therefore, despite the envisioned benefits, such highly-specialized diagnostic imaging systems adapted for intra-operative guidance seem unlikely to become standard-of-care, but rather a specialized process only available to those centers that can afford it, thereby limiting access to affordable, efficient healthcare.

The primary challenge in image-guided interventions is to provide an optimal visualization environment that spares the clinician from having to mentally fuse information from the pre- and intra-operative images during the procedure. The development of an ideal environment is not trivial: pre-procedural images are typically 3D and feature a higher quality and volume of data, while intra-operative images feature lower resolution, lower quality and are often 2D, trading the amount of information displayed for real-time visualization. While an optimal visualization paradigm is hard to accurately identify, an image guidance system is deemed successful if it provides sufficient information for accurate navigation of the therapeutic instruments to the target tissue and successful on-site target tissue manipulation while avoiding damage to critical anatomical structures. Most often, this task is achieved by integrating pre-operative images and derived models into the intra-operative setting (Cleary & Peters 2010, Yaniv & Cleary 2006).

## 3 IMAGE-GUIDANCE WORKFLOW

### 3.1 Imaging and instrument tracking

The typical computer-assisted intervention workflow (Fig. 2) includes several steps, each serving a specific scope. The pre-operative images and anatomical models provide the necessary contextual information for the surgeon to identify the region of interest and navigate the surgical instrument to the

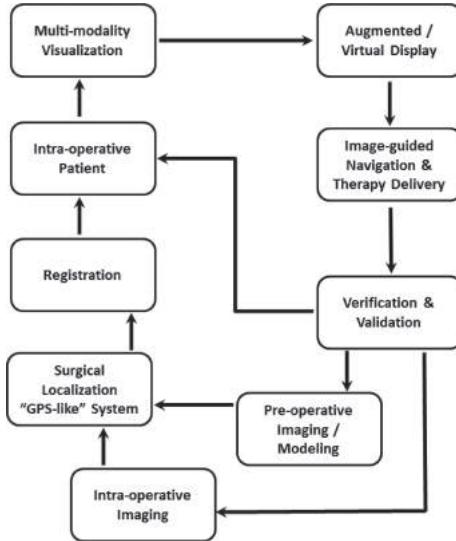


Figure 2. Components of a computer-assisted intervention workflow: pre-operative imaging, surgical instrument localization, data integration via registration, visualization and information display, surgical navigation, and intra-operative verification and evaluation.

target, serving as a road map. Provided the patient anatomy is registered to the pre-operative road map (typically using a tracking system), the surgical instrument representations can be visualized in the same coordinate system as the road map, much like a driver using a GPS navigation system to obtain real time positioning information along the route. Lastly, since pre-operative images cannot faithfully depict the intra-operative anatomy given their acquisition prior to the procedure, real-time images such as US or cone-beam CT are critical to visualize the true, intra-operative target location and confirm accurate and precise on-site therapy delivery.

### 3.2 Integration, visualization and display

All of the above components (pre- and intra-operative images and tracked instrument representations) are integrated into the same coordinate system that includes the patient via their inherent registration transformations, a mathematically complex step hidden from the surgeon, but requiring heavy engineering involvement. However, the appropriate visualization and display of the complex surgical environment is the critical stage that involves the surgeon and clinical team, as optimal presentation of the available information is paramount for procedure guidance and therapy delivery. While traditional 2D displays have been a part of the routine interventional suite

equipment, other paradigms have been explored, including the projection of virtual information on the patient's body surface either directly or using video augmentation (i.e., in the form of an augmented or virtual reality, depending on the extent of “real” and “virtual” information presented), or via the use of stereoscopic head-mounted displays or commercially available 3D displays (Linte, Davenport, Cleary, Peters, Vosburgh, Navab, Edwards, Jannin, Peters, Holmes III, & Robb 2013).

### 3.3 Targeting accuracy: How close are we?

A critical, yet often not explicit step of the image guidance workflow is “verification”: what is the uncertainty associated with the navigation, targeting and on-site therapy delivery? Typically, the true answer is provided by analyzing the intra-operative images, as these images depict the location of the surgical target most accurately, along with the instrument position and orientation. While the other components of the surgical navigation environment and their integration aim to provide minimal error, each component of the platform has its inherent uncertainty, in the end contributing to the overall targeting accuracy of the tracking system. Hence, the verification step is critical in evaluating the achieved targeting accuracy of the navigation platform, and while it is not explicitly performed during the procedure workflow, users assess accuracy at each step—navigation and positioning, by ensuring the information provided by the guidance platform indeed matches the information provided by the intra-operative images.

## 4 APPLICATIONS

### 4.1 Cardiac intervention applications

**Limitations and Challenges:** Catheter-based tissue ablation interventions for treatment of arrhythmic conditions still rely heavily on X-ray fluoroscopy guidance despite the limited visualization it provides. The intracardiac ablation path is typically planned prior to the procedure, but cannot be faithfully reproduced during the intervention. This challenge arises due to the lack of sufficient visualization and navigation information for accurate targeting, as well as the lack of information with regards to the geometry and pathophysiology of the ablation lesions. In these cases, despite the complex planning of the planned irreversible isolation of the electric pathway, patients end up with temporary edema and recurring arrhythmia requiring continued treatment and further procedural interventions.

**Tentative Solution:** Left atrial CT or MRI images enable the integration of the pre-procedural plan into the guidance data via registration of the pre- and intra-operative anatomical models

(Fig. 3). This approach facilitates the visualization and interpretation of pre-operative information in the context of the intra-operative data, augmented by real-time US imaging (Linte, Lang, Rettmann, Cho, Holmes III, Robb, & Peters 2012).

In addition, the pre-procedural images can be employed as substrate for modeling the tissue response to ablation, along with the physical and physiological tissue parameters, as well as the parameters provided by the ablation system. As a result, fast and accurate surrogate thermal models that predict the lesion geometry and lesion quality (extent of reversible and irreversible thermal damage induced in the tissue) can be developed, providing a clear visualization of the superimposed on the anatomical model (Fig. 4). While current approaches enable users to “drop” spherical glyphs of various size and colour intensity at the locations visited by

the catheter, their size and colour intensity is meaningless in terms of providing realistic information with respect to the quality (temperature, duration, thermal dose etc.) of the generated lesion. However, the proposed surrogate models can portray such information in real time, facilitating ablation lesion monitoring and providing cues with regards to the location at which subsequent ablation lesions should be delivered to avoid the presence of gaps in the ablation line, which would, in turn, lead to incomplete isolation of the arrhythmic pathway (Linte, Camp, Holmes III, Rettmann, & Robb 2013).

#### 4.2 Spine intervention applications

**Limitations and Challenges:** Spine instrumentation procedures require accurate and precise implantation of fixators (surgical screws) into the vertebral pedicle to immobilize motion at specific vertebral segments. These procedures entail detailed planning to identify the optimal implant size and trajectory to ensure firm fixation and also avoid their interference with proximal vascular or neural structures. Moreover, knowledge of the mechanical strength of the pedicle body is critical to avoid the instrumentation of weak bone regions, which would compromise fixation strength. Nevertheless, current planning approaches rely on the use of 2D radiographs or axial CT images to depict the 3D anatomy of the spine and estimate implant trajectory, despite the curvature of the spine. In addition, although 3D CT images of the patient are typically acquired several days before the procedure, the planning still occurs in the interventional suite with the patient under anesthesia, and often continues after opening up the patient, leading to prolonged intervention times.

**Tentative Solution:** The proposed application leverages the availability of the pre-operative images several days prior to the procedure, enabling the planning to be conducted by the surgeon or surgical staff prior to the intervention and allowing for additional consultations if necessary. The planning platform provides the user with full control to browse through and select virtual screws of appropriate dimension and specification from a database of virtual implants that replicate the actual implants manufactured by different spine instrumentation suppliers. The selected pedicle implants can then be virtually inserted in a 3D volume-rendered representation of the patient's spine CT data and their position and orientation can be altered as needed to avoid interference with critical neural or vascular structures (Fig. 5). Furthermore, the screw trajectory is optimized according to a surrogate fixation strength estimated based on the dimension of the selected implant and the strength (i.e., bone mineral density) of the pedicle bone traveled by the screw (Linte,

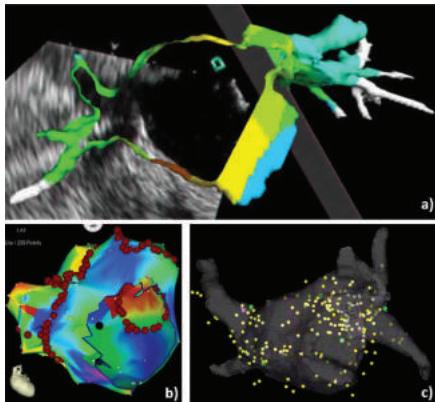


Figure 3. a) Example of a high-quality pre-operative left atrial model augmented with real-time US image; b) Electro-anatomical map of the left atrium generated intra-operatively using a tracked catheter; c) Registration of pre- and intra-operative left atrial model using point-cloud to surface registration.

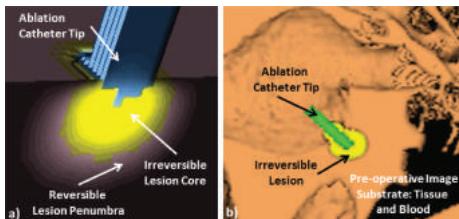


Figure 4. a) Example of modeled ablation lesion showing the irreversible lesion core surrounded by the reversible lesion penumbra; b) Volume-rendered representation of a pre-operative anatomical image thresholded into tissue and blood showing the ablation catheter tip and real-time lesion.

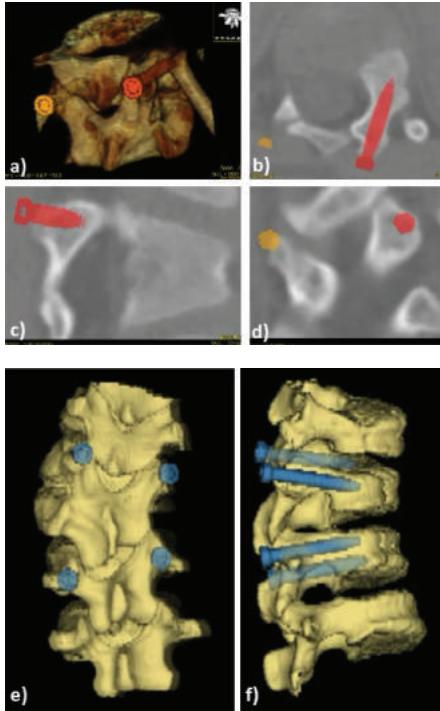


Figure 5. Snapshot of the spine intervention planning platform comprising a volume-rendered display of the CT image with implanted virtual screw templates (a); orthogonal views showing the screw templates within the pedicles (b, c, d), as well as the resulting volume-rendered representation of the instrumented spine segment featuring the virtual implants.

Camp, K., Huddleston, Robb, & Holmes III 2015). As such, provided several trajectories are identified, their associated fixation strength is estimated, enabling the user to select the trajectory that maximizes fixation strength while avoiding interference with critical anatomy. Therefore, the proposed platform provides optimal planning solutions based on objective metrics and yields automatically generated reports that prescribe the selected implant dimension, trajectory and achieved fastening strength deemed optimal at each vertebral level, enabling the full extent of the planning (and additional revisions) to be conducted prior to the procedure.

#### 4.3 Virtual and augmented visualization for simulation, teaching and training

**Limitations and Challenges:** Despite the availability of high quality 2D, 3D 4D (3D + time) and 5D (space + time + function) medical imaging data, current learning techniques still employ abstract,

2D displays of the anatomy from textbooks and slides. Moreover, the training of medical professionals also entails the use of traditional, “old school” approaches, further explaining the under-utilization of available medical imaging data, as well as visualization and display technology.

**Tentative Solution:** Here we describe two applications that leverage the power of virtual and augmented reality to explore different paradigms for presenting anatomical and functional information to viewers while exploiting the multi-dimensionalality of the data.

The first application aims to provide a virtual environment for learning the anatomy of the skull decomposed into several key bone segments. The dataset was generated using the *BodyParts3D* libraries that provide virtual representations of the anatomy of several organs and their interconnected components. The environment allows users to browse through different components of the skull anatomy while providing a brief audio description of the selected component and its adjacent, neighbouring components, guiding the user during the assembly process (Fig. 6). The virtual environment was designed in *Vizard*, a well-known software application for virtual and augmented reality implementations. The user controls the view and manipulates the components within the virtual environment using either a 3D mouse or (less intuitive) pre-defined key-board strokes. The environment can be displayed either on a 3D monitor, requiring each user to wear passive polarizing glasses to enable stereoscopic vision, or on virtual reality head-mounted displays, such as the popular and relatively inexpensive Oculus RIFT.

The second application demonstrates the use of augmented reality displays to enable the visualization of underlying information beyond the surface view of an object seen using real-time video (Potter, Besnch, Dawson-Elli, & Linte 2015). An example of such an application in medical practice would be the augmentation of laparoscopic video showing the surface of the liver surface with internal information extracted from pre-procedural CT

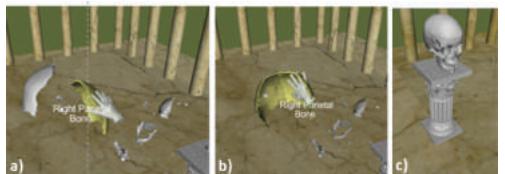


Figure 6. Example of a user-interactive virtual anatomy building platform for understanding the anatomical components and their assembly of the skull. Upon selection, each component is accompanied by a brief audio description of itself and its neighbouring components.

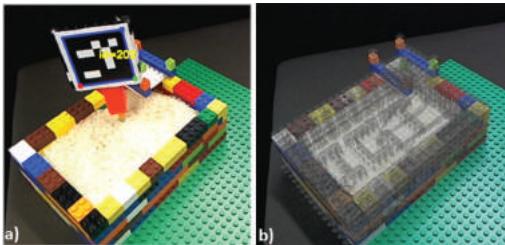


Figure 7. Example of an augmented display using the IGT Phantom. When filled with rice, the “IGT” letters inside the phantom are not visible; however, this information can be extracted from a CT image of the phantom and overlaid onto the real video, using the camera itself as a tracking device.

or MRI images, enabling the surgeon to “see” what lies beneath the surface in the attempt to optimize the path to resect the tumour.

The key constraints for these applications is the proper alignment of the real and virtual representations of the objects or organs, which entails accurate tracking of the camera and imaged scene. Several options entail the use of external spatial tracking and localization systems to determine the position and orientation of the image scene, camera and other elements. Another solution is to employ the camera itself as a tracking system and determine in real time the position and orientation of the imaged scene relative to the camera and use this information, along with the magnification information, to provide the correct position, orientation and scaling of the virtual object model when superimposed onto the real video view used as background image (Fig. 7).

## 5 IMAGE-GUIDED THERAPY: RESEARCH OR COMMODITY

Computer-assisted intervention environments provide clinicians with less invasive approaches for accurate and intuitive visualization, replacing or augmenting their direct view of the real world with computer graphics information extracted from the pre- and intra-operative medical images. While some image guidance environments feature highly complex infrastructures that often overwhelm the user, other systems empower the user, enabling him/her to carefully select and focus only on the relevant information. Moreover, although human-computer interaction in image guidance has been explored to a limited extent, for example when selecting the optimal visualization strategy for percutaneous procedures in interventional radiology (Hynynen 2011), such studies are the exception rather than the norm and additional research is yet

needed to confirm whether less information may, in fact, be more valuable.

**What does the ideal image guidance system look like?** An ideal system best positioned to make an impact in current clinical practice seamlessly blends diagnostic, planning, and guidance information into a single visualization environment displayed at the interventional site. In addition, the system provides accurate registration and tracking, high-fidelity visualization and information display, enables smooth data transfer, and fits within the traditional interventional suites without interfering with current workflows and typical equipment.

## REFERENCES

- Cleary, K. & T.M. Peters (2010). Image-guided interventions: technology review and clinical applications. *Annu. Rev. Biomed. Eng.* 12, 119–142.
- Hynynen, K. (2011). MRIGHIFU: A tool for image-guided therapeutics. *J Magn Reson Imaging.* 34(3), 482–493.
- Linte, C.A., J. Camp, A.K., P.M. Huddleston, R.A. Robb, & D.R. Holmes III (2015). Estimating pedicle screw fastening strength via a virtual templating platform for spine surgery planning: a retrospective preliminary clinical demonstration. *Comput method Biomed Eng. Imaging & Visualization In Press*, DOI:10.1080/21681163.2014.900725.
- Linte,C.A.,J.J.Camp,D.R.HolmesIII,M.E.Rettmann,& R.A. Robb (2013). Towards online modeling for lesion visualization and monitoring in cardiac ablation therapy. In *Proc. Med Image Comput Comput Assist Interv.*, Volume 8149 of *Lect Notes Comput Sci.*, pp. 9–17.
- Linte, C.A., K.P. Davenport, K. Cleary, C. Peters, K.G. Vosburgh, N. Navab, p. J. Edwards, P. Jannin, T.M. Peters, D.R. Holmes III, & R.A. Robb (2013). On mixed reality environments for minimally invasive therapy guidance: Systems architecture, successes and challenges in their implementation from laboratory to clinic. *Comp. Med. Imag. And Graph.* 37(2), 83–97.
- Linte, C.A., P. Lang, M.E. Rettmann, D.S. Cho, D.R. Holmes III, R.A. Robb, & T.M. Peters (2012). Accuracy considerations in image-guided cardiac interventions: experience and lessons learned. *Int J Comput Assist Radiol Surg.* 7, 13–25.
- Litynski, G.S. (1999). Endoscopic surgery: the history, the pioneers. *World J Surg.* 23(8), 745–753.
- Potter, M.J., A. Besnch, A. Dawson-Elli, & C.A. Linte (2015). Augmenting real-time video with virtual models for enhanced visualization for simulation, teaching, training and guidance. In *Proc. SPIE Medical Imaging: Image Perception, Observer performance and Technology Assessment*, Volume 9416, pp. 941615–8.
- Soares, G.M. & T.P. Murphy (2005). Clinical interventional radiology: Parallels with the evolution of general surgery. *Semin Intervent Radiol.* 22(1), 10–14.
- Yaniv, Z. & K. Cleary (2006, April). Image-guided procedures: A review. Technical Report CAIMR TR-2006-3, Image Science and Information Systems Center, Georgetown University.

*Contributed papers*

This page intentionally left blank

# The fast method of creating High Dynamic Range image for fluorescent microscopy applications

A. Bal

*Institute of Automatic Control, Silesian University of Technology, Gliwice, Poland*

**ABSTRACT:** In this work simple and effective method for creating of High Dynamic Range (HDR) images for use in the fluorescence microscopy applications is presented. For creating of the HDR image the Camera Response Function (CRF) is used, so the pixel values of the HDR image, in general, are proportional to the signal which have been reached the camera sensor (i.e. irradiance of sensor). Therefore the obtained HDR image can be used not only for qualitative analysis of the specimen, but also for its quantitative analysis. The proposed method has been tested on real data.

## 1 INTRODUCTION

The fluorescence microscopy is one of the most versatile and widely used imaging method in biology and medicine researches (Waters 2009). This method is mainly used for visualization of the location of cells or tissue components that have been stained with fluorescent molecules. It is also used for the quantitative assessment of fluorescent stained components. The advantages of the fluorescence microscopy are e.g. its: relative simplicity, versatility, high sensitivity and selectivity.

The usage of the fluorescence microscopy is connected with the problem of limited dynamic range of cameras used for image acquisition. This problem is particularly acute in applications in which e.g.:

- signal associated with a single object has greater dynamic than can be acquired by a camera,
- group of objects such that signal associated separately, which each object has a dynamic range lower than camera dynamic range, but the difference between signals from different objects is greater than the dynamic range of the camera.

In the second case objects can be located not only in the near neighbourhood by also in different places of one or more microscopy specimens.

In such cases the limitation of camera dynamic range leads to under or over exposition of the obtained single images. For clarity, since now, such images i.e. images obtained directly from the camera will be described as the *Low Dynamic Range* (LDR) images. The usage of LDR images leads to the loss of information contained in the analysed specimen, and therefore the analysis of such images can give false results.

An example of a case in which the camera dynamic range is too low for proper acquisition of the specimen image is presented in Figure 1. In this Figure the sequence of images, acquired with different exposure time  $t$ , of the same cell nuclei is presented. These images were obtained with the usage of fluorescence microscopy and DAPI fluorescent dye, which is used for detection of DNA. Hence, in this case, the signal intensity is related to the amount of DNA contained in different cell sites. As it can be seen the dynamic range of object signal is too high for the used camera. Therefore, there is no possibility of acquisition of a single, proper exposed, image on which, both, low and high object signals are properly reproduced.

The Figure 2 presents profiles of images (a)–(c) from Figure 1. The profiles are calculated for pixels which are lying on the line segment AB (the red line in the image 7(a)). The profiles are presented for image brightness (pixel values) normalized to the range [0; 1], which corresponds to the dynamic range of a LDR image. It can be seen in the image 7(b) that any of the images 1(a)–(c) is properly exposed—the image 1(a) is underexposed and the images 1(b) and 1(c) are overexposed.

In those applications where dynamic range of LDR image is insufficient the idea of the *High Dynamic Range* (HDR) imaging can be used (McCann & Rizzi 2012). The usage of HDR images in microscopy imaging is not a new idea. It was presented, e.g. in Bell et al. (2009) and in Eastwood and Childs (2012), but there is lack of example of usage of the HDR imaging method in fluorescence microscopy for qualitative and quantitative specimen analysis. The aim of this work was the investigation of the possibility of the usage of HDR imaging in such applications.

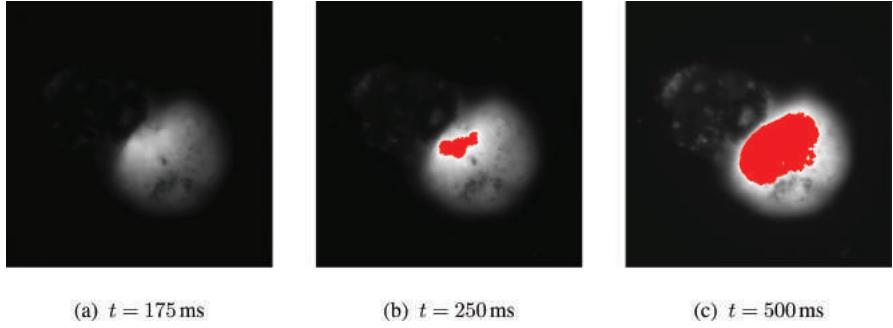


Figure 1. Images of the same cell nuclei acquisited in DAPI channel with different exposure times  $t$ ; the signal saturation ( $\tau > 0.975$ ) is indicated on red.

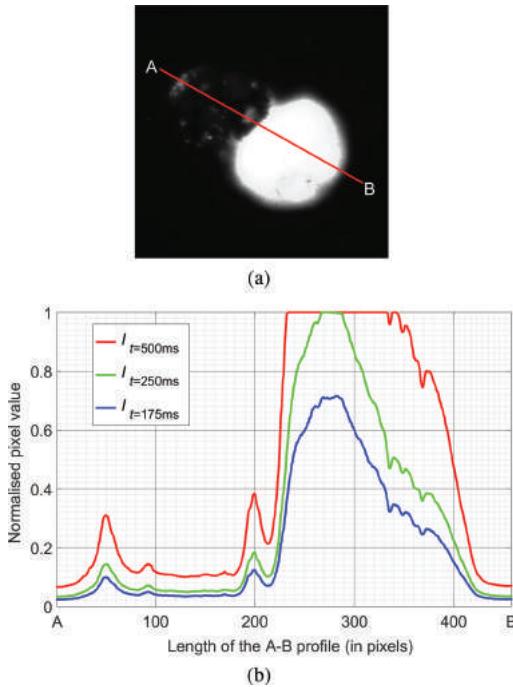


Figure 2. Profiles A-B of images 1(a)–(c) i.e. for DAPI channel: (a) cell image with marked line segment AB (red) for which the profiles have been calculated, (b) image profiles in normalized range of pixel values.

## 2 DESCRIPTION OF THE METHOD

In the proposed method the HDR image  $\mathbf{I}$  is assembled from two or more LDR images  $I_i \in \mathcal{I}$ , where  $\mathcal{I} = \{I_1, I_2, \dots, I_i, \dots, I_{i_{\max}}\}$ . The index  $i$  represents exposure time  $t_i$  of the image  $I_i$ , and images  $I_i \in \mathcal{I}$  are sorted in the descending order of its exposure time i.e.  $t_i > t_{i+1}$ . The images  $I_i$  and  $I_j$  are differ only in the exposure times  $t_i$  and  $t_j$  used for their acquisition.

All images  $I_i \in \mathcal{I}$ —except the last image  $I_{i_{\max}}$  i.e. image acquires with shortest exposure time—contain areas of overexposure (signal saturation). It was assumed that overexposure area  $\omega_i$  in image  $I_i$  satisfy following condition

$$\omega_i = \{p_i(x, y) : I_i(x, y) > \tau\}, \quad (1)$$

where  $(x, y)$  are the coordinates of the pixel,  $I_i(x, y)$  is the value of the pixel  $p_i(x, y)$  in the image  $I_i$ , and  $\tau$  is a threshold parameter. For a normalised range of image values, i.e.  $I_i(x, y) \in [0, 1]$ , the value of  $\tau$  is close to 1.

The image  $\mathbf{I}$  is created in an iterative process. Let  $\mathbf{I}_i$  is the HDR image after  $i$ th iteration, hence  $\mathbf{I} = \mathbf{I}_{i_{\max}}$ . The whole process starts form the image obtained for the longest exposure time  $t_1$  i.e.

$$\mathbf{I}_1 = I_1. \quad (2)$$

The image  $\mathbf{I}_i$  for  $i > 1$  is calculated as follows

$$\mathbf{I}_i(\omega_{i-1}) = \alpha_i \mathbf{I}_i(\omega_{i-1}) - \tau + \Delta_i, \quad (3)$$

where

$$\alpha_i = \frac{t_{i-1}}{t_i} \quad (4)$$

and

$$\Delta_i = \frac{1}{\|\mathcal{C}_{i-1}\|} \sum_{p_{i-1} \in \mathcal{C}_{i-1}} \mathbf{I}_{i-1}(x, y). \quad (5)$$

The  $\mathcal{C}(\omega_j)$  is an inner contour of region  $\omega_j$  and it is defined as follows

$$\mathcal{C}_j = \{p_j(x, y) : p_j(x, y) \in \omega_j \setminus \in (\omega_j, m)\}, \quad (6)$$

where  $\in (\omega_j, m)$  is a result of morphological erosion of region  $\omega_j$  with structuring element  $m$ , which in this case is defined as  $3 \times 3$  array of 1.

In the proposed method of HDR image creation the LDR images  $I_i$  used in this process have to be linear in the sense of values  $I_i(x, y)$ , i.e. image brightness, which are the camera response on the amount of light reaching each pixel of the sensor (sensor irradiance)  $E_i(x, y)$ . However, in case of the real camera, such response is non linear. This camera response can be described by the *Camera Response Function* (CRF) (Grossberg and Nayar 2003), and this function can be used for linearisation of non linear images obtained from a camera.

Let  $\tilde{I}_i(x, y)$  represents image value for pixel  $p_i(x, y)$  obtained directly from the camera after the acquisition process with exposure time  $t_i$ . The image  $I$ , which is linearised version of  $\tilde{I}$ , is calculated as follows

$$I(x, y) = \text{CRF}^{-1}(\tilde{I}(x, y)), \quad (7)$$

where  $\text{CRF}^{-1}$  is the inverse function of CRF, which is calculated for the camera used to capture images  $\tilde{I}$ . The images after linearisation, i.e.  $I$ , are the input data to the process of creating HDR image.

### 3 EXPERIMENT AND ITS RESULTS

The proposed method was tested on imaging of human cells specimen stained with two fluorescent stains i.e. DAPI and FITC. For image acquisition the Olympus cell<sup>R</sup> excellence system based on Olympus IX81 inverse fluorescent microscope with a Hamamatsu Orca-R2 camera was used (Fig. 3). The image acquisition was performed for each imaged cell in two channels, i.e. independently for both fluorescent dyes, and for each channel appropriate configuration (i.e. the dichroic mirror, excitation and emission filters) of the microscope was used. For each cell two series of LDR images, one in each channel, were acquired. The differences

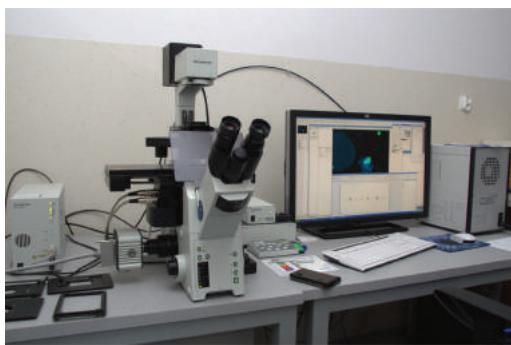


Figure 3. The Olympus cell<sup>R</sup> excellence microscope system with Hamamatsu Orca-R2 camera which was used in the experiment.

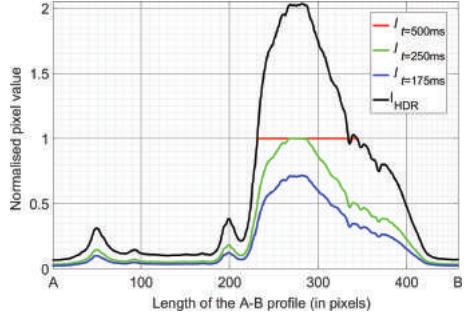


Figure 4. Comparison of HDR and LDR image profiles A-B (Fig. 7(a)) for DAPI channel: black line—profile of HDR image, coloured lines—profiles of LDR images obtained for different exposure times; pixel values are shown in normalized range, where 1 is the maximum pixel value in LDR image.

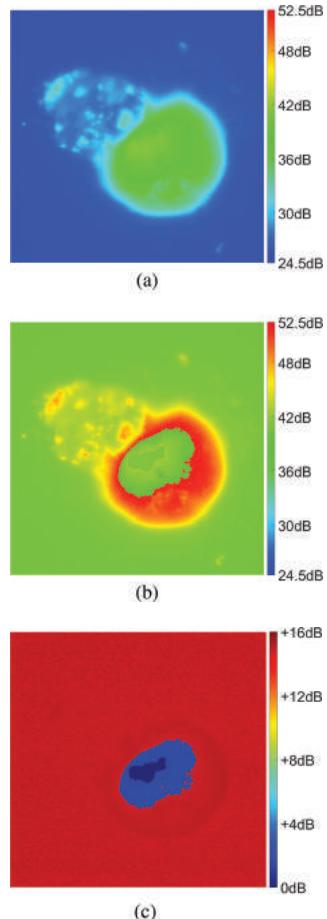


Figure 5. Maps of SNR for LDR and HDR image in DAPI channel: (a) LDR image acquired with exposure time  $t = 175$  ms, (b) HDR image; (c) difference  $\Delta\text{SNR}$  for images (a) and (b).

between images  $I_i$  in each series  $\mathcal{I}$  are only results of difference in their exposition time  $t_i$ .

The exemplary series of LDR images obtained for the same cell are presented: for DAPI channel in the Figure 1 and for FITC channel in the Figure 6. In the both figures the places in each image where the signal is saturated is marked in red.

In Figures 4 and 7 the profiles of the obtained HDR images are presented. For creation of HDR images for both channels the  $\tau = 0.975$  was used.

Additionally, in Figures 5 and 8 comparison between theoretical Signal-to-Noise Ratio (SNR) of LDR and HDR images are presented. The SNR for LDR images is calculated as follows

$$SNR_{LDR_i}(x,y) = 20 \log_{10} \frac{\psi I_i(x,y)}{\sqrt{\psi I_i(x,y) + \eta_r^2 + \eta_d^2}}, \quad (8)$$

where  $\psi$  is a full well capacity of the used camera,  $\eta_r$  is a readout noise and  $\eta_d$  is a dark noise. In case of the used camera, according to the manufacturer data (Hamamatsu Photonics):  $\psi = 18000$  e,  $\eta_r = 5.5$  e

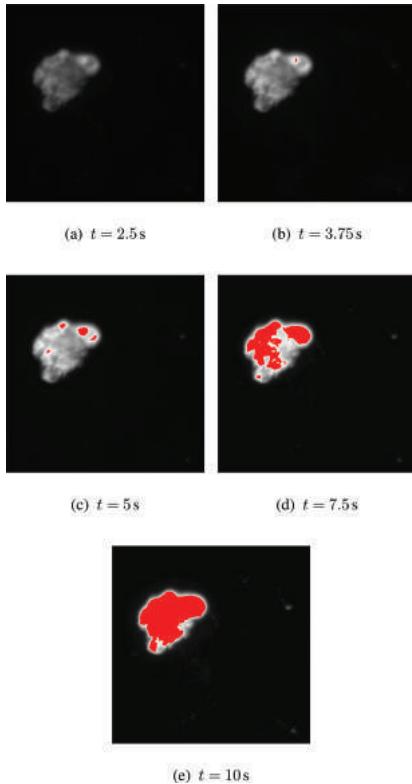
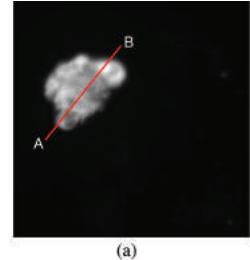
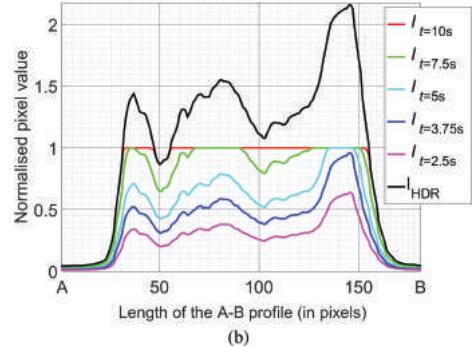


Figure 6. Images of the same cell nuclei acquired in FITC channel with different exposure times  $t$ ; the signal saturation ( $\tau > 0.975$ ) is indicated on red.



(a)



(b)

Figure 7. Comparison of image profiles for FITC channel: (a) cell image in FITC channel with marked line segment AB (red), (b) profiles A–B: black line—profile of HDR image, coloured lines—profiles of LDR images obtained for different exposure times; pixel values are shown in normalized range (1 is the maximum pixel value in LDR image).

and  $\eta_d = 5 \cdot 10^{-4}$  e/pixel/s  $\cdot t_i$ . Because maximal exposition time ( $t_i$ ) used during experiment  $\max(t_i) = 10$  s the value  $\eta_d \ll \sqrt{\psi I_i(x,y)}$  and also  $\eta_d \ll \eta_r$  therefore the value of  $\eta_d$  has not been taken into account in the calculation of SNR.

For calculation of SNR for HDR image ( $SNR_{HDR}$ ) crucial is that adding a constant to the value  $I_i(x, y)$  and multiplying the value  $I_i(x, y)$  by a constant does not change the value of  $SNR_{LDR_i}(x, y)$ . Therefore

$$SNR_{HDR}(x,y) = SNR_{LDR_i}(x,y)|_{I_i(x,y)=f(I_i(x,y))}, \quad (9)$$

where  $I_i(x,y) = f(I_i(x,y))$  means that the value of pixel  $p(x, y)$  of HDR image is obtained from the value of LDR image  $I_i(x, y)$  according to the equation (3).

Additionally, in Figures 5 and 8 the difference  $\Delta SNR$  between SNR for HDR image and SNR of the best LDR image for appropriate channel is presented (Fig. 5(c) and Fig. 8(c));  $\Delta SNR$  is defined as follows

$$\Delta SNR(x,y) = SNR_{HDR}(x,y) - SNR_{LDR}(x,y). \quad (10)$$

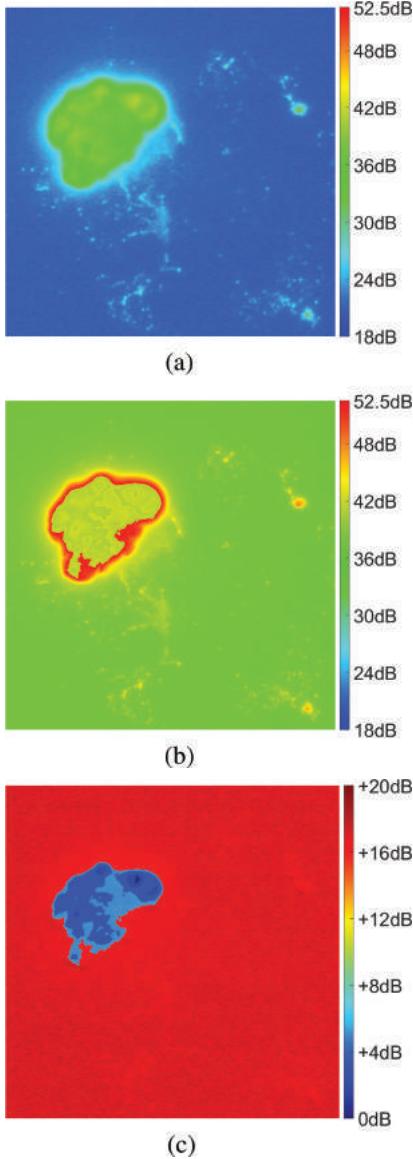


Figure 8. Maps of SNR for LDR and HDR image in FITC channel: (a) LDR image acquired with exposure time  $t = 2.5$  s, (b) HDR image; (c) difference  $\Delta\text{SNR}$  for images (a) and (b).

#### 4 CONCLUSIONS

The obtained result shows that proposed method can be useful for image acquisition in applications in which not only qualitative but also quantitative analysis have to be performed. The method of creating HDR images presented in this work is simple and effective. The obtained HDR image

contains the whole information about signals of the analysed specimen what is very important for obtaining proper results of analysis. There is no overexposure in the HDR images, therefore the information about strong signal is not loose, and the weak signals are also maintained.

A good example of this is the Figure 4 where small changes of the profile height at about 50 and 200 pixels of the A–B profile are maintained together with maintaining of the very strong signal at about 280 pixel of the profile. As can be seen on the profiles of LDR images presented in this figure, the LDR image can not maintain whole information about those two kind of signals. What is important, because of the linearisation of the images  $\tilde{I}_f$ , the pixel values of HDR image are proportional to the signal—in case of LDR images  $\tilde{I}_i$  this relationship is not true.

The proposed method can be used not only for fluorescence microscopy imaging, but also for other imaging methods used in biology and medicine. An a good example of such new application area of the proposed method are e.g. methods based on the gel electrophoresis (Corley 2005) like western blot technique (Mathews et al. 2009) or 2D gel electrophoresis (Berth et al. 2007).

As it is presented in Figures 5 and 8 the SNR values of the HDR images for both channels are much better than the SNR values of the corresponding LDR images without signal saturation. What is important SNR for HDR images is much bigger, than SNR of LDR images, for weak signals.

Of course with the creation of HDR images a lot of problems are associated. The first of them is the choice of the  $\tau$  parameter value. Differences between HDR images obtained for different values of  $\tau$  can be expressed as

$$\Delta\mathbf{I}_\tau(x, y) = \frac{\mathbf{I}_\tau(x, y) - \bar{\mathbf{I}}(x, y)}{\bar{\mathbf{I}}(x, y)}, \quad (11)$$

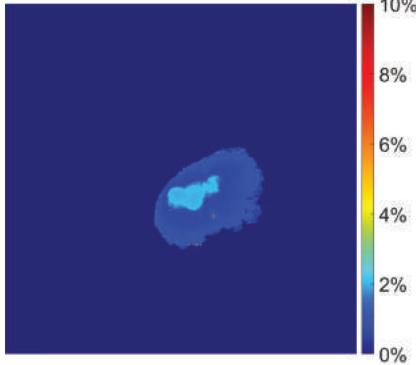
where

$$\bar{\mathbf{I}}(x, y) = \frac{1}{\|\mathcal{T}\|} \sum_{\tau \in \mathcal{T}} \mathbf{I}_\tau(x, y) \cdot 100\%, \quad (12)$$

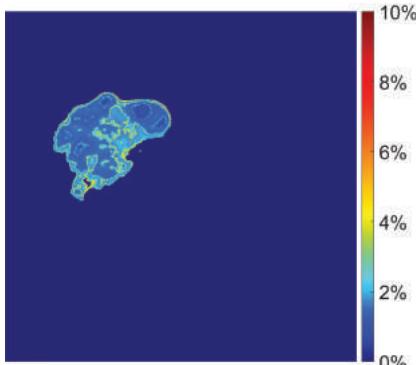
and  $\mathcal{T} = \{\tau\}$  is a set of tested  $\tau$  values. The Figure 9 presents in pseudocolour maximum absolute values of  $\Delta\mathbf{I}_\tau(x, y)$ , i.e.

$$\widehat{\Delta\mathbf{I}}(x, y) = \max_{\tau \in \mathcal{T}} (\|\Delta\mathbf{I}_\tau(x, y)\|), \quad (13)$$

for DAPI and FITC channels obtained for  $\mathcal{T} = \{0.9, 0.95, 0.975, 0.99\}$ . As it can be seen, in



(a)  $\widehat{\Delta I}(x, y)$  for DAPI channel



(b)  $\widehat{\Delta I}(x, y)$  for FTIC channel

Figure 9. The  $\widehat{\Delta I}$  values for  $T = \{0.9, 0.95, 0.975, 0.99\}$ .

both channels, for most pixels  $\widehat{\Delta I}(x, y) \leq 2.5\%$ . However, for a larger differences of between used  $\tau$  values the  $\widehat{\Delta I}(x, y)$  value is bigger, e.g. for  $T = \{0.75, 0.8, 0.85, 0.9, 0.95, 0.975, 0.99\}$  for most pixels  $\widehat{\Delta I}(x, y) \leq 5\%$ . In both cases, the error associated with the creation of HDR images is acceptable from the viewpoint of the biological purpose of the specimen analysis. However, the reason of such large relative change of  $\widehat{\Delta I}(x, y)$  value requires an explanation during further researches.

Another interesting problem connected with HDR image creation is the choice of exposure times  $t$  for acquisition of LDR images. The choice of  $t_i$  influences the quality of the obtained HDR images. Usually the usage of more LDR

images obtained for different exposure times means better HDR image. However, because of e.g. photobleaching of fluorescent stain the total exposure time

$$T(x, y) = \sum_{i=1}^{i_{\max}} t_i(x, y), \quad (14)$$

which is independently calculated for each place of the specimen, cannot be increased without limit. This problem will be one of the aim of the future research.

There are at least two important problems which have to be investigated during future researches. One of them is the calculation of CRF. This function can be treated as a camera fixed parameter. Therefore, in practice, for each camera CRF can be determined only once, and then it can be used in each experiment. Thus, there is a need for effective and accurate methods for calculation of CRF. The second issue is to develop an objective method of assessing the accuracy of the images obtained HDR. These both problems will be the aim of the future researches.

## ACKNOWLEDGEMENTS

This work was partially supported by Polish National Science Centre (grant no. NN518 283640) and from funds for statutory activities of the Institute of Automatic Control, Silesian University of Technology.

The calculations were performed using the infrastructure supported by POIG.02.03.01-24-099/13 grant: GeCONiI—Upper Silesian Centre for Computational Science and Engineering

Presented experiment was performed in the Laboratory of Digital Image Processing and Biomedical Visualization, which works in the Biotechnology Centre of the Silesian University of Technology.

## REFERENCES

- Bell, A.A., J. Brauers, J. Kaftan, D. Meyer-Ebrecht, A. Bocking, & T. Aach (2009). High dynamic range microscopy for cytopathological cancer diagnosis. *IEEE Journal of Selected Topics in Signal Processing (Special Issue on: Digital Image Processing Techniques for Oncology)* 3(1), 170–184.
- Berth, M., F.M. Moser, M. Kolbe, & J. Bernhardt (2007). The state of the art in the analysis of two-dimensional gel electrophoresis images. *Applied Microbiology and Biotechnology* 76(6), 1223–1243.

- Corley, R.B. (2005). *A Guide to Methods in the Biomedical Sciences*. The Wiley—IS&T Series in Imaging Science and Technology. Springer US.
- Eastwood, B.S. & E.C. Childs (2012, January). Image alignment for multiple camera high dynamic range microscopy. In *Proceedings IEEEWorkshop on Applications of Computer Vision*, pp. 225–232.
- Grossberg, M.D. & S.K. Nayar (2003). Determining the camera response from images: What is knowable? *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25(11), 1455–1467.
- Hamamatsu Photonics. *ORCA-R2. Digital CCD camera C10600-10B. Nechnical note*. Hamamatsu Photonics.
- Mathews, Suresh, T., P. Plaisance, Eric, & T. Kim (2009). Imaging systems for westerns: Chemiluminescence vs. infrared detection. In B.T. Kurien and R.H. Scofield (Eds.), *Protein Blotting and Detection*, Volume 536 of *Methods in Molecular Biology*, pp. 499–513. Humana Press.
- McCann, J.J. & A. Rizzi (2012). *The Art and Science of HDR Imaging*. The Wiley—IS&T Series in Imaging Science and Technology. John Wiley & Sons, Ltd.
- Waters, J.C. (2009). Accuracy and precision in quantitative fluorescence microscopy. *Journal of Cell Biology* 185(7), 1135–1148.

This page intentionally left blank

# Automatic cheek detection in digital images

M. Frackiewicz, H. Palus & K. Radlak

*Institute of Automatic Control, Silesian University of Technology, Gliwice, Poland*

**ABSTRACT:** This paper deals with some problems of cheek detection in colour images. Authors reviewed the literature in this field and proposed a new effective method for cheek detection. A characteristic feature of the proposed method is the use of 13 essential face landmarks generated by a known face landmark detector. In the paper four methods from the literature and one's own method are described in detail. All these methods have been tested experimentally using a suitable face image database PUT. Preliminary tests showed the advantages of our method. Additionally we investigated the impact of head movements on the location detected cheeks. A small number of false cheek pixels lying outside the face contour and a small colour error caused by movements of the head predetermines the proposed method for further research and applications.

## 1 INTRODUCTION

Several applications of face image processing may require detecting facial regions such the mouth, eyes, eyebrows, nose, cheeks and chin. Automatic cheek detection is an important operation in this field. The cheek colour well represents the face colour, because the cheeks are not covered by hair as sometimes the forehead or the chin and not generate highlights as the nose area. Each cheek is a consistent patch of skin. Among the typical cheek detection applications are medical diagnostics, human face beautification as well as coding of videophone sequences.

Person's emotional state generates colour changes in its facial skin (Ramirez et al. 2014). In this work the colour changes were considered and measured in three regions of the face, i.e. the forehead and both cheeks. These regions of interest limited to small squares were determined on the basis of data from the facial feature tracker. Good results are obtained from the combination of the forehead and the left cheek. Automatic cheek detection may be particularly useful in medical diagnostics. Human health evaluation based on the colour of his/her face is used in traditional oriental medicine for thousands of years (Yang et al. 2012, Seo et al. 2014, Zhao et al. 2014). The result of the latest research is a strong statistical relationship between perceptually uniform colour space CIELAB components in specific regions of the face and pathological patterns used in oriental medicine. In the paper (Yang et al. 2012) both cheek skin boxes are detected after the previous eye and mouth detection. Similarly, in the work

(Su et al. 2013) both cheeks are detected after the previous using the Viola-Jones algorithm for eye and mouth detection. Based on the colours of cheek pixels the haemoglobin and melanin components are separated, making it easier to assess the health of a person. Sometimes the colours of the face in selected locations (five blocks each of  $32 \times 32$  pixels) are determined and on the basis of these colours such liver diseases as hepatitis are diagnosed (Liu and Guo 2007). Two of these blocks are located on the surface of both cheeks through finding another facial features. Some other application of this approach is described in the paper (Kim et al. 2008), that presents an idea to use the colours of the forehead centre and the lips for cardiovascular disease detection.

An original application of the cheek detection is used to beautify the human face in photographic portraits (Lai et al. 2013). Beautification is performed by the total removal or the magnitude reduction of unwanted skin details: scars, pores, blemishes, wrinkles, etc. The developed skin changing system can, on the one of the early processing stages, automatically extract the cheeks by using the active shape modelling and the watershed algorithm.

Automatic cheek detection was also applied in the field of model-based coding of videophone sequences (Hess and Martinez 2004). The human face in such sequence can be segmented into few facial regions such as e.g. cheeks. The cheek borders were detected based on the well-known SUSAN algorithm (Smith and Brady 1997) and additional knowledge-based rules.

The main goal of this study is to compare the automatic cheek detection methods existing in the

literature with a new method proposed in this paper. This article is organized as follows. In Section 2 we shortly present four known methods and one our proposal for automatic cheek detection. The experimental results for colour face images from PUT face database are included in Section 3. Finally, the Section 4 concludes the paper.

## 2 CHEEK DETECTION METHODS

In the literature we searched the cheek detection methods which determine the cheeks by means the primary and secondary facial landmarks. There are methods based on a small number of face landmarks, e.g. 27 landmarks (Campadelli et al. 2007) and a large number of landmarks e.g. 192 landmarks (Kazemi and Sullivan 2014). Unfortunately, for our problem was found in the literature only a few solutions.

The first known method (Xian 2011) is based on the coordinates of eye centres:  $L_{eye} = (x_l, y_l)$ ,  $R_{eye} = (x_r, y_r)$  and coordinates of mouth centre:  $M = (x_m, y_m)$ . This method assumes that the face is oriented vertically:  $x_l = x_r = x_{lr}$ . The horizontal distance between the eye centres:  $B = |y_r - y_l|$  and the vertical distance between the eye centres and mouth centre:  $A = |x_m - x_{lr}|$  are calculated. Detected cheek regions are rectangles with following coordinates of the left upper corners and following lengths of sides:

$$L_{cheek} = [(x_{lr} + A/4, y_l); A/2; B/2] \quad (1)$$

$$R_{cheek} = [(x_{lr} + A/4, y_r - B/2); A/2; B/2] \quad (2)$$

The second method described in (Yang et al. 2012) determines the cheek centres:  $(x_{cl}, y_{cl})$ ,  $(x_{cr}, y_{cr})$ , where  $x_{cl} = x_{cr} = x_{clr}$ . Unfortunately, the size of the square areas around of these cheek centres has not been defined. The authors of paper present three ways of determining the cheek centres depending on the accuracy with which are known the eye and the mouth positions. In our implementation of this method assumes that the cheek regions are squares with defined centres and side lengths equal:  $B/2$ . Generator of landmarks fairly accurately locates the eyes and the mouth. Therefore, for the determination of cheek centres we adopted following formulas:

$$x_{clr} = x_{cr} = x_{clr} = (mouth_x + eye_x)/2 \quad (3)$$

$$y_{cl} = mouth_{yl} + mouth_D/3 \quad (4)$$

$$y_{cr} = mouth_{yr} - mouth_D/3 \quad (5)$$

where:  $mouth_D = |mouth_{yl} - mouth_{yr}|$ —distance between the mouth corners,  $B = |y_r - y_l|$ —distance

between the eye centres,  $mouth_x$ —mean value of  $x$  coordinates of mouth corners,  $eye_x$ —mean value of  $x$  coordinates of eye centres.

The third known method (Su et al. 2013) generates three bounding boxes: two located on the eyes and one located on the mouth. Detected region representing the left cheek is above the mouth box and reaches the left eye box and has its width. The right cheek region is also above the mouth box and reaches the right eye box preserving its width.

The fourth method described in (Hess and Martinez 2004) uses for each cheek detection the landmarks located in the mouth corners, on the lower edge of eyes and six landmarks on the face contour. Cheek regions are defined by the segments that connect these facial landmarks. The cheek region is limited from the face centre by a vertical line passing through the mouth corner, from the top by a line passing through the lower edge of the eyes and from the bottom by a line passing through a contact place between the head and shoulders.

The new cheek detection method, proposed by the authors, is based on selected facial landmarks. We decided to use open-source Davis King's Dlib library (King 2009, King 2015a), in which the faces are detected using Histograms of Oriented Gradient (HOG) descriptors (Dalal and Triggs 2005) trained with structural SVM based training algorithm (King 2015b) and facial landmarks are detected using the method proposed in (Kazemi and Sullivan 2014). This method generates 68 landmarks in the well-established landmark configuration of Multi-PIE (Gross et al. 2010) presented in Figure 1.

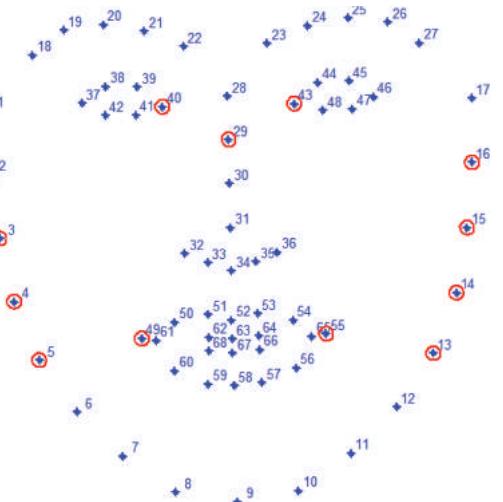


Figure 1. The 68 Multi-PIE landmarks scheme and the landmarks selected for our method marked by the circles.

In our work we additionally considered also Zhu and Ramanan face detection and landmarks localization method (Zhu and Ramanan 2012). The performance of both methods were evaluated using an error measure introduced in (Cristinacce and Cootes 2006) on a subset of 2193 near-frontal images from PUT face database for which a contour model were manually annotated. This measure is the mean value of distance between the landmarks detected by tested algorithm and manually marked points divided by the inter-ocular distance (measured as the Euclidean distance between the outer corners of the eyes). The comparison of these algorithms is presented in Figure 2. As can be observed the facial landmarks localization proposed in (Kazemi and Sullivan 2014) and implemented in the Dlib library outperforms the Zhu Ramanan method (Zhu and Ramanan 2012) and was chosen for later tests. Exemplary results of landmarks detected on PUT database and manually marked landmarks are presented in Figure 3.

Among so many landmarks our method utilizes thirteen selected points (eight points on the face contour, two mouth corners, two inner eye corners and one nose landmark). The border of the right cheek region is created by a connection of the following four landmarks (2, 3, 4 and 5) from the right face contour and is located between the ear and the chin. Last fourth landmark (pt.5) is connected to the right corner of the mouth (pt.49). Then, from the right corner of the mouth a line is passing in the direction of inner corner of the right eye (pt.40). Another line is run from the starting point on the face contour (pt.2) in the direction of the landmark located on the nose (pt.29). Both lines are running until they intersect. The result is a closed region of the cheek. The left

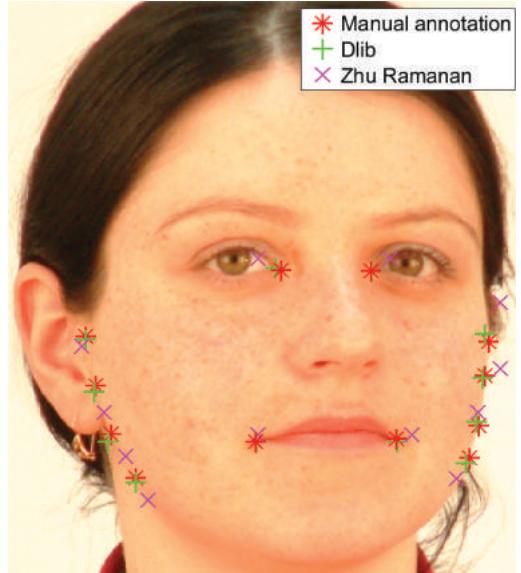


Figure 3. Exemplary results of landmarks detection and comparison with ground truth.

cheek is detected in a similar manner. In this case the landmarks from left face contour, left corner of the mouth and the inner corner of the left eye, are used. The same nose landmark (pt.29) is used for detection of both cheeks.

The described methods use different facial landmarks. The first three methods apply the horizontal and vertical lines of the mouth and eyes only and cheek regions are rectangles. It is known that rectangles inaccurately describe the cheek regions. The fourth method additionally uses facial contour points and the above mentioned lines, which causes an unnatural shape of cheeks. The proposed method is based on facial contour points and does not use those lines, which creates an opportunity for more detailed description of the shape of the cheek.

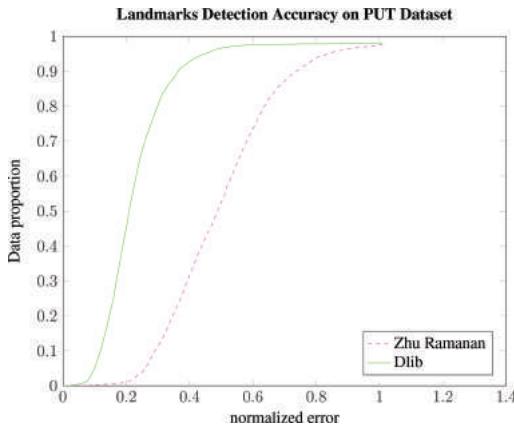


Figure 2. Cumulative error curves on the PUT database using landmarks chosen for cheek detection.

### 3 EXPERIMENTAL TESTS

The four described above methods and one new method proposed by the authors have been implemented. The detected cheek regions, as results of initial tests, are shown in the image of hairdresser styling head (Fig. 4).

In our further tests the PUT face database (Kasinski et al. 2008) was applied because of its advantages: high resolution ( $2048 \times 1536$ ) images of 100 people realized in partially controlled illumination on a uniform background.

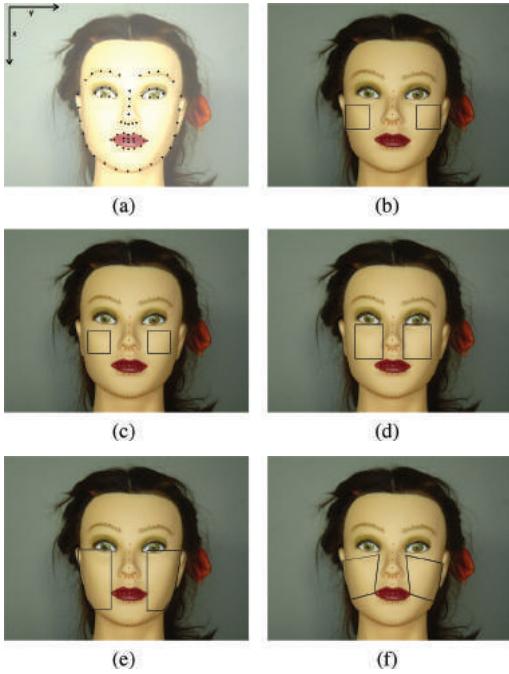


Figure 4. Detected cheeks: (a) 68 face landmarks used for cheek detection, (b) first method, (c) second method, (d) third method, (e) fourth method, (f) our method.

The first two cheek detection methods, despite their simplicity, designate the cheek areas lying partly outside the face contour (Fig. 5c) and (Fig. 6c). These methods are very sensitive to small lateral turns of the head. As a result of their action are found small, central areas of the cheeks.

The third method (Fig. 7) determines the greater cheek region than two previous methods, but often adds to the detected cheek area the pixels of nasal region. This method is also sensitive to small lateral turns of the head.

The fourth method (Fig. 8) allows detect largest part of the cheek. Detected area does not go beyond the face contour, but covers the lower lashes and an area under the eyes.

The area found by the authors method (Fig. 9) does not go beyond the face contour and safely bypasses the nasal region and an area under the eyes. In this case the detected cheek area is slightly smaller than the area found by the fourth method.

PUT image database, used in this study, includes also the face images acquired after different movements of the head: to the left, right, up, down and four diagonal movements (left-up, right-up, left-down, right-down). Figure 10 shows these eight images for the first tested person. At the centre

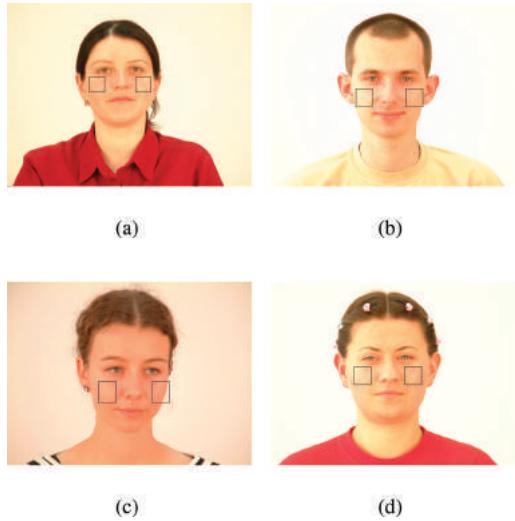


Figure 5. Cheeks detected by the first method.

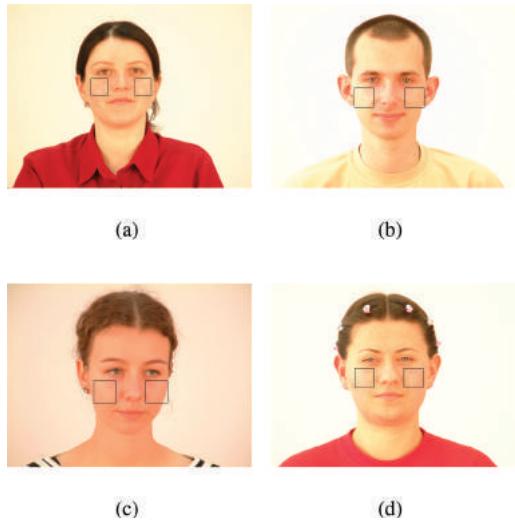
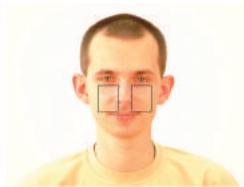


Figure 6. Cheeks detected by the second method.

of the figure is placed a face image of the head acquired in a central position. Each image contains both cheek regions marked with black line borders and found with the help of the proposed method. All images show the correct detection of cheeks. Detection results are not always as positive, especially for other methods and other persons. Such examples are shown in Figure 8, where on the magnified faces the regions belonging to the cheeks and at the same time extending beyond the face contour, have been manually marked. Pixels in these regions can be specified as false detected



(a)



(b)



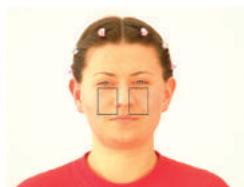
(a)



(b)



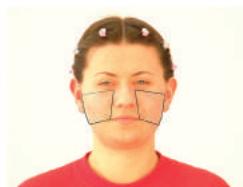
(c)



(d)



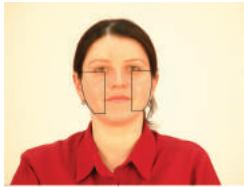
(c)



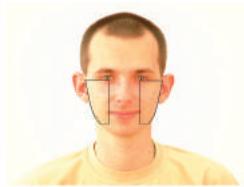
(d)

Figure 7. Cheeks detected by the third method.

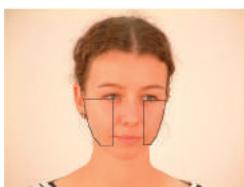
Figure 9. Cheeks detected by our method.



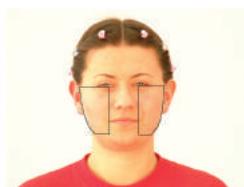
(a)



(b)



(c)



(d)

Figure 8. Cheeks detected by the fourth method.

cheek pixels. These pixels were coloured with a contrasting colour (e.g. cyan), which allows easily and automatically to count them.

Further studies were carried out for four persons and for each of them have been used the central image and eight images created as the results of head movements. Then, using five cheek detection methods for each image, were determined the cheek regions. For each person and each method was calculated an average percentage of false classified pixels in the areas designated as cheeks. The results of this experiment were recorded in Table 1.



Figure 10. Cheeks detected by our method at different head poses.

The average results computed for all persons and placed in the last row of the table, well describe the quality of each method. Methods 1–2 carry the largest errors, while the methods 4–5 characterize by the smallest errors, especially the proposed method. In this paper we did not consider a classification of pixels, belonging to the other parts of the face such as the mouth, nose and the ears, as cheek pixels.

Another method of assessing impact of head movements on the tested cheek detection methods is a comparison of average cheek colour in each of the eight images resulting from the movement of the head with the average cheek colour in the image acquired in the central position. For each person and each method of cheek detection we

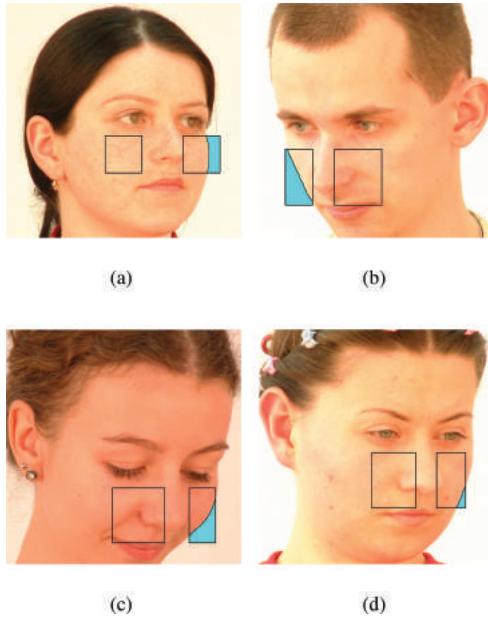


Figure 11. Examples of detected false cheek pixels.

Table 1. Average percentage of false cheek pixels.

	Meth. 1	Meth. 2	Meth. 3	Meth. 4	Meth. 5
Person 1	11.91	12.52	0.34	0.08	0.00
Person 2	17.54	17.51	3.68	0.00	0.00
Person 3	17.56	15.95	3.66	0.79	0.75
Person 4	4.13	6.43	0.25	0.02	0.00
AVG	<b>12.79</b>	<b>13.10</b>	<b>1.98</b>	<b>0.22</b>	<b>0.19</b>

Table 2. Average colour errors of cheek regions.

	Meth. 1	Meth. 2	Meth. 3	Meth. 4	Meth. 5
Person 1	6.23	6.23	3.52	4.12	4.12
Person 2	10.58	15.14	6.42	4.41	4.30
Person 3	7.18	12.14	5.82	3.98	4.02
Person 4	5.75	8.89	4.01	3.29	2.95
AVG	<b>7.44</b>	<b>10.60</b>	<b>4.94</b>	<b>3.95</b>	<b>3.85</b>

calculated an average colour error  $\Delta E_{76}$  using a standard distance formula in the perceptually uniform CIELAB colour space (Fairchild 1998):

$$\Delta E_{76} = \sqrt{(L_0 - L_i)^2 + (a_0 - a_i)^2 + (b_0 - b_i)^2} \quad (6)$$

where  $L_0, a_0, b_0$  are the components of average cheek colour for central positioned image and  $L_i, a_i, b_i$ —components of average cheek colour for

one of eight test images acquired after movement of the head.

The obtained results of the colour errors for the left and right cheeks were averaged. Table 2 shows the results of this experiment and, as in the case of Table 1, shows also the advantages of the proposed method over other methods.

#### 4 CONCLUSION

We tested several methods found in literature for automatic cheek detection using the PUT face database. Each of these methods based on previously established face landmarks shown some disadvantages. We have proposed a new method for cheek detection where the above mentioned disadvantages are minimized while almost all cheek regions are correctly detected. Based on visual and quantitative assessment, we find that our method accurately models the shape of the cheek and generally not extend it to other facial features. Automatic cheek detection will allow to evaluate the colour changes caused by different emotional states.

#### ACKNOWLEDGEMENTS

This work was supported by the Polish Ministry for Science and Higher Education under internal grant BK-227/RAu1/2015 for Institute of Automatic Control, Silesian University of Technology, Gliwice, Poland.

#### REFERENCES

- Campadelli, P., G. Lipori, & R. Lanzarotti (2007). *Automatic facial feature extraction for face recognition*. INTECH Open Access Publisher.
- Cristinacce, D. & T.F. Cootes (2006). Feature detection and tracking with constrained local models. In *Proc. BMVC*, pp. 95.1–95.10. doi:10.5244/C.20.95.
- Dalal, N. & B. Triggs (2005). Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 1, pp. 886–893.
- Fairchild, M. (1998). *Color appearance models*. Addison-Wesley.
- Gross, R., I. Matthews, J. Cohn, T. Kanade, & S. Baker (2010). Multi-pie. *Image and Vision Computing* 28(5), 807–813.
- Hess, M. & G. Martinez (2004). Facial feature extraction based on the smallest univalue segment assimilating nucleus (susan) algorithm. In *Proceedings of Picture Coding Symposium*, Volume 1, pp. 261–266.
- Kasinski, A., A. Florek, & A. Schmidt (2008). The PUT face database. *Image Processing and Communications* 13(3–4), 59–64.

- Kazemi, V. & J. Sullivan (2014). One millisecond face alignment with an ensemble of regression trees. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1867–1874.
- Kim, B.-H., S.-H. Lee, D.-U. Cho, & S.-Y. Oh (2008). A proposal of heart diseases diagnosis method using analysis of face color. In *International Conference on Advanced Language Processing and Web Information Technology*, pp. 220–225.
- King, D.E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research* 10, 1755–1758.
- King, D.E. (2015a). Dlib. <http://dlib.net>. Accessed on March, 2015.
- King, D.E. (2015b). Max-margin object detection. *CoRR abs/1502.00046*.
- Lai, Y.-R., C.-Y. Yao, H.-S. Hu, M.-T. Chi, & Y.-C. Lai (2013). A facial skin changing system. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 1–7.
- Liu, M. & Z. Guo (2007). Hepatitis diagnosis using facial color image. In *Medical Biometrics*, pp. 160–167. Springer.
- Ramirez, G.A., O. Fuentes, S.L. Crites, M. Jimenez, & J. Ordonez (2014). Color analysis of facial skin: Detection of emotional state. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 474–479.
- Seo, J.-H., Y.-B. Park, & Y.-J. Park (2014). Reliable facial color analysis using a digital camera and its relationship with pathological patterns: A pilot study. *European Journal of Integrative Medicine* 6(3), 322–327.
- Smith, S.M. & J.M. Brady (1997). SUSAN—a new approach to low level image processing. *International Journal of Computer Vision* 23(1), 45–78.
- Su, R., T.M. Dockins, & M. Huber (2013). ICA analysis of face color for health applications. In *The Twenty-Sixth International FLAIRS Conference*.
- Xian, W. (2011). Face region recognizer. B.E. Thesis, Malaysia University of Technology, Johor, Malaysia.
- Yang, Y., J. Zhang, L. Zhuo, Y. Cai, & X. Zhang (2012). Cheek region extraction method for face diagnosis of traditional chinese medicine. In *IEEE 11th International Conference on Signal Processing (ICSP)*, Volume 3, pp. 1663–1667.
- Zhao, C., G.-z. Li, F. Li, Z. Wang, & C. Liu (2014). Qualitative and quantitative analysis for facial complexion in traditional Chinese medicine. *BioMed Research International* 2014.
- Zhu, X. & D. Ramanan (2012). Face detection, pose estimation, and landmark localization in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2879–2886.

This page intentionally left blank

# A variational model for image fusion with simultaneous cartoon and texture decomposition\*

Mahdi Dodangeh & Isabel N. Figueiredo

*Department of Mathematics, CMUC, University of Coimbra, Coimbra, Portugal*

Gil Gonçalves

*INESC—Coimbra, Portugal*

**ABSTRACT:** Image fusion is a technique that merges the information of multiple images, representing the same scene, to produce a single image that should gather the major and meaningful information contained in the different images. On the other hand, cartoon+texture image decomposition is another image processing technique, that decomposes the image into the sum of a cartoon image, containing the major geometric information, *i.e.*, piece-wise smooth regions, and a textural image, containing the small details and oscillating patterns. We propose a model to perform the fusion of multiple images, relying on gradient information, that provides as well a cartoon and texture decomposition of the fused image. The model is formulated as a variational minimization problem and solved with the split Bregman method. The suitability of the proposed model is illustrated with some earth observation images and also medical images.

## 1 INTRODUCTION

Image fusion is an image processing tool, that generates a single composite image from multi-source images, of the same scene. It is particularly important in remote sensing (it makes possible, for example, to combine the information of earth observation images acquired by different sensors or satellites, for improving earth surveillance and monitoring) and in medicine (the fused image obtained from different medical imaging modalities may improve diagnosis and prognosis).

In this paper we propose a variational model for image fusion (we refer to (Fang, Li, Zhang, & Shen 2013, Piella 2009) for two related variational models) and that performs simultaneously a cartoon + texture decomposition of the fused image.

The proposed model fuses the gradients of the input images and combines this fusion with perceptual enhancement and intensity correction. At the

same time it also does the decomposition of the fused image into cartoon and textural parts, adopting the generalized denoising model of (Bresson, Esedoglu, Van derghenst, Thiran, & Osher 2007), based on total variation regularization. Finally we apply the split Bregman method (Goldstein & Osher 2009) to solve the model.

After this introduction, the rest of the paper is organized in three sections: Section 2 describes the model, Section 3 explains the application of the split Bregman method and finally in Section 4 some experimental results are shown. The paper ends with some conclusions and future work.

## 2 PROPOSED MODEL

Let  $I_i: \Omega \rightarrow [0,1]$ ,  $i=1, \dots, N$ , be (normalized) gray-scale and multisource images, where  $\Omega \subset R^2$  represents the pixel domain, and for a pixel  $x \in \Omega$   $I_i(x)$  is the intensity of  $I_i$  at  $x$ . We define  $g := \sum_{i=1}^N w_i \nabla I_i$ , a weighted average of the gradients of  $I_i$  with weights defined by  $w_i := |\nabla I_i| / \sum_{j=1}^N |\nabla I_j|$ , where  $|\cdot|$  is the Euclidean norm and  $\nabla$  the gradient operator. We also define  $I_0 := \sum_{i=1}^N z_i I_i$ , a weighted average of the input images  $I_i$  (a possible choice for these weights is  $z_i := w_i$  or  $z_i := |I_i| / \sum_{j=1}^N |I_j|$ ).

In (Fang, Li, Zhang, & Shen 2013) the fusion model corresponds to the following minimization problem

---

\*This work was partially supported by the research project PTDC/MATNAN/0593/2012 of FCT (the Portuguese national funding agency for science, research and technology) and by CMUC (the Centre for Mathematics of the University of Coimbra—UID/MAT/00324/2013), funded by the Portuguese Government through FCT/MEC and co-funded by the European Regional Development Fund through the Partnership Agreement PT2020.

$$\begin{aligned} & \min_I \int_{\Omega} \left[ |\nabla I - g| + \frac{\eta}{2} \left| I - \frac{1}{2} \right|^2 + \frac{\mu}{2} |I - I_0|^2 \right] dx \\ &= \min_I \left[ \|\nabla I - g\|_{L^1(\Omega)} + \frac{\eta}{2} \|I - \frac{1}{2}\|_{L^2(\Omega)}^2 \right. \\ &\quad \left. + \frac{\mu}{2} \|I - I_0\|_{L^2(\Omega)}^2 \right], \end{aligned}$$

where  $L^1(\Omega)$  and  $L^2(\Omega)$  are the spaces of, respectively, absolutely and square integrable functions in  $\Omega$ , with norms denoted by  $\|\cdot\|_{L^1(\Omega)}$  and  $\|\cdot\|_{L^2(\Omega)}$ , respectively. In this model, the fused image  $I$  is searched in such a way, that its gradient matches the combined gradient  $g$  (the function  $g$  merges the gradient information of all the different input images  $I_i$ ). In addition the linear combination of fitting terms  $\frac{\eta}{2} \|I - \frac{1}{2}\|_{L^2(\Omega)}^2 + \frac{\mu}{2} \|I - I_0\|_{L^2(\Omega)}^2$  (as proposed in (Bertalmio, Caselles, Provenzi, & Rizzi 2007)) makes the fused image  $I$  to be perceptually more uniform and close to the average  $I_0$  (the weighted sum of the several input images);  $\eta$  and  $\mu$  are two positive parameters that balance the influence of each term.

On the other hand, in (Bresson, Esedoglu, Vandergheynst, Thiran, & Osher 2007) the following model for image cartoon+texture decomposition is defined

$$\begin{aligned} & \min_{(u,v)} \int_{\Omega} \left[ h(x) |\nabla u| + \frac{1}{2\theta} |u+v-f|^2 + \alpha |v| \right] dx \\ &= \min_{(u,v)} \left[ TV_h(u) + \frac{1}{2\theta} \|u+v-f\|_{L^2(\Omega)}^2 + \alpha \|v\|_{L^1(\Omega)} \right] \end{aligned}$$

Here  $f : \Omega \subset R^2 \rightarrow R$  is the input (grayscale) image and the goal is to find a decomposition  $u+v$  of  $f$ , that is  $f \sim u+v$ , where the image  $u$  represents the geometric information of  $f$  (*i.e.*, the cartoon or piecewise smooth regions) and the image  $v$  captures the texture information. The function  $h$  is an edge detector, that vanishes at object boundaries (such as  $h(|\nabla f|) = 1/(1+\beta|\nabla f|^2)$ , with  $\beta$  an arbitrary positive constant),  $TV_h(u)$  is the total variation norm of the function  $u$  weighted by  $h$  and  $\theta > 0$  is a very small parameter, so that we almost have  $f = u+v$ . The terms  $TV_h(u)$  and  $\|v\|_{L^1(\Omega)}$  act as regularizers for  $u$  and  $v$ , respectively, and  $\alpha > 0$  is a regularization parameter.

By combining the ideas of these two aforementioned models, we propose in this paper a variational model for multisource image fusion that simultaneously performs a cartoon and texture decomposition of the fused image. Given the normalized grayscale multi-source images  $\{I_i\}_i^N$ , the combined gradient function  $g$  and the average image  $I_0$ , already defined, we search for a fused image in the form  $u+v$ , as the solution of the following unconstrained optimization problem (herein denoted by  $\mathcal{P}$ )

$$\begin{aligned} & \min_{(u,v)} \left[ \|\nabla(u+v)-g\|_{L^1(\Omega)} + TV_h(u) + \alpha \|v\|_{L^1(\Omega)} \right. \\ &\quad \left. + \frac{1}{2\theta} \|(u+v)-I_0\|_{L^2(\Omega)}^2 + \frac{\eta}{2} \|(u+v)-\frac{1}{2}\|_{L^2(\Omega)}^2 \right]. \end{aligned}$$

where for simplicity  $h=1$ .

### 3 NUMERICAL SOLUTION

Since the objective functional in  $\mathcal{P}$  is convex, the solution of  $\mathcal{P}$  can be computed by minimizing separately, with respect to  $u$  and  $v$ , that is by solving the cartoon and texture subproblems (denoted by  $C$  and  $T$ , respectively):

**Subproblem  $C$** — $v$  being fixed, find  $u$  solution of

$$\begin{aligned} & \min_u \left[ \|\nabla(u+v)-g\|_{L^1(\Omega)} + TV_h(u) + \frac{1}{2\theta} \|(u+v) \right. \\ &\quad \left. - I_0\|_{L^2(\Omega)}^2 + \frac{\eta}{2} \|(u+v)-\frac{1}{2}\|_{L^2(\Omega)}^2 \right]. \end{aligned}$$

**Subproblem  $T$** — $u$  being fixed, find  $v$  solution of

$$\begin{aligned} & \min_v \left[ \|\nabla(u+v)-g\|_{L^1(\Omega)} + \alpha \|v\|_{L^1(\Omega)} + \frac{1}{2\theta} \|(u+v) \right. \\ &\quad \left. - I_0\|_{L^2(\Omega)}^2 + \frac{\eta}{2} \|(u+v)-\frac{1}{2}\|_{L^2(\Omega)}^2 \right]. \end{aligned}$$

Due to the structure of subproblems  $C$  and  $T$ , involving particular  $L^1$  and  $L^2$  terms (that are convex functions) a suitable iterative method for approximating their solutions, and that we adopt in this paper, is the split Bregman method (Goldstein & Osher 2009).

Firstly, extra variables are introduced through the constraints  $d_1 = \nabla(u+v) - g$ ,  $d_2 = \nabla u$ , for subproblem  $C$ , and  $e = \nabla(u+v) - g$ ,  $z = v$  for subproblem  $T$ . The goal is to separate completely the  $L^1$  and  $L^2$  terms in each subproblem. Then we enforce these constraints with the Bregman iteration process (Bregman 1967). Therefore we have a sequence of cartoon and texture subproblems, respectively  $C^k$  and  $T^k$ , for  $k=1, 2, \dots$ , as follows:

**Subproblem  $C^k$** — $v^k$  being fixed, find  $(u^{k+1}, d_1^{k+1}, d_2^{k+1})$  solution of

$$\begin{aligned} & \min_{(u,d_1,d_2)} \left[ \|d_1\|_{L^1(\Omega)} + \|d_2\|_{L^1(\Omega)} + \frac{1}{2\theta} \|(u+v^k) - I_0\|_{L^2(\Omega)}^2 \right. \\ &\quad \left. + \frac{\eta}{2} \|(u+v^k) - \frac{1}{2}\|_{L^2(\Omega)}^2 + \frac{\lambda_1}{2} \|d_1 - \nabla(u+v^k) \right. \\ &\quad \left. + g - b_1^k\|_{L^2(\Omega)}^2 + \frac{\lambda_2}{2} \|d_2 - \nabla u - b_2^k\|_{L^2(\Omega)}^2 \right], \end{aligned}$$

$$b_1^{k+1} = b_1^k - d_1^{k+1} + \nabla(u^{k+1} + v^k) - g,$$

$$b_2^{k+1} = b_2^k - d_2^{k+1} + \nabla u^{k+1}.$$

**Subproblem**  $T^k - u^{k+1}$  being fixed, find  $(v^{k+1}, e^{k+1}, z^{k+1})$  solution of

$$\begin{cases} \min_{(v,e,z)} \left[ \|e\|_{L^1(\Omega)} + \alpha \|z\|_{L^1(\Omega)} + \frac{1}{2\theta} \|(u^{k+1} + v) - I_0\|_{L^2(\Omega)}^2 + \frac{\eta}{2} \|(u^{k+1} + v) - \frac{1}{2} \|_{L^2(\Omega)}^2 \right. \\ \quad \left. + \frac{\lambda_3}{2} \|e - \nabla(u^{k+1} + v) + g - c^k\|_{L^2(\Omega)}^2 + \frac{\lambda_4}{2} \|z - v - w^k\|_{L^2(\Omega)}^2 \right], \end{cases}$$

$$c^{k+1} = c^k - e^{k+1} + \nabla(u^{k+1} + v^{k+1}) - g, \\ w^{k+1} = w^k - z^{k+1} + v^{k+1}.$$

The constants  $\lambda_i$ , for  $i = 1, 2, 3, 4$  are the fixed penalty parameters, used in the Bregman approach.

The minimization problem in subproblem  $C^k$  is solved by iteratively minimizing with respect  $u$ ,  $d_1$  and  $d_2$ , alternatively. Similarly, in subproblem  $T^k$  the minimization problem is also solved by iteratively minimizing with respect  $v$ ,  $e$  and  $z$ , separately. Consequently we have that:

- The formula for  $u^{k+1}$  is derived from the Euler-Lagrange equation, i.e. it is the solution of the following PDE (partial differential equation) in  $\Omega$ , hereafter denoted by "cartoon PDE"

$$\begin{aligned} \left( \eta + \frac{1}{\theta} - (\lambda_1 + \lambda_2)\Delta \right) u^{k+1} &= - \left( \eta + \frac{1}{\theta} - \lambda_1 \Delta \right) v^k \\ &\quad - \lambda_1 \operatorname{div}(d_1^k + g - b_1^k) - \lambda_2 \operatorname{div}(d_2^k - b_2^k) \\ &\quad + \frac{1}{2} \eta + \frac{1}{\theta} u_0 \end{aligned}$$

where  $\Delta$  and  $\operatorname{div}$  are the Laplace and divergence operators, respectively, with the non-homogeneous Neumann boundary condition on the boundary  $\partial\Omega$  of  $\Omega$

$$\begin{aligned} \frac{\partial u^{k+1}}{\partial n} &= \frac{\lambda_1}{\lambda_1 + \lambda_2} (d_1 + g - b_1^k - \nabla v^k) \cdot n \\ &\quad + \frac{\lambda_2}{\lambda_1 + \lambda_2} (d_2 - b_2^k) \cdot n \end{aligned}$$

where  $n$  is the unit outward normal to  $\partial\Omega$  and " $\cdot$ " denotes the inner product in  $R^2$ .

- The formulas for  $d_1^{k+1}$  and  $d_2^{k+1}$  are explicit using the shrinkage operators, so at each pixel  $(i, j)$  in  $\Omega$

$$d_{1i,j}^{k+1} = \operatorname{shrink} \left( (\nabla(u^{k+1} + v^k) - g + b_1^k)_{i,j}, \frac{1}{\lambda_1} \right),$$

$$d_{2i,j}^{k+1} = \operatorname{shrink} \left( (\nabla u^{k+1} + b_2^k)_{i,j}, \frac{1}{\lambda_2} \right),$$

where for  $z, \gamma$  in  $R$

$$\operatorname{shrink}(z, \gamma) = \frac{z}{|z|} \cdot \max(|z| - \gamma, 0).$$

- The formula for  $v^{k+1}$  is derived again from the Euler-Lagrange equation, thus  $v^{k+1}$  is the solution of the following PDE in  $\Omega$ , hereafter denoted by "texture PDE"

$$\begin{aligned} \left( \eta + \frac{1}{\theta} + \lambda_4 - \lambda_3 \Delta \right) v^{k+1} &= - \left( \eta + \frac{1}{\theta} - \lambda_3 \Delta \right) u^{k+1} \\ &\quad - \lambda_3 \operatorname{div}(e^k + g - c^k) + \lambda_4 (z^k - w^k) + \frac{1}{2} \eta + \frac{1}{\theta} I_0, \end{aligned}$$

with the non-homogeneous Neumann boundary condition

$$\frac{\partial v^{k+1}}{\partial n} = (e - \nabla u^{k+1} + g - c^k) \cdot n, \quad \text{on } \partial\Omega.$$

- The formulas for  $e^{k+1}$  and  $z^{k+1}$  are also explicit using the shrinkage operators (likewise for  $d_1^{k+1}$  and  $d_2^{k+1}$ ).

Summarizing, the split Bregman method for  $\mathcal{P}$  is:

#### Algorithm

**Input**—Multisource images  $I_i$ ,  $i = 1, 2, \dots, N$ .

**Initialize**— $u^0 = I_0$ , fix  $\theta$ ,  $\eta$ ,  $\alpha$ ,  $\lambda_i$  ( $i = 1, 2, 3, 4$ ),  $tol$ , and fix  $v^0, e^0, b_1^0, b_2^0, c^0, d_1^0, d_2^0, w^0, z^0$  equal to zero.

**While**  $\max\{|u^k - u^{k-1}|, |v^k - v^{k-1}|\} > tol$

*Cartoon part* -

$$\begin{aligned} u^{k+1} &\text{ solution of the "cartoon PDE",} \\ d_1^{k+1} &= \operatorname{shrink}(\nabla(u^{k+1} + v^k) - g + b_1^k, \frac{1}{\lambda_1}), \\ d_2^{k+1} &= \operatorname{shrink}(\nabla u^{k+1} + b_2^k, \frac{1}{\lambda_2}), \\ b_1^{k+1} &= b_1^k - d_1^{k+1} + \nabla(u^{k+1} + v^k) - g, \\ b_2^{k+1} &= b_2^k - d_2^{k+1} + \nabla u^{k+1}. \end{aligned}$$

*Texture part* -

$$\begin{aligned} v^{k+1} &\text{ solution of the "texture PDE",} \\ e^{k+1} &= \operatorname{shrink}(\nabla(u^{k+1} + v^k) - g + c^k, \frac{1}{\lambda_3}), \\ z^{k+1} &= \operatorname{shrink}(v^{k+1} + w^k, \frac{\alpha}{\lambda_4}), \\ c^{k+1} &= c^k - e^{k+1} + \nabla(u^{k+1} + v^{k+1}) - g, \\ w^{k+1} &= w^k - z^{k+1} + v^{k+1}. \end{aligned}$$

**End**

**Output**—Fused image  $u^k + v^k$ , where  $u^k$  is the cartoon part and  $v^k$  the texture part.

## 4 EXPERIMENTS

In this section we apply the proposed technique to three different types of images. In the

experiments, we use the following parameters  $\eta = 10^{-5}$ ,  $\theta = 10^{-5}$ ,  $\lambda_1 = \lambda_3 = 1$ ,  $\lambda_2 = 10$ ,  $\lambda_4 = 10^4$ ,  $\alpha = 1$ , and  $tol = 10^{-4}$ .

The first example, displayed in Figure 1, shows an EO (earth observation) image, obtained with the remote sensing satellite *WorldView-2*. The EO image was downloaded freely from Digital Globe Inc. ([www.digitalglobe.com](http://www.digitalglobe.com)), and acquired on April 3rd 2011 over Sydney (Australia). For this experiment we consider the panchromatic band (with a high spatial resolution of 0.46 meters at nadir) and three visible multispectral bands (Red, Green, Blue) (with a low spatial resolution of 1.84 meters at nadir and a high spectral resolution). The goal is to fuse the high spatial and high spectral images in a single image and extract from it the cartoon and texture information. We fuse each multispectral band ( $64 \times 64$  pixels) with the panchromatic image ( $256 \times 256$  pixels), and to this end we first perform a pre-processing step, where each multispectral image is up-sampled to the panchromatic image size, by using a bi-cubic interpolation. In Figure 1 we see that our model produces a fused image with significant improvements. Also note that the cartoon is almost the same as the fused image, and the texture information is meaningful.

Figure 2 shows the results of our method applied to a medical (RGB) image (with  $536 \times 536$  pixels), acquired with the wireless capsule Pillcam Colon 2 of *Given Imaging*. It displays a colonic polyp (the reddish region) exhibiting strong texture. Here we fuse each channel with the grayscale version of the three color image. Again the results confirm that the cartoon has almost all the details of the fused image, and the textural image captures the small fea-

tures on the polyp surface. Finally, Figure 3 shows the outcome of our method for a standard test (RGB) image (with  $256 \times 256$  pixels), downloaded from the IPOL archive (<http://www.ipol.im/>).

For assessing the quality of the results, we have used the following different evaluation measures:

1. Average Gradient (AG), which for a grayscale image  $u$  is defined by (Liu, Huang, & Zhao 2006).

$$AG = \frac{1}{|\Omega|} \int_{\Omega} \sqrt{\left( \frac{\partial u}{\partial x_1} \right)^2 + \left( \frac{\partial u}{\partial x_2} \right)^2} dx,$$

and generally, a big AG indicates a good fusion result.

2. Entropy (E), which for a discrete grayscale image  $u$  is defined by (Liu, Huang, & Zhao 2006)

$$E(u) = - \sum_{i=0}^{L-1} p_i \log_2 p_i$$

where  $p_i$  is the probability of the  $i$ th gray level in the image  $u$ , and  $L$  is the gray level. Similarly, the larger the entropy is, the better the fusion result.

3. Peak Signal to Noise Ratio (PSNR), that is evaluated in decibels and is inversely proportional the mean squared error. For a grayscale restored image  $u$  and rinput image  $\bar{u}$  it is defined by the following formula (Huynh-Thu & Ghanbari 2008)

$$PSNR = 10 \log_{10} \frac{(2^n - 1) |\Omega|}{\| u - \bar{u} \|_{L^2(\Omega)}^2}$$

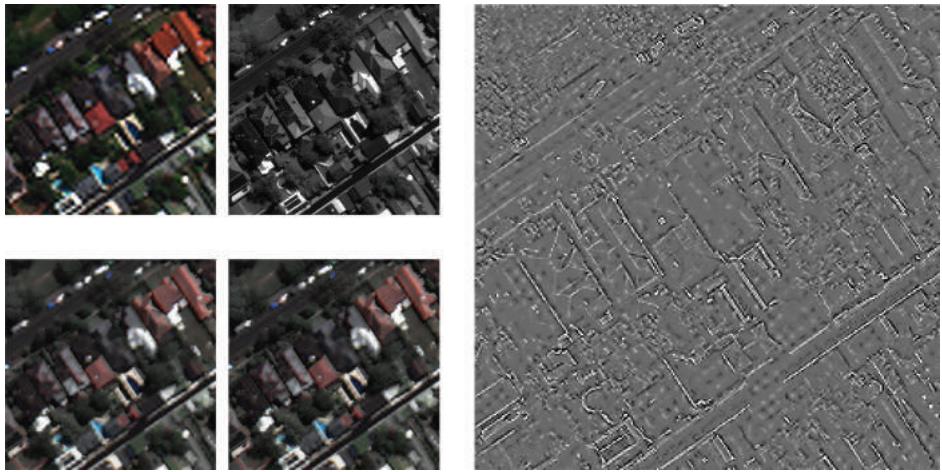


Figure 1. Top left: Multispectral (R, G, B) scene of Sydney, Australia. Top middle: Panchromatic image. Bottom left: Cartoon. Bottom middle: Fused Image. Right: Texture.

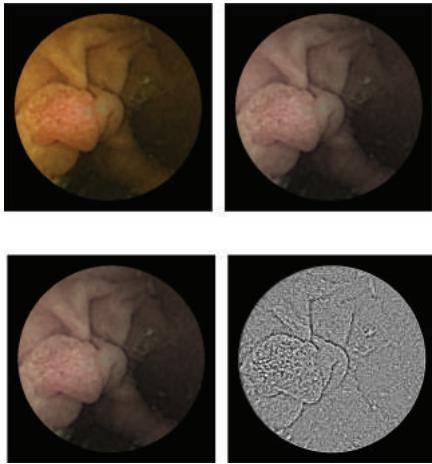


Figure 2. Top left: Polyp image obtained with Pillcam Colon 2, by courtesy of University Hospital of Coimbra, Portugal. Top right: Fused Image. Bottom left: Cartoon. Bottom right: Texture.

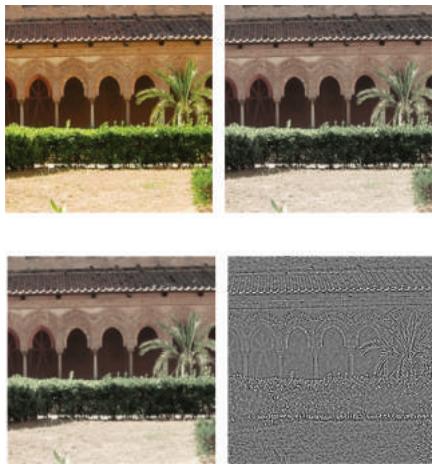


Figure 3. Top left: Garden standard test image. Top right: Fused Image. Bottom left: Cartoon. Bottom right: Texture.

- Generally a big PSNR indicates that the image reconstruction is of high quality.
- Signal-to-Noise Ratio (SNR), which is defined in decibels as the ratio of the power of a signal (meaningful information) and the power of background noise (unwanted signal):

$$\text{SNR} = 10 \log_{10} \left( \frac{P_{\text{signal}}}{P_{\text{noise}}} \right),$$

where  $P$  represents average power (Gonzalez & Woods 2006). Generally a big SNR indicates that the image reconstruction is of high quality.

The results obtained with the proposed method are reported in Tables 1 and 2, where the columns are organized as follows:

- The first column corresponds to the figure numbers.
- The second column indicates the images: Panchromatic (Pan.), multi-spectral image (RGB) and grayscale version of the multi-spectral image (Gray).
- The third column corresponds to input and output images, where
  - Input stands for the multi-spectral image.
  - Average refers to the weighted average of the multi-spectral image and the panchromatic band (using the weights  $w_p$ , based on the gradients, previously defined).
  - Fused represents the fused image.
  - Cartoon indicates the cartoon image.
- The fourth and fifth columns correspond to the measures E, AG, PSNR and SNR.

Table 1. Entropy (E) and Average Gradient (AG) for the three figures.

			E	$AG \times 10^3$
Figure 1	Pan. RGB	–	7.08	65.88
		Input	6.91	43.18
		Average	7.03	58.46
		Fused	7.03	58.39
	Gray	Cartoon	7.03	58.34
		Input	6.96	42.36
		Average	7.03	56.83
		Fused	7.03	56.77
Figure 2	Pan. RGB	Cartoon	7.03	56.72
		–	6.35	6.14
		Input	6.13	6.19
		Average	6.30	10.11
	Gray	Fused	6.30	10.10
		Cartoon	6.30	10.09
		Input	6.35	6.14
		Average	6.35	7.20
Figure 3	Pan. RGB	Fused	6.35	7.19
		Cartoon	6.35	7.19
		–	7.60	65.28
		Input	7.56	64.94
	Gray	Average	7.59	66.01
		Fused	7.59	65.94
		Cartoon	7.59	65.89
		Input	7.60	65.28
		Average	7.60	65.39
		Fused	7.60	65.31
		Cartoon	7.60	65.26

Table 2. Peak-Signal-to-Noise-Ratio (PSNR) and Signal-to-Noise-Ratio (SNR) of the three figures. The reference images to compute PSNRs and SNRs are the input images.

			PSNR	SNR
Figure 1	RGB	Fused	20.22	9.21
		Cartoon	20.23	9.21
	Gray	Fused	21.90	10.98
		Cartoon	21.90	10.98
Figure 2	RGB	Fused	24.00	13.00
		Cartoon	24.00	13.00
	Gray	Fused	45.41	34.45
		Cartoon	45.42	34.46
Figure 3	RGB	Fused	24.21	18.10
		Cartoon	24.21	18.10
	Gray	Fused	48.41	42.61
		Cartoon	48.42	42.62

Table 3. The computational time in seconds for the three figures. Columns 2–4 correspond to the execution time of the proposed method (fusion and simultaneous cartoon+texture decomposition of the fused image) of the red, green and blue bands, respectively.

	Red	Green	Blue
Figure 1	0.3276	0.3120	0.3744
Figure 2	1.7940	1.6224	1.7472
Figure 3	0.2184	0.2652	0.1560

All the numbers reported in Tables 1 and 2 are averages of the three channels when it is applicable. The displayed results confirm that the average gradient AG and entropy E for the fused images are improved when compared to AG and E of the input (see Table 1). Also, from Table 2, one can see that PSNR and SNR of the cartoon and fused images are similar.

To evaluate the computational efficiency of our model, we computed the execution time. Table 3 shows that the proposed method is quite fast.

## 5 CONCLUSIONS

We have presented a variational model for image fusion with a simultaneous cartoon and texture decomposition of the fused image. The main motivation for performing these two operations at the

same time was to have a simultaneous information in real-time. In the future we are interested in comparing our (simultaneous) approach with a two-step approach.

The numerical experiments reveal the good performance of the proposed model, in terms of appropriate quality metrics.

However, the proposed model depends on several parameters, that are fixed manually. In the tests, we have noticed that the quality of the results depends on the size of these parameters. For instance, experimentally we realize that the size of  $\lambda_4$  should be relatively large when compared to the other parameters. Note that a large  $\lambda_4$  enforces the texture constraint “ $z = v$ ” to hold. In the future, we intend to address these parameter related issues and in particular to perform a deep study to find the optimal parameters, in an automatic way.

## REFERENCES

- Bertalmio, M., V. Caselles, E. Provenzi, & A. Rizzi (2007). Perceptual color correction through variational techniques. *Image Processing, IEEE Transactions on* 16(4), 1058–1072.
- Bregman, L.M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics* 7(3), 200–217.
- Bresson, X., S. Esedoglu, P. Vandergheynst, J.-P. Thiran, & S. Osher (2007). Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and vision* 28(2), 151–167.
- Fang, F., F. Li, G. Zhang, & C. Shen (2013). A variational method for multisource remote-sensing image fusion. *International Journal of Remote Sensing* 34(7), 2470–2486.
- Goldstein, T. & S. Osher (2009). The Split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences* 2(2), 323–343.
- Gonzalez, R.C. & R.E. Woods (2006). *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.
- Huynh-Thu, Q. & M. Ghanbari (2008, June). Scope of validity of PSNR in image/video quality assessment. *Electronics Letters* 44(13), 800–801.
- Liu, W., J. Huang, & Y. Zhao (2006). Image fusion based on PCA and undecimated discrete wavelet transform. In I. King, J. Wang, L.-W. Chan, and D. Wang (Eds.), *Neural Information Processing*, Volume 4233 of *Lecture Notes in Computer Science*, pp. 481–488. Springer Berlin Heidelberg.
- Piella, G. (2009). Image fusion for enhanced visualization: a variational approach. *International Journal of Computer Vision* 83(1), 1–11.

# Image contrast enhancement using split Bregman method\*

Somayeh Gh. Bardeji, Isabel N. Figueiredo & Ercília Sousa

*Department of Mathematics, CMUC, University of Coimbra, Coimbra, Portugal*

**ABSTRACT:** In this paper we propose a variational method for image contrast enhancement, by keeping the image details and correcting the non-uniform illumination. It is a minimization problem, where the objective functional consists of two different fitting terms: a  $L^1$  term that matches the gradients of the input and reconstructed images, for preserving the image details, and a  $L^2$  term that measures the misfit between the reconstructed image and the mean value of the input image, for reducing the image variance and thus correcting the illumination. For solving this minimization problem we apply the split Bregman method, which is an efficient and fast iterative method suitable for this type of non-differentiable and convex minimization problem, involving a  $L^1$  term. Some experimental results show the effectiveness of the method.

## 1 INTRODUCTION

Image contrast enhancement is an image processing technique, whose purpose is to improve the image quality, for human interpretation of the image contents or for supplying a good input in automated image processing systems.

In the literature there exists a plethora of contrast enhancement methods, which are based, for example, on histogram equalization, edge enhancement, edge sharpening, filtering and restoration. For a detailed description of these type of methods see for example the book (Gonzalez & Woods 2008). In this paper we focus on a particular variational PDE (partial differential equation) approach for contrast enhancement. We refer for example to the book (Aubert & Kornprobst 2006) for an overview of the application of functional analysis techniques and the theory of partial differential equations to different image processing problems, such as restoration of degraded images, denoising, segmentation, inpainting, decomposition into cartoon and texture, optical flow and image classification.

In the variational approach, used herein, the objective is to minimize an appropriate energy (or functional), whose corresponding Euler-Lagrange equation involves a PDE, that can be afterwards

solved by a suitable computational method. Our energy functional is a modification of that proposed in (Morel, Petro, & Sbert 2014). It is composed of two fitting terms. One fits the gradients of the input and reconstructed images, it is measured with the  $L^1$ -norm, and aims at preserving image details. The  $L^1$ -norm is the chosen measure because it has the property of better preserving discontinuities, when compared to the  $L^2$ -norm. The second term fits the reconstructed image with the mean value of the input image, in the  $L^2$ -norm, for correcting non-uniform illumination. Due to its particular structure, we then choose the split Bregman method (Goldstein & Osher 2009) to solve this variational problem. This is a particular efficient iterative method applicable to a wide class of  $L^1$ -regularized optimization problems.

After this introduction the rest of the paper includes the description of the variational problem in Section 2, its numerical solution in Section 3, some applications in Section 4 and finally the paper ends with some conclusions and comments.

## 2 DESCRIPTION OF THE MODEL

In (Morel, Petro, & Sbert 2014) it is proposed the following variational model for contrast enhancement

$$\begin{aligned} \min_u & \left\{ \int_{\Omega} |\nabla u - \nabla f|^2 dx + \lambda \int_{\Omega} (u - \tilde{u})^2 dx \right\} \\ & = \min_u \left\{ \|\nabla u - \nabla f\|_{L^2(\Omega)}^2 + \lambda \|u - \tilde{u}\|_{L^2(\Omega)}^2 \right\}, \end{aligned} \quad (1)$$

where  $f : \Omega \rightarrow R$  is the original (grayscale) input image,  $\Omega \subset R^2$  represents the image pixel domain,

---

\*This work was partially supported by the research project PTDC/MATNAN/0593/2012 of FCT (the Portuguese national funding agency for science, research and technology) and by CMUC (the Centre for Mathematics of the University of Coimbra—UID/MAT/00324/2013), funded by the Portuguese Government through FCT/MEC and co-funded by the European Regional Development Fund through the Partnership Agreement PT2020.

$x$  is a point (*i.e.* a pixel) in  $\Omega$ ,  $\nabla$  denotes the gradient operator,  $|\cdot|$  is the Euclidean norm in  $R^2$ ,  $\tilde{u}$  is the mean value of the reconstructed image  $u : \Omega \rightarrow R$ ,  $L^2(\Omega)$  is the space of square integrable functions in  $\Omega$ , and  $\|\cdot\|_{L^2(\Omega)}$  denotes the  $L^2$ -norm. This model contains two quadratic fitting terms. The first fits the gradients of  $u$  and  $f$ , and consequently aims at preserving image details. The second term intends to reduce the effect of nonuniform illumination by fitting  $u$  and its mean value  $\tilde{u}$ , thus by decreasing the image variance. The parameter  $\lambda$  is a positive constant that balances the influence of the two fitting terms. After replacing  $\tilde{u}$  by  $\tilde{f}$ , the mean value of the input image  $f$ , it is shown that the problem does not depend on the value  $\tilde{f}$ , and the problem is solved, for  $\tilde{f} = 0$  in the Fourier domain, using the discrete Fourier transform.

In this paper we modify the model (1) by replacing the  $L^2$ -norm by the  $L^1$ -norm for measuring the misfit between  $\nabla u$  and  $\nabla f$  and we also replace the mean value  $\tilde{u}$  by  $\tilde{f}$ . This yields the following variational model

$$\begin{aligned} & \min_u \left\{ \int_{\Omega} |\nabla u - \nabla f| dx + \frac{\lambda}{2} \int_{\Omega} (u - \tilde{f})^2 dx \right\} \\ &= \min_u \left\{ \|\nabla u - \nabla f\|_{L^1(\Omega)} + \frac{\lambda}{2} \|u - \tilde{f}\|_{L^2(\Omega)}^2 \right\} \quad (2) \end{aligned}$$

where  $L^1(\Omega)$  is the space of absolutely integrable functions in  $\Omega$  and  $\|\cdot\|_{L^1(\Omega)}$  denotes the  $L^1$ -norm. The reason for replacing the  $L^2$ -norm by the  $L^1$ -norm is related to the fact that the  $L^2$ -norm of the gradient tends to smear image discontinuities, as opposed to the  $L^1$ -norm that tends to preserve the discontinuities, which in image processing corresponds to sharp edges. In addition to this advantage, the presence of the  $L^1$  fitting term in (2) permits the use of fast and effective algorithms for computing its solution. In effect, problem (2) belongs to the general class of  $L^1$ -regularized problems of the form

$$\min_u \{\|\phi(u)\|_{L^1(\Omega)} + H(u)\}$$

where both  $\|\phi(u)\|_{L^1(\Omega)}$  and  $H(u)$  are convex functions. This kind of models can be efficiently solved with the split Bregman method of (Goldstein & Osher 2009). This is an appropriate algorithm for solving non-differentiable convex minimization problems, involving  $L^1$  or  $TV$  (total variation) terms. We refer to (Fang, Li, Zhang, & Shen 2013, Getreuer 2012b, Goldstein, Bresson, & Osher 2010, Yin, Osher, Goldfarb, & Darbon 2008) for a few examples of different applications of the method.

In our case  $\phi(u) = \nabla u - \nabla f$  and  $H(u) = \lambda \|u - \tilde{f}\|_{L^2(\Omega)}$ , and in the next section we apply split Bregman method to solve (2).

### 3 NUMERICAL SOLUTION BASED ON SPLIT BREGMAN METHOD

A critical and first aspect of the split Bregman method is the separation of the  $L^1$  and  $L^2$  terms, which is achieved by introducing an auxiliary variable. Thus, we first replace (2) by the following constrained optimization problem

$$\begin{aligned} & \min_u \left\{ \|d\|_{L^1(\Omega)} + \frac{\lambda}{2} \|u - \tilde{f}\|_{L^2(\Omega)} \right\} \\ & \text{subject to } d = \nabla u - \nabla f, \end{aligned}$$

and then reformulate it as an unconstrained problem, by introducing a quadratic penalty function, that is

$$\begin{aligned} & \min_u \left\{ \|d\|_{L^1(\Omega)} + \frac{\lambda}{2} \|u - \tilde{f}\|_{L^2(\Omega)}^2 \right. \\ & \quad \left. + \frac{\alpha}{2} \|d - (\nabla u - \nabla f)\|_{L^2(\Omega)}^2 \right\}. \quad (3) \end{aligned}$$

Then, the split Bregman method consists in solving the following sequence of problems for  $k = 0, 1, 2, \dots$

$$\begin{cases} (u^{k+1}, d^{k+1}) = \arg \min_{d, u} \left\{ \|d\|_{L^1(\Omega)} + \frac{\lambda}{2} \|u - \tilde{f}\|_{L^2(\Omega)}^2 \right. \\ \quad \left. + \frac{\alpha}{2} \|d - (\nabla u - \nabla f) - b^k\|_{L^2(\Omega)}^2 \right\}, \\ b^{k+1} = b^k + \nabla u^{k+1} - \nabla f - d^{k+1}. \end{cases} \quad (4)$$

where the new vector  $b^k$  results from the Bregman iteration (Bregman 1967), that is a strategy for enforcing the constraint  $d = \nabla u - \nabla f$ , using a fixed penalty parameter  $\alpha$ . This strategy is an alternative to the conventional continuation technique to solve (3) with an increasing sequence of penalty parameters  $\alpha_1 < \alpha_2 < \dots < \alpha_n$  tending to  $\infty$ , for accurately enforcing the constraint. The minimization problem in (4) is solved by iteratively minimizing with respect  $u$  and  $d$ , alternatively, which means the following two steps are performed.

**Step 1**—Minimization with respect to  $u$  (with  $d$  fixed)

$$\begin{aligned} u^{k+1} = \arg \min_u & \left\{ \frac{\lambda}{2} \|u - \tilde{f}\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|d - (\nabla u - \nabla f) - b^k\|_{L^2(\Omega)}^2 \right\}, \end{aligned} \quad (5)$$

for which the optimality condition (derived from the Euler-Lagrange equation) is, in  $\Omega$

$$(\lambda - \alpha \Delta) u^{k+1} = \lambda \tilde{f} - \alpha \operatorname{div}(d^k + \nabla f - b^k),$$

with  $\Delta$  and  $\operatorname{div}$  denoting the Laplace and divergence operators, respectively, along with the non-homogeneous Neumann boundary condition on the boundary  $\partial\Omega$  of  $\Omega$

$$\frac{\partial u^{k+1}}{\partial n} = (d^k - \nabla f - b^k) \cdot n,$$

where “.” denotes the inner product in  $R^2$  and  $n$  is the unit outward normal to  $\partial\Omega$ .

This problem can be solved efficiently with the Gauss-Seidel method, since the system is diagonally dominant. The solution  $u_{i,j}^{k+1}$  at each pixel  $(i,j)$  in  $\Omega$  (excepting in  $\partial\Omega$ ) is defined by

$$u_{i,j}^{k+1} = \frac{1}{\lambda + 4\alpha} [\alpha U_{i,j} + \lambda \tilde{f}_{i,j} - \alpha v_{i,j}^k] \\ U = (U_{i,j}) := (u_{i-1,j}^{k+1} + u_{i+1,j}^k + u_{i,j-1}^{k+1} + u_{i,j+1}^k) \\ v^k = (v_{i,j}^k) := (\operatorname{div}(d^k + \nabla f - b^k))_{i,j}.$$

Here we use finite differences for approximating the derivatives in the gradient  $\nabla$  and divergence  $\operatorname{div}$  operators, respectively. In particular the discretization used for  $v_{i,j}^k$  is obtained by applying backward finite differences for  $\operatorname{div} d^k$  and  $\operatorname{div} b^k$  and centered finite differences for  $\Delta f$

$$v_{i,j}^k = 2d_{i,j}^k - d_{i-1,j}^k - d_{i,j-1}^k - 2b_{i,j}^k + b_{i-1,j}^k + b_{i,j-1}^k \\ + f_{i-1,j}^k + f_{i+1,j}^k + f_{i,j-1}^k + f_{i,j+1}^k - 4f_{i,j}^k.$$

The Neumann boundary condition is implicitly imposed in  $\partial\Omega$ , the boundary of the rectangular pixel domain, by using backward finite differences, in the right and top sides, and forward finite differences in the left and bottom sides.

**Step 2**—Minimization with respect to  $d$  (with  $u$  fixed)

$$d^{k+1} = \arg \min_d \left\{ \|d\|_{L^1(\Omega)} + \frac{\alpha}{2} \|d - (\nabla u^{k+1} - \nabla f) - b^k\|_{L^2(\Omega)}^2 \right\}.$$

This problem can be explicitly solved using shrinkage operation (known as well as soft thresholding) at each pixel  $(i,j)$

$$d_{i,j}^{k+1} = \operatorname{shrink} \left( (\nabla u^{k+1} - \nabla f + b^k)_{i,j}, \frac{1}{\alpha} \right)$$

where for  $z, \gamma \in R$

$$\operatorname{shrink}(z, \gamma) = \frac{z}{|z|} \times \max(|z| - \gamma, 0).$$

Summarizing, the split Bregman method for model (2) is as follows:

**Algorithm -**

**Input**—Original image  $f$ .

**Initialize**— $u^0 = f$ ,  $d^0 = b^0 = 0$ , and fix  $\lambda, \alpha, tol$ .

**While**  $|u^k - u^{k-1}| > tol$

$$u^{k+1} = \frac{1}{\lambda + 4\alpha} [\alpha U + \lambda \tilde{f} - \alpha v^k], \text{ in } \Omega,$$

$$\frac{\partial u^{k+1}}{\partial n} = (d^k - \nabla f - b^k) \cdot n, \text{ in } \partial\Omega,$$

$$d^{k+1} = \operatorname{shrink} \left( (\nabla u^{k+1} - \nabla f + b^k, \frac{1}{\alpha}) \right),$$

$$b^{k+1} = b^k + \nabla u^{k+1} - \nabla f - d^{k+1}.$$

**End**

**Output**—Image  $u^k$ .

## 4 APPLICATIONS

Some results obtained with our proposed model are shown in this section. All the experiments were implemented with the software MATLAB® R2014a (The Mathworks, Inc.)

Figure 1 shows the contrast enhancement with our method for a standard test image (a scalar image with  $512 \times 343$  pixels), downloaded from the IPOL archive (<http://www.ipol.im/>). As this figure demonstrates, the details are kept and the dark regions become more visible in the enhanced image. In addition, and as expected, when  $\lambda$  increases ( $\lambda$  is the parameter associated with the fitting term intended to reduce the non-uniform illumination) the result tends to the mean value of the input image.

Figure 2 depicts the results of our method applied to a medical (RGB—red, green, blue) image (with  $536 \times 536$  pixels), acquired with the wireless capsule Pillcam Colon 2 of *Given Imaging*. It displays a colonic polyp (the reddish region



Figure 1. Top left: Original image. Top right:  $\lambda = 0.005$ ,  $\alpha = 1$ . Bottom left:  $\lambda = 0.01$ ,  $\alpha = 1$ . Bottom right:  $\lambda = 0.05$ ,  $\alpha = 1$ .

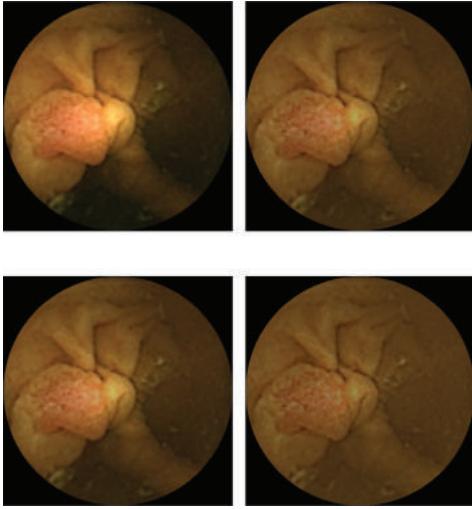


Figure 2. Top left: Polyp image obtained with Pillcam Colon 2, by courtesy of University Hospital of Coimbra, Portugal. Top right:  $\lambda = 1$ ,  $\alpha = 150$ . Bottom left:  $\lambda = 1$ ,  $\alpha = 200$ . Bottom right:  $\lambda = 2$ ,  $\alpha = 200$ .

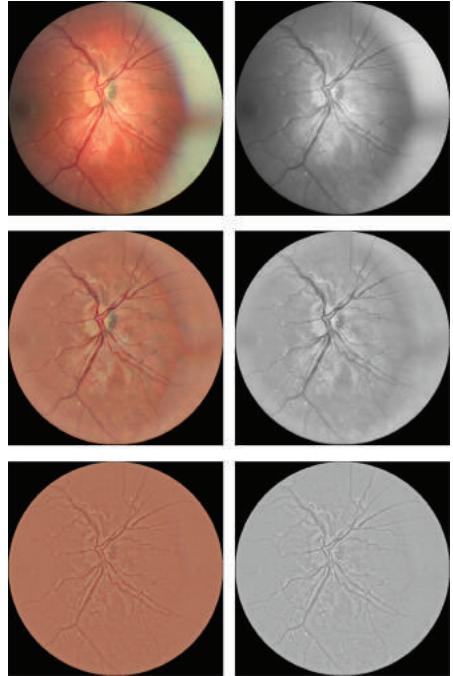


Figure 3. First column: Original (RGB) retinal fundus image and results with  $(\lambda = 1, \alpha = 5)$  in 2nd row and  $(\lambda = 5, \alpha = 5)$  in 3rd row. Second column: Grayscale version and results with  $(\lambda = 1, \alpha = 5)$  in 2nd row and  $(\lambda = 5, \alpha = 5)$  in 3rd row.

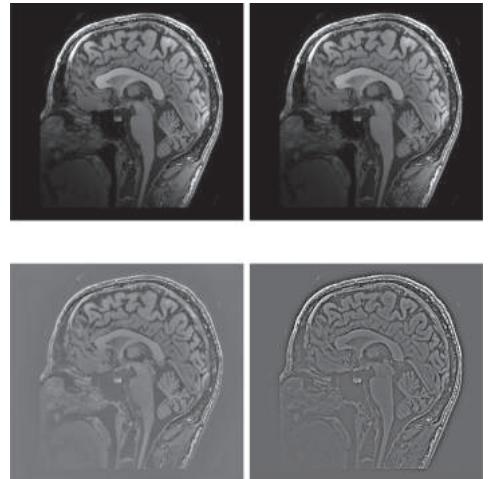


Figure 4. Top left: Original MRI and input for the proposed method. Top right: Original MRI with 2% saturation and input for the model solved with the screened Poisson equation (Morel, Petro, & Sbert 2014). Bottom left: Result for the proposed method with  $(\lambda = 1, \alpha = 10)$ . Bottom right: Result applying a simplest color balance with 2% saturation to the solution obtained with the screened Poisson equation and with  $\lambda = 0.05$ .

Morel, Petro, & Sbert 2011). We can notice that our method performs slightly better in the dark down region on the left part of the image. We also emphasize that our method doesn't use any pre or post processing methodology.

## 5 CONCLUSIONS

In this paper we propose an inverse variational model for image contrast enhancement. It is a modification of the screened Poisson equation for image contrast enhancement as proposed by (Morel, Petro, & Sbert 2014). The difference with respect to this latter model is in the term that matches the gradients of the input and reconstructed images, that is now measured in the  $L^1$ -norm (instead of the  $L^2$ -norm as in (Morel, Petro, & Sbert 2014)), because the former has the advantage of better preserving edges in image processing applications. The presence of this  $L^1$ -norm term enables the proposed model to be solved by the split Bregman technique, which is an efficient procedure that has the advantage of not requiring regularization, continuation, or the enforcement of inequality constraints.

Applications of the proposed method to different types of images show its good performance. We also remark that our methodology does not include pre and post processing steps as it happens in the methodology presented in (Morel, Petro, & Sbert 2014), where the best reconstructed images have been obtained with pre and post processing steps, involving the application of a simplest color balance technique (Limare, Lisani, Morel, Petro, & Sbert 2011) applied before and after solving the model with the screened Poisson equation.

The model we propose involves some parameters that are tuned and fixed manually. In the future an automatic or self-adapting method for choosing these parameters will be studied. Moreover an additional analysis concerning the comparison of the results obtained with the present method with

those obtained with the original model in (Morel, Petro, & Sbert 2014) will be performed, as well as a comparison with the ACE method (Getreuer 2012a), which is one of the most effective contrast enhancement methods.

## REFERENCES

- Aubert, G. & P. Kornprobst (2006). *Mathematical problems in image processing: partial differential equations and the calculus of variations*, Volume 147. Springer Science & Business Media.
- Bregman, L.M. (1967). The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Computational Mathematics and Mathematical Physics* 7(3), 200–217.
- Fang, F., F. Li, G. Zhang, & C. Shen (2013). A variational method for multisource remote-sensing image fusion. *International Journal of Remote Sensing* 34(7), 2470–2486.
- Getreuer, P. (2012a). Automatic color enhancement (ace) and its fast implementation. *Image Processing On Line* (2012).
- Getreuer, P. (2012b). Rudin-Osher-Fatemi total variation denoising using split Bregman. *Image Processing On Line* 10.
- Goldstein, T., X. Bresson, & S. Osher (2010). Geometric applications of the split bregman method: segmentation and surface reconstruction. *Journal of Scientific Computing* 45(1–3), 272–293.
- Goldstein, T. & S. Osher (2009). The Split Bregman method for L1-regularized problems. *SIAM Journal on Imaging Sciences* 2(2), 323–343.
- Gonzalez, R.C. & R.E. Woods (2008). *Digital Image Processing (3rd Edition)*. NJ, USA: Prentice-Hall, Inc.
- Limare, N., J.-L. Lisani, J.-M. Morel, A.B. Petro, & C. Sbert (2011). Simplest color balance. *Image Processing On Line* 1.
- Morel, J.-M., A.-B. Petro, & C. Sbert (2014). Screened Poisson equation for image contrast enhancement. *Image Processing On Line* 4, 16–29.
- Yin, W., S. Osher, D. Goldfarb, & J. Darbon (2008). Bregman iterative algorithms for L1-minimization with applications to compressed sensing. *SIAM Journal on Imaging Sciences* 1(1), 143–168.

This page intentionally left blank

# Extraction of spectral drowsy component from the resting electroencephalographic signal for quick, objective and direct testing of sleepiness in absolute terms

A.A. Putilov, O.G. Donskaya & E.G. Verevkin

*Research Institute for Molecular Biology and Biophysics, Novosibirsk, Russia*

**ABSTRACT:** In spite of the extraordinary progress that has been made by sleep and behavioral sciences in the last decades, the assessment of sleepiness is one of the remaining challenges. The lack of practical tools for objective sleepiness testing has become one of the critical barriers to reducing the threats of increased sleepiness to public health, productivity, and safety. The first aim of the present report was to examine whether objective sleepiness testing can be facilitated by the quantitative analysis of the resting Electroencephalographic (EEG) signal recorded at the first minute following eyes closing. The second aim was to examine applicability of a new spectral EEG index for differentiation of 4 alertness-sleepiness sub-states (alert, neither, dozy, and sleepy) in participants of two sleep deprivation experiments. In the first (exploration) study of applicability of the new index, sleepiness of 15 young adults was self-scored every second hour simultaneously with recording of the resting EEG in the course of 43–61-hour wakefulness. In another (confirmation) study, one-minute eyes closed EEG recordings were obtained with a 3-hour interval during 24-hour wakefulness of 130 adolescents and adults. The sleepiness testing was based on calculation of Spectral Drowsy Component Score (SDCS). To compute SDGS, 16 (1–16 Hz) single-Hz powers of the EEG spectra were obtained for the first one-minute interval of eyes closed EEG signal, log-transformed, ipsatized, weighted, and summed. The 16 weights were the correlation coefficients between the time course of self-scored sleepiness and the time courses of single-Hz powers. The results indicate that SDGS provided reliable criteria for distinguishing between four sub-states of alertness-sleepiness. These criteria performed even better than the traditional criteria based on latency to sleep onset. Moreover, the occipital SDGS changed their sign exactly at the boundary between neither and dozy sub-states thus providing yes-or-no criterion of onset of sleepiness state as opposed to the preceding alert state. We concluded that SDGS might be recommended for quick (app. one-minute), direct and objective testing of sleepiness in absolute terms.

## 1 INTRODUCTION

The Electroencephalogram (EEG) is becoming increasingly important measurement of brain activity. Developing and understanding signal processing methods for the analysis of EEG signals seem to be crucial in the distinct areas of fundamental and applied biomedical research (Motamedi-Fakhra et al., 2014).

The EEG signal processing seems to be of special interest for the distinguishing between different vigilance states and sub-states. In particular, the ability to monitor sleepiness is recognized as one of important research areas where signal processing methods for the analysis of EEG signals can be successfully applied (Subasi et al., 2005). Research in this area is also of practical importance due to obvious implication of the ability to monitor sleepiness for safety critical occupations and procedures, such as driving (Horne & Burley, 2010).

Although it is believed that sleep deprived individuals are able to monitor their own sleepiness, the vast majority of people seem poor in such a monitoring (Berka et al., 2004, van Dongen et al., 2004, Smith et al., 2005, Tremaine et al., 2010). For instance, the experimental results indicate that subjective sleepiness fails to predict the falling asleep event (Kaplan et al., 2007, Herrmann et al., 2010). It appears that, similar to the effects of alcohol, the negative effects of drowsiness are often subjectively underestimated (Czeisler, 2011).

The EEG signs of sleepiness can be revealed objectively by means of the Multiple Sleep Latency Test (Carskadon et al., 1986, Roehrs & Roth, 1992) or the Maintenance of Wakefulness Test (Doghramji, 1999). However, these tests are costly, time consuming, and usually administered in special settings, such as a sleep research laboratory.

Sleepiness can be also measured objectively by using other than EEG signals. For example, simple

performance tests (e.g., a test on reaction time) might be applied for this purpose, and such testing does not require the application electrodes on the scalp. Moreover, the test can be administered in less restricted environmental conditions, and the duration of testing procedure might be reduced to 5 minutes (Roach et al., 2006, Horne & Burley, 2010).

However, similar shortening of the testing procedure was also suggested for such quicker and simpler EEG tests of sleepiness as the Karolinska Drowsiness Test (Åkerstedt & Gillberg, 1990) and the Alpha Attenuation Test (Stampi et al., 1995). These tests combine estimation of decline of alpha activity in eyes closed condition with measurement of "alpha bursts" reflecting a drowsiness-related rise of alpha activity in eyes open condition.

The general disadvantage of any of the proposed tests is that the experimental and field studies of sleep-wake behavior in real life situations require even a much quicker and simpler testing procedure. It was recognized that the area of practical applications of such quick test would be very broad (Mullington et al., 2011, Quan et al., 2011).

Alloway et al. (1997) noted that the increase of alpha activity in eyes open condition did not contribute much to accuracy of distinguishing between drowsy and alert individuals compared to the contribution of the alpha rhythm decrease in eyes closed condition. Moreover, our earlier studies showed that attenuation of alpha-based EEG indexes right after closing the eyes strongly correlates with the feeling of sleepiness and shortened latency to the EEG signs of sleep onset (Donskaya et al., 2012, Putilov et al., 2012, Putilov & Donskaya, 2014b). Therefore, one can question the necessity of the EEG recordings in both conditions for accurate testing of sleepiness. If the EEG sampling in eyes open condition can be excluded from a testing procedure, this can open a new perspective of reducing duration of the procedure to a minute or so (Putilov & Donskaya, 2014b).

We also showed in several our earlier publications that an epoch of sleep onset usually coincides with zero-crossing decline of some of alpha-based EEG indexes, such as alpha-theta power difference in occipital derivation and score on the second principal component of the EEG spectrum in any (e.g., frontal, occipital, etc.) derivation (Putilov, 2011, 2014, 2015, Putilov & Donskaya, 2013a, b, 2014a). However, prior to the transition from wake to sleep state, these measures are always more or less higher than zero and, additionally, demonstrate remarkable inter-individual variation (Putilov & Donskaya, 2014b). Therefore, they cannot serve as yes-or-no criteria for distinguishing between sleepiness sub-states of still waking people. Consequently, a further step on the way of development of a quick

sleepiness test would be a search for such criteria. These criteria were recently proposed by one of us (Putilov, in press). It was hypothesized that the resting EEG signal might be used for calculating spectral alert and drowsy component scores. Weights for alert components score can be derived as a difference between spectra for the extreme (e.g., alert and sleepy) sub-states or between spectra for distant (e.g., first and forth) deprivation phases. Weights for drowsy components score can be provided by correlating the time courses of single-Hz powers with the time course of either subjective or objective sleepiness measure (Putilov, in press).

The major purpose of the present report was to examine one of such recently suggested (Putilov, in press) spectral EEG indexes of sleepiness, Spectral Drowsy Component Score (SDCS). We hypothesized that, as compared to the EEG indexes of sleep onset, the extraction of SDCS from the EEG spectrum can provide more suitable criteria for demarcation of the boundaries between several alertness-sleepiness sub-states of still waking individuals. For instance, it is expected that, in contrast to the existing sleepiness and performance tests, such criteria can allow the description of the apparent level of sleepiness in absolute numbers related to understandable terms, and, consequently, the drowsy component scoring procedure can be recommended for simple and quick (e.g., one minute) objective testing of sleepiness.

## 2 METHOD

Participants of the first study were practically healthy young males ( $N = 7$ ) and females ( $N = 8$ ). Their ages ranged from 22 to 26 and from 19 to 24 years old (average  $\pm$  standard deviation:  $24.7 \pm 1.9$  and  $21.6 \pm 2.2$ , respectively). The EEG signal was recorded with a 2-h interval in the eyes open and eyes closed conditions (2 and 5 min, respectively) starting from Friday evening (19:00 hours). The last session was scheduled on Sunday evening (19:00). Each participant could terminate his/her participation in the experiment when felt irresistible desire to sleep during one of the 2-hour intervals between the EEG recordings. All participants completed, at least, 16 recording sessions, but only four of them completed all 25 sessions. Before and after each EEG recording, the participants self-rated their alertness-sleepiness on the 9-step Karolinska Sleepiness Scale or KSS (Åkerstedt & Gillberg, 1990). Mean scores obtained by averaging scores self-reported before and after each EEG recording session were used in the present analysis as the indicators of subjective sleepiness.

Two (frontal and occipital) electrodes were used for the EEG recordings. The active electrodes were

placed in the middle from the top of the head to forehead (Fz-A2) and midway between O1 and O2 of the International 10–20 system of electrode placement (Oz-A2). To fix the electrodes, Ten20 conductive paste was used (Nicolet Biomedical, Madison, WI, USA). The exact positions of active electrodes were preliminary inked by permanent marker, and the electrodes were removed after each EEG measurement. The EEG signals were recorded via a 16-channel electroencephalograph (Neuron-Spectrum-2, Neurosoft, Ivanovo, Russia), conditioned by the high-pass, low-pass and notch filters (0.5 Hz, 35 Hz, and 50 Hz, respectively), sampled and stored on a hard disc with a frequency of 200 Hz. The EEG records were inspected visually to remove 2-s epochs containing artifacts from further analysis. Power spectra for artifact-free 2-s epochs were computed using the FFTW (the Fastest Fourier Transform in the West) package (Frigo & Johnson, 2005; see also [www.fftw.org](http://www.fftw.org)). Absolute spectral powers ( $\mu\text{V}^2$ ) were calculated for single-Hz frequency bandwidths (i.e., 0.50–1.49 Hz, 1.50–2.49 Hz, etc.). The spectra of the artifact-free 2-s epochs were averaged within each one-min interval of the eyes closed EEG records.

Data from another study were used to provide evidence for reliability of the measures obtained in the analysis of the first (internal) dataset. Participants of this confirmation study were practically healthy adolescents and adults. The mean ages of 54 male and 76 female participants were  $27.4 \pm 10.1$  (range from 15 to 55 years) and  $30.8 \pm 13.4$  (range from 16 to 66 years), respectively. In the experimental morning (between 08:00 and 8:30 o'clock), they were admitted to a research unit of the institute and remained there until approximately 11:00 the next morning. Over the 24 hour period (from 9:00 to 9:00) each study participant completed 9 3–4-min EEG recording sessions divided by 3-hour intervals. KSS was self-scored after each session.

The EEG recordings were taken at frontal and occipital scalp sites corresponding to derivations Fz and O2 referenced to the right mastoid, A2. For one half of the participants, the recordings with close eyes were made prior to the recordings with eyes open, and for another half of the participants the eyes closed recordings were following the eyes open recordings. In the present analysis, only, at least, one minute of eyes closed section of each EEG record was used for comparison with similar data obtained in the first study.

The EEG records from the first study were sorted into 4 groups in accord with the duration of wakefulness, i.e., from only wakefulness (−6 min) to sleep initiated already during the first minute with eyes closed (−1 min). Latency to Sleep onset (SL, min) was determined as a minute during which the difference between alpha (9–12 Hz) and theta

powers (5–8 Hz) changed from positive to negative. Four sub-states of alertness-sleepiness corresponding to self-perceived minimal, mild, moderate, and severe sleepiness were named Alert (SL = −6 min), Neither (SL from −5 to −3), Dozy (SL = −2 min), and Sleepy sub-states (SL = −1 min). Another (subjective) 4-sub-state classification was based on KSS score (1–4.5 for Alert, 5–6 for Neither, 6.5–7 for Dozy, and 7.5–9 for Sleepy sub-state).

The absolute power values in the frequency range from 1 to 16 Hz were converted into a natural logarithmic scale. In the present analysis, such sets of single-Hz spectral powers calculated for the first minute of eyes closed relaxation were used for computation of SDCS. In order to compute SDCS for the participants of the first study, each of 630 (Fz and Oz) sets of 16 ipsatized powers was weighted in accord with correlations between the time courses of these powers and the time course of KSS score (25 time points) and then summed:

$$\text{SDCS}_i = \sum w_j * f_{ij} \quad (1)$$

where  $\text{SDCS}_i$  is the  $i$ -th SDCS ( $i$  refers to either frontal or occipital score),  $w_j$  is a weight of the  $j$ -th frequency bin ( $j$  from 1 to 16 Hz), and  $f_{ij}$  is an ipsatized power value for a  $j$ -th power density bin of an  $i$ -th score.

The procedure of calculation of SDCS was also applied for computing scores for the participants from the second (external) sample. Additionally, the weights provided by the analysis of one sample were used for calculating SDCS for participants of another study.

Moreover, the Oz SDCS provided criteria for the third classification of alertness-sleepiness sub-states. Such (SDCS-based) sub-states were distinguished for assigning KSS scores and SL to the following 4 intervals of SDCS: below −0.2 (Alert), −0.2–0 (Neither), between 0 and 0.2 (Dozy), and above 0.2 (Sleepy sub-state).

Differences between four sub-states of alertness-sleepiness were tested by applying two-way ANOVAs with a fixed factor "Alertness-sleepiness sub-state" and a random factor "Participant" ( $N = 15$  or 130). Prior to applying these ANOVAs, normality of distribution of the EEG variables was tested by means of the exploratory analysis (SPSS) that was performed for each time point of each index. The hypothesis of normality (Kolmogorov-Smirnov criterion) was always supported.

### 3 RESULTS

Very strong correlations (Spearman's coefficients of correlation from 0.93 to 0.96,  $p < 0.001$ ) were attained between the 25-point time courses of

sleepiness measures and the corresponding time courses of SDCS (Table 1).

The occipital SDCS calculated for dozy and sleepy sub-states were positive, whereas scores calculated for alert and neither sub-states were negative. The borderline between neither and dozy sub-states was close to zero score. Thus, such scoring provided yes-or-no criteria of sleepiness onset. Namely, positive SDCS suggested sleepiness state, whereas negative SDCS corresponded to the preceding alert state.

The two-way ANOVAs detected highly significant main effect of the sub-state factor irrespective of the criteria applied for distinguishing between 4 sub-states of alertness-sleepiness, KSS-, SL-, and SDCS-based. Particularly, SL- and KSS-based classifications gave  $F(N = 15, df = 3) = 60.6$  and  $F = 22.1$ , respectively, in analysis of data on occipital SDCS. SL- and SDSC-based classifications gave  $F = 23.2$  and  $F = 23.9$  for KSS score. SDSC- and KSS-based classifications gave  $F = 59.6$  and  $F = 23.0$  for SL, min ( $p < 0.001$  for any of these F-ratios).

The results of post hoc pairwise comparisons suggested that any of 4 objectively defined sub-states demonstrated highly significantly difference from any of 3 other sub-states. Moreover,

Table 1. Correlations between the time courses of SDCS and traditional sleepiness measures.

Derivation	Source of weights	Correlation
		KSS score
Fz	Internal	0.938***/0.938***
Oz	Internal	0.950***/0.945***
Oz	External	0.944***/0.954***
		KSS sub-state
Fz	Internal	0.940***/0.940***
Oz	Internal	0.935***/0.928***
Oz	External	0.928***/0.939***
		SL sub-state
Fz	Internal	0.953***/0.953***
Oz	Internal	0.961***/0.959***
Oz	External	0.919***/0.942***
		SL, min
Fz	Internal	0.944***/0.944***
Oz	Internal	0.952***/0.950***
Oz	External	0.913***/0.931***

Note: The time courses of SDCS were correlated with the time courses of four major sleepiness measures based on KSS and SL (the first dataset). SDCSs were calculated for either frontal or occipital subset (Fz or Oz) using weights derived from either the same (internal) or another (external) dataset, and either from frontal or occipital subset, Fz/Oz or Fz/O2. Asterisks (\*\*\* denote highly significant ( $p < 0.001$ ) Spearman's coefficients of correlation ( $N = 25$ ).

Table 2. Differences between two sleepiness sub-states (dozy and sleepy) in occipital SDCS.

Weighting	Sub-state	Mean score	CI	
			-95%	+95%
<i>Dataset, derivation:</i> Internal, Oz ( $N = 15$ )				
Internal Fz	Averaged	0.002	-0.030	0.035
	Dozy	0.040	-0.034	0.115
	Sleepy	0.249	0.172	0.326
	Difference	0.209**	0.064	0.353
Internal Oz	Averaged	0.020	-0.013	0.053
	Dozy	0.061	-0.014	0.137
	Sleepy	0.269	0.190	0.347
	Difference	0.207**	0.061	0.354
<i>Dataset, derivation:</i> External, O2 ( $N = 130$ )				
Internal Fz	Averaged	0.075	0.066	0.084
	Dozy	0.081	0.058	0.104
	Sleepy	0.182	0.155	0.209
	Difference	0.101***	0.053	0.148
Internal Oz	Averaged	0.091	0.082	0.100
	Dozy	0.097	0.074	0.120
	Sleepy	0.203	0.176	0.229
	Difference	0.106***	0.059	0.153

Notes: SDCS were derived from either the first (internal) or second (external) datasets from subsets of data on occipital derivations (Oz and O2, respectively) using weights derived from the first (internal) dataset, either Fz or Oz subsets. Averaged: General mean score obtained by averaging over 4 sub-states. Difference: Difference between Sleepy and Dozy sub-states. \*\*\* $p < 0.001$ , \*\* $p < 0.01$ : Level of significance for t-value in post-hoc pairwise comparisons of the dozy and sleepy sub-states (with Bonferroni adjustment for the number of comparisons). CI: Confidence Interval.

some of the results suggested that SDSC-based classification over-performed SL-based classification in distinguishing between KSS scores assigned to two sleepiness (dozy and sleepy) sub-states.

Very similar results were obtained in the analysis of the second (external) dataset irrespective of which weights were used for calculating SDSC for the participants of this experimental study, either own or those provided by the analysis of the first (internal) dataset (Table 2).

#### 4 DISCUSSION

In spite of the extraordinary progress that has been made by sleep and behavioral sciences in the last decades, the assessment of sleepiness is one of the remaining challenges. The lack of practical tools for objective sleepiness testing has become one of the critical barriers to reducing the threats

of increased sleepiness to public health, productivity, and safety (e.g., Mullington et al., 2011). Unlike subjective (KSS) scoring, the existing objective sleepiness and performance tests do not allow quick assessment of the apparent level of sleepiness in absolute numbers related to understandable terms (Horne & Burley, 2010). The duration of any of testing procedures that were earlier proposed for objective sleepiness measurement cannot be reduced to less than 5-minute time interval (see Introduction). However, our earlier studies showed that attenuation of alpha-based EEG indexes right after closing the eyes strongly correlates with the feeling of sleepiness and shortened latency to the EEG signs of sleep onset. Therefore, such quick EEG measurement can allow the reduction of duration of the procedure to a minute or so (Putilov & Donskaya, 2014b). Data provided by such measurement can be, in particular, used for calculating such objective sleepiness index as SDCS (Putilov, in press). In the resent study the weights for calculating SDCS were 16 coefficients of correlation of the time course of subjective sleepiness with the time courses of single-Hz powers. The time course consisted of either 15 or 9 time points (the first and second studies, respectively). Irrespective of the dataset, the occipital SDCS of -0.2, 0, and 0.2 provided the borderline scores for subdivision of the alertness-sleepiness continuum into four different sub-states (alert, neither, dozy, and sleepy, respectively) in the vast majority of two study participants.

Thus, the present results suggest that the findings on SDCS were irrelevant to the source of weights for calculation of scores (internal vs. external). Such results supported reliability of the SDCS-based testing of alertness-sleepiness. They, in particular, indicated that the weights for scoring of drowsy component can be obtained from studies with rather distinct sleep deprivation protocols. Namely, these weights were found to lead to very similar scores (Table 2) in despite of remarkably different clock times chosen for subjective sleepiness evaluation. Furthermore, the weights derived from one dataset were found to be applicable for another dataset (i.e., led to similar classifications of sleepiness sub-states) in despite of additional very remarkable difference between these two sleep deprivation experiments in the finally reached levels of sleepiness.

In general, the present results indicate that i) objective sleepiness testing can be facilitated by the quantitative analysis of the resting EEG signal recorded at the first minute following eyes closing, ii) this analysis can include calculation of occipital SDCS as a zero-crossing criterion of sleepiness onset, and iii) the results on SDCS open

a new perspective of development of objective, reliable, simple and quick (app. one-minute) testing of sleepiness.

It is necessary to stress that, in despite of a clear tendency of sleep deprived participants to underestimate the negative effects of sleep loss, the comparison of the results on KSS based SDCS with the results on SL that is a traditional measure of sleepiness showed that, due to close correlation between the time courses of subjective and objective sleepiness, this score can be regarded as a reliable objective rather than subjective measure of sleepiness. In other words, the effect of underestimation of sleepiness was evident only for subjective (KSS) scoring, whereas data on SDCS and SL did not reveal this effect.

However, it is also necessary to stress that the present results cannot be regarded as one of the final steps on the way of development of a quick and simple sleepiness test. First, the present analysis included such additional procedures as manual artifact detection, ipsatization of each participant power values, averaging over several scores obtained from the same participant, etc. Second, the analysis was limited to only two derivations and did not test contribution of other derivations and such additional indicators of sleepiness as changes in coupling between activities of distant cortical areas. Third, it remains to be explored in the future research whether yes-or-no criterion of transition between alert and sleepy states can be applied without preliminary subject-specific calibration in other studies including the field studies of shift workers, clinical studies of patients with sleep disorders, studies of individuals without prominent alpha rhythms and people representing distinct ages.

## 5 CONCLUSION

The spectral component scoring of the EEG signals recorded during the first minute of eyes closed relaxation provides a possibility of quantifying sleepiness and distinguishing between alertness and sleepiness states in absolute terms, i.e., by using zero-crossing criterion of sleepiness onset.

## ACKNOWLEDGEMENT

The experimental studies were supported by the Russian Foundation for Basic Research (grants 07-06-00263-a, 10-06-00114-a, and 13-06-00042-a), and the Russian Foundation for Humanities (grants 06-06-00375-a, 12-06-18001-e, and 15-06-10403-a).

## REFERENCES

- Akerstedt, T. & Gillberg, M. 1990. Subjective and objective sleepiness in the active individual. *Int J Neurosci* 52: 29–37.
- Alloway, C.E.D., Ogilvie, R.D. & Shapiro, C.M. 1997. The alpha attenuation test: assessing excessive daytime sleepiness in narcolepsy-cataplexy. *Sleep* 20: 258–266.
- Berka, C., Levendowski, D.J., Olmstead, R.E., Popovic, M.V., Cvetinovic, M., Petrovic, M.M., Davis, G., Lumicao, M.N. & Westbrook, P. 2004. Real-time analysis of EEG indices of alertness, cognition and memory with a wireless EEG headset. *Int J Hum Comput Interact* 17: 151–170.
- Carskadon, M.A., Dement, W.C., Mitler, M., Roth, T., Westbrook, P.R. & Keenan, S. 1986. Guidelines for the Multiple Sleep Latency Test (MSLT): a standard measure of sleepiness. *Sleep* 9: 519–24.
- Czeisler, C.A. 2011. Impact of sleepiness and sleep deficiency on public health: utility of biomarkers. *J Clin Sleep Med* 7(5 Suppl): S6–S8.
- Donskaya, O.G., Verevkin, E.G., Putilov, A.A. 2012. The first and second principal components of the EEG spectrum as the indicators of sleepiness. *Somnologie* 16: 69–79.
- Doghramji, K. 1999. Maintenance of Wakefulness Test. In S. Chokroverty (ed). *Sleep disorders Medicine*: 231–236. Woburn MA: Butterworth-Heinemann.
- Frigo, M. & Johnson, S.G. 2005. The design and implementation of FFTW3. *Proc IEEE* 93: 216–231.
- Herrmann, U.S., Hess, C.W., Guggisberg, A.G., Roth, C., Gugger, M. & Mathis, J. 2010. Sleepiness is not always perceived before falling asleep in healthy, sleep-deprived subjects. *Sleep Med* 11: 747–751.
- Horne, J.A. & Burley, C.V. 2010. We know when we are sleepy: subjective versus objective measurements of moderate sleepiness in healthy adults. *Biol Psychol* 83: 266–268.
- Kaplan, K.A., Itoi, A. & Dement, W.C. 2007. Awareness of sleepiness and ability to predict sleep onset: Can drivers avoid falling asleep at the wheel? *Sleep Med* 9: 71–79.
- Motamedi-Fakhra, S., Moshrefi-Torbat, M., Hill, M., Hill, C.M., Paul, R. & White, P.R. 2014. Signal processing techniques applied to human sleep EEG signals—A review. *Biomed Signal Process Control* 10: 21–33.
- Mullington, J.M., Czeisler, C.A., Goel, N., Krueger, J.M., Balkin, T.J., Johns, M. & Shaw, P.J. 2011. Panel discussion: current status of measuring sleepiness. *J Clin Sleep Med* 7(5 Suppl): S22–S25.
- Putilov, A.A. 2011. Principal components of electroencephalographic spectrum as markers of opponent processes underlying ultradian sleep cycles. *Chronobiol Int* 28: 287–299.
- Putilov, A.A. 2014. When does this cortical region drop off? Principal component structuring of the EEG spectrum yields yes-or-no criteria of local sleep onset. *Physiology & Behavior* 133: 115–121.
- Putilov, A.A. 2015. Rapid changes in scores on principal components of the EEG spectrum do not occur in the course of “drowsy” sleep of varying length. *Clin EEG & Neurosci* 46 147–152.
- Putilov, A.A. (in press) Can we evaluate sleepiness quickly, directly, objectively, and in absolute terms? Results of scoring of alert/drowsy components of the resting EEG spectrum. *Somnologie* 2015; 19(3).
- Putilov, A.A. & Donskaya, O.G. 2013a. Rapid changes in scores on the two largest principal components of the electroencephalographic spectrum demarcate the boundaries of drowsy sleep. *Sleep Biol Rhythms* 11: 154–164.
- Putilov, A.A. & Donskaya, O.G. 2013b. Construction and validation of the EEG analogues of the Karolinska sleepiness scale based on the Karolinska drowsiness test. *Clin Neurophysiol* 124: 1346–1352.
- Putilov, A.A. & Donskaya, O.G. 2014a. Calibration of an objective alertness scale. *Int J Psychophysiol* 94: 69–75.
- Putilov, A.A. & Donskaya, O.G. 2014b. Alpha attenuation soon after closing the eyes as an objective indicator of sleepiness. *Clin Exp Pharmacol Physiol* 41: 956–964.
- Putilov, A.A., Donskaya, O.G. & Verevkin, E.G. 2012. Quantification of sleepiness through principal component analysis of the EEG spectrum. *Chronobiol Int* 29: 509–522.
- Quan, S.F., Shaw, P.J., Naidoo, N., Haeggström, E., Krueger, J.M. & Church, G.M. 2011. Panel discussion: can there be a biomarker for sleepiness? *J Clin Sleep Med* 7(5 Suppl): S45–S48.
- Roach, G.D., Dawson, D. & Lamond, N. 2006. Can a shorter psychomotor vigilance task be used as a reasonable substitute for the ten-minute psychomotor vigilance task? *Chronobiol Int* 23: 1379–1387.
- Roehrs, T. & Roth, T. 1992. Multiple Sleep Latency Test: technical aspects and normal values. *J Clin Neurophysiol* 9: 63–67.
- Smith, S., Carrington, M. & Trinder, J. 2005. Subjective and predicted sleepiness while driving in young adults. *Accid Anal Prev* 37: 1066–1073.
- Stampi, C., Stone, P. & Michimori, A. 1995. A new quantitative method for assessing sleepiness: The alpha attenuation test. *Work & Stress* 9: 368–376.
- Subasi, A., Kiymik, M.K., Akin, M., Erogul, O. 2005. Automatic recognition of vigilance state by using a wavelet-based artificial neural network. *Neural Comput Appl* 14: 45–55.
- Tremaine, R., Dorrian, J., Lack, L., Lovato, N., Ferguson, S., Zhou, X. & Roach, G. 2010. The relationship between subjective and objective sleepiness and performance during a simulated night-shift with a nap countermeasure. *Appl Ergon* 42: 52–61.
- van Dongen, H.P., Baynard, M.D., Maislin, G. & Dingus, D.F. 2004. Systematic inter-individual differences in neurobehavioral impairment from sleep loss: evidence of trait-like differential vulnerability. *Sleep* 27: 423–433.

# The illusion of a blackbody at the human ear and the human temperature measurement

António Cardoso

*Materiais e Produtos, CATIM—Centro de Apoio Tecnológico à Indústria Metalomecânica, Porto, Portugal*

**ABSTRACT:** The insertion of probes in the human ear or any other technique involving contact thermometers, or the use of the inner canthus for temperature measurement, using infrared thermal imagers, with the purpose of obtaining a reference to determine the core temperature in humans, can be compromised by a wide number of factors that may lead to faulty results.

In this paper, an innovative approach based on the use of a thermographic camera and on the characteristics of the external auditory meatus is proposed. This new approach enables to solve the problems associated with the previously proposed methods, adding new possibilities namely allowing the precise measurement of the core temperature in humans in a quick and simple way, without body contact, opening the possibility of setting self-screening stations, working without operators in fully automatic mode.

## 1 INTRODUCTION

The majority of the methods used to determine the core temperature in humans are affected by a number of factors that apart other clinical considerations may yield to faulty measurements.

One method to determine the core temperature in humans is the measuring of the temperature in the eye corner (inner canthus). The ophthalmic artery, located beneath the inner cantus, is supplied directly by the internal carotid artery, reason why screening thermography utilizes it to determine the core temperature in humans. However, and by the same reason, the superficial temperature measured at the inner cantus by the thermographic camera may be affected by other factors like some cerebrovascular diseases that affect the internal carotid artery. These diseases may cause a diminishing in the blood flow, and consequently, in the amount of heat dissipated at the inner canthus, leading to a drop in the superficial temperature. Another factor that may affect reading is the accumulation of gound, the mucous that is discharged by the eyes while sleeping, which accumulates in their corners, at the inner canthus, enhancing temperature dropping. In thermographic terms, the inner canthus area has a very well determined emissivity: 98%, which implies the need to determine the reflected apparent temperature to get an accurate measurement.

The ear thermometer is another non-contact method that, however, requires the use of a disposable plastic probe cover, since mechanical support is needed which involves physical contact with the ear. Therefore, it poses an economic and ecological problem, as each measurement requires the use of

another plastic probe for hygienic reasons, generating plastic waste. It also presents some emissivity related problems, derived from placement issues. To get an accurate measurement the ear thermometer must be directed to the ear drum or otherwise, if the reading angle is less than 45°, it will be reading the radiation emanated from the walls of the external auditory meatus. At the interior of the external auditory meatus, the emissivity has a value of 98%. When a probe at room temperature is inserted, its own reflected radiation will affect the reading, causing the disappearance of the blackbody illusion and introducing an error in the measurement, even if the probe is properly aimed at the ear drum. Only after three or four readings it produces a correct measurement value.

The insertion of a Pt100 or of a thermocouple on the human ear is not a good solution either, because the lack of air flow inside the external auditory meatus, and therefore the associated good level of insulation provided by still air, contributes to a faulty measurement. In fact, still air is an effective insulator because it eliminates convection and has low conduction. Furthermore, in this technique the metallic wires of the thermometer are another problem. Metals are good thermal conductors and thus the wires act as thermal drains, compromising the final measurement value.

A range of other methods based on contact thermometers is used in clinical scenarios. Apart hygienic measures that makes mandatory the disinfection of the thermometers to avoid the propagation of diseases like flu, hepatitis, herpes, tuberculosis and other bacterial and viral related diseases, the temperature given by contact

thermometers is the internal sensor temperature and not the patient's temperature.

Even the clinical temporal artery thermometers involve contact with the patient, and thus they are affected by the same problems. Furthermore, the measurement accuracy is affected by the positioning of the thermometer. If the thermometer is not properly positioned over the temporal artery, the reading will be inaccurate.

Contact thermometers always measure their own temperature and not the temperature of the object they are in contact with. Equation (1) represents how temperature reading evolves over time. When  $t \rightarrow \infty$  the thermometer temperature will eventually reaches the temperature of the patient.

$$T_{\text{sensor}} = (T_{\text{patient}} - T_{\text{room}}) \cdot \left( 1 - e^{-\frac{t}{\tau_{\text{sensor}}}} \right) + T_{\text{room}} \quad (1)$$

The biggest advantage of thermographic cameras is that they don't measure the temperature. They calculate it. Indeed, thermographic cameras are imaging devices where each pixel represents an amount of energy radiated towards the camera by a surface. This energy has two components: the emitted radiation, and the reflected radiation. Since the majority of the surfaces belong to opaque objects, the transmissivity component is usually considered to be zero (this is indeed the case of the application described herein). In this case, Equation (2) is used to express the relations among the relevant physical quantities.

$$W = \sigma \cdot (\varepsilon \cdot T_{\text{surface}}^4 + (1 - \varepsilon) \cdot T_{\text{reflected}}^4) \quad (2)$$

In (2)  $W$  represents the magnitude of the pointing vector in the surface,  $\sigma$  the Stefan-Boltzmann constant,  $\varepsilon$  the emissivity component and  $T_{\text{reflected}}$  the reflected apparent temperature that has to be determined (ISO 18434-1 2008).

In this paper, it is described a new thermographic method to accurately measure human body temperature. This method is based on the thermographic measuring of the external auditory meatus temperature. It is shown that this method provides accurate results, overcoming the problems faced by previous contact and contact-less measurement methods, while enabling quick temperature measurements, broadening its application possibilities.

## 2 TRADITIONAL HUMAN TEMPERATURE MEASUREMENT METHODS

### 2.1 Clinical thermometers

The clinical thermometer is a contact thermometer composed by a bulb, the sensor, and a transparent

tube with a scale, the "display". The increase in temperature forces the expansion of a liquid placed inside the tube, thus indicating the temperature of the patient's body. In the past, the liquid inside clinical thermometers was mercury, a poisonous heavy metal. Due to these concerns, mercury has been phased out in favour of alcohol-filled or galinstan-filled thermometers. Lately, these thermometers are being substituted by digital ones with thermistors as the sensing element. They can be used in cavities of the human body covered with mucous membranes, like the rectum (rectal) or mouth (oral), or in the armpit (axillary), the last one being the most usual and also the least accurate of the three.

### 2.2 Esophageal probes

Esophageal probes are commonly used when a constant monitoring is mandatory, like in Intensive Care Units (ICU) and Continuing Care Units (CCU). Usually, a Negative Temperature Coefficient (NTC) thermistor or a thermopile is used as sensing element. Like the clinical thermometers, esophageal probes are contact thermometers. Besides, its placement in the patient's esophagus causes discomfort.

### 2.3 Ear thermometers

As mentioned in the Introduction section, one of the main problems with ear thermometers is the narcissism of the probe; the amount of radiation emitted by the probe itself and reflected on the ear drum surface whose radiation is intended to be measure. Due to the scale factor and other constraints, the probe has to be pointed towards the ear drum with an angle higher than 85°, because the sensor has roughly the same size of the object measured and is placed at a distance of less than 3 cm. To get more accurate readings the measurement may not be done immediately after inserting the probe. The probe is at room temperature and after insertion it is necessary to wait a certain amount of time to allow the establishment of the thermal equilibrium between the patient and the probe.

### 2.4 Temporal artery thermometers

As already mentioned, temporal artery thermometers are contact thermometers with an additional accuracy problem related with positioning. If not properly placed over temporal artery the reading is inaccurate. Furthermore, like the ear thermometer each measurement consumes a disposable cap.

### 2.5 Inner canthus

During the epidemic of Severe Acute Respiratory Syndrome (SARS) it was observed that infected

patients experienced an increase of temperature on the eyes zone (Ring 2012). This fact, in combination with the morphology of the ophthalmic artery led to considerer the inner canthus a possible spot to estimate core temperature in human beings. Indeed, the ophthalmic artery, fed directly by the internal carotid artery, is located beneath the inner cantus where it splits in two, the nasal artery and the frontal artery (Guyton & Hall 2000). Despite these characteristics, the inner canthus has a series of snags that compromise, in thermographic terms, the accuracy of the calculations done to determine the core temperature (Wood 1964).

Human skin is known to have an emissivity of 98% (Ring 2012), regardless of the color of the skin. This feature imposes the need to estimate the reflected apparent temperature. If this is not done, a reference source with a very well-known emissivity and a very well defined temperature has to be put in the field of vision of the thermographic camera to allow for the computation of the reflected apparent temperature. As a rule, this reference source has to be on the same plane of the inner canthus, in order to get a correct reference measurement. However, filling 75% of the field of vision of the camera (ISO/TR 13154 2009) with the face of the patient to ensure a Measurement Field of View (MFOV) smaller than the inner canthus size, compromise in several ways the estimation of the reflected apparent temperature. First because the rule mentioned before cannot be kept, and second because the narcissism effect of the camera, and possibly of the operator over the scene, is huge.

Since the outer layer of the skin has no irrigation, heat flow can only be achieved in a passive way through thermal conduction. This outer layer acts as a low-pass filter and the temperature calculated is always below or above the core temperature depending on room temperature—if this is below or above core temperature. Ground accumulation at the inner canthus may amplify this effect.

Cerebrovascular diseases may also compromise readings because they may cause a diminishing in the blood flow in the ophthalmic artery and consequently in the amount of heat dissipated at the inner canthus, leading to a drop in the superficial temperature and promoting false negatives as a result (Ring 2012). The opposite effect takes place when a healthy person makes physical efforts. The increase of the muscular metabolism during exercise generates heat that needs to be dissipated in order to maintain body heat balance. To help dissipate this excess heat, capillary blood vessels dilate, increasing facial temperature and thus the possibility of producing false positives.

### 3 EXTERNAL AUDITORY MEATUS

The set formed by the external auditory meatus—the external ear canal—and the eardrum, Figure 1, is no more than a mechanical waveguide terminated by an adapted mechanical load. This condition is essential to not have echo formed within our ears. This condition is complemented by another one: the isothermal nature of the external auditory meatus cavity assures that the air inside it has an average speed equal to 0. This second condition ensures the absence of the Doppler Effect in the external auditory meatus, which can only be achieved when there are no temperature differences, thus ensuring that there are no static pressure differences. If no pressure differences are observed, no flow will exist, no mass transport from one side to the other. On the other hand, it also ensures that the speed of the sound propagation does not change along the external auditory meatus, from the moment the wave enters it until it reaches and is absorbed by the eardrum.

The canal and the eardrum have the same coating characteristics as both are coated with the same material, skin. Therefore, there are no differences on the material in contact with the wave while it travels through the waveguide and until being absorbed by the charge, the eardrum.

From an electromagnetic point of view, the geometry of the cavity creates in its opening the optical illusion of a blackbody surface. This phenomenon prevents us from seeing into the cavity, since the electromagnetic waves, after entering the canal, and after successive reflections, are completely absorbed within the cavity. This means that the incoming electromagnetic radiation gets completely trapped inside the cavity. Consequently, this makes the entrance of the external auditory meatus a dark spot.

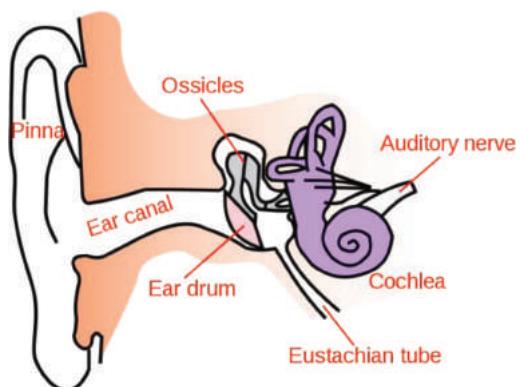


Figure 1. External auditory meatus (Gray & Lewis 1918).

However, and according to the law of Stefan-Boltzmann, all bodies with a temperature above absolute zero emit radiation. So there is radiation coming out of the external auditory canal. Since the external auditory meatus and the eardrum are at the same temperature, the set acts as a heat source with a single temperature. Taking into account that all the radiation that enters the external auditory meatus is absorbed inside, there are no reflected radiations coming from the external auditory meatus. Therefore it is possible to correlate the radiation radiated through the external auditory meatus opening with the temperature of the heat source and thus to measure the temperature of the eardrum located 3.5 cm away from the entrance of the external auditory canal.

In sum, due to its geometrical configuration and characteristics, an illusion of blackbody is always present at the interface between the external auditory meatus and the exterior.

As in any isothermal cavity, the meatus is completely surrounded by a volume, the head. The cavity has a truncated conical shape with a ratio of 3.5 between the length and the opening diameter (1 cm  $\times$  3.5 cm) (Guyton 2000), with the eardrum in the truncated end.

Like a blackbody of room temperature, all radiation entering the meatus is trapped inside. The radiation that exits the external auditory meatus is only the radiation that is radiated inside. The external auditory meatus, when observed from the outside with a thermographic camera appears as an isothermal surface. No details can be seen inside. Only a flat surface at the same temperature (same level of radiation) can be observed, as seen in Figure 2, an “optical” illusion. This virtual surface, by definition, has an emissivity of 1. No calculus or estimation of reflected apparent temperature and emissivity are needed.

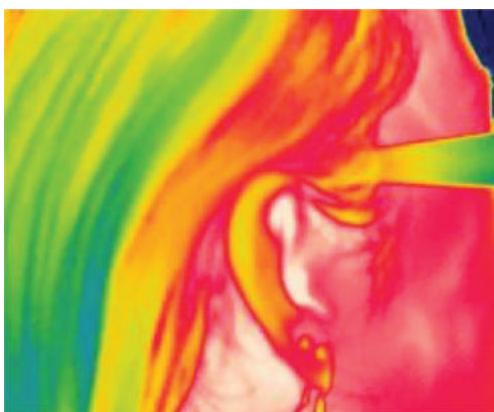


Figure 2. Thermogram of the ear.

Unlike ear thermometers which are affected by crooked external auditory meatus or wax, because they have to “see” the eardrum, thermographic cameras do not have their accuracy compromised by these issues, provided that they are not visible from the outside. If they are not seen from the outside, crooked external auditory meatus or wax accumulation are no longer a problem.

Therefore, none of the previously identified sources of error interfere with the thermographic measurement, thus guaranteeing an accurate and quick core temperature measurement. In all situations, the external auditory meatus, when observed from the outside, always have an emissivity of, or very close to, 100% (99,971% (Saunders 2009)), relieving the measurement from the determination of the reflected apparent temperature. For that reason,

$$W = \sigma \cdot T^4 \quad (3)$$

which may be derived from (2) when emissivity tends to 1.

#### 4 SOURCES OF UNCERTAINTY

The MFOV poses the greatest concern when using thermographic cameras to perform temperature measurements. MFOV corresponds to the smallest object whose temperature the camera can measure. Typically its value varies between 3 (3-by-3 pixels = 9 pixels) and 5 (5-by-5 pixels = 25 pixels) times the size of each pixel in the camera image sensor. If the MFOV is 3 then 9 pixels are need to measure the object temperature. Therefore, the minimum measurement size depends on the distance—closer the object, smaller the measurement area. However, the distance to the object may not be reduced to zero as there is the need to focus the camera. Furthermore, changing the lenses changes the MFOV and the minimum focus distance.

To be able to correctly read the temperature at the external auditory meatus the thermographic camera needs to have an MFOV smaller than the diameter of the meatus—the diameter of the external ear canal. In addition, the camera needs to have the emissivity parameter selectable to 1 and preferably the possibility to be programmed to find and measure the maximum temperature within a defined area.

The thermographic cameras available on the market have an MFOV smaller enough to allow in the limit an accurate temperature reading in adult patients. However, in pediatric practice, as seen in Figure 3, a camera with a detector of 320  $\times$  240 pixels may be necessary. Cameras with detectors of 120  $\times$  120 pixels, a field of view of



Figure 3. Child being monitoring during sleep.

25° and minimum focus distance of 12 cm, giving a minimum measurement size of 2.31 mm, are examples of ideal cameras for all types of patients (Cardoso 2012a, b).

Another source of uncertainty is the Noise Equivalent Temperature Difference (NETD), defined as the amount of infrared radiation required to produce an output signal equal to the systems own noise. In practice, it is the minimum level of thermal sensitivity exhibit by the camera. Currently, typical values for NETD are below, or equal to, 100 mK, allowing measurement resolutions of, at least, 100 mK.

In stand still stations, depth of field, out of focus and positioning of the camera in relation to the spot where persons should stand may also be sources of problems. The place where a person stands must be close enough to minimize the ratio MFOV/meatus diameter, and at the same time far apart to maximize depth of field. Measuring at a distance too close to the minimal focus distance may introduce uncertainty, because any focus error may not be compensated by the depth of field.

Further studies are needed to establish the contributions of each one of these uncertainty sources and to implement proper operation practices with the goal to reduce or compensate them. In the limit, if a proper operation practice is achieved, only the single and unavoidable NETD source of uncertainty will remain.

## 5 APPLICATIONS

The simplicity of the measurement procedure effectively broadens its number of applications. Traditional clinical uses like physicians' offices, ICUs, CCUs, Cardiac/Coronary Care Units, Emergency Rooms or in the nursing practice in patients wards are just a small number of examples where

hand held thermographic cameras can be used to, in less than 5 seconds, measure patient's temperature without annoying or waking up her/him. In cases where a tight control of the temperature is required, like in operation theatres, measurements can be done continuously during surgery without having the patient wired.

However, other no traditional applications where human control temperature may be important may also be envisaged. That's the case of airports, train stations or other wide open public spaces, like receptions, where fixed stations with or without an operator can be installed to monitor passers'-by temperature. One possibility is to add thermographic cameras to ePassport gates, automated self-service passport barriers recently installed in airports worldwide. To pass the barrier, a person needs to stand still in front of a camera for just some seconds, before being authorize to walk through the gate. At the same time, she/he may be warned to pull back her/his hair from in front of her/his ears to allow an accurate temperature measurement.

## 6 CONCLUSIONS

The approach proposed in this paper to measure human temperature based on the use of a thermographic camera and on the characteristics of the external auditory meatus has several advantages over other traditional methods. No disposable parts are needed and therefore it does not produce waste. It is a completely non-intrusive method and therefore it does not cause any discomfort to the patient. It is a predictable, more accurate method than the other previously described methods, if a set of proper practices are implemented.

Currently, the longest flight takes less than 24 hours long and this is the time needed for a H5N1, or other virus, to travel across the world. Long-haul flights are therefore a major concern for the role they can play in disease spreading. At airports, the current method implies the removal of glasses to observe the temperature in the inner canthus. Since a 98% emissivity should be expected, errors are to be consider. False positives are annoying for those spotted, but the major concern is with false negatives. In airports, where the access from the plane to the terminal is done exposed to the elements, the patient's face can chill in frosty weather and the alarm at the thermographic camera may not sound, failing a potential positive (Chan et al. 2004).

While further studies are needed to reduce uncertainty sources to a minimum and to implement proper operation practices, this method seems to be the most accurate and promising and

able to cover a broader spectrum of applications when compared to existing methods.

## REFERENCES

- Cardoso, A. 2012a. Human Temperature Measurement with Infrared Cameras. 7th International Infrared Forum infraR&D. Hannover, Germany.
- Cardoso, A. 2012b. Screening fever, a new approach. XII EAT Congress on Thermology. Porto, Portugal.
- Chan L.S., Cheung G.T., Kumana C.R. & Lauder I.J. 2004. Screening for fever by remote-sensing infrared thermographic camera. *J Travel Med.* 2004 Sep-Oct;11(5):273–9.
- Gray, H. & Lewis, W.H. 1918. Anatomy of the human body. Philadelphia: Lea & Febiger.
- Guyton, A.C. & Hall, J.E. 2000. Textbook of medical physiology London: W.B. Saunders.
- ISO 18434-1 2008. Condition monitoring and diagnostics of machines. International Organization for Standardization, Thermography—Part 1: General procedures.
- ISO/TR 13154 2009. Medical electrical equipment—Deployment, implementation and operational guidelines for identifying febrile humans using a screening thermograph. International Organization for Standardization.
- Ring, F. 2012. Fever Detection in Children with Infrared Thermal Imaging using the new ISO Recommendations. 7th International Infrared Forum infraR&D. Hannover, Germany.
- Saunders, P. 2009. Calibration of Low-Temperature Infrared Thermometers. Measurement Standards Laboratory of New Zealand. Techniqueal Guide 22.
- Wood, E.H. 1964. Thermography in the Diagnosis of Cerebrovascular Disease: Preliminary Report. *Radiology.* September 1964, 83:540–542.

# Video-based Turkish Sign Language recognition systems

M. Aktaş & E. Battini Sonmez

*Computer Engineering Department, İstanbul Bilgi University, İstanbul, Turkey*

**ABSTRACT:** Automatic recognition of sign language will be extremely beneficial for hearing impaired people, by allowing them to easily communicate with the outside world. Sign language is a non-verbal and visual language, which evolved detached from the spoken language and it is based on the simultaneous contribution of both temporal and spatial configurations. Even if it is spread all over the world, sign language is not a universal language and every country needs to build its own automatic system. This paper challenges the issue of the development of automatic video-based Turkish Sign Language (TSL) recognition systems; it presents two computer vision-based automatic systems for TSL recognition and compares their performances. Since we are using a newly released TSL database, this paper has also the side-effect to divulge the database to the research community.

**Keywords:** video-based sign recognition; Turkish Sign Language; spatial-temporal features

## 1 INTRODUCTION

Sign Language (SL) is the most important communication channel among hearing impaired people; however, since the majority of subjects outside the deaf community do not know this language, the communication between deaf and hearing societies is still too weak. Automatic systems capable to translate SL into text and/or speech and vice versa could bridge the existing gap between deaf and hearing people.

Sign language is characterized by manual and non-manual features; examples of non-manual parameters are gaze, facial expression, and mouth, head and trunk movements. Since sign language evolved and separated from the spoken language, their grammars are fundamentally different from each other.

Furthermore, to make the problem more difficult, every country has its own sign language, and, therefore, there is also the need to build different systems, one for every country.

In 1995, Waldron and Kim (Waldron & Kim 1995) published one of the first studies on automatic American Sign Language; since then, different researchers tackled this issue, but the problem is still open, because it is particularly challenging due to the complexity of translating manual and non-manual characteristics into a word or sentence.

In order to easier the recognition task, some signers wear special gloves with motion-sensors; in the following, we will refer to those methods as ‘data gloves-based systems’. In alternative ‘computer vision-based systems’ rely on classical image

processing techniques to extract informative features and classify body movements, when a deaf person makes a sign. We believe that a system designed to work in a public place must be as natural as possible and, because of this, we decided to build different computer vision-based systems and to compare their performances.

To summarize, this paper challenges the issue of the development of automatic video-based Turkish Sign Language (TSL) recognition systems; it uses the Turkish Sign Language database recently introduced by Memis and Albayrak (Memiş & Albayrak 2013).

The rest of the paper is organized as follow: Section 2 discusses the related work; Section 3 introduces the TSL database used in this paper; Section 4 gives an overview of the proposed systems; Section 5 details experiments. Finally, conclusions and future work are drawn in Section 6.

## 2 LITERATURE REVIEW

Automatic sign language recognition systems have been investigated since few decades; the most common classification techniques used in this field are Hidden Markov Models (HMM) and Neural Networks (NN).

In 1997, (Vogler & Metaxas 1997) Vogler and Metaxas used HMM for continuous American Sign Language (ASL) recognition and they were able to recognize 53 signs out of a completely unconstrained sentence; in 1999, the same authors were able to recognize ASL sentences with 22 signs

based on ASL phonemes; and, in 2004, (Vogler & Metaxas 2004), they extended their study to parallel HMM to implement a multi-channel ASL system, with sensory gloves. In 2001, (Wang, Gao, & Xuan 2001), Wang et al. used HMM to design a data gloves based system for Chinese Sign Language (CSL) recognition; in 2003, (Allen, Asselin, & Foulds 2003) Allen et al. developed an Artificial Neural Network (ANN) to recognize 24 ASL letters; in 2003, Chen et al. used HMM to implement a Chinese Sign Language recognition system, with sensory gloves; in 2004, (Wang, Sarawate, & Leu 2004), Wang et al. used ANN and HMM to design an ASL gesture recognition systems with sensory gloves; in 2005, (Mohandes & Deriche 2005), Mohandes and Deriche used HMM to design a system for Arabic Sign Language recognition; in 2007, (Munib, Habeeb, Takruri, & Al-Malik 2007), Munib et al. used Hough transform and neural networks for ASL recognition; in 2009, Lee and Tsai used 3D data and Neural Networks to recognize Taiwan sign language.

The study of the Turkish Sign Language (TSL) is still at the very beginning: in 2005 Haberdar and Albayrak (Haberdar & Albayrak 2005) used local features with a HMM based system for TSL recognition and reached 95.5% recognition rate over 50 sign words; in 2011, (Altun & Albayrak 2011), Altun and Albayrak, used Hough transform to create a Turkish finger-spelling recognition system. In 2013, (Memiş & Albayrak 2013), Abbas and Albayrak introduced to the research community a new TSL database together with a benchmark algorithm which classifies 1002 videos representing 111 words and short sentences of the Turkish Sign Language; they reached their best performance of 91.52% using both RGB and depth videos.

The aim of this paper is to contribute to the study of automatic TSL recognition systems by using the recently introduced TSL database of Abbas and Albayrak; in this initial study, we extend the benchmark experiment of Abbas et al by working with all videos of section 1.

### 3 THE TURKISH SIGN LANGUAGE DATABASE

The TSL database of Abbas et al. consists of a total of 6300 videos divided into two sections, storing 1700 and 1450 videos, respectively; all videos are in both RGB and depth image formats. The label of every video records information related to the section, the category, the words (class), and the signers.

The entire database has a total of 236 words and sentences, which are grouped into the 5 categories of “daily life” (category 1, 24 words),

“time” (category 2, 61 words), “jobs” (category 3, 58 words), “associations and places” (category 4, 39 words), and “school” (category 5, 54 words). Every word or sentence is played in a short video by a professional deaf signer; 4 different people (or actors) mimic all words.

The database is divided into 2 sections, the first section stores words and sentences belonging to category 1 (storing 75 videos of player 1, 72 videos of player 2, 72 videos of player 3 and 75 videos of player 4), category 2 (with 184 172 185–184 videos respectively played by actor 1, 2, 3 and 4) and category 3 (storing 171 172 171–167 videos respectively played by player 1, 2, 3 and 4); the second section stores words and sentences belonging to category 1 (with 72 videos of player 2 and 72 videos of player 4), category 4 (storing 119 videos of player 2 and 125 videos of player 4) and category 5 (with 163 videos of player 2 and 168 videos of player 4). To summarize section 1 stores a total of 1700 videos and section 2 contains 1450 videos, respectively.

Table 1 shows the distribution of the 1700 videos belonging to section 1 among the 4 players and the 3 categories.

Table 1. Distribution of the 1700 videos in section 1.

Section 1

Players	Category 1	Category 2	Category 3
1	75	184	171
2	72	172	172
3	72	185	171
4	75	184	167

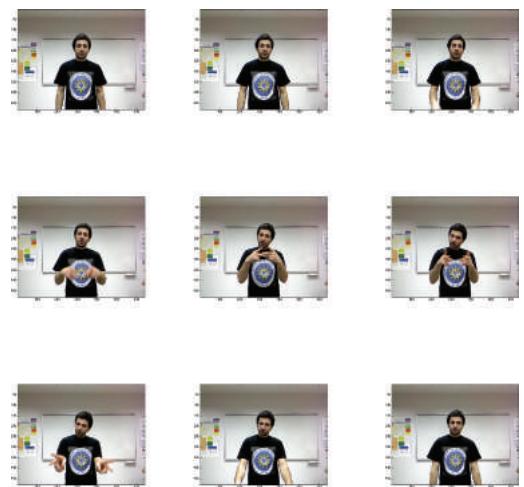


Figure 1. Selected frames of the sign ‘excuse me’.

Like Abbas et al. we worked only with section 1, which stores words and sentences of daily life's conversation (e.g. excuse me, I'll see you then, goodbye, good morning, where, when, what, etc.) as well as time's related terminology (e.g. August, month, sometimes, this week, today, in the afternoon, etc.), and job's key words (e.g. cook, soldier, lawyer, shoe repairer, hairdresser, etc.).

Figure 1 shows selected frames of the sentence 'excuse me'.

The database was recorded using the Microsoft Kinect camera, which gave the possibility to record RGB images and depth images; key landmarks of the body posture, were not recorded.

## 4 THE PROPOSED SYSTEMS

Computer vision-based systems for automatic recognition of TSL rely on classical image processing techniques to extract informative features and classify body movements, when a deaf person makes a sign.

In this paper we compare the performance of two systems, both of them are an extension of the benchmark algorithm proposed by Abbas et al. together with the new TSL database, (Memiş & Albayrak 2013).

### 4.1 Reading and preprocessing

We used all the 1700 videos of section 1, the average number of frames per video is 43, their initial size is  $640 \times 480$ . Due to out of memory problems, it has been necessary to extract and memorize only the relevant information already during the reading step.

First of all, RGB frames were converted into gray frames, after that, a naive cropping method was developed and applied on the frames of each video. The aim of this cropping is to eliminate useless background information, and keep only the player's body figure; this results in frames of size  $361 \times 427$ . The input and output frames of naive cropping function can be seen in Figure 2.

### 4.2 Feature extraction

Every two consecutive frames in a video produce an 'image of differences', which is calculated using the following formula:

$$I_{D_i} = I_i - I_{i+1}, \forall i = 1, 2, \dots, n$$

where  $I_i$  represents the  $i$ th frame of the video, and  $I_{D_i}$  stores the difference of images; hence,  $n - 1$  'image of differences' are obtained from a video with  $n$  frames. The logic behind subtracting



[a]



[b]

Figure 2. A sample gray frame before (a) and after (b) moving the useless background information.

successive frames is to be able to determine the area where the motion of hands is significant because, if the values of some pixels change from one frame to another, these form a lighter area within the difference of images,  $I_{D_i}$ .

Subsequently, the difference image is converted into a binary image by a threshold function; the value of the threshold is database depended and it has been fixed, manually, to 50. Figure 3 shows the thresholded image of differences.

Finally, all thresholded images of differences are summed up to calculate a spatial-temporal feature, named by Abbas et al. the accumulated motion image:

$$A_{v_i} = \sum_{i=1}^{n-1} I_{th_i}$$

where  $I_{th_i}$  stores the  $i$ th thresholded image of difference of the video. The accumulated image,  $A_{v_i}$ , represents the whole video in one frame; it is the spatial-temporal feature of that video in the form of a matrix.

In an alternative system, we extracted features in the frequency domain. The only difference of this



Figure 3. A sample thresholded difference image.

system is the application of 2D Discrete Cosine Transform (DCT) after obtaining the spatial-temporal feature,  $A_{v_r}$ .

### 4.3 Classification

We used the K-fold cross validation method to separate all samples into training and test sets; we fixed  $K = 3$ , and we divided the data set into three nearly equal subsets. Since the algorithm uses random selection to assign samples to folders, the whole process is repeated for 5 times and the given performance is the average of  $3 \times 5 = 15$  trials.

We used the Euclidean distance to built a  $Distance_{ij}$  matrix of dimensions  $M \times N$ , where  $M$  is the number of test samples and  $N$  is the number of training samples:

$$Distance_{i,j} = (TestVector_i - TrainVector_j)^2$$

Classification is performed by assigning every test sample to the class of the nearby training sample; that is, we used the K-Nearest Neighbor algorithm with  $K = 1$ .

## 5 EXPERIMENTAL RESULTS

We worked with 1700 RGB videos of session 1 of the Turkish Sign Language database created by Abbas et al., (Memiş & Albayrak 2013). We randomly assign an order to all samples and we repeated the benchmark algorithm of Abbas et al.; that is, we applied the 3-fold cross validation technique to divide the database into train and test sets; all experiments are repeated 5 times and the given performance is their average. Table 2 is a portion

Table 2. Confusion matrix of first 6 classes within fold 1 in trial 1.

	c1	c2	c3	c4	c5	c6
c1	1	0	0	0	0	0
c2	0	4	0	0	0	0
c3	0	0	3	1	0	0
c4	0	0	1	3	0	0
c5	0	0	0	0	5	0
c6	0	0	0	0	0	3

Table 3. The corresponding words represented in Table 2.

	Corresponding sign in Turkish	In English
c1	Afedersiniz	Excuse me
c2	Afiyet olsun	Bon appetit
c3	Gorusmek uzere	See you soon
c4	Gorusuruz	See you
c5	Gule gule	Goodbye
c6	Gunaydin	Good morning

Table 4. Accuracy table using spatial-temporal features.

	Fold 1	Fold 2	Fold 3
Trial 1	84.30	82.54	84.63
Trial 2	82.54	80.78	83.04
Trial 3	79.72	85.89	80.57
Trial 4	83.95	84.10	81.48
Trial 5	82.54	80.92	78.48

Table 5. Accuracy table using frequency domain features.

	Fold 1	Fold 2	Fold 3
Trial 1	72.13	75.84	76.50
Trial 2	78.09	75.49	79.01
Trial 3	78.66	75.44	75.49
Trial 4	76.50	76.72	79.89
Trial 5	78.80	75.66	75.49

of confusion matrix belonging to one sample fold of one trial.

Because the confusion matrix is of dimension  $C \times C$ , where  $C = 143$  is the number of classes, it is not possible to properly display the whole matrix.

Obviously, different accuracy rates were achieved for each fold within each trial; the following table shows the performances obtained with

both system 1, based on temporal features, and system 2, using frequency domain features:

The overall performance is determined by calculating the mean of all accuracies. Our best performance is achieved using spatial-temporal features; it is equal to 82.37%. We claim that it is comparable with the benchmark result given by Abbas et al. since we worked with all 1700 videos of section 1.

## 6 CONCLUSION AND FUTURE WORK

Automatic recognition of sign language helps to strength the connection between deaf and hearing societies. It is a very challenging and important research field which requires more attention.

We worked on Turkish Sign Language recognition and we challenged the given benchmark algorithm released together with the database. Current results are promising, but we are willing to do better by using the full potentiality of depth images.

## REFERENCES

- Allen, J.M., P.K. Asselin, & R. Foulds (2003). American sign language finger spelling recognition system. In *Bioengineering Conference, 2003 IEEE 29th Annual, Proceedings of*, pp. 285–286. IEEE.
- Altun, O. & S. Albayrak (2011). Turkish fingerspelling recognition system using generalized hough transform, interest regions, and local descriptors. In *Pattern Recognition Letters*, pp. 1626–1632. Springer-Verlag, Berlin Heidelberg.
- Haberdar, H. & S. Albayrak (2005). Real time isolated turkish sign language recognition from video using hidden markov models with global features. In *Computer and Information Sciences (ISCIS 2005), LNCS*, pp. 677–688. Springer-Verlag, Berlin Heidelberg.
- Memiş, A. & S. Albayrak (2013). A kinect based sign language recognition system using spatio-temporal features. In *Sixth International Conference on Machine Vision (ICMV13)*, pp. 90670X–90670X. International Society for Optics and Photonics.
- Mohandes, M. & M. Deriche (2005). Image based arabic sign language recognition. In *Signal Processing and Its Applications, 2005. Proceedings of the Eighth International Symposium on*, Volume 1, pp. 86–89. IEEE.
- Munib, Q., M. Habeeb, B. Takruri, & H.A. Al-Malik (2007). American sign language (asl) recognition based on hough transform and neural networks. *Expert systems with Applications* 32(1), 24–37.
- Vogler, C. & D. Metaxas (1997). Adapting hidden markov models for asl recognition by using three-dimensional computer vision methods. In *Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on*, Volume 1, pp. 156–161. IEEE.
- Vogler, C. & D. Metaxas (2004). Handshapes and movements: Multiple-channel american sign language recognition. In *Gesture-Based Communication in Human-Computer Interaction*, pp. 247–258. Springer.
- Waldron, M.B. & S. Kim (1995). Isolated asl sign recognition system for deaf persons. *Rehabilitation Engineering, IEEE Transactions on* 3(3), 261–271.
- Wang, C., W. Gao, & Z. Xuan (2001). A real-time large vocabulary continuous recognition system for chinese sign language. In *Advances in Multimedia Information Processing-PCM 2001*, pp. 150–157. Springer.
- Wang, H., N. Sarawate, & M. Leu (2004). Recognition of american sign language gestures with a sensory glove. In *Japan USA Symposium on Flexible Automation, Denver, CO*, pp. 102–109.

This page intentionally left blank

# A nonsmooth nonconvex sparsity-promoting variational approach for deblurring images corrupted by impulse noise

A. Lanza, S. Morigi & F. Sgallari

*Department of Mathematics, University of Bologna, Bologna, Italy*

**ABSTRACT:** We propose a new variational approach for the restoration of images simultaneously corrupted by blur and impulse noise. Promising results have been obtained by the  $\ell_1$ -TV variational model which contains a Total Variation (TV) regularization term and a nonsmooth convex  $\ell_1$ -norm data-fidelity term. We introduce a sparsity-promoting nonconvex data-fidelity term which performs better but makes the problem more difficult to handle. The main contribution of this paper is to develop a numerical algorithm based on the Alternating Direction Method of Multipliers (ADMM) strategy and on the use of proximal operators. Preliminary experiments are presented which strongly indicate that using nonconvex versus convex fidelities holds the potential for more accurate restorations at a modest computational extra-cost.

## 1 INTRODUCTION

In this paper we present a variational approach for the restoration of blurred images in the presence of impulse noise, which can be due to transmission errors, malfunctioning pixel elements in camera sensors, faulty memory locations, etc.. An important characteristic of impulse noise is that only a subset of the pixels is corrupted by noise while the rest is noise-free. In particular, let  $u \in \mathbb{R}^{d^2}$  be the unknown clean image concatenated into a column vector, and defined, without loss of generality, on a square  $d \times d$  domain  $\Omega$ , with values in  $[0,1]$ . The observed corrupted image  $g \in \mathbb{R}^{d^2}$  at pixel location  $i \in \Omega$  is modelled as follows:

$$g_i = \begin{cases} (Ku)_i & \text{for } i \in \Omega_0 \\ n_i & \text{for } i \in \Omega_1 := \Omega \setminus \Omega_0, \end{cases} \quad (1)$$

where  $K \in \mathbb{R}^{d^2 \times d^2}$  is a known linear blurring operator and  $n \in \mathbb{R}^{d^2}$  is an unknown realization of the impulsive noise process. For images corrupted by salt-and-pepper impulse noise,  $n_i$  take either 0 or 1 values, while for images corrupted by random-valued impulse noise,  $n_i \in [0,1]$  are uniformly distributed random numbers. In the discrete model (1) of the image degradation process we assume that  $\Omega_1$  is unknown.

The most popular methods to solve this image restoration problem are based on a two-stage approach which estimates the noise-corrupted region  $\Omega_1$  before estimating  $u$ . In these approaches, the second stage becomes a regular image deblurring problem, while for the accurate detection of  $\Omega_1$

several filters were proposed, e.g. Adaptive Median Filter (AMF) is used to detect salt-and-pepper impulse noise, Adaptive Center-Weighted Median Filter (ACWMF) and Rank-Ordered Logarithmic Difference (ROLD) have been utilized to detect random-valued impulse noise, see (Dong, Chan, & Xu 2007). In (Chan, Dong, & Hintermüller 2010), a two-phase image restoration method based on TV regularization combined with an  $\ell_1$ -data-fitting term for impulse noise removal and deblurring has been proposed.

Given  $K$  and  $g$ , our goal is to solve the inverse problem of recovering an as accurate as possible estimate  $u^*$  of the clean image  $u$ . Since the blurring operator is typically very ill-conditioned or even singular, some a-priori preferences are usually added in the form of a regularization term in order to get meaningful estimates.

We propose the following restoration model:

$$\min_{u \in \mathbb{R}^{d^2}} \left\{ \sum_{i=1}^{d^2} \phi((Ku - g)_i; a) + \mu TV(u) \right\}, \quad (2)$$

where  $TV(u) := \sum_{i=1}^{d^2} \|(\nabla u)_i\|_2$  denotes the discrete TV semi-norm of  $u$ , and  $\phi$  is the so-called penalty function  $\phi: \mathbb{R} \rightarrow \mathbb{R}$  with parameter  $a$ .

The cost functional reflects the need to deblur the image, recover its discontinuities and at the same time remove the impulse noise. The regularization parameter  $\mu$  controls the trade-off between the fidelity term and the regularization term. It is well known that the TV regularizer successfully allows for discontinuities while smoothing the other regions, and thus we are motivated to

consider TV norm for our proposed model. In this work we draw a particular attention to the choice of the data-fidelity term.

A popular choice for the penalty function  $\phi$  in presence of additive Gaussian noise is  $\phi(x) = x^2$ . However, this choice seems to be inadequate for the degradation model (1) since the quadratic function assigns too much weight to distant points which could be potential outliers.

Nonsmooth, convex  $\ell_1$ -norm penalty terms,  $\phi(x) = |x|$ , were shown to be more useful towards this aim, since they are robust to outliers, i.e., to impulse noise, see (Nikolova 2004) and (Chan, Dong, & Hintermüller 2010).

For the data fidelity term in (2) we propose in Sect. 2 nonconvex and nonsmooth penalty functions  $\phi$  aimed at strongly promoting sparsity of the residual. However, this leads to a nonconvex minimization problem (2) which cannot benefit from convex optimization theory and methodology to obtain robust algorithms and ensure convergence. In Sect. 3 the minimization of the proposed non-smooth nonconvex functional (2) is carried out by a suitable variant of the ADMM which allows to split the original nonconvex problem into a sequence of simple convex subproblems whose (unique) solutions can be obtained very efficiently. The numerical experiments in Sect. 4 demonstrate that the proposed method is effective in deblurring images corrupted by impulse noise.

## 2 NONCONVEX PENALTY FUNCTIONS

In the following, we denote the sets of non-negative and positive real numbers as  $\mathbb{R}_+ := \{x \in \mathbb{R} : t \geq 0\}$  and  $\mathbb{R}_+^* := \{x \in \mathbb{R} : x > 0\}$ , respectively. To induce sparsity of the residual  $Ku - g$  according to the proposed restoration model (2), we consider parameterized nonconvex penalty functions  $\phi(\cdot; a) : \mathbb{R} \rightarrow \mathbb{R}$ , with parameter  $a \in \mathbb{R}_+^*$ , satisfying the following assumptions:

- |  |   |
|--|---|
| A1) $\phi(\cdot; a) \in \mathcal{C}^0(\mathbb{R})$     | (continuous)                                |
| A2) $\phi(-x; a) = \phi(x; a)$                         | $\forall x \in \mathbb{R}_+^*$ (even)       |
| A3) $\phi(\cdot; a) \in \mathcal{C}^2(\mathbb{R}_+^*)$ |   |
| A4) $\phi'(x; a) > 0$                                  | $\forall x \in \mathbb{R}_+^*$ (increasing) |
| A5) $\phi''(x; a) \leq 0$                              | $\forall x \in \mathbb{R}_+^*$ (concave)    |
| A6) $\inf_{x \in \mathbb{R}_+^*} \phi''(x; a) = -a$    |   |

Due to the last assumption A6), we will refer to  $a$  as the *concavity* parameter. Three examples of penalty functions, referred to as  $\phi_{\log}$ ,  $\phi_{\text{rat}}$  and  $\phi_{\text{atan}}$ , which satisfy the above assumptions A1)–A6) and which have been considered, e.g., in (Chen & Selesnick 2014), (Lanza, Morigi, & Sgallari 2015), are reported in Table 1 together with some properties. We notice that for  $a \searrow 0$ , the three penalty functions tend to the convex absolute value function  $|x|$ , thus yielding the  $\ell_1$ -norm fidelity. In Figure 1 we show the plots of the three considered penalty functions for two different values  $a = 2$  and  $a = 3$  of the concavity parameter: the higher is  $a$ , the more concave is  $\phi$ .

## 3 ADMM FOR THE PROPOSED MODEL

In this section we introduce a suitable variant of the basic ADMM approach (Boyd, Parikh, Chu, Peleato, & Eckstein 2011) to solve the proposed nonconvex nonsmooth minimization problem in (2). Toward this aim, first we introduce two auxiliary variables  $t \in Q$  and  $r \in V$ , with  $V := \mathbb{R}^{d^2}$  and  $Q := \mathbb{R}^{2d^2}$ , which allow to reformulate (2) into the following linearly constrained equivalent form:

$$\begin{aligned} \{u^*, t^*, r^*\} &\leftarrow \arg \min_{u, t, r} \sum_{i=1}^{d^2} \{\phi(r_i; a) + \mu \|t_i\|_2\} \\ \text{s.t.: } t &= Du, \quad r = Ku - g, \end{aligned} \quad (3)$$

where  $D := (D_h; D_v) \in \mathbb{R}^{2d^2 \times d^2}$ ,  $D_h, D_v \in \mathbb{R}^{d^2}$  are linear operators representing the discretization of the first-order horizontal and vertical partial derivatives,

Table 1. Three examples of sparsity-inducing penalty functions.

	$\phi_{\log}$	$\phi_{\text{rat}}$	$\phi_{\text{atan}}$
$\phi(x; a)$	$\frac{1}{a} \log(1 + a x )$	$\frac{ x }{1 + a x /2}$	$\frac{\text{atan}\left(\frac{1+2a x }{\sqrt{3}}\right) - \frac{\pi}{6}}{a\sqrt{3}/2}$
$\phi(0; a)$	0	0	0
$\phi'(0^\pm; a)$	$\pm 1$	$\pm 1$	$\pm 1$
$\phi''(0^\pm; a)$	$-a$	$-a$	$-a$
$\lim_{x \rightarrow \pm\infty} \phi(x; a)$	$+\infty$	$\frac{2}{a}$	$\frac{2\pi\sqrt{3}}{a^2/9}$

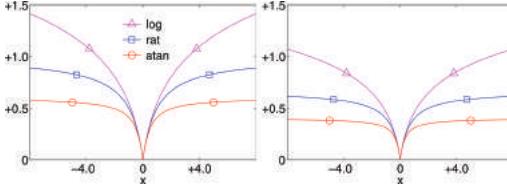


Figure 1. Plots of the penalty functions  $\phi_{\log}(x; a)$ ,  $\phi_{\text{rat}}(x; a)$ ,  $\phi_{\text{atan}}(x; a)$  defined in Table 1 for two different values  $a = 2$  (left) and  $a = 3$  (right) of the concavity parameter.

respectively, and  $t_i = ((D_h u)_i, (D_v u)_i) \in \mathbb{R}^2$  represents the discrete gradient of image  $u$  at pixel  $i$ .

To solve (3), we define the augmented Lagrangian functional

$$\begin{aligned} \mathcal{L}(u, t, r; \lambda_t, \lambda_r) = & \sum_{i=1}^{d^2} \{\phi(r_i; a) + \mu \|t_i\|_2\} \\ & - \langle \lambda_t, t - Du \rangle + \frac{\beta_t}{2} \|t - Du\|_2^2 \\ & - \langle \lambda_r, r - (Ku - g) \rangle + \frac{\beta_r}{2} \|r - (Ku - g)\|_2^2, \end{aligned} \quad (4)$$

where  $\beta_t, \beta_r > 0$  are scalar penalty parameters and  $\lambda_t \in Q$ ,  $\lambda_r \in V$  are the vectors of Lagrange multipliers.

Solving (3) is thus equivalent to search for the solutions of the following saddle-point problem:

$$\begin{aligned} \text{Find } (x^*; \lambda^*) & \in X \times \Lambda \\ \text{s.t. } \mathcal{L}(x^*; \lambda) & \leq \mathcal{L}(x^*; \lambda^*) \leq \mathcal{L}(x; \lambda^*) \\ \forall (x; \lambda) & \in X \times \Lambda, \end{aligned} \quad (5)$$

with  $\mathcal{L}$  defined in (4) and where, for simplicity of notations, we set  $x = (u, t, r)$ ,  $\lambda = (\lambda_t, \lambda_r)$ ,  $X = V \times Q \times V$  and  $\Lambda = Q \times V$ .

The  $k$ th iteration of the ADMM iterative scheme applied to the solution of (3) reads as follows:

$$t^{k+1} \leftarrow \arg \min_{t \in Q} \mathcal{L}(u^k, t, r^k; \lambda_t^k, \lambda_r^k) \quad (6)$$

$$r^{k+1} \leftarrow \arg \min_{r \in V} \mathcal{L}(u^k, t^{k+1}, r; \lambda_t^k, \lambda_r^k) \quad (7)$$

$$u^{k+1} \leftarrow \arg \min_{u \in V} \mathcal{L}(u, t^{k+1}, r^{k+1}; \lambda_t^k, \lambda_r^k) \quad (8)$$

$$\begin{cases} \lambda_t^{k+1} \leftarrow \lambda_t^k - \gamma \beta_t (t^{k+1} - Du^{k+1}) \\ \lambda_r^{k+1} \leftarrow \lambda_r^k - \gamma \beta_r (r^{k+1} - (Ku^{k+1} - g)) \end{cases} \quad (9)$$

where  $\gamma$  is a relaxation parameter chosen in the interval  $(0, (\sqrt{5} + 1)/2)$ , as analyzed in (Chan, Tao, & Yuan 2013).

In the following subsections we show in detail how to solve the three sub-problems (6)–(8) for the variables  $t$ ,  $r$  and  $u$ , respectively, then we present the overall iterative ADMM-based minimization algorithm.

### 3.1 Solving the sub-problem for $t$

Recalling the definition of the augmented Lagrangian functional  $\mathcal{L}$  in (4), after some simple manipulations the sub-problem for  $t$  in (6) can be written as follows:

$$t^{k+1} \leftarrow \arg \min_{t \in Q} \sum_{i=1}^{d^2} \left\{ \|t_i\|_2 + \frac{\beta_t}{2\mu} \|t_i - q_i^k\|_2^2 \right\}, \quad (10)$$

with the constant (w.r.t.  $t$ ) vectors  $q_i^k \in \mathbb{R}^2$  given by

$$q_i^k = \left( Du^k \right)_i + \frac{1}{\beta_t} \left( \lambda_t^k \right)_i, \quad i = 1, \dots, d^2, \quad (11)$$

and where  $(Du^k)_i, (\lambda_t^k)_i \in \mathbb{R}^2$  denote the discrete gradient and the Lagrange multipliers at pixel  $i$ , respectively. The minimization in (10) is thus equivalent to the following  $d^2$  independent 2-dimensional problems:

$$t_i^{k+1} \leftarrow \arg \min_{t_i \in \mathbb{R}^2} \left\{ \|t_i\|_2 + \frac{\beta_t}{2\mu} \|t_i - q_i^k\|_2^2 \right\}, \quad (12)$$

which, according to (Wang, Yang, Yin, & Zhang 2008), admit closed-form solutions given by the following soft-thresholding operators:

$$t_i^{k+1} = \max \left\{ \|q_i^k\| - \frac{\mu}{\beta_t}, 0 \right\} \frac{q_i^k}{\|q_i^k\|}, \quad (13)$$

where  $0 \cdot (0/0) = 0$  is assumed. The computational cost of this sub-problem is linear in the number of pixels  $d^2$ .

### 3.2 Solving the sub-problem for $r$

Given the definition of  $\mathcal{L}$  in (4), after some manipulations the sub-problem for  $r$  in (7) reduces to:

$$r^{k+1} \leftarrow \arg \min_{r \in V} \sum_{i=1}^{d^2} \left\{ \phi(r_i; a) + \frac{\beta_r}{2} (r_i - z_i^k)^2 \right\}, \quad (14)$$

where  $z_i^k \in \mathbb{R}$  denotes the  $i$ th entry of the constant (w.r.t.  $r$ ) vector  $z^k \in \mathbb{R}^{d^2}$  defined as follows:

$$z^k = Ku^k - g + \frac{1}{\beta_r} \lambda_r^k. \quad (15)$$

The minimization in (14) is thus equivalent to the following  $d^2$  independent scalar problems:

$$r_i^{k+1} \leftarrow \arg \min_{r_i \in \mathbb{R}} \left\{ \phi(r_i; a) + \frac{\beta_r}{2} (r_i - z_i^k)^2 \right\}. \quad (16)$$

To the purpose of imposing that (16) are convex problems, we report Proposition 1 below, whose proof can be found in (Chen & Selesnick 2014), Proposition 2. After noticing that the minimized functions in (16) have the same form of function  $f$  defined in (21), we can conclude that (16) are strictly convex problems if and only if the following condition holds:

$$\beta_r > a. \quad (17)$$

In case that (17) is satisfied, the unique solutions of the strictly convex problems in (16) can be computed based on the proximity maps introduced in Proposition 2 below. In particular, according to Proposition 2, for a generic penalty function  $\phi$  the solutions of (16) can be obtained as follows:

$$r_i^{k+1} = \begin{cases} 0 & \text{if } |z_i^k| \leq 1/\beta_r \\ \text{sign}(z_i^k)r^* & \text{if } |z_i^k| > 1/\beta_r \end{cases}, \quad (18)$$

where  $r^*$  denotes the (unique) solution of the following constrained nonlinear equation:

$$r^* + \frac{1}{\beta_r} \phi'(r^*; a) - |z_i^k| = 0, \quad 0 < r^* < |z_i^k|. \quad (19)$$

According to (26), the solution  $r^*$  of (19) can be obtained as the limit point of the following (quadratically convergent) Newton-Raphson iteration:

$$\begin{aligned} r^0 &= z_i^k \\ r^{k+1} &= \frac{r^k \phi''(r^k; a) - \phi'(r^k; a) + z_i^k \beta_r}{\phi''(r^k; a) + \beta_r}. \end{aligned} \quad (20)$$

For what concerns computational efficiency of (18)–(20), first we notice that iterations (20) has to be carried out only for the subset of indexes  $i$  for which the quantity  $|z_i^k|$  is above the threshold in (18). Then, the quantity  $z_i^k \beta_r$  in the numerator of (20) is constant during iterations and can be precomputed. Last but not the least, very few iterations (20) are sufficient to compute accurate estimates of  $r^*$ , as it will be demonstrated in the experimental section.

*Proposition 1.* Let  $\phi: \mathbb{R} \rightarrow \mathbb{R}$  be a function satisfying assumptions A1)–A6) and  $a, \eta \in \mathbb{R}_+^*$ ,  $b \in \mathbb{R}$  be given constants. Then, the function  $f: \mathbb{R} \rightarrow \mathbb{R}$  defined as

$$f(y) := \frac{1}{2}(y - b)^2 + \eta \phi(y; a) \quad (21)$$

is strictly convex if and only if

$$\eta a < 1. \quad (22)$$

*Proposition 2.* Let  $\phi: \mathbb{R} \rightarrow \mathbb{R}$  be a function satisfying assumptions A1)–A6),  $a, \eta \in \mathbb{R}_+^*$  be given constants satisfying condition (22) and  $f: \mathbb{R} \rightarrow \mathbb{R}$  be the function defined in (21). Then, the proximity map  $\text{prox}_{\phi}^{\eta}: \mathbb{R} \rightarrow \mathbb{R}$  of function  $\phi$  defined as:

$$\text{prox}_{\phi}^{\eta}(b) := \arg \min_{y \in \mathbb{R}} f(y), \quad b \in \mathbb{R}, \quad (23)$$

is given by

$$\text{prox}_{\phi}^{\eta}(b) = \begin{cases} 0 & \text{if } |b| \leq \eta \\ \text{sign}(b)y^* & \text{if } |b| > \eta, \end{cases} \quad (24)$$

where  $y^*$  is the unique solution of the following constrained nonlinear equation:

$$y^* + \eta \phi'(y^*; a) - |b| = 0, \quad 0 < y^* < |b|. \quad (25)$$

Moreover, in case that  $|b| > \eta$ , the Newton-Raphson iterative scheme defined by:

$$\begin{aligned} y^0 &= b \\ y^{k+1} &= \frac{y^k \phi''(y^k; a) - \phi'(y^k; a) + b/\eta}{\phi''(y^k; a) + 1/\eta}, \end{aligned} \quad (26)$$

converges with quadratic rate to the unique solution  $y^*$  of the constrained nonlinear equation in (25).

*Proof* First, we notice that for any penalty function  $\phi$  satisfying assumptions A1)–A6), the proximity map  $\text{prox}_{\phi}^{\eta}$  in (23) has the following properties:

$$\text{prox}_{\phi}^{\eta}(0) = 0, \quad \text{prox}_{\phi}^{\eta}(b) = -\text{prox}_{\phi}^{\eta}(-b) \quad \forall b \in \mathbb{R}, \quad (27)$$

which follow immediately from the definitions in (21) and (23). Since according to (27)  $\text{prox}_{\phi}^{\eta}$  is a odd function, it follows that:

$$\text{prox}_{\phi}^{\eta}(b) = \text{sign}(b) \text{prox}_{\phi}^{\eta}(|b|) \quad \forall b \in \mathbb{R}. \quad (28)$$

Therefore, in the rest of the proof we focus on the case  $b \in \mathbb{R}_+^*$  and then apply (28) to derive the expression of  $\text{prox}_{\phi}^{\eta}$  for  $b \in \mathbb{R}$ .

For any  $a, \eta \in \mathbb{R}_+^*$  satisfying condition (22) and any  $b \in \mathbb{R}_+^*$ , the function  $f$  in (21) is strictly convex and, in particular, has the following properties:

$$\begin{cases} f'(y) = y - b + \eta \phi'(y; a) < 0 & \forall y \in \mathbb{R}_-^* \\ f'(0^-) = -b - \eta < 0 \\ f'(0^+) = -b + \eta \geq 0 & \forall b \in ]0, \eta] \\ f''(y) = 1 + \eta \phi''(y; a) > 0 & \forall y \in \mathbb{R}_+^* \end{cases} \quad (29)$$

Hence, in case that  $b \in ]0, \eta]$ , the function  $f$  has its unique global minimizer at  $y=0$ , so that (28) allows to conclude that the first part of statement (24) holds.

In case that  $b \in ]\eta, +\infty]$ , all the properties in (29) hold true but the third one which becomes  $f'(0^+) < 0$ . Hence, in this case the global minimizer of function  $f$  coincides with its unique (positive) stationary point, thus concluding the proof of (24).

### 3.3 Solving the sub-problem for $u$

The sub-problem for  $u$  in (8) is a quadratic minimization problem, whose first-order optimality conditions lead to the following linear system:

$$\begin{aligned} Au^{k+1} &= b^k, \text{ with } A = D^T D + \frac{\beta_r}{\beta_t} K^T K, \\ b^k &= D^T \left( t^{k+1} - \frac{\lambda_t^k}{\beta_t} \right) + \frac{\beta_r}{\beta_t} K^T \left( r^{k+1} - \frac{\lambda_r^k}{\beta_r} + g \right) \end{aligned} \quad (30)$$

The  $d^2 \times d^2$  linear system in (30) is solvable if the coefficient matrix has full-rank, that is if the following condition holds:

$$\text{Ker}(D^T D) \cap \text{Ker}(K^T K) = \{0\}, \quad (31)$$

where  $\text{Ker}(M)$  denotes the null space of matrix  $M$  and 0 indicates the  $d^2$ -dimensional null vector.

---

**Algorithm 1** ADMM-based iterative scheme for the solution of the nonconvex nonsmooth problem in (2)

---

**inputs:**  $g \in \mathbb{R}^{d^2}$ ,  $K \in \mathbb{R}^{d^2 \times d^2}$ ,  $\mu, a > 0$ ,  $\phi(\cdot; a): \mathbb{R} \rightarrow \mathbb{R}$   
**output:** approximate solution  $u^* \in \mathbb{R}^{d^2}$  of (2)  
**params:**  $\beta_t > 0$ ,  $\beta_r > a$ ,  $0 < \gamma < \frac{1+\sqrt{5}}{2}$

1. Initialization:  $u^0 = g$ ,  $\lambda_t^0 = 0$ ,  $\lambda_r^0 = 0$
  2. **for**  $k = 0, 1, 2, \dots$  until convergence **do**:
    - compute  $t^{k+1}$  by (11), (13)
    - compute  $r^{k+1}$  by (18)–(20)
    - compute  $u^{k+1}$  by solving (30)
    - compute  $\lambda_t^{k+1}, \lambda_r^{k+1}$  by (9)  - end for**
  3.  $u^* = u^{k+1}$
- 

In our case, condition (31) is satisfied. In fact,  $K$  represents a blurring operator, which is a low-pass filter, whereas the regularization matrix  $D$  is a finite difference operator and, hence, is a high-pass filter.

In case that (31) is satisfied, the coefficient matrix in (30) is symmetric positive definite and, typically, is highly sparse. Hence, the linear system in (30) can be solved quite efficiently by the iterative (preconditioned) Conjugate Gradient method. Moreover, under appropriate assumptions about the solution  $u^{k+1}$  near the image boundary, the linear system can be solved even more efficiently. For example, under *periodic*, *reflective* or *anti-reflective* boundary conditions for  $u^{k+1}$ , the coefficient matrix in (30) can be diagonalized by the 2D discrete Fourier transform (FFT implementation), 2D discrete cosine transform (FCT implementation) or 2D discrete sine transform (FST implementation), respectively. Provided that the penalty parameters  $\beta_t$  and  $\beta_r$  are kept fixed for all the ADMM iterations, the coefficient matrix in (30) does not change and it can be diagonalized once for all at the beginning. Therefore, at each ADMM iteration the linear system (30) can be solved by one forward FFT/FCT/FST and one inverse FFT/FCT/FST, each at a cost of  $O(d^2 \log d)$ .

### 3.4 ADMM-based minimization algorithm

To summarize previous results, in Alg. 1 we report the main computational steps of the proposed ADMM-based iterative scheme used to solve the nonconvex nonsmooth minimization problem in (2).

## 4 EXPERIMENTAL RESULTS

We present some preliminary experiments assessing the performance of the proposed nonsmooth nonconvex model (2), referred to as  $\phi$ -TV, when applied to the restoration of the brain and cameraman test images depicted in Figure 2(a) and Figure 2(e), respectively. The images are synthetically corrupted by Gaussian blur and impulse noise of both salt & pepper (s&p) and random-valued (rand) type with different noise levels. We recall that the level of an impulse noise corruption, expressed in percent, represents the fraction of image pixels which are affected by the noise.

To provide evidence of the advantages of using nonconvex versus convex fidelity terms, we compare the proposed  $\phi$ -TV model with the popular nonsmooth but convex  $\ell_1$ -TV model defined by (2) with  $\phi(x) = |x|$ , see (Yang, Zhang, & Yin 2009). Accuracy of the obtained restored images  $u^*$  is evaluated by the Improved Signal-to-Noise Ratio (ISNR) defined as

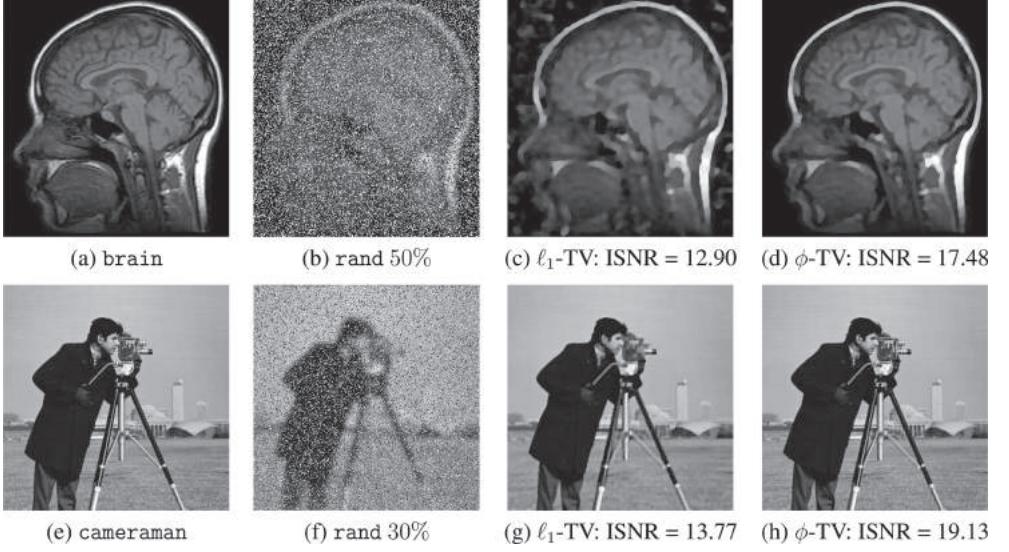


Figure 2. Restoration results for brain (top row) and cameraman (bottom row) test images.

$$ISNR(u^*, g, u) = 10 \log_{10} \frac{\|g - u\|_2^2}{\|u^* - u\|_2^2} \text{ dB},$$

where  $u$  denotes the uncorrupted image.

For both the compared models and for all the examples, we used the iterative minimization ADMM scheme reported in Section 3 with the same parameter values  $\beta_t = 1/100$ ,  $\beta_r = 100$ ,  $\gamma = 1.6$ , and taking as restored image  $u^*$  the iterate  $u^k$  obtained as soon as the following stopping condition is satisfied:

$$\|u^k - u^{k-1}\|_2 / \|u^{k-1}\|_2 < 10^{-5}.$$

For all the experiments, the regularization parameter  $\mu$  is tuned independently for the  $\phi$ -TV and the  $\ell_1$ -TV models so as to get the highest possible ISNR value of the restored image.

In our  $\phi$ -TV model, we used a value  $a = 10$  of the concavity parameter, such that condition (17) is satisfied and the sub-problem for the variable  $r$  is convex. Moreover, in the regularization term of our  $\phi$ -TV model we used the penalty function  $\phi_{\text{atan}}$  defined in Table 1 and we performed only one iteration of the Newton-Raphson scheme in (20) for the solution of the variable  $r$  sub-problem. To support these choices, in Table 2 we report the ISNR values obtained by running our  $\phi$ -TV model on the brain test image corrupted by Gaussian blur and impulse noise of level 50%, with different penalty functions  $\phi$  and different number of Newton-Raphson iterations for the solution of the

Table 2. ISNR values obtained by  $\phi$ -TV on the brain image.

its	s & p 50%			rand 50%		
	$\phi_{\text{log}}$	$\phi_{\text{rat}}$	$\phi_{\text{atan}}$	$\phi_{\text{log}}$	$\phi_{\text{rat}}$	$\phi_{\text{atan}}$
1	25.68	25.69	25.78	17.40	17.43	17.48
5	25.67	25.70	25.79	17.40	17.44	17.48

$r$  variable sub-problem. We notice that the penalty function  $\phi_{\text{atan}}$  provides slightly better results than  $\phi_{\text{log}}$  and  $\phi_{\text{rat}}$  and that performing only one iteration of the Newton-Raphson scheme in (20) does not affect the accuracy of the restorations.

In Table 3 and Table 4 we report the ISNR values obtained by  $\ell_1$ -TV and  $\phi$ -TV models on the cameraman and brain test images corrupted by Gaussian blur (band = 11, sigma = 2.5) and impulse noise of increasing levels 30%, 50%, 70%. The benefits of using the nonconvex fidelity term in the proposed  $\phi$ -TV model with respect to the convex fidelity term of the  $\ell_1$ -TV model are well represented by the results in Tables 3–4, where the highest ISNR values are marked in boldface. This is confirmed by the restored images reported in Figure 2. The nonconvex data-fitting allows us to weight more the deblurring term, which leads to better quality solutions.

These preliminary experiments strongly indicate that using nonconvex versus convex fidelities allows for more accurate restorations at a modest

Table 3. ISNR values obtained for the brain test image.

	30%		50%		70%	
	s&p	rand	s&p	rand	s&p	rand
$\ell_1$ -TV	23.71	18.80	22.36	12.90	17.93	9.04
$\phi$ -TV	<b>27.68</b>	<b>23.52</b>	<b>25.78</b>	<b>17.48</b>	<b>18.67</b>	<b>10.43</b>

Table 4. ISNR values obtained for the cameraman test image.

	30%		50%		70%	
	s&p	rand	s&p	rand	s&p	rand
$\ell_1$ -TV	20.13	13.77	17.90	11.39	15.99	8.70
$\phi$ -TV	<b>25.00</b>	<b>19.13</b>	<b>19.27</b>	<b>14.50</b>	<b>17.40</b>	<b>10.17</b>

computational extra-cost which can be approximately evaluated as a 30% overhead.

## 5 CONCLUSIONS

In this paper, we presented a nonsmooth nonconvex variational model for non-blind deblurring of images corrupted by impulse noise of both salt-and-pepper and random-valued type. In the model, a classical TV regularization term is coupled with a nonsmooth nonconvex fidelity term which allows to promote sparsity of the residual image more strongly than the convex  $\ell_1$ -norm fidelity commonly used for this kind of noise. Difficulties arising from non-smoothness and non-convexity of the proposed model are dealt with by an efficient minimization algorithm based on a suitable variant of the popular ADMM procedure which allows to split the nonconvex problem into a sequence of strictly convex sub-problems.

Some preliminary experiments demonstrate the effectiveness of the proposed approach. In

particular, results of the experimental comparison with the  $\ell_1$ -TV model strongly indicate that using nonconvex versus convex fidelity terms holds the potential for more accurate restorations at a modest computational extra-cost.

In future work we will investigate the use of sparsity-promoting penalty functions for both the fidelity term and the regularization term.

## REFERENCES

- Boyd, S., N. Parikh, E. Chu, B. Peleato, & J. Eckstein (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning* 3(1), 1–122.
- Chan, R., Y. Dong, & M. Hintermuller (2010, July). An efficient two-phase  $\ell_1$ -tv method for restoring blurred images with impulse noise. *Image Processing, IEEE Transactions on* 19(7), 1731–1739.
- Chan, R., M. Tao, & X. Yuan (2013). Constrained total variational deblurring models and fast algorithms based on alternating direction method of multipliers. *SIAM J. Imag. Sci.* 6, 680–697.
- Chen, P.-Y. & I. Selesnick (2014, July). Group-sparse signal denoising: Non-convex regularization, convex optimization. *Signal Processing, IEEE Transactions on* 62(13), 3464–3478.
- Dong, Y., R. Chan, & S. Xu (2007, April). A detection statistic for random-valued impulse noise. *Image Processing, IEEE Transactions on* 16(4), 1112–1120.
- Lanza, A., S. Mori, & F. Sgallari (2015). Convex image denoising via non-convex regularization. In *Scale Space and Variational Methods in Computer Vision*, Volume 9087, pp. 666–677.
- Nikolova, M. (2004). A variational approach to remove outliers and impulse noise. *Journal of Mathematical Imaging and Vision* 20(1–2), 99–120.
- Wang, Y., J. Yang, W. Yin, & Y. Zhang (2008). A new alternating minimization algorithm for total variation image reconstruction. *SIAM J. IMAGING SCI*, 248–272.
- Yang, J., Y. Zhang, & W. Yin (2009). An efficient tvll algorithm for deblurring multichannel images corrupted by impulsive noise. *SIAM Journal on Scientific Computing* 31(4), 2842–2865.

This page intentionally left blank

# Classification-based blood vessel segmentation in retinal images

J. Odstrcilik, R. Kolar & V. Harabis

*St. Anne's University Hospital—International Clinical Research Center (ICRC), Brno, Czech Republic  
Department of Biomedical Engineering, Faculty of Electrical Engineering and Communication,  
Brno University of Technology, Brno, Czech Republic*

R.P. Tornow

*University of Erlangen–Nuremberg, Department of Ophthalmology, Erlangen, Germany*

**ABSTRACT:** Automatic and precise segmentation of retinal blood vessels can help in computer aided assessment of various retinal diseases, especially diseases related to cardiovascular system or glaucoma. Hence, an accurate detection of retinal vascular structures is one of the most addressed topic in the field of retinal image processing today. Most of the methods are designed directly for utilization with a special dataset only or have problems to segment blurry and noisy images. In this study, we showed that combination of three standard approaches—matched filtering, Hessian-based approach, and morphological processing, together with support vector machine classification technique, can be satisfactorily used as a universal segmentation approach in retinal images acquired using entirely different devices.

## 1 INTRODUCTION

An accurate segmentation of retinal vasculature is a necessary step preceding higher-level analysis of fundus images. It can be consequently used for detection of various diseases, especially diseases related to cardiovascular system (Bhuiyan 2014) or glaucoma (Wang 2007). Other goals to segment vasculatures in retinal images can be related to biometric applications as well (Lajevardi 2013) or as a preprocessing step for image registration (Kolar 2013, Kolar 2013). Therefore, segmentation of retinal vessels is one of the most addressed topic in retinal image processing today (Annunziata 2015, Rathod 2015, Cheng 2014, Hannink 2014, Budai 2013, Yu 2012 & Fraz 2012). Most of the methods are designed directly for one dataset or have problems to accurately segment blurry and noisy images.

In this study, we showed that combination of three standard and computationally simple approaches—Matched Filtering (MF), Hessian-based Approach (HA), and Morphological Processing (MP), together with Support Vector Machine (SVM) classification technique, can be used successfully as a universal segmentation approach in different kinds of retinal images with various resolution and image quality. Especially, a proper segmentation of vasculature in retinal video-sequences, acquired non-mydriatically under the low illumination conditions, can be highly useful as a preprocessing step for extraction of

different kinds of diagnostically important parameters, e.g. for glaucoma diagnosis (Tornow 2015 & Odstrcilik 2014).

## 2 IMAGE DATASETS

Two datasets of standard static color fundus images and one dataset of dynamic low-quality video-sequences with different resolution and different Field of Views (FOV) were used for development and testing of the proposed methodology: A. HRF (high-resolution fundus) database (color images, size of  $3504 \times 2336$  px, 60-degree FOV), B. local patient fundus image dataset (color images, size of  $2048 \times 1536$  px, 45-degree FOV) acquired in St. Anne's Faculty Hospital, Brno, Czech Republic, and C. dataset of experimental dynamic fundus video-sequences (gray-level video format, size of  $640 \times 480$  px, 15-degree FOV, framerate: 11 to 39 fps) acquired at collaborating ophthalmology clinic at University of Erlangen, Erlangen-Nuremberg, Germany. Figure 1 shows respective examples of images from particular datasets.

## 3 METHODS

Three methods, namely Matched Filtering (MF), Hessian-based Approach (HA), and Morphological Processing (MP) were used for generation of image features to segment blood vessels. Supervised

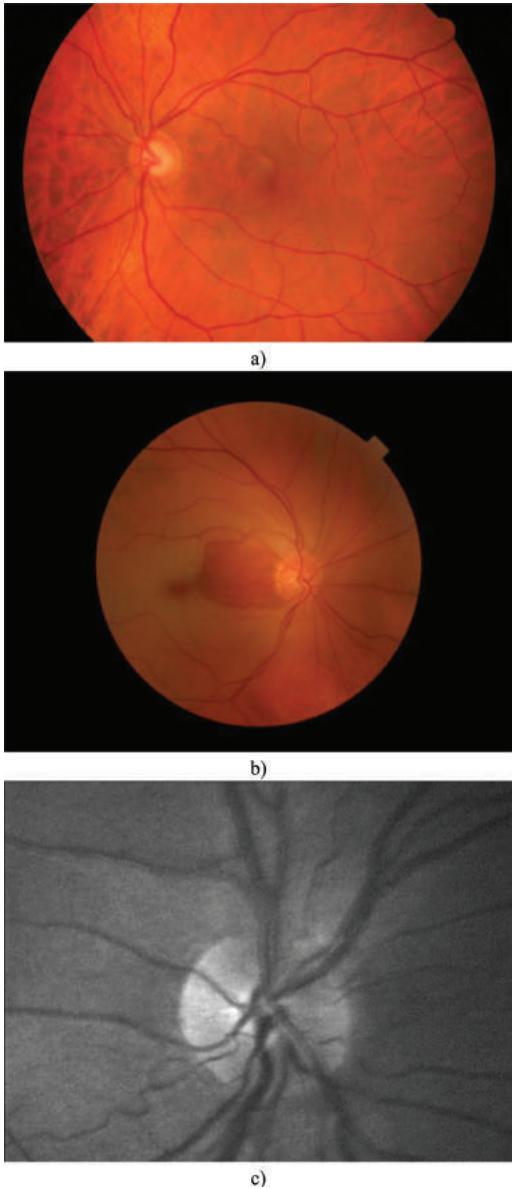


Figure 1. Examples of images from a) HRF database, local patient fundus image dataset, and c) one frame of a low-quality video-sequence.

Support Vector Machine (SVM) was further utilized for pixel-wise classification to get a binary image.

### 3.1 Feature extraction

#### 3.1.1 Matched filtering

According to our previous methodology (Odstrcilik 2013), 2D filtration kernels were designed with

respect to the appearance of typical blood vessel segments of different width (diameter) and orientation. These kernels were created according to measurement of numerous perpendicular cross-sectional profiles of retinal vessels in standard images of the HRF database. The cross-sectional profiles were then heuristically classified into five classes (indexed as  $k = 0, 1, \dots, 4$ ) of different blood vessel thicknesses to achieve a reliable and precise detection of all possible blood vessel segments with an acceptable width resolution. Particular kernels were rotated into the angular direction  $\varphi = 0^\circ, 15^\circ, \dots, 165^\circ$  in order to cover sufficiently all possible orientations of the blood vessel segments. A square shape of the obtained masks is assumed. Thus, a number of 12 differently oriented kernels  $h_{k,\varphi}$  for each of the width classes is obtained. Furthermore, each kernel is convolved with the input image  $I$ :

$$MFR_{k,\varphi} = I * h_{k,\varphi}. \quad (1)$$

In this way, a number of 60 ( $5 \times 12$ ) parametric images ( $MFR_{k,\varphi}$ —matched filter responses) is obtained, related to the corresponding width class  $k$  and the orientation of blood vessel segment  $\varphi$ . A joint parametric image  $O_{MFR}$  is then computed by selection of a maximum response from the set of resulted parametric images in pixel-wise manner (Fig. 2a):

$$O_{MFR} = \max_{k,\varphi} \{MFR_{k,\varphi}\}. \quad (2)$$

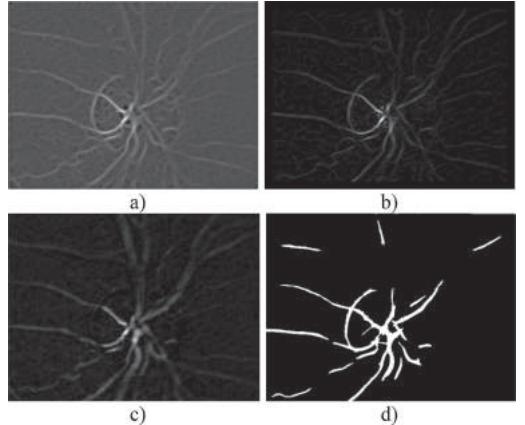


Figure 2. Example of parametric images obtained via (a) MF, (b) HA, (c) MP approaches, and (d) resulted binary representation of blood vessel tree classified via SVM classifier. Particular images represent an output of processing of one frame of a low-quality video-sequence, which is shown in Figure 1c.

The five kernels designed can be used for vessel segmentation in retinal images with high resolution. In case of low-quality video-sequences, the size of large and medium blood vessels ( $\sim 5\text{--}10$  pixels) in central and peripheral part of the images (frames) is appropriately comparable with the thinnest blood vessels in the images of HRF database. This means that only the kernel  $k = 0$  can be used for vessel segmentation in studied video-sequences.

### 3.1.2 Hessian-based approach

The HA is used to segment blood vessels (Odstrcilik 2014). First, the input image  $I$  is smoothed by Gaussian kernel to suppress the noise:

$$f = I * g_\sigma \quad (3)$$

The size of kernel  $g_\sigma$  is set to  $22 \times 22$  px for images in datasets A and B and  $9 \times 9$  px for dataset C. Standard deviation is set  $\sigma = 8$  (datasets A, B) and  $\sigma = 3$  (dataset C). Both parameters were identified in a heuristic manner with respect to the size of blood vessels in images of particular datasets. Since the blood vessels appear mainly as the valleys in the image (darker than background), their position can be determined using Hessian matrix. This matrix of the second-order partial derivatives is defined as follows:

$$H_{ij} = \begin{pmatrix} \partial_{xx} f_{ij} & \partial_{xy} f_{ij} \\ \partial_{yx} f_{ij} & \partial_{yy} f_{ij} \end{pmatrix}, \quad (4)$$

where  $f_{ij}$  represents a grey-scale image and  $i, j$  are spatial coordinates along the axes  $x$  and  $y$ . Since  $\partial_{yx} f_{ij} = \partial_{xy} f_{ij}$ , Hessian  $H_{ij}$  has two real eigenvalues  $\lambda_{1,2}$ .  $H_{ij}$  and corresponding eigenvalues  $\lambda_{1,2}$  are computed in a pixel-wise manner using Sobel difference kernel. Then, only the higher positive eigenvalue representing image valleys is taken into account. In this way, a parametric image  $O_{\text{HA}}$  is computed (Fig. 2b).

### 3.1.3 Morphological processing

The third feature to segment blood vessels is obtained via morphological bottom-hat filtering of the grayscale input image  $I$  with predefined structuring element (SE):

$$B_{\text{hat}}(I) = (I \bullet SE) - I, \quad (5)$$

where the operation  $\bullet$  represents morphological closing (Soille 2010). Three disk-shaped structuring elements with radii  $R = 7, 15$ , and  $30$  were defined to segment blood vessels in images from HRF database (dataset A). The elements' size was estimated to best fit the diameter of blood vessels in these high-resolution images. For data

from local patient fundus image database (dataset B), only the elements with  $R = 7$  and  $R = 15$  were utilized. In case of low-resolution fundus video-sequences (dataset C), utilization of SE with  $R = 7$  is sufficient.

The final parametric image  $O_{\text{MP}}$  is then obtained in the same way as in case of MF via selection of maximum response from the set of individual parametric images in pixel-wise manner (Fig. 2c).

### 3.2 Classification

Support Vector Machines (SVMs) have become very popular machine learning technique that are used in many areas for classification as well as regression, and other learning tasks (Chang 2011).

Standard version of C-SVM classifier with RBF (radial basis function) kernel was used to classify blood vessels (Fig. 2d). The value of the regularization parameter  $C = 100$  has been identified experimentally via cross-validation. All other parameters of the SVM classifier have been left with default values as proposed in (Chang 2011). The classifier was trained on image features extracted from images of HRF database. A number of 498 square-shaped ( $5 \times 5$  pixels) image regions of interest (ROIs) were manually extracted from regions of blood vessels of different diameters in 5 randomly chosen images from HRF database, which contains corresponding ground truth segmentations. Next, a number of 468 ROIs were extracted from the image background. For each ROI, the mean value of proposed features was computed. Thus, a dataset of instances of two groups ("vessels" and "background") was established, which can be used for training the SVM classifier. Figure 3 shows a depiction of instances in three-dimensional parametric space. It can be

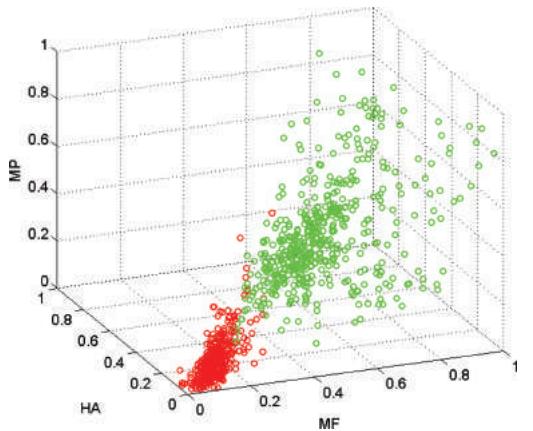


Figure 3. A depiction of instances (green—vessels, red—background) in three-dimensional parametric space.

noticed that both groups can be quite well separated. Occasional outliers of background instances are most probably caused by noise or artefacts in parametric images.

Finally (after the classification), morphological cleaning is applied by deleting non-connected objects with size less than  $\sim 500$  and  $1000\text{ px}$  in case of datasets A and B, and dataset C, respectively.

## 4 RESULTS AND DISCUSSION

### 4.1 Classification performance

In the first step, the classification performance was evaluated via the repeated random sub-sampling cross-validation procedure using extracted instances of vessels and background. Seventy and thirty percent of randomly selected instances were used for training and testing the classifier in each run of cross-validation, respectively. The repetition of random sub-sampling procedure was set to 100. The results of cross-validation procedure (Table 1) signalize rather high performance of the method in term of sensitivity (SE), specificity (SP), and classification accuracy (ACC), which corresponds with visual findings in Figure 3. The improvement was also achieved in comparison with our previous approach for vessel segmentation in low-quality video-sequences (Odstrcilik 2014). All quantitative results of cross-validation (SE, SP, and ACC) are a bit higher in case of the proposed method (by 2–5%).

### 4.2 Segmentation of static fundus images

In the second step, the proposed method was tested on static fundus images in HRF database—dataset A (Fig. 4). The HRF database contains three sets of fundus images: one of healthy retinas, one of glaucomatous, and one of Diabetic Retinopathy (DR) retinas. Each set is composed of 15 color fundus images with corresponding gold standard segmentation made by a human observer. Since

Table 1. Classification performance.

Evaluation parameters	Average results
$SE = TP / (TP + FN)$	$96.86 \pm 1.66\%$
$SP = TN / (TN + FP)$	$98.76 \pm 0.89\%$
$ACC = (TP + TN) / (TP + FN + TN + FP)$	$97.78 \pm 0.81\%$

TP—number of ROIs correctly classified as the blood vessels, FN—number of ROIs incorrectly classified as the background, TN—number of ROIs correctly detected as the background, FP—number of ROIs incorrectly classified as the blood vessels.

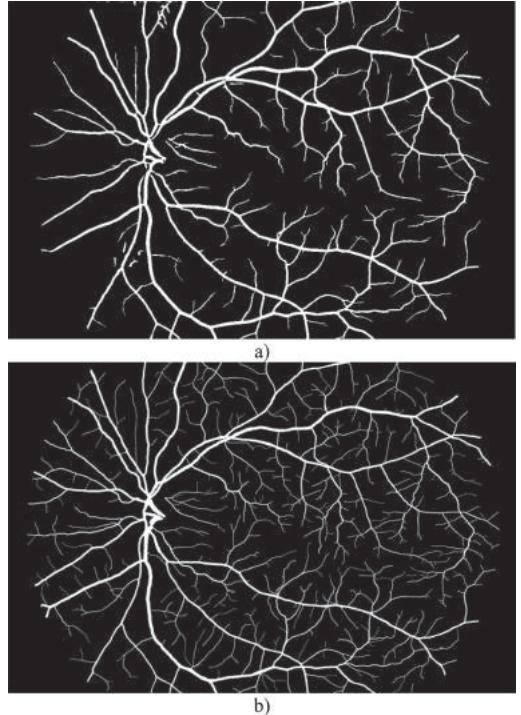


Figure 4. a) Example of one segmentation result of image from HRF database and b) corresponding ground truth segmentation. The image corresponds with the image in Figure 1a.

ground truth segmentations are available, quantitative evaluation of the proposed method has been performed via computation of SE, SP, and ACC parameters (Table 2). Values of other authors are also presented to compare the proposed approach with recent state-of-the-art methods, which were evaluated on HRF dataset. It is shown the proposed method is well comparable with other methods.

Next, the proposed approach was tested on static patient fundus image dataset—dataset B. (Fig. 5). The dataset contains 45 images of patients with various diseases—DR, glaucoma, and macular degeneration. In this case, the method was evaluated only qualitatively, since no gold standard segmentation is available. The original method was developed especially for images of HRF database. However, the results revealed the utilization for other dataset is possible.

### 4.3 Segmentation of dynamic video-sequences

In the third step, the proposed method was tested to segment blood vessels on dataset of dynamic low-quality video-sequences acquired by an

Table 2. Evaluation of the method on HRF database.

		SE	SP	ACC
Proposed method, 2015	H	0.7887	0.9931	0.9581
	G	0.7982	0.9850	0.9554
	DR	0.7588	0.9480	0.9554
Budai et al., 2013	H	0.6620	0.9920	0.9610
	G	0.6870	0.9860	0.9650
	DR	0.6580	0.9770	0.9550
Odstrcilik et al., 2013	H	0.7860	0.9750	0.9530
	G	0.7900	0.9640	0.9490
	DR	0.7460	0.9610	0.9440
Annunziata et al., 2015	H	0.6820	0.9935	0.9587
	G	0.7566	0.9785	0.9603
	DR	0.6997	0.9787	0.9554
Cheng et al., 2014	H	0.7889	0.9865	0.9647
	G	0.6921	0.9875	0.9634
	DR	0.6313	0.9851	0.9561
Yu et al., 2012	H	0.7938	0.9767	0.9566
	G	0.7890	0.9662	0.9518
	DR	0.7604	0.9625	0.9460

H—dataset of healthy images, G—dataset of glaucomatous images, DR—dataset of diabetic retinopathy images.

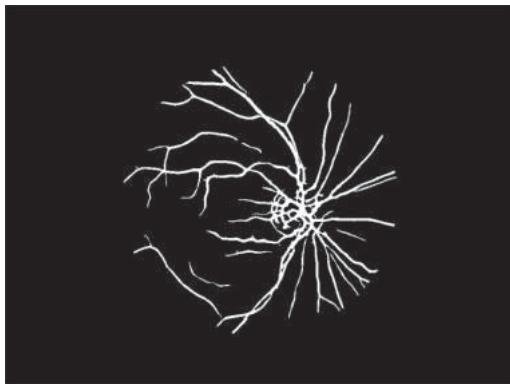


Figure 5. Example of segmentation result of one image of patient with DR from local patient fundus image dataset. The image corresponds with the image in Figure 1b.

experimental video ophthalmoscope—dataset C. The dataset contains 15 video-sequences of various acquisition time  $\sim 16\text{--}50$  seconds. Since the experimental dataset has no ground truth segmentation, evaluation of the proposed methodology was performed in a heuristic manner: the area covered by segmented blood vessels ( $A_{vessel} [\%]$ ) was computed for each frame in the sequence and its variation over the entire sequence have been investigated (Table 3). Figure 6 shows variations of  $A_{vessel}$  for each sequence in a dataset via statistical boxplot diagrams.

Table 3. Evaluation of the method using dynamic low-quality fundus video-sequences. A mean value along with standard deviation of  $A_{vessel}$  is provided.

Sequence no.	Video frame rate [fps]	Number of frames [-]	Acquisition time [s]	$A_{vessel} [\%]$
1	12	189	16	$2.96 \pm 0.90$
2	39	836	21	$2.89 \pm 0.25$
3	12	191	16	$6.49 \pm 0.77$
4	12	189	16	$0.41 \pm 0.37$
5	24	376	16	$1.51 \pm 0.26$
6	35	878	25	$2.47 \pm 0.33$
7	11	189	17	$2.50 \pm 0.35$
8	21	403	19	$6.69 \pm 0.88$
9	12	189	16	$1.14 \pm 0.28$
10	22	389	18	$4.04 \pm 0.69$
11	20	838	42	$0.76 \pm 0.11$
12	12	189	16	$1.38 \pm 0.64$
13	24	376	16	$2.24 \pm 0.47$
14	19	964	51	$2.37 \pm 0.53$
15	20	400	20	$2.37 \pm 0.53$

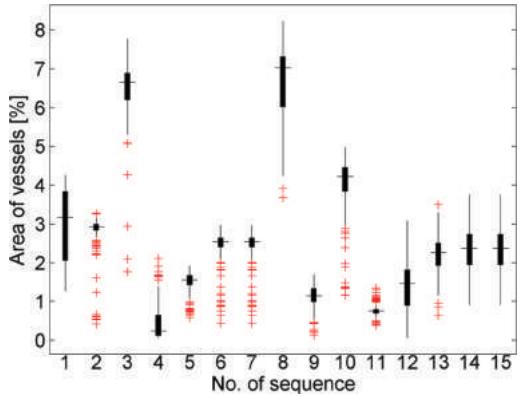


Figure 6. Variations of the area covered by blood vessels expressed by statistical boxplot diagrams. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data points not considered outliers. Outliers are then plotted individually by '+' symbols.

The results in Table 3 and Figure 6 show rather high stability of the segmentation across the sequences ( $\sigma < 1\%$  in all cases). The area variation as well as outliers (Fig. 6) can be caused by inter-frame variability of the image quality across the sequence as well as by fast eye movements of the subject during data acquisition, which finally results in blurred frames. Comparing the results with earlier approach (Odstrcilik 2014), when  $A_{vessel}$  fluctuated more ( $\sigma \sim 1\text{--}3\%$ ), these of the new

proposed technique reveal the method is more stable and robust. As can be seen from an example of resulted segmentation of one frame of a video-sequence in Figure 2d, mainly the blood vessels inside the Optic Disc (OD) region are segmented. It is because these vessels are more contrasted in our low-quality video-sequences. However, this can be further utilized as an advantage for further processing towards evaluation of cardiac cycle induced changes in fundus reflection in the OD region (Tornow 2015).

## 5 CONCLUSIONS

The paper presents an ongoing work on the blood vessel segmentation, which can be used universally for different types of fundus image data. The experiments showed that the fusion of three known approaches—MF, HA, and MP can be used successfully for accurate and precise detection of blood vessels in combination with supervised classification technique given by SVM classifier. It has been also shown that the SVM model trained on standard database can be used for segmentation of blood vessels in completely different data with different image quality (with different resolution and FOV). Especially, the method reached good results in case of experimental and noisy fundus video-sequences with low resolution. For further development, testing on other datasets and possibly combination with other types of features can be considered. In addition, also enhancement and restoration techniques for preprocessing of low-intensity and noisy video-sequence data could be utilized to improve the segmentation.

## ACKNOWLEDGEMENT

This work has been supported by European Regional Development Fund—Project FNUSA-ICRC (No. CZ.1.05/1.1.00/02.0123).

## REFERENCES

- Annunziata R., A. Garzelli, L. Ballerini, et al., “Leveraging Multiscale Hessian-Based Enhancement with a Novel Exudate Inpainting Technique for Retinal Vessel Segmentation,” *IEEE Journal of Biomedical and Health Informatics*, vol. 00, no. 00 (2015).
- Bhuiyan A., R. Kawasaki, E. Lamoureux, et al., “Retinal Vascular Features for Cardio Vascular Disease Prediction: Review,” *Recent Patents Comput. Sci.*, vol. 3, no. 3, pp. 164–175 (2014).
- Budai A., R. Bock, A. Maier, et al., “Robust Vessel Segmentation in Fundus Images,” *International Journal of Biomedical Imaging*, vol. 2013, pp. 11 (2013).
- Chang C.C., C.J. Lin, “LIBSVM: a library for support vector machines,” *ACM Trans Int Sys Tech*, vol. 27, no. 2, pp. 31–39 (2011).
- Cheng E., L. Du, Y. Wu, et al., “Discriminative vessel segmentation in retinal images by fusing context-aware hybrid features,” *Machine Vision and Applications*, vol. 25, no. 7, pp. 1779–1792 (2014).
- Fraz M.M., P. Remagnino, A. Hoppe, “Blood vessel segmentation methodologies in retinal images—A Survey,” *Computer methods and Programs in Biomedicine*, vol. 108, no. 1, pp. 407–433 (2012).
- Hannink J., R. Duits, E. Bekkers, “Crossing-preserving multi-scale vesselness,” *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2014*, vol. 8674, pp. 603–610 (2014).
- Kolar R., B. Hoerer, J. Odstrcilik et al., “Registration of Image Sequences from Experimental Low-Cost Fundus Camera,” *Lecture Notes in Computer Science, Proc. of Workshop on Biomedical Image Registration*, pp. 174–183 (2014).
- Kolar R., V. Harabis, J. Odstrcilik, “Hybrid retinal image registration using phase correlation,” *Imaging Science Journal*, vol. 61, no. 4, pp. 269–384 (2013).
- Lajevardi S.M., A. Arakala, S.A. Davis, et al., “Retina Verification System Based on Biometric Graph Matching Image Processing,” *IEEE Transactions on Image Processing*, vol. 22, no. 9, pp. 3625–3635 (2013).
- Odstrcilik J., R. Kolar, J. Jan, R.P. Tornow, et al., “Blood vessel segmentation in video-sequences from the human retina,” *IEEE International Conference on Imaging Systems and Techniques (IST 2014)*, pp. 129–133 (2014).
- Odstrcilik J., R. Kolar, T. Kubena, et al., “Retinal vessel segmentation by improved matched filtering: evaluation on a new high resolution fundus image database,” *IET Image Process.*, vol. 7, no. 4, pp. 373–383 (2013).
- Rathod D., R.R. Manza, Y.M. Rajput, “Retinal Blood Vessels Extraction using Matched Filter on High Resolution Fundus Image Database,” *International Journal of Computer Applications* (2015).
- Soille P., “Morphological image analysis: principles and applications,” Springer, p. 392 (2010).
- Tornow R.P., R. Kolar, J. Odstrcilik, “Non-mydriatic video ophthalmoscope to measure fast temporal changes of the human retina,” *Proc. SPIE 9540, Novel Biophotonics Techniques and Applications III* (2015).
- Wang S., L. Xu, Y. Wang, et al., “Retinal vessel diameter in normal and glaucomatous eyes: the Beijing eye study,” *Clin. Exp. Ophth.*, vol. 35, no. 9, pp. 800–807 (2007).
- Yu H., S. Barriga, C. Agurto, et al., “Fast vessel segmentation in retinal images using multiscale enhancement and second-order local entropy,” *Proc. SPIE 8315, Medical Imaging 2012: Computer-Aided Diagnosis, 83151B* (2012).

# Line extraction via phase congruency with a novel adaptive scale selection for Poisson noisy images

Vladimir A. Krylov

*Department of Electrical, Electronic and Telecommunication Engineering and Naval Architecture,  
University of Genoa, Genoa, Italy*

James D.B. Nelson

*Department of Statistical Science, University College London, London, UK*

**ABSTRACT:** In this paper we address the problem of extracting curvilinear structures from images in the presence of Poisson-distributed noise that is typically found in low-quality, low-contrast data, as well as in numerous medical imaging modalities. A contrast-robust phase congruency estimation method is proposed such that the optimal range of scales is selected in an adaptive manner for each location in the image. The adaptive regime is driven by a statistically well-principled variance test statistic which is based upon a local measure of the dispersion index. Experimental validation confirms that the adaptive scheme delivers superior line extraction results on mammographic imagery when compared to other recent attempts including the non-adaptive phase congruency.

## 1 INTRODUCTION

The extraction of lines and curvilinear structures (CLS) from noisy and cluttered imagery is one of the fundamental problems in pattern recognition. Myriad applications rely on accurate extraction of lines or line-like features either as an end goal in itself or as a means to support other image processing tasks. Curvilinear extraction can pose serious challenges when the analysed imagery is affected by strong noise or when the image is cluttered by objects that do not contain useful line features (Lee and Kweon 1997). A particularly difficult line detection problem arises when the imagery contains CLS at various scales and when these features overlap. This can happen as the result of 3D-data projection, which is often the case of medical scans. These images typically contain multiple occlusion and superposition artefacts, due to overlaps of features at different layers of tissue.

Recent decades have seen much, and continued, interest in the design of line and CLS detection methods and various different line-models and assumptions have evolved. Notable multi-scale techniques have been designed based on directionally adaptive transforms such as steerable filters (Jacob and Unser 2004), contourlets (da Cunha et al. 2006), feature-adapted beamlets (Berlemont and Olivo-Marin 2010), and curvelets (Ma and Plonka 2010). In contrast stochastic geometry approaches proceed by randomly sampling line

elements and propagating them in such a way as to take into account the line-object interactions, e.g. in (Verdié and Lafarge 2014). At a steep computational complexity price, this allows one to take a far more adaptive and flexible approach which can accommodate various CLS properties by suitable design of prior energy terms. Other approaches combine elements of various extraction techniques, like isotropic non-linear filtering (Liu et al. 2007), space-scale line profile analysis (Steger 1998), skeleton extraction (Jang and Hong 2002), local Radon maximization (Krylov and Nelson 2014). Specialized line detection techniques have been adopted in various application domains, such as remote sensing (Verdié and Lafarge 2014), medical imaging (Obara et al. 2012), biometrical applications (Huang et al. 2008). Often CLS detection is a necessary preprocessing stage to extract complex objects that are comprised of distinct combinations of linear features, such as, e.g., spicule patterns that often surround malignant masses in mammographic images (Sampat et al. 2008, Muralidhar et al. 2010).

In this paper we focus on the scenario of images affected by Poisson-distributed noise, whereas the majority of existing state-of-the-art methods explicitly or implicitly operate under the noise Gaussianity assumption. The Poisson distribution is characterized by one parameter, referred to as intensity, and both the distribution's mean and variance are equal to its value. The corresponding noising

model is particularly appropriate for discretized, positive-value data and characterizes several widely used image modalities, including low-light and low-quality sensor data, astronomy imagery, and many kinds of medical imagery (Gravel et al. 2004), such as, mammographic, X-ray, PET, SPECT, fluorescent confocal microscopy imaging, etc.

Along with the degradation of the analysed imagery by the Poisson noise, we also assume the presence of CLS at different scales, with possible occlusions and superpositions. In order to address the line extraction on such imagery, we will assume that the CLS of interest manifest as image ridges, i.e. smooth elongated curves, whose points are local maxima in at least one direction. Typically, these maxima are identified by resorting to local gradient information (Canny 1986, Lindeberg 1998). In this work we will, instead, employ the phase congruency characteristic, which evaluates the arrangement of phases of the signal at different frequencies and has an advantage of being stable under translation, geometric deformations and contrast variations (Kovesi 2003). This advantage is of significance for various types of imagery, especially in case when some of the relevant CLS demonstrate very weak contrast. In order to extract the phases, we will employ the directional Gabor filters, see (Kovesi 2003).

Since we assume that the imagery is inherently multiscale, the linear features can be superimposed one on top of the other. As a result, this can significantly interfere with the phase congruency estimation which requires non-cluttered line features. To tackle this problem we adaptively identify the optimal range of scales. Unlike (Schenk and Brady 2003), we propose an automatic adaptive scheme that is statistically well-founded in that it is based upon the dispersion index—an oft-utilised goodness-of-fit measure in statistical estimation of Poisson distributions (Karlis and Xekalaki 2000). The dispersion index allows to evaluate the degree of homogeneity of the Poisson-noisy image, and, thus, helps to identify the scale at which the local Gabor filter encounters an edge. Note that the employed adaptive scale selection approach is consistent with the scale-space edge concept defined in (Lindeberg 1998), according to which, the features represent connected sets of points characterized by some measure that persists over a consecutive range of scales.

The contribution of this work lies in the design of dispersion index-based automatic algorithm for adaptive optimal scales identification to extract CLS using phase congruency on Poisson-noisy images. The experimental validation is performed on mammographic images, where CLS extraction supports an array of important tasks such as coregistration or inference regarding tissue malformation or pathology (Muralidhar et al. 2010).

The rest of the paper is arranged as follows. In Section 2 we recall the employed concepts of phase congruency and Gabor filtering. In Section 3 we design a novel scale selection approach for the Poisson-noisy imagery. In Section 4 we report experiments, and in Section 5 we draw the conclusions.

## 2 PHASE CONGRUENCY AND GABOR FILTERING

Phase congruency exploits the expected behaviour of the image line- and edge-like features in the frequency domain (Kovesi 2003). In particular, the frequency components of such features share similar phases—e.g., observe the behaviour of the square wave frequency components at points  $\{-1, 0, 1\}$  corresponding to the edges in Figure 1(a). The phase congruency concept disregards the amplitudes of the frequency components and, similarly to coherence, analyses solely the phase information.

The phase congruency measurement procedure at a signal point is exemplified geometrically in Figure 1(b). The Fourier components in the analysed signal have amplitudes  $C_n$  and phase angles  $\gamma_n$ ; the magnitude of the sum vector is the local energy  $E$ . The phase congruency at point  $z$  is classically defined as  $PC(z) = |E(z)|/\sum_n C_n(z)$ . In this paper we will employ a slightly different definition of phase congruency that allows a better localization (Kovesi 2003):

$$PC = \frac{\sum_n C_n(z) \max[0, \alpha_n(z)]}{\sum_n C_n(z)}, \quad (1)$$

where  $\alpha_n(z) = \cos(\gamma_n(z) - \bar{\gamma}(x)) - |\sin(\gamma_n(z) - \bar{\gamma}(x))|$ . Instead of resorting to the Fourier transform, the local frequency information is obtained via banks

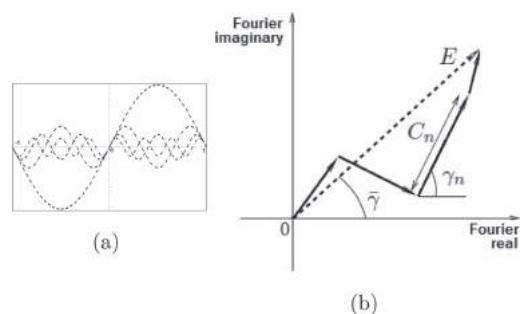


Figure 1. (a) First four components in the Fourier series of a square wave. (b) Fourier components of a signal plotted as a sum:  $C_n$ ,  $\gamma_n$  are the components' amplitudes and phase angles,  $\bar{\gamma}$ —weighted mean phase, and  $E$ —the total magnitude.

of Gabor filters tuned to different spatial frequencies (Kovesi 2003).

Gabor filters have received considerable attention because they have an analytical expression in both time and frequency, are easy to implement, and are highly tunable. It has also been argued that the filters resemble the behaviour of certain cells in the visual cortex of mammals (Serre et al. 2005). We use these filters here as a means to estimate the time-frequency coefficients necessary for the estimation of phase congruency. Other wavelet bases could be used in their place instead, however the choice of wavelet is beyond the scope of this work. A Gabor filter is defined as providing an impulse response equal to a harmonic function multiplied by a Gaussian function. In the frequency domain its filtering function can be written as

$$\Psi(u, v) = \frac{1}{2\pi\sigma_u\sigma_v} \exp\left\{-\frac{(u-u_0)^2}{2\sigma_u^2} - \frac{(v-v_0)^2}{2\sigma_v^2}\right\},$$

with  $\sigma_u = 1/(2\pi\sigma_x)$ ,  $\sigma_u = 1/(2\pi\sigma_x)$ , and  $\sigma_x, \sigma_y$ —the standard deviations in the time domain. The Gabor filter can be thought of as a Gaussian function shifted in frequency to position  $(u_0, v_0)$  at orientation  $\tan^{-1}\left(\frac{u_0}{v_0}\right)$ , see Figure 2. In this paper the Gabor filters are convolved with the signal at  $S$  distinct scales and  $\mathcal{O}$  uniformly spread orientations.

The Gabor filters have been successfully used for various image processing applications (Kovesi 2003, Serre et al. 2005, Teuner et al. 1995). In particular, in case of Poisson-noisy images in (Buciu and Gacsadi 2011) the Gabor wavelets were reported as an efficient tool for directional feature analysis for mammographic imagery. Nevertheless, one of the conclusions in the study was that high frequency image components are rather sensitive to the Poisson noise.

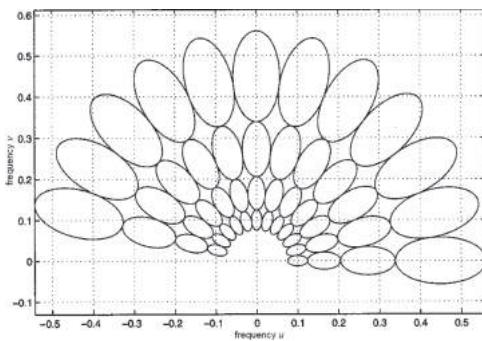


Figure 2. Gabor filters in the frequency domain. Each of the ellipses represents the range of the corresponding filter response in the squared magnitude values from 0.5 to 1.0.

### 3 ADAPTIVE SCALE SELECTION

Consider imagery affected by Poisson noise. Unlike additive Gaussian noise (which is the limit case for Poisson noise (Lehmann and Romano 2005)), it is signal-dependent. This can make separating signal from noise a difficult task. Assume that some image area depicts a constant signal affected by noise—this means that the data will be Poisson-distributed with intensity  $\lambda$  equal to the pure signal intensity (Gravel et al. 2004). The purpose of adaptive scale selection is at every image point to identify the resolution at which this point is no longer contained inside a homogeneous area. In other words, to identify the minimum scale  $s^*$  where the neighbourhood of the given pixel includes the Poisson-distributed values with distinct intensity parameters, i.e., multiple sample populations are represented in this neighbourhood characterised by distinct Poisson distributions.

To estimate  $s^*$  we employ the dispersion index, which is a widely used tool to construct goodness-of-fit tests and analyse the homogeneity property of the Poisson distribution (Karlis and Xekalaki 2000). The dispersion index, also called variance-to-mean ratio, is defined as  $D = \sigma^2/\mu$ , with variance  $\sigma^2$  and mean value  $\mu$ . It is immediate, that the Poisson distribution is characterised by  $D \equiv 1$ . In order to employ the dispersion index to test the homogeneity of a sample of  $N$  random observations  $\{X_i\}_{i=1}^N$ , we employ the Variance Test ( $VT$ ) statistic (Karlis and Xekalaki 2000) defined as

$$VT = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{\bar{X}\sqrt{2N}} - \sqrt{\frac{N}{2}}, \quad (2)$$

with the sample mean  $\bar{X} = \sum_{i=1}^N X_i/N$ . This statistic constitutes the locally most powerful unbiased test for the Poisson homogeneity hypothesis  $\{H_0 : \lambda_1 = \dots = \lambda_N\}$  against the mixed Poisson alternative  $\{H_A : \exists m : 1 < m \leq N, \lambda_1 \neq \lambda_m\}$ . It is immediate that this alternative hypothesis definition corresponds exactly to the specified above region homogeneity specification. The value of  $VT$  is a slight modification of the dispersion index, which is obtained by introducing a correction with respect to the sample size  $N$ . The values of the  $VT$  statistic are standard normal distributed (Karlis and Xekalaki 2000) under  $H_0$ .

In this paper we employ the Gabor filters to evaluate the phase congruency, and, therefore, we will use the  $VT$  on the samples originating from inside the ellipses constructed according to the levels of filter responses in square magnitude—from 0.5 to 1, see Figure 2. Thus, for every scale  $s = 1, \dots, S$  and orientation  $o = 1, \dots, \mathcal{O}$  we retrieve a data sample of size  $N_{so}$ . For each of these samples

we calculate the corresponding value of the sample dispersion index  $VT_{so}$ . Our target is to select the optimal scale  $s^*$  such that the deviation of the  $VT_{so}$  from the standard normal distribution is meaningful. At every scale we calculate  $\mathcal{O}$  distinct values of the variance test and then average over the different orientations because (i) we do not aim to identify the optimal orientation in this procedure, and (ii) the averaging improves robustness with respect to the signal fluctuations. Summarizing the above, the optimal scale can be identified as

$$s^* = \arg \max_{s=1,\dots,S} \sum_{o=1}^{\mathcal{O}} |VT_{so}| / \mathcal{O}.$$

In the experimental study we do not assign statistical test meaning to the  $VT$  values, due to a high variability of confidence levels obtained with the test. This observed variability can be explained by the employed two simplifying assumptions: (i) pixel independence and (ii) constant-intensity signal in the homogeneous zones. Nevertheless, the experiments suggest that at a certain scale a substantial increase of  $VT$  is typically observed, and we will use this behaviour as an indicator to obtain the scale estimates. Specifically, we will set

$$s^* = \arg \max_{s=2,\dots,S} \sum_{o=1}^{\mathcal{O}} |VT_{so}| / \sum_{o=1}^{\mathcal{O}} |VT_{s-1,o}|. \quad (3)$$

After having identified for every pixel the optimal scale  $s^*$ , we perform phase congruency-based edge extraction as in Eq. (1) using scales  $\{s, \dots, \min(s + \Delta s, S)\}$ . It is worth noting that in order to benefit from adaptive scale selection,  $\Delta s$  should be set with a small value, we suggest using  $\Delta s \leq 4$ . Once the features are extracted, it is necessary to pick the bright ridges from all the line- and edge-like features. To do so we propose to multiply the extracted  $PC$  values by the factor  $\max^2[0, 2\bar{\gamma}/\pi]$ , where  $\bar{\gamma}$  is the pixel weighted mean phase (corresponding to the magnitude  $E$ , see Figure 1(b)). This manipulation removes all the dark lines and edges (with  $-\pi/2 \leq \bar{\gamma} \leq 0$ ), and penalizes the less distinct non-symmetric bright line features ( $0 < \bar{\gamma} < \pi/2$ ). Finally, a hard thresholding could be employed to obtain a binary CLS map.

#### 4 EXPERIMENTAL VALIDATION

In this section we report experiments on mammographic images from the Digital Database for Screening Mammography (Heath et al. 2001). We have selected the number of scales  $S = 8$  and consider  $\mathcal{O} = 16$  orientations. The choice of the latter parameter is not significant as long as it is big

enough ( $\mathcal{O} > 8$ ), the former is selected with respect to the CLS present, but in the general case should not be small. The scale range parameter  $\Delta s$  was set equal to 3.

The results on three typical mammographic regions of interest are demonstrated in Figure 3. Note that except for thresholding by the mean  $PC$  value for every set of pixels extracted at the same scale  $s^*$ , we report the raw results, without resorting to image morphology tools, that can improve the visual quality, nor do we discard any scales, which is an option if the CLS are *a priori* expected at some particular scales. We observe that the scale maps are informative on their own and contain the CLS, which are visible as regions with the adaptive scale different from the surrounding background's scale. By construction the designed method is local and can be successfully applied to non-stationary imagery that contains regions characterized by distinct statistical properties. The scale maps (after appropriate smoothing) can be employed as an indicator to distinguish between different layers of the low-fibroglandular tissue on the mammograms.

The designed method demonstrates an ability to automatically adapt the scale such that, on the one hand, the blob in the top image in Figure 3 is separated from some of the CLS but, on the other hand, the higher scale (red) line on the right half of the image is clearly distinguished from the neighbouring low scale CLS. Likewise, in the second image the finer line structures are extracted against the slightly coarser scale content of the fibroglandular tissue more clearly than by the other methods. The third mammogram demonstrates a high variety of CLS at different scales and various levels of occlusion in the tissue. The designed method performs well with higher scale (finer) features but confuses some low scale features.

In Figure 3 we present results obtained with three benchmark methods. Comparison between the standard phase congruency extractor (based on global manual setting of scales) and the proposed method demonstrates a substantially stronger capacity of the latter to extract multiscale features. Comparing with the Radon-based enhancement method (Sampat et al. 2008) reveals the contrast dependence of the latter and its weaker performance in modeling different scale features. As compared with a primitive-based monoscale method (Krylov and Nelson 2014), the proposed approach produces noisier but also more CLS-sensitive maps. The designed method can be used to reduce the parametric initialization in (Krylov and Nelson 2014) and address the multiscale line-features.

In Figure 4 we demonstrate the performance of the scale selection based on  $s^*$  at several image points and visualize the process of the final map

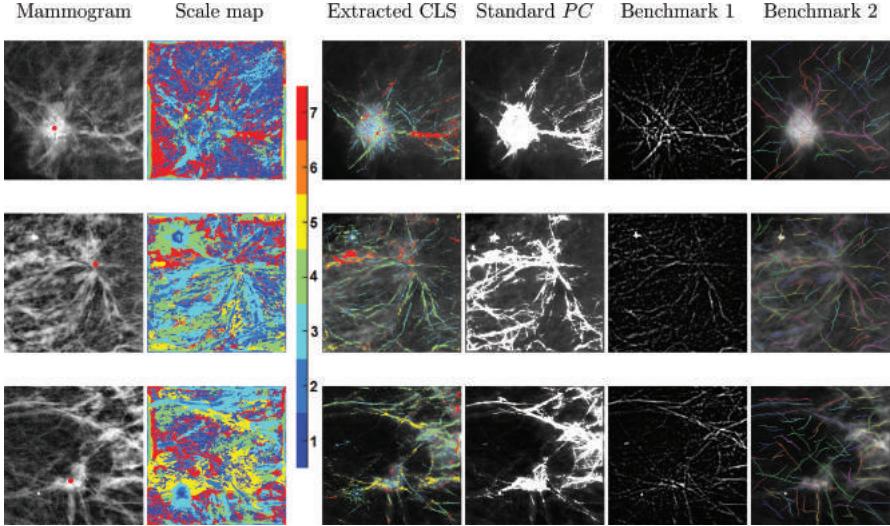


Figure 3. CLS extraction on mammograms containing architectural distortions (with their centers marked by red dots) and surrounded by CLS. The proposed method's results are coloured according to the optimal scale at which the CLS are extracted. The comparisons are with Radon-based method (Sampat et al. 2008) (Benchmark 1) and CLS extractor (Krylov and Nelson 2014) (Benchmark 2).

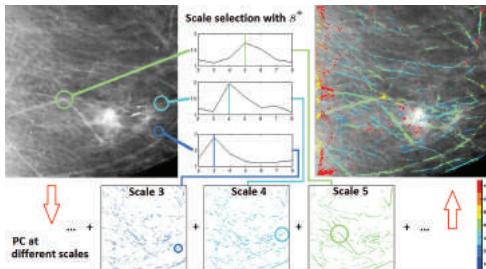


Figure 4. Example of scale selection by the proposed method and construction of line map from different scale components.

construction as a sum of phase congruency components extracted at adaptively estimated scales. It can be seen how CLS of different scales contribute to the final line-detection map.

## 5 CONCLUSIONS

We have proposed a novel automatic approach to address the problem of multiscale CLS extraction from Poisson-noisy images. This approach employs the phase congruency principle and performs the adaptive scale selection for the phase extraction. The latter is performed in a statistically well-founded way by employing the variation

test statistic based on the dispersion index value. We thus formulate an application-independent Poisson-noise specific technique, which in this respect is unlike the majority of existing state-of-the-art methods that typically target the Gaussian noise. The proposed line detection method has been experimentally validated and compared with benchmark methods on mammographic images.

The principal direction for future development is in designing of primitive-based multiscale CLS extraction techniques that could benefit from the pixel-based result and scale selection procedure designed in this study, e.g., by integration with local Radon transforms (Krylov and Nelson 2014), active contours (Muralidhar et al. 2010), etc.

## REFERENCES

- Berlemont, S. & J.-C. Olivo-Marin (2010). Combining local filtering and multiscale analysis for edge, ridge, and curvilinear objects detection. *IEEE Trans. Image Process.* 19(1), 74–84.
- Buciu, I. & A. Gacsadi (2011). Directional features for automatic tumor classification of mammogram images. *Biomed Signal Process Control* 6(4), 370–378.
- Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell. PAMI-8(6)*, 679–698.
- da Cunha, A.L., J. Zhou, & M. Do (2006, Oct). The non-subsampled contourlet transform: Theory, design, and applications. *IEEE Trans. Image Process.* 15(10), 3089–3101.

- Gravel, P., G. Beaudoin, & J.A. De Guise (2004). A method for modeling noise in medical images. *IEEE Trans. Med. Imaging* 23(10), 1221–1232.
- Heath, M., K. Bowyer, D. Kopans, R. Moore, & W. Kegelmeyer (2001). The digital database for screening mammography. In M. Yaffe (Ed.), *Proc. of the Fifth Int. Workshop on Digital Mammography*, pp. 212–218. Medical Physics Publ. available online at <http://marathon.csee.usf.edu/Mammography/Database.html>.
- Huang, D.-S., W. Jia, & D. Zhang (2008). Palmprint verification based on principal lines. *Pattern Recogn.* 41(4), 1316–1328.
- Jacob, M. & M. Unser (2004). Design of steerable filters for feature detection using Canny-like criteria. *IEEE Trans. Pattern Anal. Mach. Intell.* 26(8), 1007–1019.
- Jang, J.-H. & K.-S. Hong (2002). Detection of curvilinear structures and reconstruction of their regions in gray-scale images. *Pattern Recogn.* 35(4), 807–824.
- Karlis, D. & E. Xekalaki (2000). A simulation comparison of several procedures for testing the Poisson assumption. *The Statistician* 49(3), 355–382.
- Kovesi, P. (2003). Phase congruency detects corners and edges. In *Proc. Australian Pattern Recogn. Soc. Conf. (DICTA)*.
- Krylov, V.A. & J.D.B. Nelson (2014). Stochastic extraction of elongated curvilinear structures with applications. *IEEE Trans. Image Process.* 23(12), 5360–5373.
- Lee, J.-W. & I.-S. Kweon (1997). Extraction of line features in a noisy image. *Pattern Recognition* 30 (10), 1651–1660.
- Lehmann, E. & J. Romano (2005). *Testing statistical hypotheses*. London, UK: Springer.
- Lindeberg, T. (1998). Feature detection with automatic scale selection. *Int. J. Comput. Vision* 30(2), 79–116.
- Liu, L., D. Zhang, & J. You (2007). Detecting wide lines using isotropic nonlinear filtering. *IEEE Trans. Image Process.* 16(6), 1584–1595.
- Ma, J. & G. Plonka (2010). The curvelet transform. *IEEE Signal Process. Magazine* 27(2), 118–133.
- Muralidhar, G., A. Bovik, J. Giese, M. Sampat, G. Whitman, T. Haygood, T. Stephens, & M. Markey (2010). Snakules: A model-based active contour algorithm for the annotation of spicules on mammography. *IEEE Trans. Med. Imaging* 29(10), 1768–1780.
- Obara, B., M. Fricker, D. Gavaghan, & V. Grau (2012, May). Contrast-independent curvilinear structure detection in biomedical images. *IEEE Trans. Image Process.* 21(5), 2572–2581.
- Sampat, M., G. Whitman, A. Bovik, & M. Markey (2008). Comparison of algorithms to enhance spicules of spiculated masses on mammography. *J. Digit. Imaging* 21(1), 9–17.
- Schenk, V.U.B. & M. Brady (2003). Improving phase-congruency based feature detection through automatic scale-selection. In *Progress in Pattern Recognition, Speech and Image Analysis*, pp. 121–128.
- Serre, T., L. Wolf, & T. Poggio (2005). Object recognition with features inspired by visual cortex. In *Proc. Comp. Vision Pattern Recogn. (CVPR)*, Volume 2, pp. 994–1000.
- Steger, C. (1998). An unbiased detector of curvilinear structures. *IEEE Trans. Pattern Anal. Mach. Intell.* 20(2), 113–125.
- Teuner, A., O. Pichler, & B.J. Hosticka (1995). Unsupervised texture segmentation of images using tuned matched Gabor filters. *IEEE Trans. Image Process.* 4(6), 863–870.
- Verdié, Y. & F. Lafarge (2014). Detecting parametric objects in large scenes by Monte Carlo sampling. *Int. J. Comput. Vision* 106(1), 57–75.

# Diagnosis of human intestinal parasites by deep learning

A.Z. Peixinho, S.B. Martins, J.E. Vargas & A.X. Falcão

*Institute of Computing, University of Campinas, Campinas, Brazil*

J.F. Gomes

*Institute of Biology and Institute of Computing, University of Campinas, Campinas, Brazil*

C.T.N. Suzuki

*Immunocamp Science and Technology LTDA, Brazil*

**ABSTRACT:** Intestinal parasitic infections can cause serious health problems, especially in children and immunodeficient adults. In order to make the diagnosis of intestinal parasites fast and effective, we have developed an automated system based on optical microscopy image analysis. This work presents a deep learning approach to discover more effective parasite image features from a small training set. We also discuss how to prepare the training set in order to cope with object scale and pose variations. By using random kernels, our approach considerably simplifies the learning task of a suitable convolutional network architecture for feature extraction. The results demonstrate significant accuracy gains in classification of the 15 most common species of human intestinal parasites in Brazil with respect to our previous state-of-the-art solution.

## 1 INTRODUCTION

Infections by human intestinal parasites can cause death, physical and mental disorders in children and immunodeficient adults (Popejoy 2012). The detection of parasites in optical microscopy images obtained from fecal samples is determinant for the diagnosis. However, this procedure is not effective when executed by humans. In order to circumvent the problem, automated image analysis methods have been investigated (Dogantekin et al. 2008, Avci and Varol 2009, Al-Sameraai et al. 2013, Flores-Quispe et al. 2014), but they are usually limited to a few species of parasites and do not take into account automatic image acquisition problems, such as focus adjustment, object motion, overlapping with impurity components, etc. In (Suzuki et al. 2013a), we proposed a first solution to automate the diagnosis of the 15 most common species of parasites (protozoa and helminths) in Brazil. Subsequently, the effectiveness and efficiency of this solution was improved in (Suzuki et al. 2013b).

Inspired by the recent success of deep learning techniques (Krizhevsky et al. 2012, Chiachia et al. 014a, Menotti et al. 2015), we present a method for parasite feature learning from optical microscopy images. Our approach uses convolutional networks (ConvNets) with random kernels to considerably simplify the learning task of the ConvNet

architecture. We discuss how to prepare the training set in order to cope with class object scale and pose variations. The method is validated in comparison with our previous state-of-the-art solution in (Suzuki et al. 2013b), by using a dataset with 16,437 images whose automatic annotation was confirmed/corrected by experts in Parasitology.

Section 2 shortly describes the automated system for the diagnosis of human intestinal parasites and the previous feature extraction and classification methods, while Section 3 presents the proposed ConvNet-based approach for feature extraction and classification. Experimental results are presented in Section 4 and Section 5 draws the conclusive remarks.

## 2 AUTOMATED SYSTEM FOR THE DIAGNOSIS OF INTESTINAL PARASITES

The automated system for the diagnosis of human intestinal parasites consists of an optical microscope with motorized stage, focus driver, and digital camera, all controlled by a computer for automated image acquisition and analysis. The acquired images are first segmented into objects (single connected components), which can be either impurities or parasites. These objects are located through the color difference between the

image background, with blue color, and candidate objects, with brown color. Objects with area outside the range of parasites are eliminated, and the remaining ones are used as markers for a fine delineation via the IFT-watershed algorithm (Falcão et al. 2004). For image analysis, the parasites images are divided into three different groups: (a) protozoan cysts; (b) helminth eggs; and (c) helminth larvae. The three groups can be easily distinguished by simple shape measures, being helminth eggs and larvae about 20 times bigger than protozoa. Hence, the pipeline extracts more complex features and classify the objects within each group into one out of the 15 species of interest or as impurity. The main challenge is to learn the most effective features to distinguish similar objects from different classes (Fig. 1).

In (Suzuki et al. 2013b), we defined features based on our experience: color histograms of texture and flat regions, object area, perimeter, symmetry, major and minor axes of the best fit ellipse

within the object, the difference between the ellipse and the object, energy, entropy, variance, and homogeneity of the co-occurrence matrix. These features were used with different weights in a single feature vector. Some features are more relevant than others to differentiate the classes (e.g., major and minor axes of the best fit ellipse, perimeter, area and histograms of texture and flat regions), and the weight of each feature was found by optimization in order to maximize classification accuracy within the training set by cross-validation. The dissimilarity between objects was measured by the Euclidean distance of their feature vectors, considering the weight of each feature. We also used optimum-path forest classification (Papa et al. 2009), since it outperformed Support Vector Machines (SVMs) with this feature set.

In this work, we learn the ConvNet architecture that produces effective object feature vectors for SVM-based classification. In this case, the resulting high-dimensional feature space favored linear SVMs.

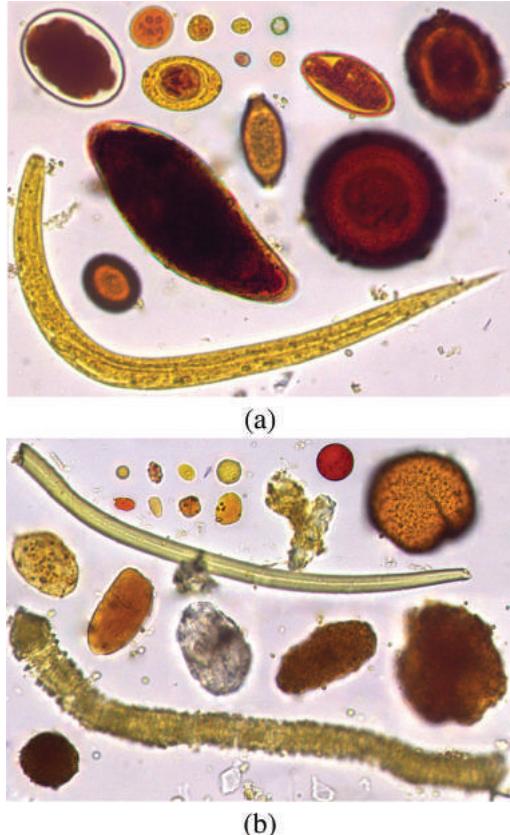


Figure 1. (a) The 15 most common species of human intestinal parasites in Brazil and (b) similar impurity objects.

### 3 CONVOLUTIONAL NETWORK

Our ConvNet-based feature learning approach relies on random kernels to considerably simplify the hyper-parameter optimization of the ConvNet architecture (Pinto and Cox 2011).

ConvNet models are biologically inspired models, that includes feature detecting layers, each of which alternated with sub-sampling layers to obtain properties leading to some translation and deformation invariance (Matsugu et al. 2003).

Basically, a ConvNet is composed of linear and non-linear image processing operations. When stacked, these operations essentially extract higher level representations, named *multiband images*, whose pixel attributes are concatenated into a high-dimensional feature vector per input image for subsequent classification.

We consider a *layer* as the combination of well-known operations (i) *convolution* with a bank of filters, (ii) *activation*, (iii) spatial *pooling*, and (iv) *local normalization*. Figure 3 (a) shows how these operations are stacked within a layer.

This section describes ConvNets from the image processing perspective. Let  $\hat{I} = (D_I, \vec{I})$  be a multiband image, where  $D_I \subset Z^2$  is the image domain and  $\vec{I}(p) = \{I_1(p), I_2(p), \dots, I_m(p)\}$  is the attribute vector of a pixel  $p = (x_p, y_p) \in D_I$ .

#### 3.1 Filter bank convolution

Let  $\Phi_i = (\mathcal{A}, \vec{W}_i)$  be a multiband filter with weight vector  $\vec{W}_i(q) = \{w_{i,1}(q), w_{i,2}(q), \dots, w_{i,m}(q)\}$ , where  $q \in \mathcal{A}(p)$  is adjacent to the origin  $p$  of the filter.

A multiband filter bank  $\Phi = \{\Phi_1, \Phi_2, \dots, \Phi_n\}$  is a set of filters  $\Phi_i = (\mathcal{A}, \bar{W}_i)$ ,  $i = \{1, 2, \dots, n\}$ . The weights in each  $\bar{W}_i(q)$  are randomly generated from a uniform distribution and normalized to zero mean and unit norm.

The convolution between an input image  $\hat{I}$  and a filter  $\Phi_i$  produces a band  $i$  of the filtered image  $\hat{J} = (D_J, \bar{J})$ , where  $D_J \subset D_I$  and  $\bar{J} = (J_1, J_2, \dots, J_n)$  as shown in Figure 2.

### 3.2 Activation

The activation operation simply creates an image  $\hat{K} = (D_J, \bar{K})$  (a sparse code since we use zero mean and unit norm kernels) by  $K_i(p) = \max(J_i(p), 0)$ , where  $p \in D_J$  are pixels of the image, and  $i = \{1, 2, \dots, n\}$  are the image bands.

### 3.3 Spatial pooling

Spatial pooling aims at bringing small translational invariance to the features by aggregating activations from the same filter in a given region (LeCun et al. 1998). We adopted the max-pooling operation.

Let  $D_M = D_K/s$  be a regular subsampling of every  $s \geq 1$  pixels in an adjacency  $\mathcal{B}$ . We call  $s$  the *stride* of the pooling operation which results into a spatial resolution reduction for  $s > 1$ . The pooling operation creates the image  $\hat{M} = (D_M, \bar{M})$ ,

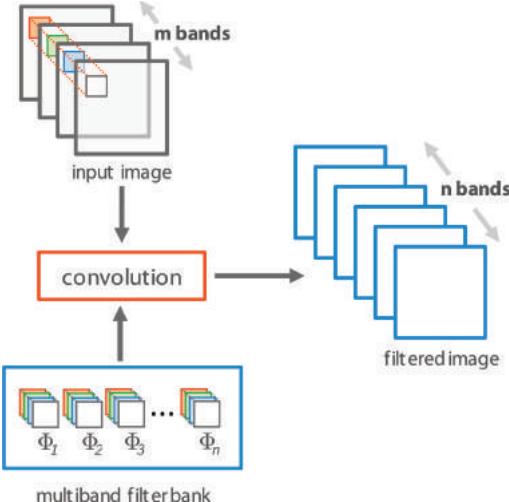


Figure 2. The convolution of an  $m$ -band with the  $m$ -band filter bank. For a better understanding of the operation, we have assigned a different color for each filter band and highlighted a specific region in the input image that follows the same sequence of colors. (Martins and Falcão 2015).

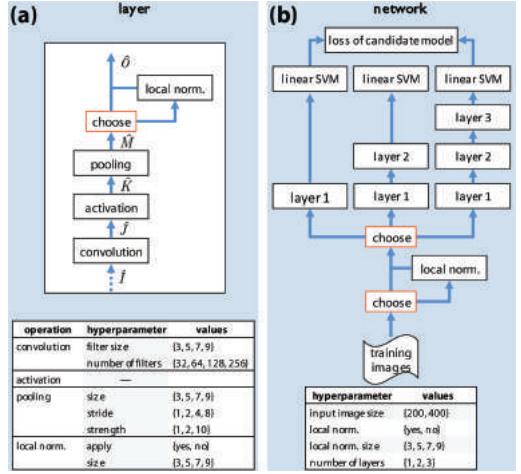


Figure 3. Schematic diagram of the considered framework that illustrates how the fundamental operations are combined in a layer (a) and how the ConvNet is instantiated and evaluated according to the hyper-parameter values (b). The tables show the adopted values for all considered hyper-parameters of each layer (a) and global hyper-parameters (b). This figure was adapted from (Menotti et al. 2015).

where  $\vec{M}(p) = \{M_1(p), M_2(p), \dots, M_n(p)\}$  and  $M_i(p) = \max_{q \in \mathcal{B}(p)} \{M_i(q)\}$ ,  $i = 1, 2, \dots, n$ .

### 3.4 Divisive normalization

The last operation is the divisive normalization, which is based on gain control mechanisms found in cortical neurons (Geisler and Albrecht 1992). It creates an output image  $\hat{O} = (D_O, \bar{O})$ ,  $D_O \subset D_M$ , where  $\bar{O}(p) = \{O_1(p), O_2(p), \dots, O_n(p)\}$  and

$$O_i(p) = \frac{M_i(p)}{\sqrt{\sum_{j=1}^n \sum_{q \in \mathcal{C}(p)} M_j(q) M_j(q)}} \quad (1)$$

for some adjacency  $\mathcal{C}$ . It promotes a competition among pooled filter bands, such that high responses predominate over low ones. The output feature vector results from the concatenation of  $\hat{O}(p)$  for all  $p \in D_O$  when  $\hat{O}$  is the output of the last layer.

### 3.5 Hyper-parameter optimization

As presented in Figure 3 (a), each layer has seven hyper-parameters (a considerable reduction in comparison with ConvNets that learn kernel weights).

The seven hyper-parameters determine the local features to be highlighted in each layer.

1. Convolution window size: neighborhood considered for convolution with the filter bank. It determines the scale of texture details to be extracted from the input.
2. Number of filters in the bank: higher is the number of filters more information is captured from the input.
3. Pooling window size: neighborhood considered in the pooling step. It determines how much translational variation is discarded.
4. Pooling stride: it determines the reduction factor applied to the input resolution.
5. Pooling strength: it determines how susceptible the pooling is to the highest response in the input.
6. Local normalization: it indicates whether or not local normalization must be applied to its input.
7. Local normalization window size: neighborhood considered for local normalization. The highest response within this neighborhood prevails over the lower ones.

There are also four network hyper-parameters.

1. Input image size: it must reveal enough texture details about the objects of interest.
2. Local normalization: it indicates whether or not local normalization must be applied to its input.
3. Local normalization window size: neighborhood considered for local normalization. The highest response within this neighborhood prevails over the lower ones.
4. Number of layers in the network: each layer is defined by the aforementioned layer hyper-parameters.

For this application, it was enough to learn deep representations up to three layers (Fig. 3 (b)). The hyper-parameters above define the network architecture and their possible values define a *search space* for optimization.

We adopted the strategy in (Menotti et al. 2015), which uses *random search* of the possible discrete values for all hyper-parameters to find a ConvNet that produces the highest classification accuracy, using multiple random subsampling in the training set.

Additionally, we evaluated the impact of the input image resolution and class imbalance, since helminth eggs and larvae are considerably larger than protozoa, the number of impurities is considerably higher than the number of parasites, and some species are less frequent than others.

## 4 EXPERIMENTS AND RESULTS

The experiments used a dataset with 16,437 objects (segmented images with their corresponding class as confirmed by experts), including examples from the 15 most common species of human intestinal parasites in Brazil and impurities.

According to shape size and symmetry measures, candidate objects were automatically divided into three groups, *protozoa/impurities*, *helminth eggs/impurities*, and *larvae/impurities*, with 5,112, 9,568, and 1,757 images, respectively.

Each group of images was randomly divided into three subsets,  $Z_1$ ,  $Z_2$ , and  $Z_3$ , with 10%, 45%, and 45% of the images from each class, respectively, such that  $Z_1 \cap Z_2 \cap Z_3 = \emptyset$ . The ConvNet architecture (hyper-parameters) was learned from set  $Z_1$ , by using the software developed in our group Chiachia (2014b). Then the ConvNet-based features were extracted from all images in  $Z_2$  and  $Z_3$ , set  $Z_2$  was used to train a linear SVM classifier with a precomputed kernel and a hard margin ( $C = 10^5$ ). The classification accuracy (Papa et al. 2009) was obtained on set  $Z_3$ . Training from  $Z_2$  and test on  $Z_3$  was repeated 10 times with random sets,  $Z_2$  and  $Z_3$ , for statistical validation.

In order to evaluate the impact of input image resolution (object scale) for architecture learning, we evaluated two types of input image resolutions:  $200 \times 200$  and  $400 \times 400$  pixels. The impact of class imbalance in architecture learning was evaluated by cross-validation in  $Z_1$ , when features were learned from the actual  $Z_1$  and from a modified version of  $Z_1$  with a same number of samples per class.

### 4.1 Feature learning

Considering the discrete hyper-parameter search space in Figure 3, the candidate ConvNets were evaluated by training a SVM classifier on 50% of the samples in  $Z_1$  and testing it on the remaining samples. This process was also repeated for 50%–50% random splits of  $Z_1$  in order to select the best architecture and evaluate the impact of input image resolution and class imbalance.

Table 1 shows the accuracy results of the best ConvNet on 50% of  $Z_1$  for the cases of balanced and unbalanced versions of it.

Class imbalance is not a main issue for feature learning in the cases of *protozoa* and *larvae*, but it considerably affects the learning process in the case of helminth eggs. Therefore, the final ConvNet architecture for each group was learned by forcing a same number of samples per class in a modified version of  $Z_1$ .

Table 1. Accuracy results of the best ConvNet on 50% of  $Z_1$ , using the balanced and unbalanced versions of it.

	Balanced	Unbalanced
Protozoa	$0.8288 \pm 0.0150$	$0.8171 \pm 0.0190$
Eggs	<b><math>0.8994 \pm 0.0212</math></b>	$0.8310 \pm 0.0358$
Larvae	$0.8735 \pm 0.0778$	$0.8688 \pm 0.0240$

Table 2 shows the accuracy results of the best ConvNet on 50% of the balanced version of  $Z_1$ , for two input image resolutions.

This experiment shows that ConvNets cannot deal with large object scale variations in a same dataset. Therefore, we fixed the input image resolution in  $400 \times 400$  pixels for *protozoa* feature learning and in  $200 \times 200$  pixels for eggs and *larvae* feature learning.

#### 4.2 Feature vector evaluation

Given the best ConvNet architecture for each group, we extract the ConvNet features from all images in  $Z_2$  and  $Z_3$ , use  $Z_2$  to train a SVM classifier, and test its accuracy on  $Z_3$ , for each 50% – 50% random split of  $Z_2 \cup Z_3$  into  $Z_2$  and  $Z_3$ .

Simard et al. (2003) points out the benefits of applying transformations in the training set to increase the ConvNet effectiveness. In order to cope with possible object variations in pose, which often occurs for parasites and impurities in the real scenario, we also evaluated the classifier when  $Z_2$  is extended by sample rotations of 90°, 180° and 270° for training. Table 3 shows the average accuracy of the classifier over  $Z_3$  when features are learned without and with additional rotated samples in  $Z_2$ .

A slight improvement can be observed when adding rotated samples to  $Z_2$ , except for the case of *helminth* eggs.

Table 2. Accuracy results of the best ConvNet on 50% of the balanced version of  $Z_1$ , for two input image resolutions.

	$200 \times 200$	$400 \times 400$
Protozoa	$0.8288 \pm 0.0150$	<b><math>0.8970 \pm 0.0583</math></b>
Eggs	<b><math>0.8994 \pm 0.0212</math></b>	$0.8983 \pm 0.0207$
Larvae	<b><math>0.8735 \pm 0.0778</math></b>	$0.8568 \pm 0.0467$

Table 3. Average accuracy over  $Z_3$ , when features are learned without and with additional sample rotation in  $Z_2$ .

Dataset	Without rotation	Using rotation
Protozoa	$0.9559 \pm 0.0074$	<b><math>0.9649 \pm 0.0043</math></b>
Eggs	$0.9646 \pm 0.0048$	$0.9679 \pm 0.0033$
Larvae	$0.9511 \pm 0.0090$	<b><math>0.9570 \pm 0.0079</math></b>

Table 4. Average accuracy over  $Z_3$  of the ConvNet-based approach and the method in (Suzuki et al. 2013b).

Dataset	Suzuki et al. (2013b)	ConvNet
Protozoa	$0.9028 \pm 0.0038$	<b><math>0.9649 \pm 0.0043</math></b>
Eggs	$0.9316 \pm 0.0074$	<b><math>0.9679 \pm 0.0033</math></b>
Larvae	$0.9536 \pm 0.0062$	$0.9570 \pm 0.0079$

Table 5. Average recall and precision of the proposed approach on each class of parasite and impurity.

Class	Recall	Precision
<i>Entamoeba coli</i>	0.9867	0.9753
<i>Entamoeba histolytica/E. dispar</i>	0.8868	0.8594
<i>Endolimax nana</i>	0.9433	0.9175
<i>Giardia duodenalis</i>	0.9391	0.9364
<i>Iodamoeba butschlii</i>	0.9521	0.9532
<i>Blastocystis hominis</i>	0.8833	0.9164
Protozoa impurities	0.9740	0.9782
<i>Hymenolepis nana</i>	0.9946	0.9856
<i>Hymenolepis diminuta</i>	0.9258	0.9345
Hookworms	0.8510	0.9581
<i>Enterobius vermicularis</i>	0.9313	0.9679
<i>Ascaris lumbricoides</i>	0.9043	0.9199
<i>Trichuris trichiura</i>	0.9950	0.9888
<i>Schistosoma mansoni</i>	0.9044	0.9153
<i>Taenia</i> spp.	0.9904	0.9734
Helminth eggs impurities	0.9827	0.9765
<i>Strongyloides stercoralis</i>	0.9269	0.9116
Helminth larvae impurities	0.9871	0.9894

Table 6. Hyperparameters defining the best ConvNet architecture for each parasite group.

Parameters	Larvae	Proto	Eggs
<b>Network</b>			
Number of layers	3	2	3
Input size	$200 \times 200$	$400 \times 400$	$200 \times 200$
Input normalization	No	Yes	Yes
Input norm. window	–	5	7
<b>Layer #1</b>			
Number of filters	32	128	32
Convolution size	7	5	7
Pooling size	7	5	9
Pooling stride	4	2	1
Pooling strength	1	2	1
Normalization	No	No	No
Normalization size	–	–	–
<b>Layer #2</b>			
Number of filters	128	256	256
Convolution size	3	9	9
Pooling size	5	7	5
Pooling stride	1	4	4
Pooling strength	2	1	10
Normalization	Yes	Yes	Yes
Normalization size	9	7	3
<b>Layer #3</b>			
Number of filters	256	–	256
Convolution size	5	–	3
Pooling size	9	–	9
Pooling stride	8	–	4
Pooling strength	10	–	10
Normalization	No	–	No
Normalization size	–	–	–

The same procedure involving  $Z_2$  and  $Z_3$  was executed with the original feature vector (which is not based on deep learning) and the optimum-path forest classifier, as proposed in (Suzuki et al. 2013b). Finally, we compare the results obtained by the ConvNet-based approach with those of the method in (Suzuki et al. 2013b) (see Table 4).

Except for the case of *larvae*, the proposed approach can considerably improve effectiveness of our automated system for the diagnosis of human intestinal parasites. This can be explained by the uniformly good performance in classification over all 15 species of parasites (see Table 5).

The best architecture for each parasite group is shown in Table 6.

## 5 CONCLUSION AND FUTURE WORK

We proposed a ConvNet-based approach for the diagnosis of human intestinal parasites and presented a strategy to deal with class imbalance and considerable object scale variations in the dataset. Our approach has demonstrated the ability to learn features from a small dataset and obtain significant gains in classification accuracy for *protozoa* and *helminth* eggs, maintaining the same good performance for larvae, as compared to our previous state-of-the-art solution (Suzuki et al. 2013b).

The ConvNet-based feature extraction can be efficiently implemented in parallel and the linear SVM classification is a simple inner product between feature vectors and the normal vector of the hyper-plane of each class. Therefore, we may conclude that our ConvNet-based approach can be an efficient and effective solution for the automated diagnosis of human intestinal parasites. We intend to further investigate ConvNet architectures for the specific least accurate cases in Table 5 and extend the work to the diagnosis of parasites in animals.

## REFERENCES

- Al-Sameraai, R.S.H., G.K. Hawari, & M. Zeehaida (2013). Automated system for diagnosis intestinal parasites by computerized image analysis. *Modern Applied Science* 7(5), 98–114.
- Avcı, D. & A. Varol (2009). An expert diagnosis system for classification of human parasite eggs based on multi-class svm. *Expert Systems with Applications* 36(1), 43–48.
- Chiachia, G. (2014b). Hyperparameter Optimization made Simple.
- Chiachia, G., A.X. Falcão, N. Pinto, A. Rocha, & D. Cox (2014a, Dec). Learning person-specific representations from faces in the wild. *IEEE Transactions on Information Forensics and Security (TIFS)* 9(12), 2089–2099.
- Dogantekin, E., M. Yilmaz, A. Dogantekin, E. Avci, & A. Sengur (2008). A robust technique based on invariant moments—anfis for recognition of human parasite eggs in microscopic images. *Expert Systems with Applications* 35(3), 728–738.
- Falcão, A., J. Stolfi, & R. Lotufo (2004). The image foresting transform theory, algorithms, and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(1), 19–29.
- Flores-Quispe, R., R.E.P. Escarzina, Y. Velazco-Paredes, & C.A.B. Castanon (2014). Classification of human parasite eggs based on enhanced multitexton histogram. In *IEEE Colombian Conference on Communications and Computing (COLCOM)*, pp. 1–6.
- Geisler, W.S. & D.G. Albrecht (1992). Cortical neurons: Isolation of contrast gain control. *Vision Research* 32(8), 1409–1410.
- Krizhevsky, A., I. Sutskever, & G. Hinton (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 1097–1105.
- LeCun, Y., L. Bottou, Y. Bengio, & P. Haffner (1998). Gradient based learning applied to document recognition. *Proceedings of the IEEE* 86(11), 2278–2324.
- Martins, S.B. & A.X. Falcão (2015). A fast and robust negative mining approach for user enrollment in face recognition systems. Master's thesis, Universidade Estadual de Campinas, Campinas, São Paulo—Brazil.
- Matsugu, M., K. Mori, Y. Mitari, & Y. Kaneda (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks* 16(5–6), 555–559. Advances in Neural Networks Research: {IJCNN} '03.
- Menotti, D., G. Chiachia, A. Pinto, W. Schwartz, H. Pedrini, A.X. Falcão, & A. Rocha (2015). Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Transactions on Information Forensics and Security (TIFS) PP(99)*, 1–1.
- Papa, J.P., A.X. Falcão, & C.T.N. Suzuki (2009). Supervised pattern classification based on optimum-path forest. *International Journal of Imaging Systems and Technology* 19(2), 120–131.
- Pinto, N. & D. Cox (2011). Beyond simple features: A largescale feature search approach to unconstrained face recognition. In *IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG)*, pp. 8–15.
- Popejoy, M.W. (2012). Working to overcome the global impact of neglected tropical diseases. *Perspectives in Public Health* 132(4), 192.
- Simard, P.Y., D. Steinkraus, & J.C. Platt (2003). Best practices for convolutional neural networks applied to visual document analysis. In *International Conference on Document Analysis and Recognition*, pp. 958–963.
- Suzuki, C.T.N., J.F. Gomes, A.X. Falcão, J.P. Papa, & S.H. Shimizu (2013a). Automatic segmentation and classification of human intestinal parasites from microscopy images. *IEEE Transactions on Biomedical Engineering* 60(3), 803–812.
- Suzuki, C.T.N., J.F. Gomes, A.X. Falcão, S.H. Shimizu, & J.P. Papa (2013b). Automated diagnosis of human intestinal parasites using optical microscopy images. In *IEEE International Symposium on Biomedical Imaging*, pp. 460–463.

## Texture-energy features for microaneurysms detection

D. Veiga

*Enermeter, Parque Industrial de Celeirós, Lugar de Gaião, Braga, Portugal  
University of Minho, Centro Algoritmi, Azurém, Guimarães, Portugal*

N. Martins

*Centro de Computação Gráfica, Azurém, Guimarães, Portugal  
Enermeter, Parque Industrial de Celeirós, Lugar de Gaião, Braga, Portugal*

C. Pereira & M. Ferreira

*Enermeter, Parque Industrial de Celeirós, Lugar de Gaião, Braga, Portugal*

J. Monteiro

*University of Minho, Centro Algoritmi, Azurém, Guimarães, Portugal*

**ABSTRACT:** The appearance of microaneurysms is the first clinical sign in Diabetic retinopathy. These lesions appear as small red dots in color fundus images. Their detection in a preliminary phase could prevent disease progression which leads to blindness. This paper addresses the problem of automatic microaneurysms detection in color fundus images. Laws texture-energy measures were selected as feature extractor and descriptor. The suitability of these measures were investigated by testing several kernel sizes followed by support vector machine classification.

Tests were made on the LaTIM (laboratoire de Traitement de l'Information Médicale) database and results are evaluated in terms of the Free-Response Receiver-Operating Characteristic (FROC) curve. The best performance was achieved with  $31 \times 31$  kernel size.

To our knowledge, this is the first time that Laws local texture-energy transform is used to segment microaneurysms. Results are promising and a next step to analyse each microaneurysm candidate is needed.

### 1 INTRODUCTION

Detection of microaneurysms (MAs) provides insight about Diabetic Retinopathy (DR) presence, a frequent complication of diabetes mellitus. These lesions are often an early manifestation of the disease and their premature detection can delay the pathology evolution by induction of a tighter metabolic control. In addition, if the disease is detected on time, ophthalmologists can act therapeutically using laser treatment before irreversible damage occurs in the retina. MAs appear as small red dots in the superficial retinal layers and can be observed in digital color fundus images (Neubauer & Ulbig 2007).

Published methods on automatic segmentation of microaneurysms can be categorized in three groups: mathematical morphology (Walter, Massin, Erginay, Ordonez, Jeulin, & Klein 2007, Cree 2009, Niemeijer, van Ginneken, Staal, Suttorp-Schulten, & Abrámooff 2005), filter-based (Niemeijer, Van Ginneken, Cree, Mizutani, Quellec, Sánchez, Zhang, Hornero, Lamard, Muramatsu, et al. 2010, Quellec, Lamard, Josselin, Cazuguel, Cochener, & Roux 2008) match lesions templates in subbands of wavelets transformed images. In learning-based

Cazuguel, Cochener, & Roux 2008, Zhang, Karray, Li, & Zhang 2012) and supervised learning-based (García, Sánchez, Lopez, Diez, & Hornero 2008, Niemeijer, van Ginneken, Staal, Suttorp-Schulten, & Abrámooff 2005). The morphological top hat transform is the most common morphological operation used to detect MAs. Here, a morphological opening of the green channel image is performed with linear structuring elements at different orientations, short enough to fit inside curved vessels, but, at the same time long enough to not fit inside MAs. Larger objects like blood vessels are consequently detected and removed, leaving only MA candidates. This approach has been modified and used by other authors (Walter, Massin, Erginay, Ordonez, Jeulin, & Klein 2007, Niemeijer, van Ginneken, Staal, Suttorp-Schulten, & Abrámooff 2005). Filter-based algorithms aim to produce an appropriate kernel to match MAs. Wavelet and Gaussian kernels are the most common. Quellec et al. (Quellec, Lamard, Josselin, Cazuguel, Cochener, & Roux 2008) match lesions templates in subbands of wavelets transformed images. In learning-based

methods, classifiers like artificial neural networks, support vector machines, k-NN and pixel classification have been reported. Usually these algorithms start with a preprocessing step, candidate extraction and feature selection for classification. An ensemble-based system to detect MAs has recently been proposed which consists in the selection of the optimal combination of preprocessing and candidate extractor methods (Antal & Hajdu 2012). This method achieved the best results on ROC dataset. Despite the undoubtedly evolution of the reported approaches, there is still the opportunity to contribute with more advanced and effective methods.

In this paper, Laws texture-energy measurements for MAs detection are studied. Texture consists of spatially extended patterns, of more or less accurate repetitions of some basic texture element, called textels. These are continuous set of pixels with some tonal and/or regional property. To be able to find changes in the texture of an image, a simple strategy is to perform texture measurements in a sliding window. Then, window size and texture features need to be selected. To detect MAs we decided to use a higher order statistics method such as Laws texture masks. First order statistics from intensity histogram like mean and variance, and second order statistics like gray-level co-occurrence matrices are popular texture description methods due to their simplicity. However, since the MA detection problem presents various challenges, a higher order statistics method might be a good strategy for this purpose. Its advantages over first and second order statistics are that they can describe geometry and context. These features have been used in other biomedical images studies such as (Elnemr 2013) and (Dheeba & Selvi 2011). Five onedimensional kernels were chosen as in (Dheeba & Selvi 2011) to create the two-dimensional masks. Following, a SVM classifier performs pixel-based classification. The results of several kernel's size are analysed in 36 images from LaTIM database and performance is evaluated by FROC curves for each kernel size. New insights are given related with detection of MAs in color fundus images.

This paper is organized as follows. Section 2 presents the database utilized and details of the proposed approach. Section 3 discusses the obtained results on LaTIM database. Finally, major conclusions are presented in Section 4.

## 2 MATERIALS AND METHODS

### 2.1 Database

In this study the LaTIM (Laboratoire de Traitement de l'Information Médicale) database was used (QuelQuellec, Lamard, Josselin, Cazuguel,

Cochener, & Roux 2008). This database contains 36 manually annotated eye fundus color photographs from Brest University Hospital. Images have  $2240 \times 1488$  pixels and were stored in tiff file format. Images were cropped in order to remove most of the black background, characteristic of retinal images.

The performance of the proposed method is evaluated in terms of the FROC curve, that plots the sensitivity against the average number of False Positives (FP) per image.

The proposed method has a training phase. A total of 93 window samples were manually selected from 10 normalized images of the LaTIM database, with the following distribution: 63 windows containing MAs, 22 windows containing vessels and 8 windows containing only background. From these 93 images, train was performed with 5-fold cross-validation.

Tests to the developed method were made on the remaining 26 images from LaTIM.

### 2.2 Preprocessing

Preprocessing aims to attenuate retinal images imperfections such as non-uniform illumination, poor contrast and noise as well as retina's variability from different persons.

The background intensity variation can be eliminated by the background estimation and subsequent subtraction from the original green plane image. The background image is determined by applying a median filter with a kernel size corresponding to 1/30 the image height as in (Giancardo, Meriaudeau, Karnowski, Li, Garg, Tobin, & Chaum 2012). The obtained shaded corrected image shows a characteristic gray level distribution with dark structures belonging to the histogram left tail with negative values. At this point, and since all normalized images have similar histogram characteristics, it is possible to eliminate all bright structures such as the optic disc and exudates by putting these elements equal to zero. Then, only the absolute value of all pixels belonging to the dark structures is kept.

### 2.3 Microaneurysms detection

The automatic detection of microaneurysms in digital fundus images remains a challenge in image processing. These small red dots have typically low contrast and they may be hard to distinguish from variations in the background. Texture is a key component of the human visual perception, along with color and shape. Here, we will focus on texture, and the Laws masks are proposed to calculate the description features of MA for their texture discrimination. A classification step is then

performed to find the MA candidates. A schematic representation of the proposed approach is shown in Figure 1.

Laws texture energy transform combines predetermined one-dimensional convolution kernels into various convolution masks with moving window updating. Figure 2 shows the vectors that are convolved with their own transposes to generate twenty-five  $5 \times 5$  weighted masks (Shapiro & Stockman 2001, Dheeba & Selvi 2011). The names of the vectors describe their purposes. The L5 vector gives a center-weighted local average, the E5 vector detects edges, the S5 detects spots, the R5

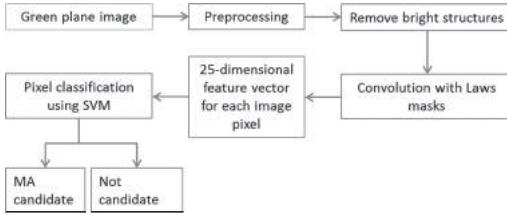


Figure 1. Proposed approach schematic.

$$L5 = [1 \ 4 \ 6 \ 4 \ 1]$$

$$E5 = [-1 \ -2 \ 0 \ 2 \ 1]$$

$$S5 = [-1 \ 0 \ 2 \ 0 \ -1]$$

$$R5 = [1 \ -4 \ 6 \ -4 \ 1]$$

$$W5 = [-1 \ 2 \ 0 \ -2 \ 1]$$

(a)

$$E5L5 = \begin{bmatrix} E(1) * L(1) & \dots & E(1) * L(5) \\ E(2) * L(1) & \dots & E(2) * L(5) \\ E(3) * L(1) & \dots & E(3) * L(5) \\ E(4) * L(1) & \dots & E(4) * L(5) \\ E(5) * L(1) & \dots & E(5) * L(5) \end{bmatrix}$$

$$E5L5 = \begin{bmatrix} -1 & -4 & -6 & -4 & -1 \\ -2 & -8 & -12 & -8 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 2 & 8 & 12 & 8 & 2 \\ 1 & 4 & 6 & 5 & 1 \end{bmatrix}$$

(b)

Figure 2. (a)  $5 \times 5$  texture vectors used to create two-dimensional texture masks. L, E, S, W and R stand for level, edge, spot, ripple and wave. (b) The convolution of E5 and L5 vectors.

detects ripples and the W5 vector detects waves. Consequently, the weighted masks will measure content related with each vector used to generate the mask. For example, E5L5 measures horizontal edge content, and L5E5 measures vertical edge content. The filters are binomially weighted masks that enhance image spots, edges, or high-frequency components and measure the spatial frequencies or texture structures (Laws 1985, Lee & Schenk 1993). A particular texture-energy measure thus depends on the spatial filters used, the window size, and the method of measuring average variation within the window.

The first step is to filter the preprocessed images with the convolution masks. Several kernel sizes were tested in order to find out which one achieved better performance in MA identification. The following sizes were selected to test:  $5 \times 5$ ,  $11 \times 11$ ,  $15 \times 15$ ,  $21 \times 21$ ,  $25 \times 25$ ,  $31 \times 31$ ,  $35 \times 35$ ,  $41 \times 41$  and  $45 \times 45$ . The comparison method was based on FROC curves. The different kernel sizes were obtained using bilinear interpolation of the  $5 \times 5$  kernels.

After convolving each preprocessed image with the 25 convolution kernels, a 25-dimensional feature vector is obtained for each pixel. Assuming that  $F_k[i, j]$  is the result of filtering with the  $k$ th mask at pixel  $[i, j]$ . Then, after the convolution process with Laws masks, the texture energy map  $E_k$  for filter  $k = L5L5$  is defined by

$$E_k[r, c] = \sum_{j=c-w}^{c+w} \sum_{i=r-w}^{r+w} |F_k[i, j]|, \quad (1)$$

with,

$$w = \frac{\text{KernelSize} \times 3}{2} - 1, \quad (2)$$

The texture energy map consists in a moving window operation, where every pixel in the image is replaced by the sum of pixels (magnitudes) in its neighbourhood. In literature (Laws 1985), the window operation is applied to all filtered images,  $F_k$ . However, this operation may smooth small lesions or even be responsible for their disappearance and thus we consider it to be more adequate for macrotexture features detection. Consequently, and since the main purpose is to discriminate MAs from noise and background, we decided to apply the window operation only in the  $F_{L5L5}$  which is then used in the contrast normalization step. The unique texture energy map is obtained by filtering the preprocessed image with Level kernel ( $L5L5$ ),  $F_{L5L5}$ , followed by the window operation. Windows with three times the kernel size were used. The texture energy map has the same size of the original image.

Filtered images from convolution with Laws masks are then normalized for contrast by dividing each resultant filtered image with the  $E_{LSL5}$  map.

MA segmentation is then performed by pixel-classification using SVM. Each pixel from the final texture energy transform image consists in a 25-dimensional vector and corresponds to the input of the classifier which then determines if that pixel is a candidate to MA or not. The OpenCV function *train\_auto* was used to train the SVM model automatically by choosing the optimal parameters C, gamma, p, nu, coef0 and degree. Parameters were selected using a grid search and a cross-validation with 5 folds. Parameters with the best results were saved and used in the final tests.

### 3 RESULTS AND DISCUSSION

The main objective of this paper was to detect microaneurysms in color fundus images with texture analysis. For this purpose, Laws texture energy values were computed in 36 retinal images. Nine kernel sizes were tested and processed images classified with a SVM classifier. The SVM was trained using manually selected regions from images. Using the entire image for training would be time consuming since the number of MAs is very small compared to the non-MAs regions. Due to this fact, and to reduce class imbalance, 93 small representative regions of 10 images were selected with the different regions present in retina (blood vessels and background). In addition, this image selection for train aims to reduce redundant information that could be introduced if the entire image was used. The training algorithm used applied 5-fold cross-validation and data was split equally. By minimizing the samples used, overtraining is avoided. Nine trainings were computed, one for each tested size. Figure 3 shows some examples of training images. For better visualization, images are contrast enhanced.

Performance was evaluated in terms of the FROC curve. This seemed to be the best choice to quantitatively evaluate the MAs location identification. In the FROC plot, sensitivity values are plotted against the average number of false positives per image. Sensitivity refers to MAs correctly

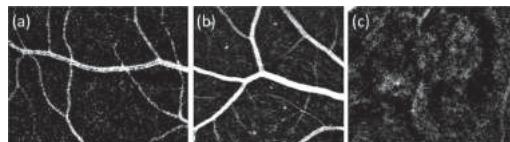


Figure 3. Example of manually selected images to train SVM: (a) image containing vessels; (b) image containing MAs and vessels; (c) image containing only background.

detected by the algorithm, while false positives are the number of non-MAs identified as MAs. To construct the FROC curve, the real SVM output values were kept and then, a threshold with values between [-1 1] was applied to that gray level image. Figure 4 displays the binary output of the SVM pixel classification for six kernel sizes. Kernels larger than  $31 \times 31$  are not shown since they are not much different from  $31 \times 31$  kernel. It is possible to observe that MAs are detected, as well as many other objects. As expected, the kernel dimension interferes in the identification of MAs and, with larger kernels the number of

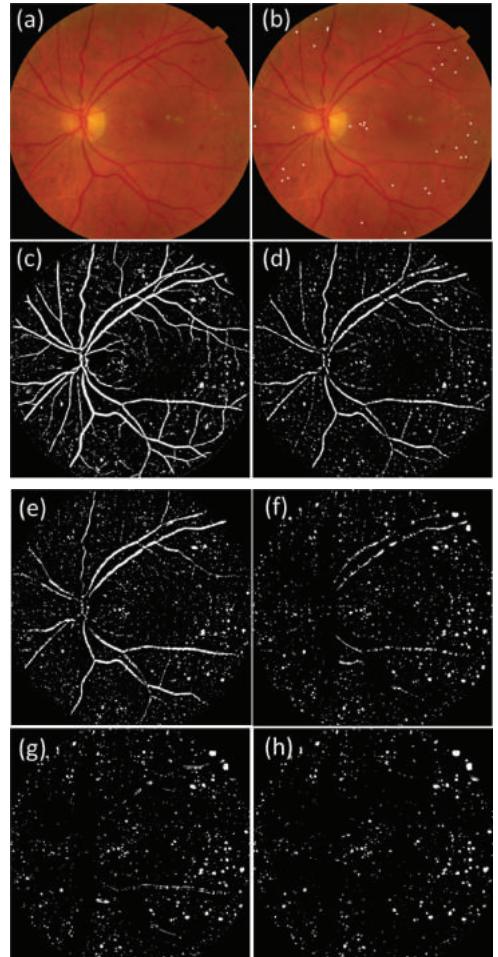


Figure 4. Images resulting from the proposed method: (a) Original color fundus image; (b) ground truth; (c) MA candidates with kernel  $5 \times 5$ ; (d) MA candidates with kernel  $11 \times 11$ ; (e) MA candidates with kernel  $15 \times 15$ ; (f) MA candidates with kernel  $21 \times 21$ ; (g) MA candidates with kernel  $25 \times 25$ ; (h) MA candidates with kernel  $31 \times 31$ .

outliers diminishes. Moreover, these texture features proved to be very good in vessel segmentation with small kernels. In Figure 4c, vessels are clearly visible with  $5 \times 5$  kernel. This result could be compared with other vessel segmentation approaches in the future. By looking at images in Figure 4 is possible to conclude that microtexture filters detects patterns of texture energy features, and with different kernels sizes different retinal structures are discriminated. An adequate size of filter must be selected.

Figure 5 and 6 shows the FROC curves for each kernel dimension. For less than 30 false positives all kernels are below 60% sensitivity. Above that number of FP, sensitivity values start to increase and for 50 False Positives (FP), approximately 70% sensitivity is reached with kernels  $31 \times 31$ ,  $35 \times 35$  and  $41 \times 41$ . In addition, sensitivities of 80% are achieved for 100 FP, and 90% for 200 FP, with kernels  $31 \times 31$  and  $35 \times 35$ . These two kernels show very close performances but for more than 100 FP kernel size  $31 \times 31$  appears to be above kernel  $35 \times 35$ . The performance of the classifier does not improve for kernels  $41 \times 41$  and  $45 \times 45$ . The worst performance was achieved with kernel  $15 \times 15$ . We presume this size is probably very close to the MAs dimension, and consequently not sufficiently high

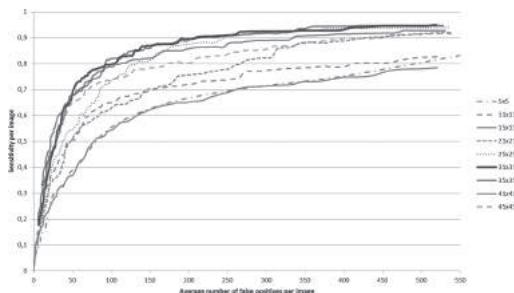


Figure 5. FROC curves of the proposed approach in LaTIM images.

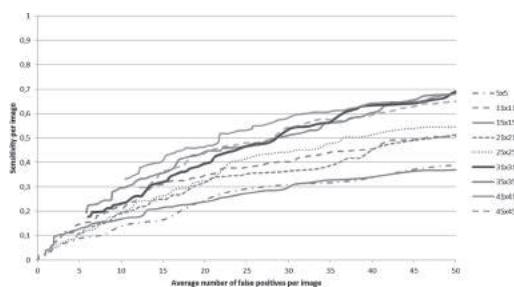


Figure 6. FROC curves in more detail between 0 and 50 false positives.

to include them, but at the same time, not enough small only to embrace MA pixels. Since the number of false positives per image is still very high for a real application, we concluded that further processing of each MA candidate is needed.

## 4 CONCLUSION

In this paper we present a MA detection method based on Laws texture-energy measurements. As far as we know these features were never reported to MAs detection in fundus images. MAs are extracted using only these features in an SVM. Nine kernels sizes were tested to analyse the impact on MAs identification. Results showed that Laws texture features can be used to find MA candidates, and  $31 \times 31$  kernel size achieved best results. However, the number of false positives is still an issue. Consequently, this method may be considered to find MA candidates but further processing should be performed. Furthermore, it was concluded that small kernels like  $5 \times 5$  are very good in retinal vessels segmentation. Finally, an advantage of this approach is its simplicity, since no vessel or optic nerve segmentation was needed. As future work results using Laws texture features will be compared with some other form of texture quantification method, such as Haralick features.

## ACKNOWLEDGMENTS

Work supported by FEDER funds through the Programa Operacional Factores de Competitividade COMPETE and by national funds by FCT Fundação para a Ciência e a Tecnologia. D. Veiga thanks the FCT for the SFRH/BDE/51824/2012 and ENERMETER.

## REFERENCES

- Antal, B. & A. Hajdu (2012, June). An ensemble-based system for microaneurysm detection and diabetic retinopathy grading. *Biomedical Engineering, IEEE Transactions on* 59(6), 1720–1726.
- Cree, M.J. (2009). Automated microaneurysm detection for screening. *Automated Image Detection of Retinal Pathology*, 155–184.
- Dheeba, J. & S. Selvi (2011). Classification of malignant and benign microcalcification using svm classifier. In *Emerging Trends in Electrical and Computer Technology (ICETECT), 2011 International Conference on*, pp. 686–690. IEEE.
- Elnemr, H.A. (2013). Statistical analysis of laws mask texture features for cancer and water lung detection. *International Journal of Computer Science Issues (IJCSI)* 10(6).

- García, M., C.I. Sánchez, M.I. Lopez, A. Diez, & R. Hornero (2008). Automatic detection of red lesions in retinal images using a multilayer perceptron neural network. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pp. 5425–5428. IEEE.
- Giancardo, L., F. Meriaudeau, T.P. Karnowski, Y. Li, S. Garg, K.W. Tobin, & E. Chaum (2012). Exudate-based diabetic macular edema detection in fundus images using publicly available datasets. *Medical Image Analysis* 16(1), 216–226.
- Laws, K.I. (1985). Goal-directed textured-image segmentation. In *1985 Technical Symposium East*, pp. 19–26. International Society for Optics and Photonics.
- Lee, D.-C. & T. Schenck (1993). Image segmentation from texture measurement. *International Archives of Photogrammetry and Remote Sensing* 29, 195–195.
- Neubauer, A.S. & M.W. Ulbig (2007). Laser treatment in diabetic retinopathy. *Ophthalmologica* 221(2), 95–102.
- Niemeijer, M., B. Van Ginneken, M.J. Cree, A. Mizutani, G. Quellec, C.I. Sánchez, B. Zhang, R. Hornero, M. Lamard, C. Muramatsu, et al. (2010). Retinopathy online challenge: automatic detection of microaneurysms in digital color fundus photographs. *Medical Imaging, IEEE Transactions on* 29(1), 185–195.
- Niemeijer, M., B. van Ginneken, J. Staal, M.S. Suttorp-Schulten, & M.D. Abrámoff (2005). Automatic detection of red lesions in digital color fundus photographs. *Medical Imaging, IEEE Transactions on* 24(5), 584–592.
- Quellec, G., M. Lamard, P.M. Josselin, G. Cazuguel, B. Cochener, & C. Roux (2008). Optimal wavelet transform for the detection of microaneurysms in retina photographs. *Medical Imaging, IEEE Transactions on* 27(9), 1230–1241.
- Shapiro, L. & G. Stockman (2001). *Computer Vision*. Paperback.
- Walter, T., P. Massin, A. Erginay, R. Ordóñez, C. Jeulin, & J.-C. Klein (2007). Automatic detection of microaneurysms in color fundus images. *Medical image analysis* 11(6), 555–566.
- Zhang, B., F. Karray, Q. Li, & L. Zhang (2012). Sparse representation classifier for microaneurysm detection and retinal blood vessel extraction. *Information Sciences* 200, 78–90.

# An iterative algorithm for Total Variation minimization in DBT imaging

A.M. Mota, N. Matela, N. Oliveira & P. Almeida

*Instituto de Biofísica e Engenharia Biomédica, Faculdade de Ciências, Universidade de Lisboa, Lisboa, Portugal*

**ABSTRACT:** Digital Breast Tomosynthesis (DBT) is a new technique that takes several projection images with different angles and combines them to reconstruct 3D image using approximately the same dose of two mammograms. Currently, handling noise is a major problem in DBT. Total Variation (TV) based minimization algorithms have achieved great success due to its virtue of preserving edges while reducing image noise. Because TV denoising depends on noise model, we tested an algorithm for Poisson noise in real DBT images.

For unfiltered images we have a Contrast to Noise Ratio (CNR), a spatial resolution and TV of 1.110, 0.483 mm and 2.319E+08, respectively. After filter application, the CNR improves up to 5.3%, spatial resolution value doesn't suffer degradation (variations up to 0.4%) and TV decrease up to 40.4%.

We conclude that application of TV denoising to DBT images is a promising technique and allows a wide field of research in the future.

## 1 AIM

Digital Breast Tomosynthesis (DBT) is a new x-ray imaging technique for breast cancer detection (Jähne 2005, U.S. Food and Drug Administration (FDA) 2011, Sechopoulos 2013) that has the ability to overcome Digital Mammography (DM) limitations on tissue superposition (Sechopoulos, 2013). It acquires a set of 2D projections from different source positions, within a limited angular range, and combines them to reconstruct 3D image using approximately the same dose of two mammograms (Timberg et al. 2008).

Currently, handling noise level is a major problem in DBT images mainly due to low dose projections and incomplete sampling of the projection frequency space due to the limited angular range used.

Image denoising is a challenging issue in image processing field and Total Variation (TV) based minimization algorithms have achieved great success among several imaging techniques due to its virtue of preserving edges while reducing image noise (Ertas et al. 2014, Feng et al. 2009, Jonsson et al. 1998, Joshi et al. 2009, Mota et al. 2015, Nett et al. 2008, Sidky et al. 2011, Zhiqiang et al. 2013). TV is a quantity that characterizes how smoothly the intensity of an image is changing and it increases significantly in the presence of noise. TV minimization algorithm was originally developed by *Rudin, Osher and Fatemi* (ROF model) (Rudin et al. 1992), where the TV of an image is minimized subject to constraints involving Gaussian noise statistics.

In the present work our aim was to implement a method to reduce DBT images noise. For this we

tested an iterative algorithm based on ROF model but for Poisson noise (Jonsson et al. 1998, Le et al. 2007, Sawatzky et al. 2009). We applied it to DBT images and carried out visual and quantitative analysis of our method.

## 2 MATERIALS AND METHODS

### 2.1 Acquisition and reconstruction

A Mammographic Accreditation Phantom Model 156 by Gammex was acquired to perform image quality assessment. It was used a Siemens Mammat Inspiration that makes 25 low dose acquisition within a limited angular range from approximately  $-25^\circ$  and  $+25^\circ$ . The equipment that performed the acquisition was installed in a clinical facility that makes DBT examinations in a daily basis.

The acquired data was reconstructed with the image reconstruction algorithm Algebraic Reconstruction Technique (ART), with a constant relaxation coefficient of 0.20 (this choice was made based on previous studies—scientific paper under approval in the international Journal of Digital Imaging) and finalized after convergence iteration (7th), using IDL from *Exelis* with a pixel size of 0.34 mm and a slice thickness of 1.0 mm.

### 2.2 Image analysis

For image analysis, quantitative and qualitative comparisons were performed between unfiltered and filtered reconstructed images. Two Figures of Merit were used for quantitative analysis of image

quality: Contrast to Noise Ratio (CNR) and Full Width at Half Maximum (FWHM) of one microcalcification. For CNR, a Region of Interest (ROI) over the lesion (mass) and another over the background were drawn. This latter was obtained as a square centered in the lesion, excluding the lesion. CNR was calculated with Equation 1, where  $\mu_L$  and  $\mu_B$  stand for mean pixel values in ROI over the lesion and background, respectively; and  $\sigma_B$  stands for standard deviation in background ROI. Signal Difference was calculated as the difference between mean pixel values in the lesion and in the background.

$$CNR = \frac{\mu_L - \mu_B}{\sigma_B} \quad (1)$$

FWHM of a Gaussian curve fitted to the Point Spread Function of one microcalcification was considered as an indicator of spatial resolution. CNR and FWHM values were calculated before and after filter application for the two scenarios.

### 2.3 Formulation of TV minimization algorithm

With this TV minimization algorithm, starting from a constrained minimization problem, it is obtained an unconstrained problem that is composed by a fidelity term corresponding to the assumed noise model and a regularization term that is the TV function.

Our unconstrained TV minimization problem, based on ROF model but for Poisson noise, can be formulated as Equation 2 (Le et al. 2007, Sawatzky et al. 2009):

$$\min_u \{TV(u) + \lambda(u - f \ln u)\} \quad (2)$$

where  $f$  and  $u$  are the original and denoised image, respectively, and  $TV(u)$  is the TV of denoised image. TV values were determined according to Equation 3, considering the selected 2D slice.

$$TV(u) = \sum_{i=1}^m \sum_{j=1}^n \sqrt{(\Delta_x u_{i,j})^2 + (\Delta_y u_{i,j})^2} \quad (3)$$

In Equation 3,  $u_{i,j}$  is the intensity value of pixel  $(i, j)$ , with  $i = \{1, \dots, m\}$ ,  $j = \{1, \dots, n\}$  and  $m \times n$  the image dimensions.  $\Delta_x$  and  $\Delta_y$  is the discretization of the horizontal and vertical derivatives, respectively.

To solve the problem of Equation 2, for Gaussian noise, ROF proposed the gradient descent method (Rudin et al. 1992), a procedure based on a parabolic equation with an artificial time parameter. Since the problem is convex, the steady

state solution of gradient descent is the minimizer of the problem. Considering Poisson noise, this means that we solve:

$$\begin{aligned} & \frac{u_{i,j}^{n+1} - u_{i,j}^n}{\Delta t} \\ &= \frac{f_{i,j}}{1 - \frac{1}{\lambda} \left[ \frac{1}{h^2} \left( \Delta_x^- \frac{\Delta_x^+ u_{i,j}}{\sqrt{\varepsilon^2 + (\Delta_x^+ u_{i,j})^2 + (\Delta_x^0 u_{i,j})^2}} \right. \right.} \\ & \quad \left. \left. + \Delta_y^- \frac{\Delta_y^+ u_{i,j}}{\sqrt{\varepsilon^2 + (\Delta_y^0 u_{i,j})^2 + (\Delta_y^+ u_{i,j})^2}} \right) \right]} - u_{i,j}^{n+1} \end{aligned} \quad (4)$$

Here  $\Delta_x^+$ ,  $\Delta_x^-$ ,  $\Delta_y^+$  and  $\Delta_y^-$  denote forward and backward one-sided differences in  $x$  and  $y$  directions, respectively; and  $\Delta_x^0$  and  $\Delta_y^0$  indicate central difference in  $x$  and  $y$  directions, respectively. This formulation is valid only for interior points, and Boundary conditions were defined as in Ref Rudin et al. 1992. To solve the problem with these boundary conditions for  $u$ , a routine was implemented and integrated in the image reconstruction software implemented in IDL.

$\Delta t$  should be a small number that forces the algorithm to make small jumps. Its optimal value depends on the function. If it is too large, the algorithm will diverge and if it is too small, it will take a long time to converge.

$\lambda$  is the Lagrange multiplier, also called as regularization or fitting parameter since it controls image regularization, between removing noise and preserving edges.

In a first phase, results for  $\Delta t$  ranging between 0.05 and 0.25 segs, with a fixed  $\lambda = 60$  were analyzed (from previous studies we know that this is an acceptable value for this level of noise. See (Mota et al., 2015)). It is considered a good  $\Delta t$  the one that is effective in minimizing the TV and that can achieve convergence with a reasonable number of iterations. In the next stage, taking into account  $\Delta t$  values chosen, different  $\lambda$  values were compared. Finally, with the most appropriate  $\Delta t$  and  $\lambda$ , filtered images were quantitatively and qualitatively analyzed.

## 3 RESULTS

In Figure 1 we can observe TV values for  $\Delta t$  ranging between 0.05 and 0.50 segs, with a fixed  $\lambda = 60$ .

As  $\Delta t = 0.05$  segs is too small to converge in 20 iterations and values from 0.30 to 0.50 diverge, we choose to continue the study with  $\Delta t = 0.25$  segs. This value is the one which reaches convergence faster (iteration number 10), also allowing a high

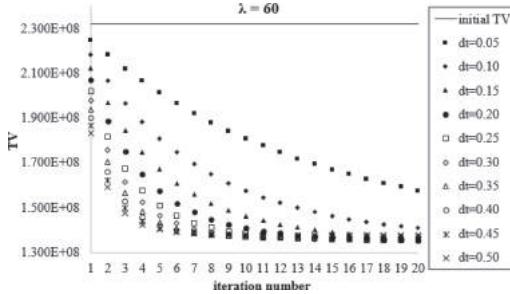


Figure 1. TV values for phantom unfiltered and filtered reconstructed images plotted as a function of iteration number and  $\Delta t$  values for  $\lambda = 60$ .

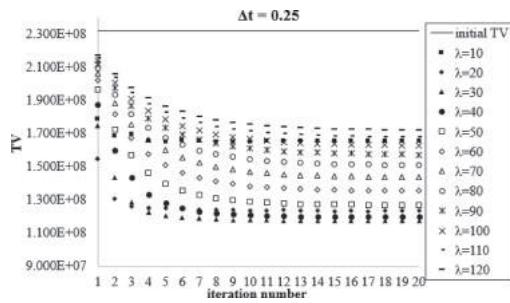


Figure 2. TV values for phantom unfiltered and filtered reconstructed images plotted as a function of iteration number and  $\lambda$  values for  $\Delta t = 0.25$  segs.

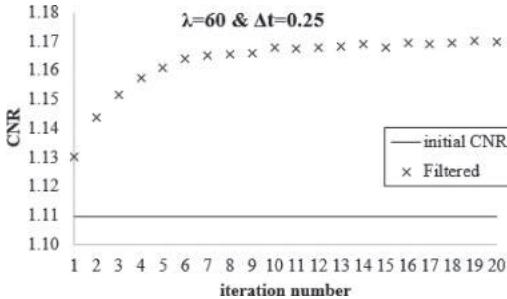


Figure 3. CNR values for phantom unfiltered and filtered reconstructed images plotted as a function of iteration number, for  $\lambda = 60$  and  $\Delta t = 0.25$  segs.

TV minimization. The results obtained for different  $\lambda$  values, taking into account  $\Delta t = 0.25$  are presented in Figure 2.

For the image quantitative and qualitative analysis,  $\lambda = 60$  was the value chosen and the results obtained for CNR and FWHM are shown in Figure 3 and Figure 4. For a visual inspection images of a mass and a cluster of microcalcifications of the phantom are presented in Figure 5 and Figure 6.

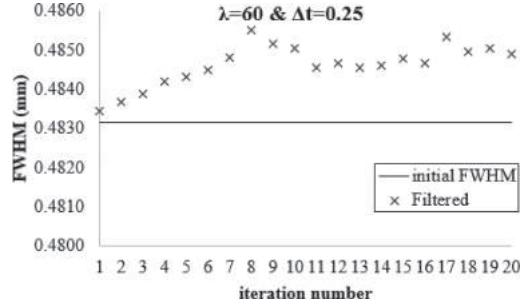


Figure 4. FWHM values for phantom unfiltered and filtered reconstructed images plotted as a function of iteration number, for  $\lambda = 60$  and  $\Delta t = 0.25$  segs.

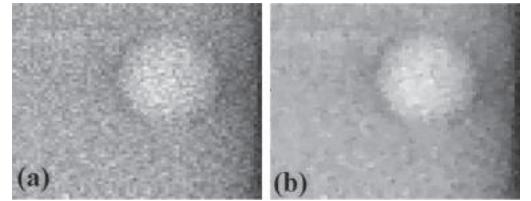


Figure 5. Images of a mass obtained for the unfiltered (a) and filtered (b) images with  $\Delta = 0.25$  segs and  $\lambda = 60$  (iteration 10).

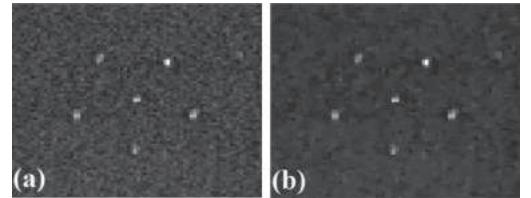


Figure 6. Images of a cluster of microcalcifications obtained for the unfiltered (a) and filtered (b) images with  $\Delta = 0.25$  segs and  $\lambda = 60$  (iteration 10).

#### 4 DISCUSSION

We presented here improved results that can be optimized in the sense of achieve an objective correlation between the parameters and image noise level. The combination  $\Delta = 0.25$  segs and  $\lambda = 60$  works well and can achieve convergence in iteration number 10.  $\lambda = 10$  doesn't ensure the fidelity term has enough weight to keep the fundamental information of the initial image and  $\lambda$  between 30 and 50 diverge. In this way those values were automatically excluded. From  $\lambda = 60$  to 120 we choose  $\lambda = 60$  because it allows the better CNR in 10 iterations.

We can see the a decrease in image TV of 40.4% as well as the fact that this method improves CNR

(up to 5.3%) without losing spatial resolution (variation up to 0.40%) which means an important progress in handling noise in these images. Further optimizations (about  $\Delta t$ ,  $\lambda$  and number of iterations needed to convergence) can greatly reduce the computation time which, for this particular case and for 10 iterations, was approximately 7 minutes. Also, as we know,  $\lambda$  depends on the image noise. As the image noise is different at the end of each iterations, we think that a  $\lambda$  variable over iterations may produce even better results.

## 5 CONCLUSIONS

Taking into account the results presented, we conclude that application of TV denoising to DBT images is a very promising technique and allows a wide field of research in the future.

## ACKNOWLEDGEMENTS

This work was supported in part by Fundação para a Ciência e a Tecnologia—Portugal (grant number SFRH/BD/43376/2008 and projects Pest-OE/SAU/IU0645/2013, Pest-OE/SAU/IU0645/2014 and PTDC/BBB-IMG/3310/2012).

## REFERENCES

- Ertas, M., Yildirim, I., Kamasak, M. & Akan, A. 2014. An iterative tomosynthesis reconstruction using total variation combined with non-local means filtering. *BioMedical Engineering OnLine*, 13, 65.
- Feng, Y., Liu, P., Li, B., Yu, L. & Chen, W. 2009. Improved reconstruction of non-Cartesian magnetic resonance imaging data through total variation minimization and POCS optimization. *Conf Proc IEEE Eng Med Biol Soc*, 2009, 2676–9.
- Food And Drug Administration. Available on: <http://www.fda.gov/Radiation-EmittingProducts/MammographyQualityStandardsActandProgram/FacilityCertificationandInspection/ucm114148.htm>. Last accessed July 24, 2014.
- Jonsson, E., Huang, S.C. & Chan, T. 1998. *Total variation regularization in positron emission tomography*.
- Joshi, S.H., Marquina, A., Osher, S.J., Dinov, I., Van Horn, J.D. & Toga, A.W. 2009. MRI Resolution Enhancement Using Total Variation Regularization. *Proc IEEE Int Symp Biomed Imaging*, 2009, 161–164.
- Le, T., Chartrand, R. & Asaki, T. 2007. A Variational Approach to Reconstructing Images Corrupted by Poisson Noise. *Journal of Mathematical Imaging and Vision*, 27, 257–263.
- Mota, A.M., Matela, N., Oliveira, N. & Almeida, P. 2015. Total variation minimization filter for DBT imaging. *Medical Physics*, 42 (6), June.
- Nett, B., Tang, J., Leng, S. & Chen, G.H. 2008. Tomosynthesis via Total Variation Minimization Reconstruction and Prior Image Constrained Compressed Sensing (PICCS) on a C-arm System. *Proc Soc Photo Opt Instrum Eng*, 6913, nihpa92672.
- Niklason, L.T., Christian, B.T., Niklason, L.E., Kopans, D.B., Castleberry, D.E., Opsahl-Ong, B.H., Landberg, C.E., Slanetz, P.J., Giardino, A.A., Moore, R., Albagli, D., Dejule, M.C., Fitzgerald, P.F., Fobare, D.F., Giambattista, B.W., Kwasnick, R.F., Liu, J., Lubowski, S.J., Possin, G.E., Richotte, J.F., Wei, C.Y. & Wirth, R.F. 1997. Digital tomosynthesis in breast imaging. *Radiology*, 205, 399–406.
- Rudin, L.I., Osher, S. & Fatemi, E. 1992. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60, 259–268.
- Sawatzky, A., Brune, C., Müller, J. & Burger, M. 2009. Total Variation Processing of Images with Poisson Statistics. *Computer Analysis of Images and Patterns*, 5702, 533–540.
- Sechopoulos, I. 2013. A review of breast tomosynthesis. Part I. The image acquisition process. *Med Phys*, 40, 014301.
- Sidky, E.Y., Duchin, Y., Pan, X. & Ullberg, C. 2011. A constrained, total-variation minimization algorithm for low-intensity x-ray CT. *Med Phys*, 38 Suppl 1, S117.
- Timberg, P., Båth, M., Andersson, I., Svahn, T., Ruschin, M., Hemdal, B., Mattsson, S. & Tingberg, A. Impact of dose on observer performance in breast tomosynthesis using breast specimens. *Proc. SPIE 6913, Medical Imaging: Physics of Medical Imaging* (2008).
- Zhiqiang, C., Xin, J., Liang, L. & Ge, W. 2013. A limited-angle CT reconstruction method based on anisotropic TV minimization. *Physics in Medicine and Biology*, 58, 2119.

## No-reference wavelet based Retinal Image Quality Assessment

L.S. Abdel Hamid

*Electronics and Communications Department, Faculty of Engineering, Misr International University, Cairo, Egypt*

A. El-Rafei

*Engineering Physics and Mathematics Department, Faculty of Engineering, Ain Shams University, Cairo, Egypt*

S. El-Ramly

*Electronics and Communication Engineering Department, Faculty of Engineering, Ain Shams University, Cairo, Egypt*

G. Michelson

*Department of Ophthalmology, Friedrich-Alexander University, Erlangen-Nuremberg, Germany*

J. Hornegger

*Pattern Recognition Laboratory, Department of Computer Science, Friedrich-Alexander University, Erlangen-Nuremberg, Germany*

**ABSTRACT:** Retinal Image Quality Assessment (RIQA) is an essential preliminarily step in Automatic Retinal Screening Systems (ARSS) to avoid misdiagnosis of retinal disease. In this work, a no-reference wavelet based RIQA algorithm suitable for real-time ARSS is introduced. Retinal images are classified into good and bad quality images using two proposed metrics. These measures are extracted from the wavelet decomposition subbands of the image to reflect sharpness and contrast. Wavelet level and regional analyses are performed to determine the most relevant levels and retinal region to be used in the RIQA. The algorithm is tested on two publicly available datasets with different resolutions. Area under the ROC curve (AUC) of 1 and 0.985 are obtained for the low and high resolution datasets respectively with run-time ranging between 0.1 and 2 seconds depending on image resolution. Comparison with other RIQA algorithms shows superiority of the proposed algorithm in both performance and time.

**Keywords:** Retinal Image Quality Assessment; wavelet transform; real-time system

### 1 INTRODUCTION

According to the World Health Organization (WHO), there is an estimate of 285 million people in 2014 suffering from blindness or low vision worldwide due to different retinal diseases such as glaucoma and diabetic retinopathy. Most of these vision impairments can be avoided if the retinal disease is early on diagnosed and properly treated. ARSS can facilitate early retinal disease detection by being available in public places, like airports or malls, and immediately advising patients to seek further medical assistance if signs of a retinal disease are present.

Substantial research effort was put in the creation of reliable ARSS that are able to provide accurate diagnosis in real-time settings. Results were found promising, yet strongly reliant on the quality of the image being processed (Patton et al. 2006).

Several factors could affect the quality of retinal images including patient movement, loss of focus and non-dilated pupil. Studies have shown that the percentage of inadequate images for screening systems is about 10% and 20.8% for dilated and non-dilated pupils respectively (Yu et al. 2012). Accordingly, RIQA is considered an essential preliminarily step in ARSS used to ensure that the retinal image being processed, is of adequate quality to allow for correct medical diagnosis.

### 2 RELATED WORK

RIQA approaches can be broadly divided into generic and segmentation based approaches. Generic methods were initially based on the intuition that good retinal images had similar histograms. Early works compared the histogram

(Lee et al. 1999) or edge magnitude histogram (Lalonde et al. 2001) of the observed image to that of a template histogram created from a set of good images. More recent generic approaches use textural features, such as Haralick features (Davis et al. 2009) or Run-Length-Matrix (RLM) features (Fasih et al. 2014), along with histogram (statistical) features extracted from global or local regions in retinal images.

Segmentation methods are based on the intuition that good quality retinal images tend to have more separable anatomical structures, specifically the blood vessels. Segmentation methods use either vessels segmented from the entire image (Usher et al. 2003, Giancardo et al. 2008, Thomas et al. 2013) or vessels segmented from the macula region (Fleming et al. 2006, Hunter et al. 2011). The later argue that macular vessels are more diagnostically important as well as being the first to become blurred with any slight image degradation whereas thick blood vessels can still be clearly visible, even when the retinal images are significantly blurred.

Generic methods have the advantage of being simple yet they completely ignore any retinal structures while segmentation methods are slow and prone to error. In order to improve the performance, hybrid approaches were introduced that combined both generic and segmentation techniques by using features extracted from segmented blood vessels along with histogram and textural features (Paulus et al. 2010, Yu et al. 2012, Sevik et al. 2014).

Nevertheless, all previous methods dealt with images in the spatial domain. Recently, the transform domain has been gaining great popularity in image quality assessment as it is more consistent with the Human Visual System (HVS), which states that the eye decomposes images into low and high subbands and processes each differently (Gao et al. 2008).

The Wavelet Transform (WT) is a multiresolution technique having the advantage of being localized in both time and frequency. It is computed by applying low (L) and high (H) pass filtering along the rows then the columns of an image giving four subbands LL, LH, HL and HH referred to as the approximate, vertical, horizontal and diagonal subbands respectively. The horizontal, vertical and diagonal subbands together are known as the detail subbands of the image holding its high frequency (edge) information as opposed to the approximate subband being a smoothed version of the original image. In order to capture finer image details, further decomposition of the approximate subbands can be performed.

Several WT based general image quality assessment algorithms have been introduced

in literature (Ferzli et al. 2005, Vu et al. 2012, Zhao et al. 2013), yet WT based algorithms have scarcely been adopted for RIQA. In context of full-reference RIQA, a wavelet based full reference distortion measure, has been introduced to quantify diagnostic information loss due to image compression (Nirmala et al. 2010). In the measure introduced by Nirmala et al., different weights were given to different wavelet levels based on an innovative analysis performed on the relation between retinal blood vessels and the different wavelet levels.

In the context of no-reference RIQA, a wavelet based focus measure defined as the mean value of the sum of detail coefficients in Level (L1) was used along with moment and statistical measures (Veiga et al. 2014). Veiga's algorithm achieved high classification results, however the system was rather slow taking approximately 33 seconds to process a new image making it unsuitable for real-time applications. Furthermore, the wavelet measure used by Veiga only used L1 subbands thus not fully exploiting the WT multiresolution potential.

A wavelet based RIQA algorithm that does not require the presence of a reference image is introduced in this paper. It is designed to distinguish between good and bad quality retinal images and to be suitable for real-time screening applications. A couple of metrics computed from wavelet features are proposed based on the intuition that good quality retinal images have visibly clear sharp blood vessels and adequate contrast. Two types of analyses are conducted using the proposed metrics. First, level analysis is performed in order to determine the most relevant wavelet level for feature extraction and the effect of image resolution on the level selection. Then, the significance of using global features, extracted from the entire image, and local features, extracted from the diagnostically significant center region containing the macular vessels, is analyzed. Results from these two analyses are concatenated to identify the best set of features. Finally, results are compared to some of the existing RIQA algorithms from both performance and run-time perspectives.

### 3 MATERIALS AND METHODS

Several retinal images datasets exist in literature, however only few have quality information. In this work, two publicly available macula centered datasets having different resolutions are used to test the proposed algorithm:

- a. Diabetic Retinopathy Image Database (DRIMDB) (Sevik et al. 2014): which includes 69 bad and 125 good quality retinal images. All images were obtained with a Canon CF-60UVi

- Fundus Camera using 60° FOV and stored in JPEG files at  $570 \times 760$  pixels. Most bad images in DRIMDB suffer from severe blur where thin vessels are invisible and thick blood vessels are either severely smeared or completely invisible.
- b. High Resolution Fundus (HRF) Image Database (Thomas et al. 2013): which includes 18 bad and 18 good quality retinal images taken using a Canon CR-1 fundus camera with a 45° FOV having resolutions varying from  $2592 \times 3888$  to  $3456 \times 5184$  pixels. The bad images in HRF have varying degree of blur. Some images are only slightly distorted having blurred thin vessels yet clear thick vessels, while other images suffering more severe blur have smeared thick vessels.

### 3.1 Preprocessing

Two preprocessing steps are performed prior to feature extraction from the retinal images. Initially, the dark regions around the retina are cropped to remove useless information and reduce processing time. Then, a  $3 \times 3$  median filter is used to reduce noise from the cropped images while preserving edges.

The green and red channels of the preprocessed retinal images are used for feature extraction. However, the blue channel is excluded from the analysis since it is under-illuminated and scarcely contributing any relevant information.

### 3.2 WT features computation

Good retinal images are characterized by having high sharpness and contrast making retinal structures and possible disease lesions clearly visible and easily separable. In order to measure retinal image sharpness and contrast, two wavelet based metrics are introduced. The green and red channels of the image are decomposed into five levels using daubechies4 ‘db4’ wavelet. Then, the measures are calculated from the coefficients of the horizontal and vertical detail subbands. These subbands are found to carry more significant information related to the retinal vessels than the diagonal subband.

#### 3.2.1 Sharpness measure

Sharp blood vessels in good quality images correspond to more information present in the detail subbands compared to bad quality images. An information theoretic metric that reflects the information content in the detail subbands is proposed to be used with retinal images. This metric is the wavelet Shannon entropy which relates to the retinal image sharpness. The entropy of each level is calculated as the average of the horizontal

entropy ( $Entr_{Horiz}$ ) and vertical entropy ( $Entr_{Vert}$ ) as given by the following equations:

$$Entr_{level} = \frac{1}{2N}(Entr_{Horiz} + Entr_{Vert}) \quad (1)$$

$$Entr_{subband} = -\sum_{i=1}^N \|C_i\|^2 \log \|C_i\|^2 \quad (2)$$

where  $N$  is the number of coefficients in the wavelet subband,  $C_i$  is the wavelet coefficient having index  $i$  and  $subband$  is the horizontal or vertical subband.

#### 3.2.2 Contrast measure

Contrast of an image can be defined as the ratio of high to low frequencies (Pu 2000). In this work, contrast is calculated for a specific level as the ratio between the average of the summation of the horizontal ( $S_{Horiz}$ ) and vertical ( $S_{Vert}$ ) subband coefficients to the summation of the approximate ( $S_{Approx}$ ) subband coefficients as given by the following equations:

$$Contrast_{level} = \frac{(S_{Horiz} + S_{Vert})/2}{S_{Approx}} \quad (3)$$

$$S_{subband} = \frac{\sum_{i=1}^N |C_i|}{N} \quad (4)$$

where  $N$  is the number of coefficients in the wavelet subband,  $C_i$  is the wavelet coefficient having index  $i$  and  $subband$  is the horizontal or vertical subband.

### 3.3 Wavelet level and region analyses

In their work, Nirmala et al. have shown a relation between different wavelet levels and retinal blood vessels. In order to find the most significant wavelet level to be used for feature extraction the aforementioned analysis is re-implemented where small sections from different retinal images, having either thick or thin vessels, are extracted and coefficients of subsequent detail subbands are zeroed. Then, the Mean Structural Similarity Index Measure (Wang et al. 2004) is calculated to measure which level have more effect on the retinal vessels. Another experiment is performed to validate the results from the previous analysis in which the presented sharpness and contrast features from each wavelet level are used for classification and results from classification using each of the different levels are compared.

Moreover, global features are calculated from the entire image while local features are computed from the center region. The center region is chosen to include the macular vessels which are diagnostically important as well as being the first to be affected by any slight blur (Hunter et al. 2011).

This center region is defined by dividing the image horizontally into four equal sections then regarding the two center sections together as the center region. The choice of a relatively wide center region allows taking into account any slight vertical shift in the macula centered retinal image. Results from using global and local features in classification are compared.

For all analyses in this subsection, the K-Nearest Neighbor (KNN) classifier is utilized as it can be easily tuned and found to provide consistent results. Conclusions from the feature analyses are used to determine the most relevant wavelet level and type (global or local) of the features. Finally, the classification using the selected set of features is fine tuned by the SVM classifier to improve the results.

## 4 RESULTS

In this section, the results from features analyses for both DRIMDB and HRF datasets are demonstrated. Performance and run-time of the proposed algorithm are evaluated and compared to other RIQA techniques in literature.

Feature extraction computations were performed under the MATLAB(Mathworks, Inc., Natick, MA, USA) development environment. Classification results were obtained using 10-fold cross validation in the Weka platform (Hall et al. 2004).

### 4.1 DRIMDB classification results

Performing analysis similar to that in (Nirmala et al. 2010) on several DRIMDB images showed that the levels most relevant to thin and thick vessels were L2 and L3 respectively. Classification using proposed sharpness and contrast features for separate levels gave best results for L3 (AUC = 0.998) followed closely by L2 (AUC = 0.987) and L4 (AUC = 0.981). Results for classification using global against local features were very close for all levels with a maximum AUC difference of 0.02. Results from the DRIMDB dataset are summarized in Table 1.

Feature analysis results using the DRIMDB dataset, showed that high classification results were obtained regardless of the selected level and type of features (global or local). Final results from

Table 1. Area under the receiver operating characteristic curve (AUC) for the DRIMDB database.

	L1	L2	L3	L4	L5
Global	0.973	0.987	0.998	0.981	0.876
Local	0.963	0.995	0.998	0.998	0.895

the SVM classifier gave an AUC larger than 0.999 for the individual levels L2, L3, or L4 and for both global and local features. A perfect differentiation between good and bad images (AUC = 1) was achieved when using L3 (for both global and local features). In (Sevik et al. 2014), the authors used 144 hybrid features for RIQA of the DRIMDB images resulting in F1-score of 0.996, 0.965 and 0.842 for classifying the images into good, bad and outliers (images having no retinal information) respectively.

### 4.2 HRF classification results

The most significant levels to thin and thick blood vessels were found to be L3 and L4 respectively by applying the method in (Nirmala et al. 2010) on several HRF images. Classification using the sharpness and contrast features from different levels led to best results for L3 (AUC = 0.917) followed closely by L4 (AUC = 0.907). Unlike the results from the DRIMDB images, classification using global and local features demonstrated better accuracy for using local features, for example L3 local features gave an AUC of 0.944 while global features resulted in AUC of 0.917. Results from the HRF dataset are summarized in Table 2.

As evident from the feature analysis results, the feature choice was more critical for the HRF dataset than the DRIMDB dataset. For final classification, wavelet level L3 was selected as it had the largest correlation to the thin retinal vessels. In addition, features from local region were utilized to prevent the relatively clear thick vessels from outweighing the noticeably blurred thin vessels in the slightly blurred images. The SVM classifier with local features from L3 resulted in an AUC of 0.985.

Results from the proposed algorithm on the HRF dataset were compared to other RIQA techniques from literature and results are shown in Table 3.

Table 2. Area under the receiver operating characteristic curve (AUC) for the HRF database.

	L1	L2	L3	L4	L5
Global	0.799	0.778	0.917	0.907	0.62
Local	0.938	0.87	0.944	0.904	0.799

Table 3. Comparison of the proposed algorithm with existing RIQA techniques.

	Proposed Algorithm	Fasih 2014	Thomas 2013	Davis 2009
AUC	0.985	0.796	0.883	0.852

Two generic RIQA algorithms relying on local features were re-implemented by the authors for the sake of comparison, namely those proposed in (Davis et al. 2009) and (Fasih et al. 2014). Davis et al. calculated statistical, textural and sharpness features from seven local regions covering the entire retinal image. Fasih et al. used textural and sharpness features extracted only from the optic disc and macula regions. The results reported from the segmentation based approach in (Thomas et al. 2013) applied to the HRF dataset were also included in the comparison. Thomas et al. computed a global quality score based on the segmented blood vessel tree.

#### 4.3 Run-time results

The proposed method is intended for real-time ARSS, therefore immediate results reporting is essential. The time required to process a new image using the proposed algorithm was recorded and found to be less than 0.1 second for small resolution images (DRIMDB of  $570 \times 760$  pixels), reaching a maximum of ~2 seconds for high resolution HRF images ( $3456 \times 5184$  pixels).

The run-time of the re-implementation of the algorithms of Davis et al. and Fasih et al. was found to be 7–10 seconds and 9–20 seconds per image respectively for the HRF dataset depending on the image size. All the previous time analysis was computed on the same PC using MATLAB.

## 5 DISCUSSION

A no-reference RIQA algorithm using wavelet based features is proposed. Sharpness and contrast measures extracted from the detail wavelet subbands were used for classification of retinal images into good and bad quality images. Evaluation of the introduced method showed high efficiency in determining the quality of the images regardless of their resolution and degree of blur.

In order to find the most relevant features to be used in classification, wavelet level and feature type (global or local) analyses were performed. Level analysis showed that there is a correlation between image resolution and level choice. Nirmala et al. have shown that for  $512 \times 512$  pixel retinal images, L2 and L3 detail subband coefficients had more influence on thin and thick blood vessel respectively. Re-implementation of Nirmala et al.'s procedure for the DRIMDB dataset ( $570 \times 760$  pixels) gave similar results. On the other hand, for the high resolution HRF dataset L3 and L4 detail subband coefficients were found to have more influence on thin and thick blood vessel respectively. Accordingly, it can be concluded from the level analysis results that for higher resolution images,

higher order wavelet decomposition levels become more significant comparison to lower resolution images.

In order to further explore this conclusion, the HRF dataset was resized to  $1024 \times 1024$  and  $512 \times 512$  pixels and level analysis was repeated for both cases. Classification results, summarized in Table 3 and Table 4, confirmed the previous conclusion. For example, when using local features from L3, the AUC was 0.944. However, the AUC decreased to 0.923 and further to 0.764 for the resized images at  $1024 \times 1024$  and  $512 \times 512$  pixels respectively.

Feature analysis also indicated that the relevance of wavelet levels and type of feature (global or local) was dependent on the type of blur in the bad quality retinal images. For images with severe blur, as in the DRIMDB dataset, high classification accuracy was achieved regardless of the wavelet level and type of feature used. An AUC larger than 0.96 was obtained from levels L1 to L4. The difference in AUC between the usage of global and local features was less than 0.02 for all mentioned levels. These values could be explained by the severe blur in this dataset. Both the thick and thin vessels are affected by the high degree of blur in these images. This in turn is reflected on the different wavelet levels carrying information related to vessels of varying thickness. Moreover, since both thin and thick vessels are significantly affected by the severe blur, both global and local features were equally significant.

On the other hand, for slightly blurred images in the HRF dataset, classification results were found to be dependent on the wavelet level and feature type (global or local). Best results were achieved when using L3 local features (AUC = 0.985). The slight blur in the bad images mainly affected the thin vessels present in the local macular region. Wavelet level L3 was found to have the highest sensitivity to

Table 4. AUC for the HRF dataset resized to  $1024 \times 1024$  pixels.

	L1	L2	L3	L4	L5
Global	0.781	0.892	0.944	0.92	0.787
Local	0.96	0.954	0.923	0.79	0.778

Table 5. AUC for the HRF dataset resized to  $512 \times 512$  pixels.

	L1	L2	L3	L4	L5
Global	0.907	0.886	0.836	0.611	0.423
Local	0.96	0.926	0.765	0.472	0.481

slight blur as the analysis showed that it is the most related level to thin vessels. Additionally, the small percentage of the thin vessels to the total vessel network makes it easily outweighed by the visible thick vessels. Thus, global features could lead to wrong classification in lightly blurred images. This explains the higher accuracy associated with the usage of local features for the high resolution dataset.

RIQA of the HRF dataset is considered specifically challenging due two reasons: 1) The slight blur affecting only the fine details (thin macular vessels) of the retinal images are hard to capture using traditional RIQA algorithms. 2) The high resolution of the HRF dataset images with some images having a size of  $3456 \times 5184$  pixels (which is almost 6 times the size of images from DRIMDB) make feature calculations computationally expensive. Nevertheless, the proposed sharpness and contrast wavelet based measures extracted from L3 of the local central region were found to efficiently capture the slight blur affecting the thin macular vessels in the bad images. Thus, achieving results that surpassed both generic and segmentation based techniques implemented for the HRF dataset. The accuracy of the wavelet based algorithm exceeded the other RIQA methods by at least 10%.

The reduced number of features and the simplicity of their calculation from the wavelet subbands had a significant effect on the algorithm's speed. The introduced RIQA algorithm was found to be 5 times and 10 times faster than the methods by Davis et al. and Fasih et al. respectively for higher resolution images. This speed ratio largely increased in case of lower resolution images. This can be attributed to the computationally extensive textural feature used in these methods.

## 6 CONCLUSIONS

Automatic screening systems for retina diseases should provide the users with an immediate accurate report. RIQA associated with ARSS is crucial and should operate in real-time and perform efficiently while taking into account the lack of a reference image. In this paper, a wavelet based algorithm for retinal image quality grading is designed to satisfy the aforementioned requirements imposed by the screening application. The algorithm exploited the relation between the clarity of the retinal blood vessels and the information content within the image wavelet levels. Metrics are proposed to address the sharpness and contrast of the images. These measures are calculated directly from the multi-resolution decomposition and used as features for classifications. Therefore, no reference is required.

Three comprehensive analyses were performed. In the first analysis, it was shown that the most

relevant wavelet level for classification was dependent on the image resolution. The resizing of high resolution images changed the order of the significant wavelet levels as found from the second analysis. Then, it was indicated in the third analysis that local features extracted from regions with thin vessels performed better than global features for slightly blurred images.

The conclusions from the previous analyses were incorporated in the algorithm to decide on the final set of features. The proposed algorithm was tested against other RIQA methods showing superior classification performance. Also, as a result of the very limited number of computationally inexpensive features that were carefully selected, the proposed algorithm was distinguishably faster than the other RIQA algorithms in comparison for both low and high resolution images.

Wavelet domain's consistency with the HVS method of image analysis makes it a valuable tool for image quality evaluation. Specifically, RIQA wavelet based algorithms have the potential to overcome the limitations of both generic and segmentation based algorithms while achieving superior results.

## REFERENCES

- Davis, H., S. Russell, et al. (2009). Vision-based, real-time retinal image quality assessment. *Proc. of IEEE International Symposium on Computer-Based Medical Systems (CMBS)*, pp. 1–6.
- Fasih, M., J.M. Langlois, et al. (2014). Retinal image quality assessment using generic features. *Proc. of SPIE 9035, Medical Imaging, International Society for Optics and Photonics*, pp. 90352Z–90352Z–7.
- Fleming, A. D., S. Philip, et al. (2006). Automated assessment of diabetic retinal image quality based on clarity and field definition. *Investigative Ophthalmology & Visual Science*, 47(3): 1120–1125.
- Ferzli, R., L.J. Karam, et al. (2005). A robust image sharpness metric based on kurtosis measurement of wavelet coefficients. *Proc. of Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics*.
- Gao, Z. & Y.F. Zheng. (2008). DWT-Based Image Quality Metric. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(7): 910–922.
- Giancardo, L., E. Chaum, et al. (2008). Elliptical Local Vessel Density: A Fast and Robust Quality Metric for Fundus Images. *Proc. of International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pp. 3534–3537.
- Hall, M., E. Frank, et al. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1): 10–18.
- Hunter, A., J.A. Lowell, et al. (2011). An automated retinal image quality grading algorithm. *Proc. of International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5955–5958.

- Lalonde, M., L. Gagnon, et al. (2001). Automatic visual quality assessment in optical fundus images. *Proc. of Vision Interface*, Vol. 32, pp. 259–264.
- Lee, S.C., & Y. Wang (1999). Automatic retinal image quality assessment and enhancement, *Medical Imaging International Society for Optics & Photonics*, pp. 1581–1590.
- Nirmala, S.R., S. Dandapat, et al. (2010). Wavelet weighted blood vessel distortion measure for retinal images. *Biomedical Signal Processing and Control*, 5(4): 282–291.
- Patton, N., T.M. Aslam, et al. (2006). Retinal image analysis: concepts, applications and potential. *Progress in Retinal and Eye Research*, 25(1): 99–127.
- Paulus, J., J. Meier, et al. (2010). Automated quality assessment of retinal fundus photos. *International journal of computer assisted radiology and surgery*, 5(6): 557–564.
- Pu, T., & G. Ni (2000). Contrast-based image fusion using the discrete wavelet transform. *Optical Engineering*, 39(8): 2075–2082.
- Şevik, U., C. Köse, et al. (2014). Identification of suitable fundus images using automated quality assessment methods. *Journal of Biomedical Optics*, 19(4): 046006–046006.
- Thomas, K., A. Budai, et al. (2013). Automatic No-Reference Quality Assessment for Retinal Fundus Images Using Vessel Segmentation. *Proc. of IEEE International Symposium on Computer-Based Medical Systems (CMBS)*, pp. 95–100.
- Usher D.B., M. Himaga, et al., (2003). Automated assessment of digital fundus image quality using detected vessel area. *Proc. of Medical Image Understanding & Analysis*: British Machine Vision Association (BMVA), pp. 81–84.
- Veiga, D., C. Pereira., et al. (2014). Quality evaluation of digital fundus images through combined measures. *Journal of Medical Imaging*, 1(1):014001–014001–9.
- Vu, P.V. and D.M. Chandler (2012). A Fast Wavelet-Based Algorithm for Global and Local Image Sharpness Estimation. *IEEE Signal Processing Letters*, 19(7): 423–426.
- Wang, Z., A.C. Bovik, et al. (2004). Image Quality Assessment : From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4): 600–612.
- Yu, H., C. Agurto, et al. (2012). Automated image quality evaluation of retinal fundus photographs in diabetic retinopathy screening. *Proc. of IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 125–128.
- Zhao, H., B. Fang., et al. (2013). A no-reference image sharpness estimation based on expectation of wavelet transform coefficients. *Proc. of IEEE International Conference Image Processing (ICIP)*, pp. 374–378.

This page intentionally left blank

# Vessel segmentation of retinal images with fuzzy morphology

P. Bibiloni, M. González-Hidalgo & S. Massanet

Department of Mathematics and Computer Science, University of the Balearic Islands, Palma de Mallorca, Spain

**ABSTRACT:** Distinguishing vessels in a retinal image is a solid first step towards a computer-aided diagnosis of several medical conditions. Mathematical morphology provides simple yet powerful transformations to process images. Extended with fuzzy techniques, it achieves the ability to handle the uncertainty present in medical imaging, both in the images and in the expert diagnosis. Within this paradigm, we propose a method that leverages the geometric and photometric features that characterize vessels as such. It is tested with two databases, DRIVE and STARE. The first promising results outperform some of the state of the art methods with which it was compared, positioning it among the most competitive techniques.

## 1 INTRODUCTION

Vessel segmentation in retinal images consists on specifying which pixels of an image would be considered by an expert as part of a vessel. These images include part of the retina (light-sensitive tissue), the optic disc (termination of the optic nerve), the macula (area of the retina with high-density of cone cells) and the blood vessels in this region, besides a number of manifestations that may indicate a medical condition.

Given their nature, these images may present a variety of particular artifacts and noise. Although an expert can easily distinguish most artifacts from actual manifestations, it is a computationally complex task. The image as a whole should be processed to convey as much information as possible. For instance, the position of the vessel tree may help to flag a dark region as the macula, which is located somewhat in the middle of that structure, as shown in Figure 1.

Some diseases may be detected by means of examining an eye-fundus retinal image. Given the

segmentation of the vessel tree, some features of the veins and arteries can be estimated (e.g. their width and tortuosity) and other manifestations can be segmented out more easily. The automatic segmentation of blood vessels is, therefore, the first step to infer more information towards a computer-aided diagnosis of a condition with noticeable effects on the retina. A more accurate and systematic diagnosis of diabetic retinopathy, glaucoma and other conditions may be possible due to image processing of eye-fundus images.

Vessel segmentation in eye-fundus images is a problem that researchers have faced with a wide variety of approaches [5]. Thus, a number of techniques to segment vessels have been already published. In 1999, Martínez-Pérez *et al.* designed a method that extracted features from the second-derivative of the pixel's intensity, obtained as the convolution of the image with derivatives of Gaussian functions. They merged these features obtained at several scales to then apply a local region-growing procedure. [10]. Later in 2000, Hoover *et al.* published a vessel segmentation method using the database they created, STARE, which is publicly available [8]. They used a matched filter (correlation of the image with a template to detect its presence) and a local thresholding process. Chanwimaluang and Guoliang presented a method that also uses matched filtering, to later apply a threshold based on local entropy [3]. In 2004, the also publicly-available database DRIVE was presented by Niemeijer *et al.*, including a method in which pixel-wise features were obtained as the output of linear filters (Gaussian and its derivatives), and further processed with a kNN classifier [11]. Extending this work, Staal *et al.* presented another method in which additional features are considered after detecting the main ridges of the image [13].

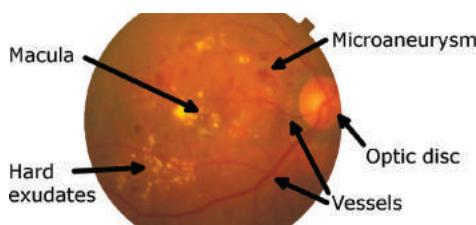


Figure 1. Image number 14 of the STARE database, with annotations. The macula is hardly visible, whereas the illumination is noticeably non-uniform.

Using fuzzy techniques, Ayala *et al.* published a method that uses averages of fuzzy sets to project grayscale images into binary estimations [1]. Likewise, Sun *et al.* presented a method based on multi-scale morphology, fuzzy filtering and the watershed transform [14].

Although some methods use fuzzy techniques and others are based on morphological operators, there are no studies involving *eye-fundus vessel segmentation with fuzzy mathematical morphology*, as far as we are aware. Such a paradigm perfectly suits the need of handling uncertainty in medical images, present in both the pixels' value and the expert diagnosis annotations.

## 2 VESSEL SEGMENTATION METHOD

In this section, the method is presented following the description of the vessels' features. Firstly, we give an overview of fuzzy mathematical morphology, which constitutes the core of our algorithm.

### 2.1 Fuzzy mathematical morphology

Mathematical morphology is a paradigm to process images by means of basic yet powerful transformations. It is based on simple and local operations that, when properly combined, transform images to extract borders, skeletons or more complex features.

The mathematical morphology paradigm that we use is the one introduced by De Baets [4], which is based on fuzzy operators. A *conjunction* is defined as a function  $C : [0,1] \times [0,1] \rightarrow [0,1]$  increasing in both variables such that  $C(0,0) = C(1,0) = C(0,1) = 0$  and  $C(1,1) = 1$ , whereas an *implicator*  $I : [0,1] \times [0,1] \rightarrow [0,1]$  is monotonically decreasing in the first variable and increasing in the second one, and fulfills  $I(0,0) = I(0,1) = I(1,1) = 1$  and  $I(1,0) = 0$  [2]. Let  $B$  be a two-dimensional structuring element (i.e. a template), and  $A$  an image. Then, the dilation  $\mathcal{D}$  and the erosion  $\mathcal{E}$  are defined as follows:

$$\begin{aligned}\mathcal{D}_C(A, B)(y) &= \sup_x C(B(x-y), A(x)), \\ \mathcal{E}_I(A, B)(y) &= \inf_x I(B(x-y), A(x)),\end{aligned}$$

where the variable  $x$  ranges over the values in which both  $A(x)$  and  $B(x-y)$  are defined.

From these two basic operations and the reflected structuring element,  $\bar{B}(x) = B(-x)$ , we can construct the opening  $\mathcal{O}$ , the closing  $\mathcal{C}$  and the top-hat transformations by opening  $\mathcal{THO}$  and by closing  $\mathcal{THC}$ :

$$\begin{aligned}\mathcal{O}_{C,I}(A, B) &= \mathcal{D}_C(\mathcal{E}_I(A, B), \bar{B}), \\ \mathcal{C}_{C,I}(A, B) &= \mathcal{E}_I(\mathcal{D}_C(A, B), \bar{B}), \\ \mathcal{THO}_{C,I}(A, B) &= A - \mathcal{O}_{C,I}(A, B), \\ \mathcal{THC}_{C,I}(A, B) &= \mathcal{C}_{C,I}(A, B) - A.\end{aligned}$$

As expected, this definition of fuzzy dilation expands the foreground object, whereas the erosion diminishes it. The opening removes sharp or thin regions of the foreground object, whereas the closing operation enlarges the foreground object by filling sharp or thin areas of the background. The Top-Hat transformations enhance, therefore, sharp or thin regions. These last operators are of interest due to their characteristic behaviour—they have already been used to process noisy biomedical images [6]. Specifically, we use a Top-Hat by Closing, to extract the dark, thin objects that represent vessels in eye-fundus images.

We will consider that  $B(0) = 1$ . Besides, our conjunctive  $C$  will be a t-norm and  $I$  its residuated implication, so the following inclusions will hold [9]:

$$\begin{aligned}\mathcal{E}_I(A, B) &\subset \mathcal{O}_{C,I}(A, B) \subset A \\ &\subset \mathcal{C}_{C,I}(A, B) \subset \mathcal{D}_C(A, B).\end{aligned}$$

Depending on the choice of the conjunctive—which would also affect its residuated implicator—different results are obtained [7]. One of the focus of our study is pinpointing which are the best settings regarding the problem of vessel segmentation.

### 2.2 Modelization of vessels

In order to construct a method that segments vessels out we should identify which *features characterize a pixel as part of a vessel*. We include both geometric and photometric features—so we can infer information regarding the shape of objects and their colour components. Likewise, it is important to *describe the noise* found in these images so it can be properly faced.

Vessels in eye-fundus images, like the one shown in Figure 1, are *thin curvilinear structures*. They are darker regions than the background surrounding them. Besides, they form a tree-like structure: a *connected region with several bifurcations*. Their width is monotonic, and tends to grow towards a central region, in which the optic disk is found. These features are considerably good indicators in order to discriminate vessels across a great number of images.

There are distinct types of elements that may interfere with the detection of the vessels. Insufficient or non-uniform illumination always causes *fluctuations in the image's bright*, and sometimes

*big rounded shapes* may also appear. Big vessels, specially arteries, present another illumination problem, *central vessel reflex*, consisting on a lighter area along the interior of the vessel. The *optic disk* is a very bright area, roughly circular. Moreover, several pathologies may also appear: *microaneurysms* and other types of *hemorrhages* (darker reddish areas), *hard exudates* (yellow points caused by lipid residues), *cotton wool spots* (yellow/gray regions produced by damage to nerve fibers) and several others.

### 2.3 High-level description

Our method leverages the features that characterize a vessel as such to create an automatic segmentation. It is composed of several stages, each one of them with a well-defined behaviour towards taking advantage of one or more of those features. The steps of our algorithm, which can be visually observed in Figure 2, work as follows:

1. **Green channel.** From the original RGB image, only the third channel is considered, since it is the one which conveys the most information about the vessels' location.
2. **Fuzzification.** Each pixel, originally within the set  $\{0, 1, \dots, 255\}$  is transformed to the interval  $[0, 1]$

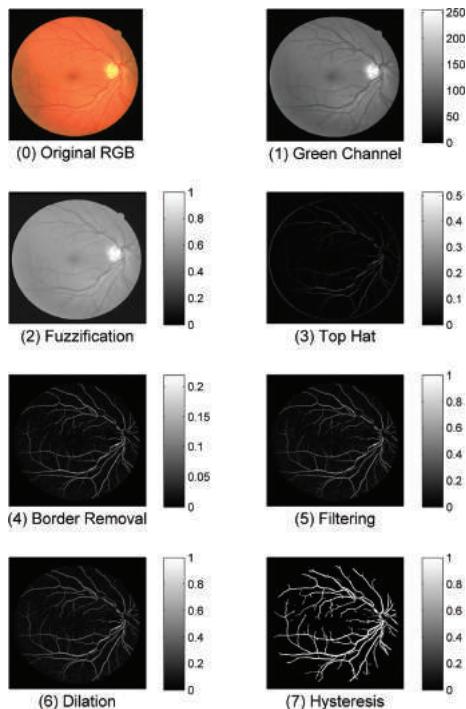


Figure 2. Step by step visualization of our method, with the second image of the DRIVE database as input.

through  $f(x) = (x/255)^k$ . The equalization factor  $k$  is chosen so that all pixels are adequately distributed in  $[0, 1]$ , providing a well contrasted image.

3. **Top hat by closing.** To take advantage of the fact that vessels are thin regions which are darker than their surroundings, a top hat by closing *THC* is the perfect choice to extract them from both the background and the big artifacts present on the image.
4. **Border removal.** The exterior of the original image can not contain any vessel, but the top hat transform may have produced some false positives there. This background area is estimated and then removed.
5. **Filtering.** Although big artifacts have been rejected, small ones may still be present. To remove *isolated points*, a filtering step is applied. Given a point  $x_0$ , its 8-neighbourhood  $x_1, \dots, x_8$  is considered. The filter then outputs:  $H(x) = \min(x_0, \max(x_1, \dots, x_8))$ . After applying this operation, the estimation is normalized to the range  $[0, 1]$ .
6. **Dilation.** The top-hat transform slightly reduces the width of the vessels. A dilation step is thus needed to cancel this effect out.
7. **Hysteresis.** To further leverage the model of the vessels, a hysteresis step is applied so that connected components are more likely detected. Two thresholds are specified,  $0 < t_{low} \leq t_{high} \leq 1$ . A pixel will be classified as “vessel” as long as it is placed in a connected region of pixels greater or equal than  $t_{low}$  such that at least one of them is also greater or equal than  $t_{high}$ .

## 3 EXPERIMENTATION RESULTS

### 3.1 Input images

The two eye-fundus databases used are DRIVE and STARE, which have *low levels of noise but a wide range of manifestations*. DRIVE [13] is a database that contains 40 colour images of mostly healthy patients, each of them with a resolution of  $565 \times 584$  pixels, along with one or two expert segmentations. The STARE database [8] contains 397 colour images of  $605 \times 700$  pixels, 20 of which have been manually segmented twice. They are very appropriate to test and optimize vessel segmentation methods: the low noise that they present reduces the difference between distinct manual segmentations, which is a recurrent problem to robustly verify the correctness of a method.

### 3.2 Implementation details

The method has been fully implemented in MATLAB® 2014a and in the C programming language.

There are several parameters that can be adjusted to obtain the optimum result—including the structuring element and the t-norm used to perform the Top Hat transform or the hysteresis thresholds  $t_{low}, t_{high}$ . They have been optimized for each of the databases independently, since each of them presents different characteristics. Besides, a batch optimization procedure has been performed to avoid overfitting.

The best *equalization factor*  $k$  varies from 0.51 to 0.61, depending on the database. The *Top-Hat transform* uses a  $7 \times 7$  flat structuring element and the Mayor-Torrens t-norm with  $\lambda = 0.7$ :

$$T_{MT}^{\lambda}(x, y) = \begin{cases} \max(x + y - \lambda, 0), & \text{if } (x, y) \in [0, \lambda]^2, \\ \max(x + y, 0), & \text{otherwise.} \end{cases}$$

The *dilation*, on the other hand, is instantiated with a  $3 \times 3$  gaussian-shaped structuring element (with  $\sigma = 3/4$ ) and the Product t-norm,  $TP(x, y) = x \cdot y$ . Lastly, the *hysteresis thresholds* also vary:  $t_{low} = 0.12$  and  $t_{high} \in [0.24, 0.26]$ .

### 3.3 Comparison and objective metrics

Our method has been tested and compared with other state-of-the-art algorithms. A visual comparison can be observed in Figure 3, in which the results of several methods are shown for comparison. The image chosen is the 255-th of the STARE database, which corresponds to a healthy patient [8] and presents uniformly distributed vessels. The good contrast that it presents highlights thin vessels, which are therefore tagged in the ground truth—in contrast with high-noise images, in which only the widest vessels can be identified.

Figure 3 remarks one of the strengths of our method: it is very sensitive to small and thin vessels, which are usually the most difficult ones to detect.

However, in order to properly compare methods, *objective indicators* of their correctness are essential. The ground truth is a classification of each pixel as either “vessel” (positive) or “non vessel” (negative). An automatic vessel segmentation aims to mimic this classification. We will consider two indicators based on these two pixel-wise masks: the  $F_1$  *Score* and the *Accuracy*.

The  $F_1$  *Score* is an appropriate metric due to its good behaviour against different deviations from the ground truth. It is the harmonic mean of precision (rate of selected pixels that correspond to a vessel) and recall (rate of vessel pixels that are actually selected). It is computed as:

$$F_1(x) = \frac{2 \cdot TP(x)}{2 \cdot TP(x) + FP(x) + FN(x)}.$$

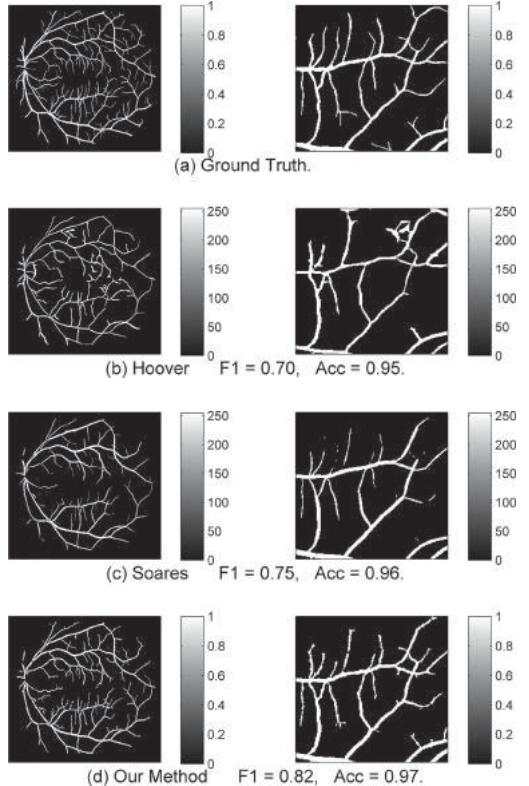


Figure 3. Comparison of different methods with the 255-th sample of the STARE database. The detail corresponds to the positions  $(x, y) \in [300, 500]^2$ .

where  $TP(x)$  are the true positives of our estimation  $x$ ,  $FP(x)$  the false positives and  $FN(x)$  the false negatives. It is sensitive to both types of errors, unlike other error metrics such as precision—which is not affected by false negatives—or recall—which is independent of the number of false positives.

In Figure 4, it is depicted a comparison of the distributions of the  $F_1$  *Score* by different methods throughout all the test images of the DRIVE database.

The second metric that we use is *Accuracy* due to its extensive use in the related literature, which helps comparison between different methods, and due to its simplicity, which provides an intuitive insight regarding the behaviour of different methods. It is defined as the rate of pixels correctly classified:

$$Acc(x) = \frac{TP(x) + TN(x)}{\#\{x\}}.$$

where  $\#\{x\}$  is the total number of pixels.

**Boxplots F1-score on DRIVE**

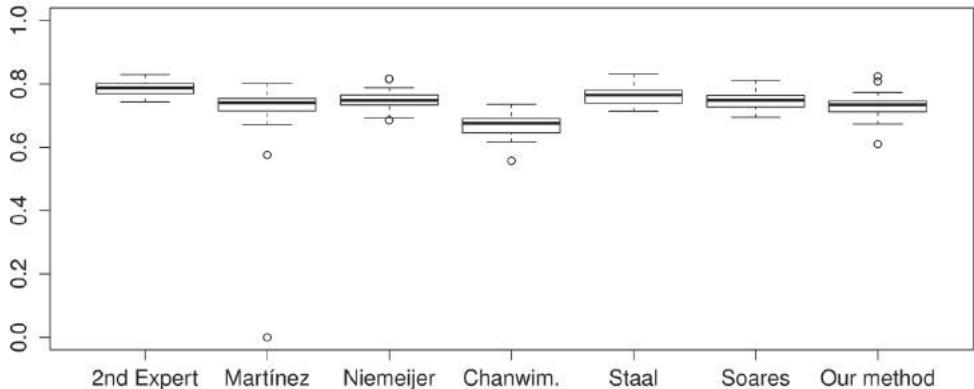


Figure 4. Boxplot showing the distributions of different methods with the 20 images of test of the DRIVE dataset.

We remark that only around 10%–15% are “vessel” pixels, so the trivial estimation that considers each pixel as “non-vessel” would obtain an *Accuracy* of 0.85–0.90.

A statistical test is needed to compare two methods with each other. Given  $n$  images, each method  $M_i$  provides a series of results  $r_i(r_i^{(1)}, \dots, r_i^{(n)})$ , which can be interpreted as samples of a probability distribution. We use the *Wilcoxon signed-rank test* to contrast whether the mean of two paired distributions differ or not. It is robust against non normal data, which is our case. The significance level is set to  $\alpha = 0.05$ .

For each database, the method has been executed and tested against the ground truth provided. In the DRIVE database, the 20 test images have been considered for experimenting purposes, using the other 20 images as training samples when needed. The 20 manually segmented images of the STARE database are considered, using the strategy leave-one-out<sup>1</sup> when training is required. The mean of the metrics of all samples is showed in Table 1.

Along the 20 test images of the DRIVE database, the method provides a mean  $F_1$  *Score* of 0.733 and an *Accuracy* of 0.958. The 20 STARE images yield an average  $F_1$  *Score* of 0.689 and an *Accuracy* of 0.960.

As we can see in Table 1, the metrics of our method are in all cases similar to the best ones. With respect to the DRIVE database, our method is statistically better than the one presented by Chanwimaluang, although the method presented by Staal *et al.* has significative better results than those

Table 1. Comparison of the average of the  $F_1$  *Score* and *Accuracy* of all possible images in both databases, where both the results and the ground truth were available.

	DRIVE		STARE	
	$F_1$	Acc	$F_1$	Acc
2nd expert	0.788	0.964	0.740	0.952
Martínez [10]	0.697	0.950	—	—
Niemeijer [11]	0.750	0.960	—	—
Hoover [8]	—	—	0.651	0.946
Chanwim. [3]	0.670	0.942	0.638	0.935
Staal [13]	<b>0.764</b>	<b>0.962</b>	—	—
Soares [12]	0.748	0.960	<b>0.698</b>	0.957
<b>Our method</b>	0.733	0.958	0.689	<b>0.960</b>

of our method. However, with the STARE database, which presents images with more noise and lower contrast, our method obtains extremely good segmentations, with a  $F_1$  *Score* only improved—without statistical significance—by Soares *et al.* It also yields the best *Accuracy* result, which is significantly better than all the other methods except for the results by Soares *et al.* Surprisingly, the *Accuracy* result is even better than that of the second expert segmentation—being the only method providing a better result than the second expert. This is due to the fact that our method and the second ground truth differ from the first manual segmentation in different ways: the  $F_1$  *Score* of the latter is clearly above the one of the former.

The methods by Chanwimaluang and Guoliang [3] and Soares [12] were executed using the source code published by their authors. Hoover *et al.* [8] published the segmentations of their method with the STARE database, and Staal *et al.* did so

<sup>1</sup>Leave-one-out consists on computing each estimation by training a new classifier with the remaining samples.

with the rest of the methods compared, including theirs.

In a computer with an Intel®i5 at 3.10 Ghz with 16 GB of RAM, our method lasts an average of 0.48 seconds after having been optimized. In contrast, the method by Chanwimaluang needs 1.76 seconds in average, whereas the one by Soares uses 22.59 seconds per sample after a learning stage of 391.12 seconds. The rest of segmentations were not computed on the same machine, so their time measurements are not comparable. Just for the sake of completeness, and remarking that this measure is not comparable to the others, the method by Staal *et al.* lasts around 15 minutes on an Intel® Pentium III PC at 3.10 Ghz with 1 GB of RAM [13].

## 4 CONCLUSIONS AND FUTURE WORK

Emerging paradigms and approaches can be applied to design finer-grained methods, so their performance can potentially outstand current state-of-the-art solutions. Although the performance of our method is promising, some more work can be done to take full advantage of the features of the vessel structure. A *line-enhancement* procedure could be implemented, and *noise filters* may be used so that the method is even more sound against highly-interfered images.

Taking advantage of technological improvements, like better-quality eye-fundus cameras or faster processors, is of paramount importance. Optimizing our method to process high-resolution images is the next step to verify that our methodology can face a variety of situations.

## ACKNOWLEDGMENTS

This work was partially supported by the project TIN 2013-42795-P. P. Bibiloni also benefited from the fellowship FPI/1645/2014 of the *Conselleria d'Educació, Cultura i Universitats of the Govern de les Illes Balears* under an operational program co-financed by the European Social Fund.

## REFERENCES

- [1] G. Ayala, T. León, and V. Zapater. Different averages of a fuzzy set with an application to vessel segmentation. *IEEE Transactions on Fuzzy Systems*, 13(3):384–393, 2005.
- [2] G. Beliakov, A. Pradera, and T. Calvo. *Aggregation functions: A guide for practitioners*, volume 361. Springer, 2007.
- [3] T. Chanwimaluang and G. Fan. An efficient blood vessel detection algorithm for retinal images using local entropy thresholding. In *Proceedings of the 2003 International Symposium on Circuits and Systems, ISCAS'03*, volume 5, pages V–21. IEEE, 2003.
- [4] B. De Baets. A fuzzy morphology: a logical approach. In *Uncertainty Analysis in Engineering and Sciences: Fuzzy Logic, Statistics, and Neural Network Approach*, pages 53–67. Springer, 1998.
- [5] M.M. Fraz, P. Remagnino, A. Hoppe, B. Uyyanonvara, A.R. Rudnicka, C.G. Owen, and S.A. Barman. Blood vessel segmentation methodologies in retinal images—a survey. *Computer methods and programs in biomedicine*, 108(1):407–433, 2012.
- [6] M. González-Hidalgo and S. Massanet. A fuzzy mathematical morphology based on discrete tnorms: fundamentals and applications to image processing. *Soft Computing*, 18(11):2297–2311, 2014.
- [7] M. González-Hidalgo, S. Massanet, A. Mir, and D. Ruiz-Aguilera. On the choice of the pair conjunction-implication into the fuzzy morphological edge detector. *IEEE Transactions on Fuzzy Systems*, 2013.
- [8] A. Hoover, V. Kouznetsova, and M. Goldbaum. Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transactions on Medical Imaging*, 19(3):203–210, 2000.
- [9] E.E. Kerre and M. Nachtegael. *Fuzzy techniques in image processing*, volume 52. Springer Science & Business Media, 2000.
- [10] M.E. Martínez-Pérez, A.D. Hughes, A.V. Stanton, S.A. Thom, A.A. Bharath, and K.H. Parker. Retinal blood vessel segmentation by means of scale-space analysis and region growing. In *Medical Image Computing and Computer-Assisted Intervention-MICCAI'99*, pages 90–97. Springer, 1999.
- [11] M. Niemeijer, J. Staal, B. van Ginneken, M. Loog, and M.D. Abramoff. Comparative study of retinal vessel segmentation methods on a new publicly available database. In *Medical Imaging 2004*, pages 648–656. International Society for Optics and Photonics, 2004.
- [12] J.V. Soares, J.J. Leandro, R.M. Cesar Jr, H.F. Jelinek, and M.J. Cree. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *Medical Imaging, IEEE Transactions on*, 25(9):1214–1222, 2006.
- [13] J. Staal, M.D. Abrámoff, M. Niemeijer, M.A. Viergever, and B. van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4):501–509, 2004.
- [14] K. Sun, Z. Chen, S. Jiang, and Y. Wang. Morphological multiscale enhancement, fuzzy filter and watershed for vascular tree extraction in angiogram. *Journal of medical systems*, 35(5):811–824, 2011.

# Identification of subendocardial infarction—a feasibility study using synthetic ultrasonic image data of a left ventricular model

J. Żmigrodzki, S. Cygan, B. Leśniak-Plewińska & K. Kałużyński

*Institute of Metrology and Biomedical Engineering, Faculty of Mechatronics, Warsaw University of Technology, Warszawa, Poland*

**ABSTRACT:** The distinction between a transmural and subendocardial infarction is important for both therapeutic and prognostic reasons. To investigate the possibility to differentiate these two cases, a numerical model of a LV phantom with uniform wall and wall with inclusion simulating subendocardial infarction was created. Using meshes of these phantoms synthetic ultrasonic data was generated using the Field II software. Block matching based estimation of displacements was carried out using the 2-dimensional cross correlation coefficient. Eulerian strain tensor and principal strain components were obtained using displacement gradient tensors. The components of the principal strain, particularly the first component, show strong differences in the affected wall as compared with the homogenous wall. The methodology has the potential for identification of non-transmural changes in the left ventricle wall.

## 1 INTRODUCTION

Contrary to the transmural infarction, in the case of the subendocardial infarction the viability of the cardiac muscle is preserved. The distinction between these two types of infarction is important for both therapeutic and prognostic reasons. One can expect, that after revascularization the cardiac contractility will improve. A method allowing for quantitative evaluation of the transmural distribution of cardiac deformation may help in estimating the disease extent and is currently sought by the cardiologists. Studies have been conducted to measure the transmural distribution of myocardial deformation in animal models (Hashimoto et al., 2003; Ishizu et al., 2010; Waldman et al., 1985). Most of them used invasive techniques, and their conclusions may be biased because of potential alteration of the cardiac function, resulting from opening of the pericardium (Belenkie et al., 2004) or the implantation of microdevices. Noninvasive studies in humans indicated the possibility to distinguish transmural and non-transmural infarction (Chan et al., 2006; Zhang et al., 2005).

The attention of cardiologists is concentrating on noninvasive imaging of 2D strain on the basis of speckle tracking, which appears to be highly reproducible and demonstrated good correlation with results of tissue strain estimation using sonomicrometry or tagged MRI (Cho et al., 2006; Heyde et al., 2012). The speckle tracking determines the local displacement of speckle patterns between sequential images, usually using a block-matching algorithm applied to the RF (Radio Frequency)

signals or baseband signals. Numerous similarity measures can be used, eg. cross-correlation function, cross-correlation coefficient (Kaluzynski et al., 2001; de Korte et al., 1997; Langeland et al., 2003; Lee & Konofagou, 2008; Luo & Konofagou, 2010), zero crossing of the phase of the cross-correlation function (Pesavento & Ermert, 1998), sum of absolute differences and sum of squared differences (Langeland et al., 2003). Speckle tracking has also been shown to enable selective assessment of epicardial, midwall, and endocardial function in both animal (Sakurai et al., 2014; Tee et al., 2015) and humans (Abate et al., 2014; Adamu et al., 2009; Becker et al., 2009; Leitman et al., 2010). This layer-specific analysis of myocardial deformation parameters has not been validated in an experimental model.

This study investigates the possibility of distinguishing a wall with nontransmural inclusion, simulating subendocardial infarction, and a uniform wall in Left Ventricular (LV) model, using synthetic ultrasonic data. The strain distribution within the model wall is used to assess the effect of the inclusion. This study is the part of an ongoing project which will comprise also experiments in Polivinyl Alcohol (PVA) cryogel ventricular phantoms. It is considered as both source of reference data and introductory phase to these experiments.

## 2 MATERIALS AND METHODS

The numerical model of the LV phantom was designed as a half of an ellipsoid with uniform

15 mm wall thickness (Fig. 1). Proposed geometry resulted in 52 ml end-systolic volume. This geometry is based on previously studied ventricular models (Cygan et al., 2014; Heyde et al., 2012; Lesniak-Plewinska et al., 2010). A tubular collar was added for mounting physical phantoms in the laboratory setup. A homogenous phantom and a phantom with an inclusion approximately 2.5 times stiffer than the rest of the phantom material, with a diameter of 20 mm and 10 mm thickness were considered. The inclusion was created as an intersection of a cylinder with the thick-walled ellipsoid of the phantom. The main axis of inclusion is normal to the surface of the phantom. This axis intersects the wall 40 mm from the base (Fig. 1).

Numerical model of the phantom geometry, created using the Autodesk Inventor 2012 software (Autodesk Inc., USA) was exported in CAD format to the Finite Element Method (FEM) software—Abaqus 6.13-3 (Simulia, USA) (Cygan et al., 2015). The phantom was meshed into linear hexahedral elements (type C3D8H) with 1.5 mm seed. In the simulations both the internal and external surfaces of the mounting collar of the phantom were immobilized. Deformation was forced by applying pressure load to the inner surface of the elliptical part of the phantom. Shape of the

pressure amplitude curve corresponded to the physiological LV volume change (Fukuta and Little, 2008) with magnitude reaching 36 kPa. Sets of deformed meshes of the numerical models obtained from FEM have been created with the time-step corresponding to the desired frame rate of the generation of synthetic ultrasonic data and were used to define the shape of the phantom during this generation.

Because the results obtained using the FEM model will serve as a reference to studies of physical phantoms, the FEM model should take into account the properties of the material. PVA gel is often assumed to be linearly elastic (Jia et al., 2006), but it has been shown to be hyperelastic or even visco-hyperelastic (Karimi et al., 2014). If the behavior of the physical phantom is to be adequately modeled, the properties of the material model should come from physical measurements. The PVA cryogel was modeled as a hyperelastic material using strain-stress relation obtained from measurements of the material samples. The hyperelastic material was defined as isotropic, with 3-rd order Yeoh strain energy potential definition and Poisson's ratio of 0.45 (Cygan et al., 2015).

## 2.1 Synthetic data generation

The FIELD II package was used in the present study (Jensen, 1996; Jensen & Svendsen, 1992). It enables incorporating realistic transducer features into the simulation (Touil et al., 2010).

The concentration of the scattering elements within the resolution cell volume was set to 3, resulting in a speckle amplitude distribution close to the Rayleigh distribution. Increasing this number did not reduce significantly the discrepancy between these distributions and incurred a heavy computational cost. The synthetic ultrasonic data were generated using parameters corresponding to the Ultrasonix Sonix Touch RP scanner and phased array transducer SA4-2/24, featuring central frequency of 2.8 MHz and 65% bandwidth. A Long Axis View (LAX) of the single wall of the phantom was simulated (Fig. 1). Scan parameters were as follows: excitation frequency 2.3 MHz, sector angle 35°, 126 lines, frame rate 100, transmit focus at 8 cm, dynamic focusing on receive. Resolution cell dimensions ( $-6$  dB FWHM) were  $x=3.12$  mm,  $y=4.24$  mm,  $z=0.39$  mm, resolution cell volume— $5.24$  mm $^3$ , where 0 xz is the imaging plane and the z axis is perpendicular to the probe section (Fig. 1).

The RF lines were generated at a sampling frequency of 80 MHz. Each RF line was then filtered with bandpass filter. Lowpass section of the filter has a cutoff frequency of 9.69 MHz and an attenuation slope equal to  $-24$  dB/octave. Highpass section of the filter has a cutoff frequency of 2.05 MHz

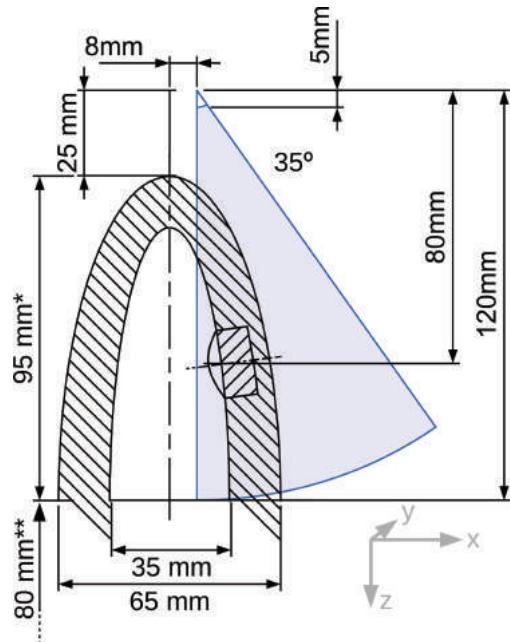


Figure 1. Section of the phantom. \* marks the LV and \*\* the fixing collar. LAX view sector shadowed, coordinate system in gray.

and attenuation slope equal to  $-22$  dB/octave. This mimics the properties of the frequency response of the analog front-end section of the Ultrasonix Sonix Touch RP scanner. Prior to the displacement estimation the RF lines were decimated by a factor of 6.

## 2.2 Displacement and strain estimation

The displacement fields were estimated using the 2D cross-correlation coefficient as the measure of similarity (Langeland et al., 2003). The size of the kernel and search region were  $7 \times 9$  [samples  $\times$  lines] and  $65 \times 23$  [samples  $\times$  lines]. The kernel size corresponded to the average speckle size. The resolution

of displacement estimation was increased by interpolating the cross-correlation coefficient by a factor of 5 in the radial direction (wave propagation direction) and by 10 in the angular direction. The outliers of displacement estimation were detected by analyzing regional values of the mean and standard deviation and corrected (Garcia, 2010; Wang et al., 2012). Subsequently displacement fields were smoothed using two-dimensional median and averaging filters, of size  $30 \times 15$  [samples  $\times$  lines] and  $10 \times 5$  [samples  $\times$  lines] in the case of the radial displacement map. In the case of the angular displacement map, the dimensions of the filters were respectively  $10 \times 10$  [samples  $\times$  lines] and  $30 \times 30$  [samples  $\times$  lines].

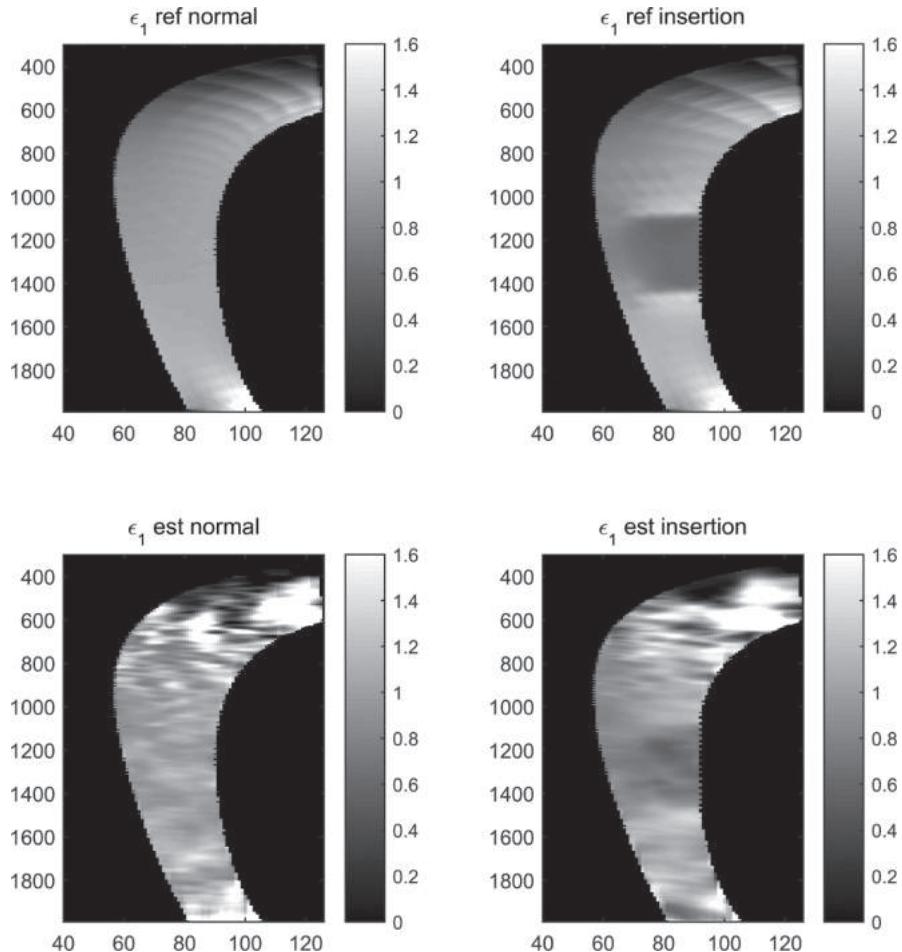


Figure 2. Maps of the first component of the principal strain, LAX view, data without scan conversion. FE simulation reference data—homogenous model (top-left), model with insertion (top-right). Estimation results—homogenous model (bottom-left), model with insertion (bottom-right). The effect of the non-transmural insertion can be seen in the (top-right) and (bottom-right) parts. Abscissa is scaled in scan line number, ordinate in RF sample number, strain nondimensionalised [%].

Displacement gradient tensors were obtained in the polar coordinate system (the centre of the ultrasonic probe was in the centre of this system) as directional derivatives of the displacement fields (radial and angular). The values of the derivatives were estimated as a value of the slope of the line fitted to N points around the point of interest. For derivatives in the radial direction, N was equal to 150 points and for angular direction N was equal to 10 points. The Eulerian strain tensor and principal strain components were obtained using displacement gradient tensors (Mase et al., 2009), a methodology similar to that presented in (Fung-Kee-Fung et al., 2005).

The FEM data were used as the reference to the results obtained from the synthetic ultrasonic data. The displacements exported from Abaqus were submitted to the same procedure of deformation

gradient tensor estimation as the displacements estimated from synthetic data, but using different derivative lengths: in the radial direction N equaled 50 points and in the angular direction N equaled 5 points.

### 3 RESULTS AND DISCUSSION

The principal strains were analyzed for two consecutive images from the part of the deformation cycle featuring the highest slope of the internal pressure curve (corresponding to the early diastolic phase). The maps of the first component of the principal strain obtained for the normal wall model and the affected wall model differ. Both maps correspond well to the reference maps obtained from the FEM data (Fig. 2).

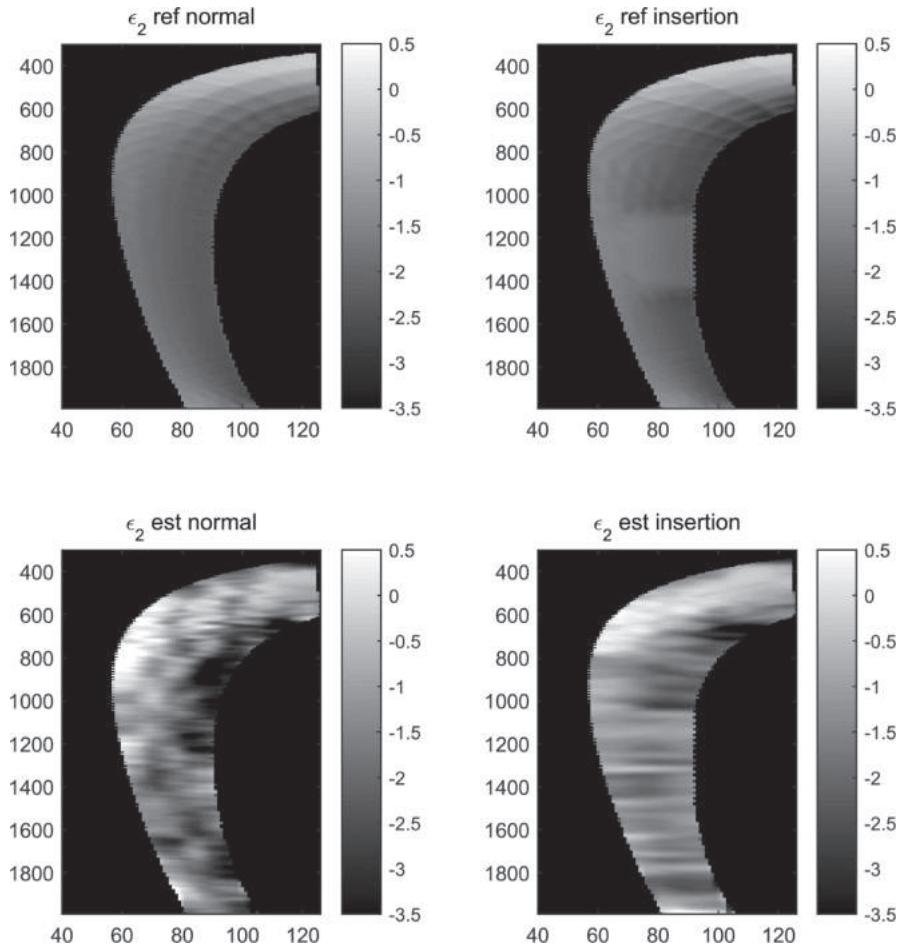


Figure 3. Maps of the second component of the principal strain, LAX view, data without scan conversion. Organization as in Figure 2.

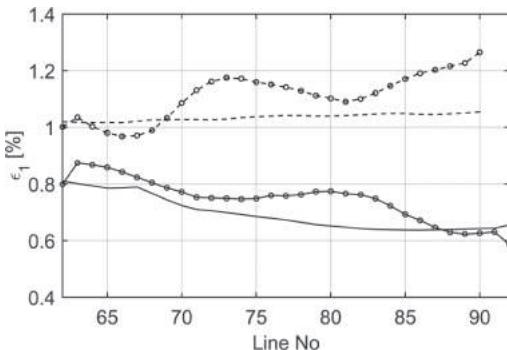


Figure 4. Plots of the first component of principal strain: estimated across the affected area (continuous circled line); reference across the affected area (continuous line); estimated across the corresponding area in the normal model (broken circled line); reference across the corresponding area in the normal model (broken line).

The maps of the second component of the principal strain also show differences between the normal wall model and the affected wall model, but due to more noise the differences are less visible. In the major part of the corresponding maps, the estimated values of the strain are close to the reference (Fig. 3). The biggest errors are observed in lower third part of the model. The maps of direction (angle) of both principal strain components very well correspond to reference maps, however, the difference between the normal wall model and the affected wall model is very subtle.

The slope of the first component of the principal strain in the normal wall and the affected wall at the place of the inclusion show different sign (Fig. 4). This means that the presence of the inclusion strongly affects the strain distribution within the model wall and also the transmural gradient of the first component of the principal strain, changing its sign. This indicates that the strain gradient may be considered as an indicator of presence of a stiffer area in the vicinity of the inner wall of the model.

#### 4 CONCLUSIONS

This preliminary study indicates, that the identification of the nontransmural inclusion in a numerical LV phantom, based on simulated synthetic ultrasonic data, is possible by analyzing estimated principal strain distribution within the wall and potentially also its gradient. Further work will include diversification of the size and material properties of the inclusion and experiments with physical LV phantoms.

#### ACKNOWLEDGEMENTS

This study was supported by the Polish National Science Centre by the decision DEC-2012/07/B/ST7/01441.

#### REFERENCES

- Abate, E., Hoogslag, G.E., Leong, D.P., Bertini, M., Antoni, M.L., Nucifora, G., Joyce, E., Holman, E.R., Siebelink, H.-M.J., Schalij, M.J., et al. (2014). Association between Multilayer Left Ventricular Rotational Mechanics and the Development of Left Ventricular Remodeling after Acute Myocardial Infarction. *J. Am. Soc. Echocardiogr.* 27: 239–248.
- Adamu, U., Schmitz, F., Becker, M., Kelm, M., and Hoffmann, R. (2009). Advanced speckle tracking echocardiography allowing a three-myocardial layer-specific analysis of deformation parameters. *Eur. J. Echocardiogr. J. Work. Group Echocardiogr. Eur. Soc. Cardiol.* 10: 303–308.
- Becker, M., Ocklenburg, C., Altio, E., Füting, A., Balzer, J., Krombach, G., Lysyansky, M., Kühl, H., Krings, R., Kelm, M., et al. (2009). Impact of infarct transmurality on layer-specific impairment of myocardial function: a myocardial deformation imaging study. *Eur. Heart J.* 30: 1467–1476.
- Belenkie, I., Sas, R., Mitchell, J., Smith, E.R., and Tyberg, J.V. (2004). Opening the pericardium during pulmonary artery constriction improves cardiac function. *J. Appl. Physiol. Bethesda Md 1985* 96: 917–922.
- Chan, J., Hanekom, L., Wong, C., Leano, R., Cho, G.-Y., and Marwick, T.H. (2006). Differentiation of Subendocardial and Transmural Infarction Using Two-Dimensional Strain Rate Imaging to Assess Short-Axis and Long-Axis Myocardial Function. *J. Am. Coll. Cardiol.* 48: 2026–2033.
- Cho, G.-Y., Chan, J., Leano, R., Strudwick, M., and Marwick, T.H. (2006). Comparison of two-dimensional speckle and tissue velocity based strain and validation with harmonic phase magnetic resonance imaging. *Am. J. Cardiol.* 97: 1661–1666.
- Cygan, S., Werys, K., Błaszczyk, Ł., Kubik, T., and Kałużyński, K. (2014). Left ventricle phantom and experimental setup for MRI and echocardiography—Preliminary results of data acquisitions. *Biocybern. Biomed. Eng.* 34: 19–24.
- Cygan, S., Źmigrodzki, J., Leśniak-Plewińska, B., Karny, M., Pakieła, Z., and Kałużyński, K. (2015). Influence of Polyvinylalcohol Cryogel Material Model in FEM Simulations on Deformation of LV Phantom. In H. van Assen, P. Bovendeerd, & T. Delhaas, eds. *Functional Imaging and Modeling of the Heart*, (Springer International Publishing), 313–320.
- De Korte, C.L., van der Steen, A.F., Dijkman, B.H., and Lancee, C.T. (1997). Performance of time delay estimation methods for small time shifts in ultrasonic signals. *Ultrasonics* 35: 263–274.
- Fukuta, H., and Little, W.C. (2008). The Cardiac Cycle and the Physiological Basis of Left Ventricular Contraction, Ejection, Relaxation, and Filling. *Heart Fail. Clin.* 4: 1–11.

- Fung-Kee-Fung, S.D., Lee, W.-N., Ingrassia, C.M., Costa, K.D., and Konofagou, E.E. (2005). Angle-independent strain mapping in myocardial elastography 2D strain tensor characterization and principal component imaging. In *2005 IEEE Ultrasonics Symposium*, 516–519.
- Garcia, D. (2010). Robust smoothing of gridded data in one and higher dimensions with missing values. *Comput. Stat. App. Data Anal.* 54: 1167–1178.
- Hashimoto, I., Li, X., Hejmadi Bhat, A., Jones, M., Zetts, A.D., and Sahn, D.J. (2003). Myocardial strain rate is a superior method for evaluation of left ventricular subendocardial function compared with tissue Doppler imaging. *J. Am. Coll. Cardiol.* 42: 1574–1583.
- Heyde, B., Cygan, S., Choi, H.F., Lesniak-Plewinska, B., Barbosa, D., Elen, A., Claus, P., Loeckx, D., Kaluzynski, K., and D’hooge, J. (2012). Regional cardiac motion and strain estimation in three-dimensional echocardiography: a validation study in thick-walled univentricular phantoms. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 59: 668–682.
- Ishizu, T., Seo, Y., Enomoto, Y., Sugimori, H., Yamamoto, M., Machino, T., Kawamura, R., and Aonuma, K. (2010). Experimental validation of left ventricular transmural strain gradient with echocardiographic two-dimensional speckle tracking imaging. *Eur. J. Echocardiogr. J. Work. Group Echocardiogr. Eur. Soc. Cardiol.* 11: 377–385.
- Jensen, J.A. (1996). FIELD: A Program for Simulating Ultrasound Systems. In *10th Nordicbaltic Conference on Biomedical Imaging*, Vol. 4, Supplement 1, Part 1: 351–353.
- Jensen, J.A., and Svendsen, N.B. (1992). Calculation of pressure fields from arbitrarily shaped, apodized, and excited ultrasound transducers. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 39: 262–267.
- Jia, C., Kim, K., Weitzel, W.F., Jia, C., Rubin, J.M., and Kolas, T.J. (2006). P1E-1 Left Ventricular Phantom with Pulsatile Circulation for Ultrasound Strain Rate Imaging. In *IEEE Ultrasonics Symposium, 2006*, 1317–1320.
- Kaluzynski, K., Chen, X., Emelianov, S.Y., Skovoroda, A.R., and O’Donnell, M. (2001). Strain rate imaging using two-dimensional speckle tracking. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, 48: 1111–1123.
- Karimi, A., Navidbakhsh, M., and Beigzadeh, B. (2014). A visco-hyperelastic constitutive approach for modeling polyvinyl alcohol sponge. *Tissue Cell* 46: 97–102.
- Langeland, S., D’hooge, J., Torp, H., Bijnens, B., and Suetens, P. (2003). Comparison of time-domain displacement estimators for two-dimensional RF tracking. *Ultrasound Med. Biol.* 29: 1177–1186.
- Lee, W.-N., and Konofagou, E.E. (2008). Angle-independent and multi-dimensional myocardial elastography—from theory to clinical validation. *Ultrasonics*, 48, 563–567.
- Leitman, M., Lysiansky, M., Lysyansky, P., Friedman, Z., Tyomkin, V., Fuchs, T., Adam, D., Krakover, R., and Vered, Z. (2010). Circumferential and Longitudinal Strain in 3 Myocardial Layers in Normal Subjects and in Patients with Regional Left Ventricular Dysfunction. *J. Am. Soc. Echocardiogr.* 23: 64–70.
- Lesniak-Plewinska, B., Cygan, S., Kaluzynski, K., D’hooge, J., Zmigrodzki, J., Kowalik, E., Kordybach, M., and Kowalski, M. (2010). A dual-chamber, thick-walled cardiac phantom for use in cardiac motion and deformation imaging by ultrasound. *Ultrasound Med. Biol.* 36: 1145–1156.
- Luo, J., and Konofagou, E.E. (2010). A fast normalized cross-correlation calculation method for motion estimation. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* 57, 1347–1357.
- Mase, G.T., Smelser, R.E., and Mase, G.E. (2009). *Continuum Mechanics for Engineers*, Third Edition (Boca Raton: CRC Press).
- Pesavento, A., and Ermert, H. (1998). Time-efficient and exact algorithms for adaptive temporal stretching and 2D-correlation for elastographic imaging using phase information. In, *1998 IEEE Ultrasonics Symposium, 1998*, vol.2: 1765–1768.
- Sakurai, D., Asanuma, T., Masuda, K., Hioki, A., and Nakatani, S. (2014). Myocardial layer-specific analysis of ischemic memory using speckle tracking echocardiography. *Int. J. Cardiovasc. Imaging*, 30: 739–748.
- Tee, N., Gu, Y., Murni, and Shim, W. (2015). Comparative Myocardial Deformation in 3 Myocardial Layers in Mice by Speckle Tracking Echocardiography. *BioMed Res. Int.* 2015, e148501.
- Touil, B., Basarab, A., Delachartre, P., Bernard, O., and Friboulet, D. (2010). Analysis of motion tracking in echocardiographic image sequences: influence of system geometry and point-spread function. *Ultrasound* 50: 373–386.
- Waldman, L.K., Fung, Y.C., and Covell, J.W. (1985). Transmural myocardial deformation in the canine left ventricle. Normal in vivo three-dimensional finite strains. *Circ. Res.* 57: 152–163.
- Wang, G., Garcia, D., Liu, Y., de Jeu, R., and Johannes Dolman, A. (2012). A three-dimensional gap filling method for large geophysical datasets: Application to global satellite soil moisture observations. *Environ. Model. Softw.* 30: 139–142.
- Zhang, Y., Chan, A.K.Y., Yu, C.-M., Yip, G.W.K., Fung, J.W.H., Lam, W.W.M., So, N.M.C., Wang, M., Wu, E.B., Wong, J.T., et al. (2005). Strain Rate Imaging Differentiates Transmural From Non-Transmural Myocardial Infarction: A Validation Study Using Delayed-Enhancement Magnetic Resonance Imaging. *J. Am. Coll. Cardiol.* 46: 864–871.

# Microcalcification segmentation in full field digital mammography

N. Martins

*Centro de Computação Gráfica, Azurém, Guimarães, Portugal*

*Enermeter, Parque Industrial de Celeirós, Lugar de Gaião, Braga, Portugal*

D. Veiga, C. Pereira & M. Ferreira

*Enermeter, Parque Industrial de Celeirós, Lugar de Gaião, Braga, Portugal*

*University of Minho, Centro Algoritmi, Azurém, Guimarães, Portugal*

N. Alves

*Centro de Computação Gráfica, Azurém, Guimarães, Portugal*

M. Delgado

*Dr Campos Costa, Imagiologia Clínica, Rua de Avis, Porto, Portugal*

**ABSTRACT:** Breast Cancer is one of the most common types of cancer in Women. Mammography is widely used as a screening method because it can detect pathologies in an early stage. There are some Computer Aided Diagnosis (CAD) systems, but their performance can be improved with new and better image processing algorithms. In this paper, we present a new microcalcification detection method for Full Field Digital Mammograms (FFDM), based on the wavelet transform, Expectation Maximization with Markov Random Fields bootstrap and a final SVM classifier. Tests were made in a public database (InBreast) and results shown that our approach achieved better False Positive (FP) rates and sensitivities compared to the literature. Features were also presented and the problems with the train addressed.

## 1 INTRODUCTION

Breast cancer is the most common type of cancer in woman. In the US, it was estimated 226,870 new cases of invasive cancer, and around 40.000 deaths, making it a serious public health problem (Jiemin Ma 2013). The most used and globally accepted method for breast cancer detection is the Mammography. With this technique is possible to find breast cancer clues, like Microcalcification (MC) clusters in an early stage. Is known that an early detection is crucial for an efficient treatment, for instance, when diagnosed in an early stage, the success rates are about 98.6% and with much less invasive treatments. When the diagnosis is in a late stage, the success rate drops to about 23.3% and in most cases needs a surgical intervention to remove part of the breast (mastectomy) (Jiemin Ma 2013). This kind of treatment is painful and is psychologic traumatic due to the removal of an important organ for the woman self esteem. Given the importance of an early detection, several algorithms have been proposed in the literature to help physicians in the diagnosis. MCs are one of the hardest clues to find in a mammogram, they consist in very small white dots, resulting from the necrotic activities

of cells. In digital images, they can be as small as 1–2 pixels, and with very low signal to noise ratio. This represents a challenge to both, physicians and image processing algorithms and that's why it still is an open problem (A. Bria 2014).

### 1.1 Related word

The number of publications in this field is extremely large. In (Karthikeyan Ganesan 2013, Afsaneh Jalalian 2013) is presented a recent review of the stat-of the-art. The problem, is that most of them use screen films with less quality than FFDM images. In fact, to our knowledge, there is only one public FFDM database which is InBreast (Domingues 2011).

Some recent and important papers, regarding MC detection, are presented next. Supervised algorithms have been used more often than the unsupervised, due to the complexity of the problem. In (E. Zhang & Bai 2014) an SVM was used to reduce the False Positives (FP), after a contrast enhancement technique based on morphologic filters and a dual threshold technique. In (A. Bria 2014) a cascade of classifiers was used to classify unbalanced data, first they used haar-like features to detect

MCs and then a new set of graph based features to remove non-malign clusters, they claim results comparable with existent commercial systems (R2CAD), but the database used was not public, and so no comparisons are possible. In (S.-Y. Shin & Yun 2014), a deep learning technique was used. They started with a peak detector, followed by a manual classification to distinguish between calcification and non-calcification, to teach the algorithm the characteristics of the calcifications. After that, 33 unknown images are presented to the classifiers and a ROC of 0.83 was achieved. Once again the database was private and the testing set small, making the results hard to replicate. In (Domingues & Cardoso 2014), they propose a Bayesian surprise method to detect MCs, they compare their approach with three other common algorithms and they claim that they achieve a sensitivity of 60.3% with a FPR per image of 108. In this work they used their own public database (inBreast), so the present work will use the same database and metrics in order to make the comparisons fair.

## 2 PROPOSED WORK

Our work consists in three main steps. First a pre-processing is applied to enhance the MCs, then a segmentation step where the MCs are extracted from the background and finally a classification step to reduce false positives. In the next sections, all techniques will be explained in more detail.

### 2.1 Pre-processing

In order to enhance the signal to noise ratio of the MC, a wavelet-based enhancement technique was applied. The technique presented in (Martins, Pereira, & Ferreira 2014) is used, more precisely the Wavelet Contrast method. Next an explanation of this method is done and the parameters chosen for the given images.

The Wavelet Transform is a well known signal transform tool. It uses wavelet filters to decompose the image in details and approximations (high and low frequencies) in multiscale. These characteristics make it interesting to use in mammography, because MCs can be seen as high frequencies in these images. Moreover, their size is variable which the multiscale property can naturally deal with. A very simple way to express the wavelet transform is shown in equation (1 and 2)

$$\{A_{s+1}, D_{s+1}^h, D_{s+1}^v, D_{s+1}^d\} = DWT(A_s) \quad (1)$$

$$\{A_{s-1}\} = IDWT(A_s, D_s^h, D_s^v, D_s^d) \quad (2)$$

where,  $DWT$  is the discrete wavelet transform and  $IDWT$  is the inverse discrete wavelet transform.  $A$  and  $D$  represent the approximation and detail coefficients,  $s$  is the scale, which varies between 0 and  $S$ , where 0 is the original image and  $S$  the coarser scale. At last,  $h$ ,  $v$ ,  $d$  represent respectively the horizontal, vertical and diagonal orientations.

In (Martins, Pereira, & Ferreira 2014), contrast is defined as the ratio between the details and approximations coefficients in the wavelet domain. After some transformations, the enhancement is achieved using equation (3)

$$\overline{D}_s^k(m,n) = G_s^k(m,n) \frac{\overline{A}_s(m,n)}{A_s(m,n)} D_s^k(m,n) \quad (3)$$

where,  $\overline{D}_s^k$  are the enhanced detail coefficients,  $G_s^k$  the gain matrix and  $\overline{A}_s$  the enhanced approximation coefficients.

In order to use equation (3) we must find  $\overline{A}_s$ . This step is achieved using the  $IDWT$  with the enhanced detail coefficients of the scale above ( $s + 1$ ). So, the enhancement starts at the coarse scale,  $S$ , until the scale 0 is reached. At each scale the information from the scale above is used. At the beginning, we define  $\overline{A}_S$  as  $A_S$ , because no scale above exists and the Gain matrix,  $G_S^k$ , is used.

The parameters selected in our approach are based on the image resolution. So a maximum scale,  $S$ , of 4 was chosen and the Gain matrix was simplified to a fixed value, 1.5. The wavelet filter used was the “biorthogonal 1.5” because it resembles the MC shape. The value of  $S = 4$  was chosen, because this is the last scale where MCs are expected to show up. In Figure 1 is shown the result of the pre-processing.

### 2.2 Segmentation

Two main segmentation techniques were used. The first is the segmentation of the breast tissues in glandular and adipose. This is done with an Expectation Maximization (EM) algorithm with an MRF bootstrap. The second is a wavelet based contrast segmentation, for each tissue. The segmentation is only computed in the breast area and, to detect it, the OTSU algorithm was used, followed by an erosion to refine the contours of the resulting mask.

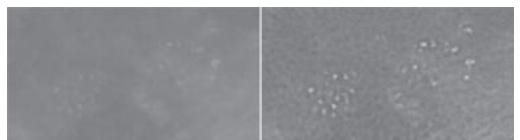


Figure 1. Microcalcifications enhancement result.

### 2.2.1 Breast tissue segmentation

Looking at the histogram of a Mammogram, is possible to find two overlapping Gaussians Figure 2. So, knowing that the glandular tissue is brighter than the adipose, we assume that the two overlapping Gaussians represent the adipose and glandular tissues.

In order to separate them, we use the EM algorithm to calculate the parameters of both gaussians and the MRF to give a spatial sensitivity to the algorithm as shown in (Zhang, Brady, & Smith 2001). We use the k-means algorithm to initialize the segmentation and then the EM-MRF to refine it. The idea of using the MRF is to enhance the separation of the tissues in the overlapping area. If a pixel is classified as glandular tissue, but is surrounded by adipose tissue, odds are that the pixel is also adipose tissue, and the other way around. This technique results in a much smoother segmentation and robust against noise.

### 2.2.2 Microcalcifications segmentation

After the tissue segmentation, the algorithm looks for MCs in each tissue. This approach follows the idea that the noise levels are different in the adipose and glandular tissue. So they are processed in separate to avoid over segmentation in one and under segmentation in the other. This was expected in our approach due to the use of a statistical contrast approach. The segmentation step starts with the image decomposition using the *DWT*. After that, the coarser approximation coefficients are set to zero, and then the *ISWT* is used to obtain the contrast image. The idea is to remove the low pass band of the image, which correspond to the tissues. Higher frequencies are considered to be possible MCs. The contrast image is then separated, using the tissue segmentation, in adipose and glandular component. Both tissues are assumed to have a gaussian like histogram (Fig. 3) and so a Z-normalization of the intensities is performed followed by a subtraction of 1 from all pixels. This step aims to reduce and uniform the quantum

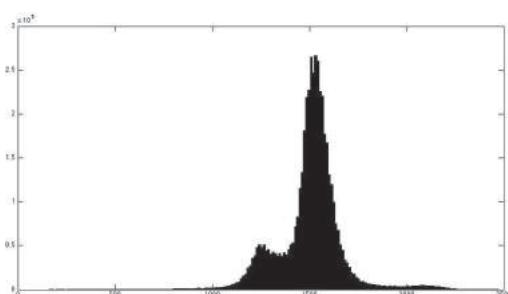


Figure 2. Histogram of a mammogram.

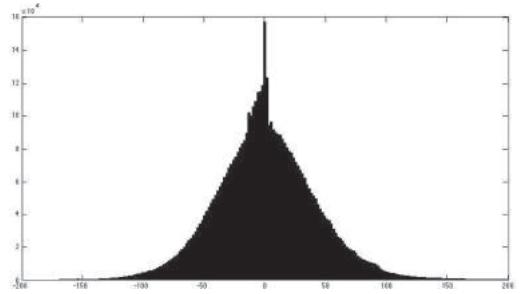


Figure 3. Histogram of the contrast image.

Table 1. Features extracted from each candidate.

Features	Description
Shape based	Area, perimeter, circularity
Intensity based	Mean, std, min, max
Zernike moments	Radial degree 0 to 4
Hu moments	Invariant to rotation and size
Fractal based	Box counting method
Gaussian similarity	8 directions std, skew, kurt

noise in the image. The subtraction of 1, removes around 84% of the pixels with less contrast and the remaining 16% are considered as candidates to the next phase.

Next the images are merged and the resulting image is once more normalized, dividing the intensities by the maximum value between the maximum intensity of the image and 20. The value 20 was obtain after observation of the images without MCs. In these cases the maximum value drops, because no high frequencies are present in the image, which would force the algorithm to always find MCs that are not present. Consequently the number of false positives would drastically increases. After that a simple linear threshold is used to differentiate the MCs from the background.

### 2.3 False positive reduction

To reduce the FP rate, an SVM was used. In this step, a set of features are extracted from each MC candidate. The selected features are presented in Table 1.

First, the candidates are extracted with the segmentation step. A small threshold was used in order to get a high sensitivity. Next, each candidate is compared with the GT and separated between correct and incorrect detections. The number of incorrect detections is much higher than the number of the correct ones. Because of that, a training strategy similar to (A. Bria 2014) was used. 75% of the

correct detected MCs were selected to train the SVM and the false group was randomly picked from the incorrect detected MCs, so that the same number of correct and incorrect detected MCs is obtained. After that, a 5-fold cross-validation was used to train the SVM and tests were made in the remaining MC candidates (not used in the train). This process is repeated until a desired SVM score is achieved. High sensitivities are desired, and so a minimum sensitivity of 85% was used and the SVM with the best specificity was kept.

#### 2.4 Metrics

The performance metric used was the same as in (Domingues & Cardoso 2014), not only because it allows a direct comparison with their results, but also, because it is based in a mathematical proven formulation. The final results are presented in a FROC curve. The FROC curve test several different thresholds to create a curve of False Positives per image vs sensitivity. In our case, 2 tests were made, one before the SVM and the other after the SVM.

### 3 RESULTS

In the Figure 4 is shown the FROC curve of our approach.

As expected, the SVM reduce the false positive rates considerably. For a sensitivity of 70% the reduction was from 75 FP to less than 45, and for a sensitivity of 80% the reduction was from 400 to 170 FP.

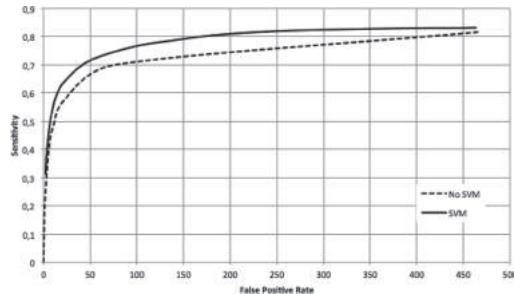


Figure 4. FROC of the proposed method.

Table 2. Operation points of some published algorithms.

Method	Sensitivity (%)	FP
Fixed threshold	34.5	164
Outlier detection	45.8	60
Mathematical morphology	40.3	225
Bayesian surprise 1	60.3	108

In Table 2 is shown the published results of other methods.

Comparing with the Bayesian Surprise, our approach performed considerable better. In their operation point (FP = 108), our method, achieved a sensitivity of 77.2% with the SVM and 69.97% before the SVM, in comparison with a sensitivity of 60.3% obtain by the Bayesian Surprise. Regarding the other detection methods, our method outperformed all.

### 4 CONCLUSIONS

A fully automatic segmentation method was proposed for MC segmentation. Results shown that, for the InBreast database, our approach achieved higher sensitivity and lower FP per image than the existing methods in the literature. Results still have a long way to go, higher sensitivities and lower FP rates are necessary to make the system more trustful. Regarding the proposal itself, we highlight the modularity, each step is independent from the other. For instance, the preprocessing could help improve existing methods and the segmentation could be used as a starting point to other methods. The Segmentation also have the advantage of being simple and that no training is needed, so it can be used, in theory, in any kind of mammography image with only small parameters adjustments. The SVM performed as expected reducing considerably the number of FPs, proving that the features chosen correctly, making it a valuable contribution of this work.

### ACKNOWLEDGMENTS

Work supported by European Union funds through the FUNDO EUROPEU DE DESENVOLVIMENTO REGIONAL and by national funds by Quadro de Referência Estratégico Nacional. N. Martins thanks the Centro de Computação Gráfica, Dr Campos Costa and ENERMETER.

### REFERENCES

- A. Bria, N.K. (2014). Learning from unbalanced data: A cascade-based approach for detecting clustered microcalcifications. *Medical Image Analysis* 18, 241–252.
- Afsaneh Jalalian, S.B.M. (2013). Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *ELSEVIER—Clinical Imaging*.
- Domingues, I., C.J. (2011). Inbreast: Toward a full-field digital mammographic database inbreast: Toward a full-field digital mammographic database. *Technical Report: Acad Radiology*.

- Domingues, I. & J. Cardoso. (2014, Aug). Using bayesian surprise to detect calcifications in mammogram images. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pp. 1091–1094.
- Jiemin Ma, Ahmedin Jemal Jiemin Ma, A.J.J.M.A.J. (2013). *Breast Cancer Metastasis and Drug Resistance*. Springer-Verlag New York Springer-Verlag New York.
- Karthikeyan Ganesan, U.R.A. (2013). Computer-aided breast cancer detection using mammograms: A review. *IEEE REVIEWS IN BIOMEDICAL ENGINEERING* 6.
- Martins, N., C. Pereira, & M. Ferreira. (2014, June). Digital mammograms contrast enhancement using wavelets—a comparative study. In *Medical Measurements and Applications (MeMeA), 2014 IEEE International Symposium on*, pp. 1–6.
- Shin, S.-Y., S.L. & I.D. Yun. (2014). Classification based microcalcification detection using discriminative restricted boltzmann machine in digitized mammograms. *SPIE Medical Imaging* 115.
- Zhang, Y., M. Brady, & S. Smith (2001, Jan). Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm. *Medical Imaging, IEEE Transactions on* 20(1), 45–57.
- Zhang, E., F. Wang, Y.L. & X. Bai. (2014). Automatic detection of microcalcifications using mathematical morphology and a support vector machine. *Bio-medical materials and engineering* 24(1), 53–59.

This page intentionally left blank

# Unsupervised delineation of the vessel tree in retinal fundus images

N. Strisciuglio

*Department of Computer Engineering and Electrical Engineering and Applied Mathematics, University of Salerno, Salerno, Italy*

*Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, Groningen, The Netherlands*

M. Vento

*Department of Computer Engineering and Electrical Engineering and Applied Mathematics, University of Salerno, Salerno, Italy*

G. Azzopardi

*Department of Intelligent Computer Systems, University of Malta, Msida, Malta*

*Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, Groningen, The Netherlands*

N. Petkov

*Johann Bernoulli Institute for Mathematics and Computer Science, University of Groningen, Groningen, The Netherlands*

**ABSTRACT:** Retinal imaging has gained particular popularity as it provides an opportunity to diagnose various medical pathologies in a non-invasive way. One of the basic and very important steps in the analysis of such images is the delineation of the vessel tree from the background. Such segmentation facilitates the investigation of the morphological characteristics of the vessel tree and the analysis of any lesions in the background, which are both indicators for various pathologies. We propose a novel method called *B-COSFIRE* for the delineation of the vessel tree. It is based on the classic COSFIRE approach, which is a trainable nonlinear filtering method. RE filter to be configured by the automatic analysis of any given vessel-like pattern. The responses of a *B-COSFIRE* filter is achieved by combining the responses of difference-of-Gaussians filters whose areas of support are determined in an automatic configuration step. We configure two types of *B-COSFIRE* filters, one that responds selectively along vessels and another that is selective to vessel endings. The segmentation of the vessel tree is achieved by summing up the response maps of both types of filters followed by thresholding. We demonstrate high effectiveness of the proposed approach by performing experiments on four public data sets, namely DRIVE, STARE, CHASE\_DB1 and HRF. The delineation approach that we propose also has lower time complexity than existing methods.

**Keywords:** delineation; retinal image analysis; trainable filters; vessel segmentation

## 1 INTRODUCTION

Retinal fundus imaging is a non-invasive technique that has gained popularity for the diagnosis of several pathologies, including diabetic retinopathy, glaucoma, hypertension, arteriosclerosis and multiple sclerosis. The automatic analysis of retinal images may be used to assist medical specialists to diagnose such pathologies.

The separation of the blood vessels from the rest of the image is an important segmentation step, which facilitates the subsequent analysis. For instance, the detection of vessel bifurcations, whose progression may provide signs for cardiovascular diseases, may be better detected in the segmented image. Moreover, subtracting the delineated vessel tree from the background provide better opportunity for the automatic detection of any retinal lesions.

There are two categories of methods that have been proposed so far, namely supervised (Staal et al. 2004, Soares et al. 2006, Ricci and Perfetti 2007, Marin et al. 2011, Fraz et al. 2012) and unsupervised (Hoover et al. 2000, Jiang and Mojon 2003, Mendonça and Campilho 2006, Al-Rawi et al. 2007) approaches. Supervised approaches form pixel-wise feature vectors and learn a binary classifier that labels pixels as vessel or non-vessel. On the other hand, unsupervised approaches, typically rely on tracking, filtering or morphological operations followed by thresholding. Supervised methods require reasonably-sized data sets of training images and corresponding ground truth, which is usually very difficult and expensive to obtain.

We propose a novel unsupervised delineation algorithm that is based on the Combination of Receptive Fields (CORF) computational model of

a simple cell in visual cortex (Azzopardi and Petkov 2012, Azzopardi et al. 2014) and its implementation called Combination Of Shifted Filter Responses (COSFIRE) (Azzopardi and Petkov 2013). It uses two types of COSFIRE filters, one that gives a response along the vessels and the other one that selectively responds to vessel endings. In general, the two COSFIRE filters that we use respond to bar structures and thus we call them bar-selective COSFIRE filters or *B*-COSFIRE for brevity.

The rest of the paper is organized as follows: in Section 2 we present the *B*-COSFIRE filters; in Section 3, we present the experimental protocol and the results achieved on four public data sets of retinal fundus images, which we compare with state-of-the-art methods in Section 4. Finally, we draw conclusions in Section 5.

## 2 PROPOSED METHOD

### 2.1 *B*-COSFIRE filter

A *B*-COSFIRE filter is trainable, in that its selectivity is determined by an automatic configuration process that analyzes a given prototype pattern. Figure 1a shows a synthetic bar structure that we use as prototype. The configuration procedure involves the convolution of a center-on Difference-of-Gaussians (DoG) filter—with a given standard deviation  $\sigma$  of the outer Gaussian function<sup>1</sup>—and the synthetic bar pattern followed by the determination of local maximum DoG responses along a number of concentric circles, Figure 1b. For each local maximum point  $i$  we extract the polar coordinates  $(\rho_i, \phi_i)$  with respect to the center of the prototype pattern. This location is indicated by the ‘x’ marker in Figure 1b.

This configuration procedure results in a set  $S = \{(\sigma_i, \rho_i, \phi_i) | i=1, \dots, n\}$  of 3-tuples. Here we

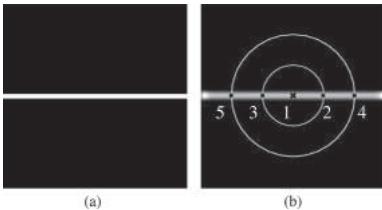


Figure 1. (a) A synthetic vessel-like structure used to configure a *B*-COSFIRE filter. (b) The output of a center-on DoG filter to the image in (a). The cross marker is the center of support of the *B*-COSFIRE filter while the enumerated spots represent the positions at which the strongest DoG responses are achieved.

<sup>1</sup>The standard deviation of the inner Gaussian function is half the standard deviation of the outer Gaussian function.

use one DoG filter and thus the  $\sigma$  values of all tuples are the same. This design decision allows the use of multi-scale DoG filtering to a prototype pattern of an irregular shape. As an example, the filter model automatically determined from the prototype pattern in Figure 1b is:

$$S = \left\{ \begin{array}{l} (\sigma_1 = 2.6, \rho_1 = 0, \phi_1 = 0), \\ (\sigma_2 = 2.6, \rho_2 = 2, \phi_2 = 0), \\ (\sigma_3 = 2.6, \rho_3 = 2, \phi_3 = 3.14), \\ (\sigma_4 = 2.6, \rho_4 = 4, \phi_4 = 0), \\ (\sigma_5 = 2.6, \rho_5 = 4, \phi_5 = 3.14) \end{array} \right\} \quad (1)$$

We compute the response of a *B*-COSFIRE filter as the geometric mean of the responses of the concerned *DoG* filters at the positions determined in the configuration step:

$$r_S(x, y) \stackrel{\text{def}}{=} \left( \prod_{i=1}^{|S|} \left( s_{\sigma_i, \rho_i, \phi_i}(x, y) \right) \right)^{1/|S|} \quad (2)$$

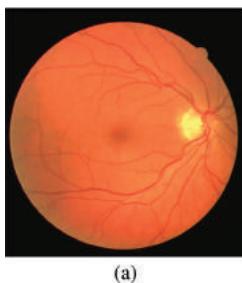
where  $s_{\sigma_i, \rho_i, \phi_i}(x, y)$  is the blurred response of a DoG filter with  $\sigma_i$ . Blurring is achieved by taking the maximum DoG response in a local neighbourhood weighted by a Gaussian function. The Gaussian function is centered around the polar coordinates  $(\rho_i, \phi_i)$  with respect to the center of the support of the *B*-COSFIRE filter (i.e. the pixel under consideration). It has a standard deviation  $\sigma'_i$  that is a linear function of the distance  $\rho_i$ :  $\sigma'_i = \sigma'_0 + \alpha \rho_i$ . Further details about the parameters  $\sigma'_0$  and  $\alpha$  are provided in Section 3.3. The blurring step is important as it allows for some tolerance with respect to the preferred positions of the concerned DoG responses.

In Azzopardi et al. (2014), we also demonstrate that the *AND*-type function (geometric mean) that we use is more robust to noise than an *OR*-type function (e.g. arithmetic mean).

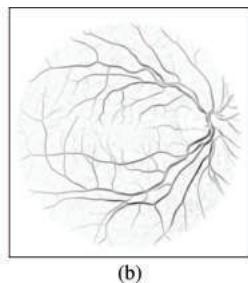
### 2.2 Tolerance to rotations

The *B*-COSFIRE filter configured above is selective to horizontal vessels. By using a rotated prototype pattern we can configure a filter that has preference to a different orientation. Alternatively, we create a new set  $R_\psi(S) = \{(\sigma_i, \rho_i, \phi_i + \psi) | i=1, \dots, n\}$  that represents a new *B*-COSFIRE filter with an orientation preference that is offset by  $\psi$  radians from that of the original filter  $S$ . We denote by  $\hat{r}_S(x, y)$  the rotation-tolerant response at location  $(x, y)$  that is achieved by taking the maximum value of various *B*-COSFIRE filters with different orientation preferences:

$$\hat{r}_S(x, y) \stackrel{\text{def}}{=} \max_{\psi \in \Psi} \{r_{R_\psi(S)}(x, y)\} \quad (3)$$



(a)

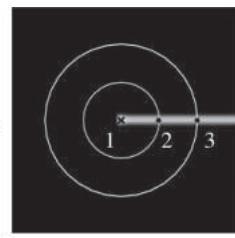


(b)

Figure 2. (a) Input retinal image and (b) the (inverted) output of a rotation-tolerant *B*-COSFIRE operator using 12 orientation preferences.

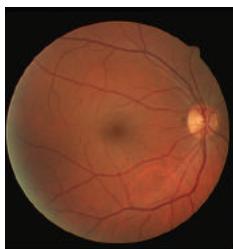


(a)

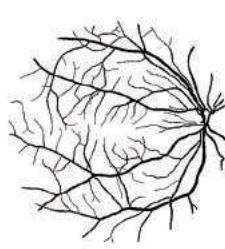


(b)

Figure 3. (b) Example of the configuration of an asymmetric *B*-COSFIRE filter by a prototype vessel-ending (a). The point of interest of the *B*-COSFIRE filter is indicated by the cross marker and lies on the end of the prototype.



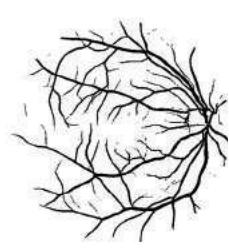
(a)



(b)



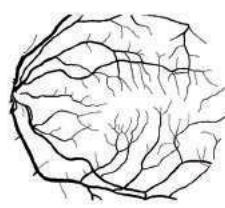
(c)



(d)



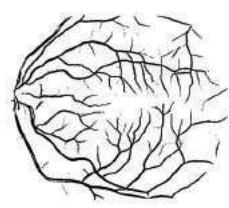
(e)



(f)



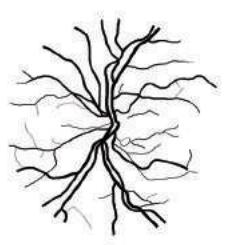
(g)



(h)



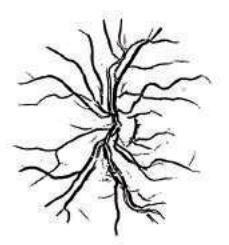
(i)



(j)



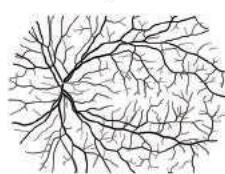
(k)



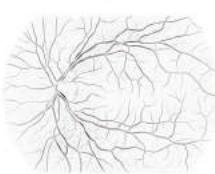
(l)



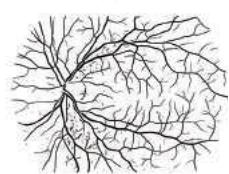
(m)



(n)



(o)



(p)

Figure 4. Examples of retinal fundus images (first column) from the four concerned data sets (DRIVE, STARE, CHASE DB1 and HRF) together with the manually segmented ground truth (second column). The output of the delineation algorithm (third column) is thresholded in order to obtain the final segmented vessel tree (fourth column).

where  $\Psi = \{0, \frac{\pi}{12}, \frac{\pi}{6}, \dots, \frac{11\pi}{12}\}$ . Figure 2b shows the rotation-tolerant output of a *B*-COSFIRE filter when applied to the green channel of the image in Figure 2a.

### 2.3 Delineation algorithm

The delineation algorithm that we propose uses two types of *B*-COSFIRE filters that we refer to them as symmetric and asymmetric. A symmetric *B*-COSFIRE filter is of the type configured above which considers the same number of DoG responses on both sides of the support center of the filter. Such a filter gives strong responses along vessels but responds weakly on the vessel endings. An asymmetric *B*-COSFIRE filter is one which achieves strong responses on vessel endings and it is configured by the prototype pattern shown in Figure 3a. Delineation is then achieved by thresholding the sum of the response maps of the symmetric and asymmetric *B*-COSFIRE filters. In Figure 4, we show examples of the output of the delineation algorithm (third column) and its thresholded version (fourth column) for each of the four data sets, together with the manually segmented ground truth (second column).

## 3 EXPERIMENTS AND RESULTS

### 3.1 Data sets

We use four data sets to evaluate the proposed approach: DRIVE (Staal et al. 2004), STARE (Hoover et al. 2000), CHASE\_DB1 (Owen et al. 2009) and HRF (Odstrcilik et al. 2013).

The DRIVE data set is composed of 40 images (20 for training and 20 for testing) with resolution 768×584 pixels. For each image, a mask of the Field Of View (FOV) of the retina is provided together with the binary manual segmentation of the blood-vessel tree. The images in the training set have been manually segmented by a single observer, while the images in the test set have been segmented by two other observers.

The STARE data set consists of 20 images (700×605 pixels), 10 of which contain signs of pathologies. Each image is distributed together with two manually segmented images provided by two different observers.

The CHASE\_DB1 data set contains 28 colour images (999×960 pixels) from 14 patients in the program Child Heart And Health Study in England. The data set contains two groups of manually segmented images provided by two observers. For the three dataset the first set of segmented images is used as ground truth.

The HRF data set consists of images with a resolution (3504×2336 pixels) that is substantially

higher than images of the former data sets. It comprises of three sets of 15 images each collected from healthy people, patients with signs of glaucoma and patients with signs of diabetic retinopathy. Each image is coupled with one manually segmented image used as ground truth.

### 3.2 Pre-processing

For our experiments we consider only the green channel of the retinal fundus images as it provides the best contrast between the vessels and the background. Furthermore, we apply the pre-processing algorithm proposed by Soares et al. (2006) to smoothen the pixels that are close to the circumference of the FOV. This is achieved in an iterative way by increasing the radius of the FOV by one pixel in each iteration. The value of a new pixel is the average of the neighbouring pixels within the FOV that is determined in the previous iteration. The initial FOV masks for the images of the DRIVE and HRF data sets are provided with the ground truth data. For the other data sets this information is, however, missing and we obtain the initial FOV masks by thresholding<sup>2</sup> the luminosity channel of the CIELab version of the original RGB image. For the HRF data set, we downsampled the original images by a factor of 2.

### 3.3 Selection of parameters

The *B*-COSFIRE filters that we propose have four parameters, namely the standard deviation  $\sigma$  of the outer Gaussian function of the afferent DoG filter, a set of  $\rho$  values that are the radii of the concentric circles used in the configuration, together with  $\sigma'_0$  and  $\alpha$  that are used by the blurring function.

For each data set we consider a validation set of images and use a grid search to first estimate the parameter values of the symmetric *B*-COSFIRE filter and then to estimate the parameter values of the asymmetric *B*-COSFIRE filter that best improves on the performance of the symmetrical one. For DRIVE we use the provided 20 training images as the validation set, and for the other three data sets we form validation sets with the first half of the images.

In Table 1 we report the estimated parameter values of the two *B*-COSFIRE filters that we use to process the images in the four considered data sets.

### 3.4 Evaluation method

We binarize the summation of the response maps of the two *B*-COSFIRE filters by using a threshold

<sup>2</sup>The thresholds are 0.5 and 0.1 for the STARE and CHASE\_DB1, respectively.

Table 1. Estimated parameter values of the *B*-COSFIRE filters used in four public data sets.

Data set	Filter	$\sigma$	$\rho$	$\sigma_0$	$\alpha$
DRIVE	Symmetric	2.4	{0, 2, ..., 8}	3	0.7
	Asymmetric	1.8	{0, 2, ..., 22}	2	0.1
STARE	Symmetric	2.7	{0, 2, ..., 12}	1	0.6
	Asymmetric	2.1	{0, 2, ..., 24}	1	0.1
CHASE_DB	Symmetric	4.8	{0, 2, ..., 18}	3	0.2
	Asymmetric	4.3	{0, 2, ..., 34}	1	0.1
HRF	Symmetric	3.2	{0, 2, ..., 14}	3	0.3
	Asymmetric	2.7	{0, 2, ..., 24}	1	0.1

Table 2. Experimental results of the proposed delineation algorithm on public data sets.

Data set	Se	Sp	AUC	Acc
DRIVE	0.7655	0.9704	0.9614	0.9442
STARE	0.7763	0.9695	0.9555	0.9496
CHASE_DB1	0.7699	0.9476	0.9497	0.9305
HRF (Healthy)	0.7467	0.9850	0.9557	0.9665
HRF (DR)	0.7620	0.9604	0.9413	0.9663
HRF (Glaucoma)	0.7446	0.9781	0.9561	0.9603

value as a fraction of the maximum response. In this way, the pixels whose sum of the two *B*-COSFIRE filters is greater than the threshold are classified as *vessels* and the remaining as *non-vessels*.

We use a systematic set of threshold values between 0 and 1 in intervals of 0.01 and match every resulting binary image with the corresponding ground truth image. For each threshold value we compute the number of true positives, false positives, true negatives and false negatives, and subsequently use them to compute the Matthews Correlation Coefficient (MCC) as a performance indicator. The MCC is a suitable measurement for unbalanced classes and thus it is appropriate for the problem at hand where the background pixels outnumber the vessel pixels approximately by a factor of 7.

Finally, we choose the threshold that provides the maximum average MCC value over a validation data set, and report the Accuracy (Acc), Sensitivity (Se) and Specificity (Sp) for that threshold. For the sake of completeness we also compute the ROC curve and its underlying area (AUC). In Table 2 we report the results that we achieve for the concerned four data sets.

#### 4 DISCUSSION

In Table 4, we report the results achieved by the proposed method on the four considered data sets, in comparison with the ones achieved by other

published methods. The results of these methods are achieved by thresholding the response of a filter or the score output of a classifier. However, the evaluation of algorithms for vessels segmentation in retinal fundus images is complex. The accuracy is not a suitable metric because it is biased by the high number of non-vessel pixels (background). Thus, in order to compare the proposed method with other published methods we move along the ROC curves. For the same specificity values reported in state-of-the-art studies, we compare the value of the sensitivity. For the DRIVE data set and for the same specificity ( $Sp = 0.9764$ ) reported by (Mendonca and Campilho 2006) we achieve a sensitivity of 0.7376, which is marginally better. Similarly, for the STARE data set and for the same specificity reported by (Mendonca and Campilho 2006) and (Al-Diri et al. 2009) ( $Sp = 0.9730$  and  $Sp = 0.9681$ ) we achieve a sensitivity of 0.7554 and 0.7848 respectively, which is a significantly better result. We also achieve the best AUC value for the DRIVE data set with respect to all other unsupervised approaches ( $AUC = 0.9435$ , Al-Rawi et al. 2007 –  $AUC = 0.9558$ , Ricci and Perfetti 2007 –  $AUC = 0.9407$ , Cinsdikici and Aydin 2009 –  $AUC = 0.9614$ , Lam et al. 2010).

For the HRF data set we achieve better results than the other published method (Odstrcilik et al. 2013). Indeed, for the same specificity achieved by Odstrcilik et al. 2013 for the healthy, diabetic retinopathy and glaucoma sets, we achieve sensitivity values of 0.8063, 0.7528 and 0.8168, respectively. Such values are higher than the ones reported by Odstrcilik et al. 2013 (0.7861, 0.7463 and 0.79). As for the CHASE\_DB1 there are no other unsupervised methods to compare with. The only other state-of-the art method is a complex supervised approach (Fraz et al. 2012) based on an ensemble of classifiers.

As regards the time required to process an image (from DRIVE and STARE data sets), the proposed method is the most efficient with respect to other methods from the state of the art (Table 3).

Table 3. Comparison of the time required to process an image from the DRIVE and STARE data sets.

Method	Processing time
<b><i>B</i>-COSFIRE</b>	<b>10 seconds</b>
(Jiang and Mojon 2003)	20 seconds
(Staal et al. 2004)	15 minutes
(Mendonca and Campilho 2006)	2.5 minutes
(Soares et al. 2006)	3 minutes
(Lam et al. 2010)	13 minutes
(Marin et al. 2011)	1.5 minutes
(Fraz et al. 2012)	2 minutes

Table 4. Performance results of the proposed unsupervised *B*-COSFIRE filter approach on the concerned data sets compared to other methods from the state of the art.

	Method	DRIVE				STARE			
		Se	Sp	AUC	Acc	Se	Sp	AUC	Acc
Unsupervised	<b><i>B</i>-COSFIRE</b>	<b>0.7655</b>	<b>0.9704</b>	<b>0.9614</b>	<b>0.9442</b>	<b>0.7716</b>	<b>0.9701</b>	<b>0.9563</b>	<b>0.9497</b>
	(Mendonca and Campilho 2006)	0.7344	0.9764	—	0.9463	0.6996	0.9730	—	0.9479
	(Martinez-Pérez et al. 2007)	0.7246	0.9655	—	0.9344	0.7506	0.9569	—	0.9410
	(Al-Rawi et al. 2007)	—	—	0.9435	0.9535	—	—	0.9467	0.9090
	(Ricci and Perfetti 2007)	—	—	0.9558	0.9563	—	—	0.9602	0.9584
	(Lam et al. 2010)	—	—	0.9614	0.9472	—	—	0.9739	0.9567
Supervised	(Niemeijer et al. 2004)	—	—	0.9294	0.9416	—	—	—	—
	(Staal et al. 2004)	—	—	0.9520	0.9441	—	—	0.9614	0.9516
	(Soares et al. 2006)	0.7332	0.9782	0.9614	0.9466	0.7207	0.9747	0.9671	0.9480
	(Ricci and Perfetti 2007)	—	—	0.9633	0.9595	—	—	0.9680	0.9646
	(Marin et al. 2011)	0.7067	0.9801	0.9588	0.9452	0.6944	0.9819	0.9769	0.9526
	(Fraz et al. 2012)	0.7406	0.9807	0.9747	0.9480	0.7548	0.9763	0.9768	0.9534
CHASE_DB1									
	Method	Se	Sp	AUC	Acc	Se	Sp	AUC	Acc
Unsup.	<b><i>B</i>-COSFIRE</b>	<b>0.7585</b>	<b>0.9587</b>	<b>0.9487</b>	<b>0.9387</b>	<b>0.7511</b>	<b>0.9745</b>	<b>0.9510</b>	<b>0.9644</b>
	(Odstrcilik et al. 2013)	—	—	—	—	0.7740	0.9669	0.9493	0.9668
Sup.	(Fraz et al. 2012)	0.7224	0.9711	0.9712	0.9469	—	—	—	—

## 5 CONCLUSIONS

The results that we achieve on DRIVE, STARE, CHASE\_DB1 and HRF data sets demonstrate the high effectiveness of the proposed method. *B*-COSFIRE filters are versatile as they can be automatically configured to be selective for any given vessel-like pattern, including bifurcations and crossovers. We have made the Matlab implementation available online<sup>3</sup>. The processing of a *B*-COSFIRE filter is also very efficient. In fact, the proposed method is the most time-efficient algorithm for vessels delineation in retinal fundus images published so far.

## REFERENCES

- Al-Diri, B., A. Hunter, & D. Steel (2009, Sept). An active contour model for segmenting and measuring retinal vessels. *Medical Imaging, IEEE Transactions on* 28(9), 1488–1497.
- Al-Rawi, M., M. Qutaishat, & M. Arrar (2007). An improved matched filter for blood vessel detection of digital retinal images. *Computer in biology and medicine* 37(2), 262–267.
- Azzopardi, G. & N. Petkov (2012). A CORF computational model of a simple cell that relies on LGN input outperforms the Gabor function model. *Biol Cybern* 106(3), 177–189.
- Azzopardi, G. & N. Petkov (2013). Trainable COSFIRE filters for keypoint detection and pattern recognition. *IEEE Trans Pattern Anal Mach Intell* 35, 490–503.
- Azzopardi, G., A. Rodriguez-Sánchez, J. Piater, & N. Petkov (2014, 07). A push-pull corf model of a simple cell with antiphase inhibition improves snr and contour detection. *PLoS ONE* 9(7), e98424.
- Azzopardi, G., N. Strisciuglio, M. Vento, & N. Petkov (2014). Trainable cosfire filters for vessel delineation with application to retinal images. *Medical Image Analysis* 19(1), 4657.
- Cinsdikici, M.G. & D. Aydin (2009). Detection of blood vessels in ophthalmoscope images using mfl/ant (matched filter/ant colony) algorithm. *Computer Methods and Programs in Biomedicine*, 85–95.
- Fraz, M., P. Remagnino, A. Hoppe, B. Uyyanonvara, A. Rudnicka, C. Owen, & S. Barman (2012). An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Trans Biomed Eng* 59(9), 2538–2548.
- Hoover, A., V. Kouznetsova, & M. Goldbaum (2000). Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Trans Med Imaging* 19(3), 203–210.
- Jiang, X. & D. Mojon (2003). Adaptive local thresholding by verification-based multithreshold probing with application to vessel detection in retinal images. *IEEE Trans. Pattern Anal. Mach. Intell.* 25(1), 131–137.

<sup>3</sup><http://www.mathworks.com/matlabcentral/fileexchange/49172>.

- Lam, B., Y. Gao, & A.-C. Liew (2010). General retinal vessel segmentation using regularization-based multiconcavity modeling. *IEEE Trans on Med Imaging* 29(7), 1369–1381.
- Marin, D., A. Aquino, M. Emilio Gegundez-Arias, & J. Manuel Bravo (2011). A New Supervised Method for Blood Vessel Segmentation in Retinal Images by Using Gray-Level and Moment Invariants-Based Features. *IEEE Trans Med Imaging* 30(1), 146–158.
- Martinez-Pérez, M.E., A.D. Hughes, S.A. Thom, A.A. Bharath, & K. H. Parker (2007). Segmentation of blood vessels from red-free and fluorescein retinal images. *Medical Image Analysis* 11(1), 47–61.
- Mendonca, A.M. & A. Campilho (2006). Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Trans Med Imaging* 25(9), 1200–1213.
- Niemeijer, M., J. Staal, B. van Ginneken, M. Loog, & M. Abramoff (2004). Comparative study of retinal vessel segmentation methods on a new publicly available database. In *Proc. of the SPIE—The International Society for Optical Engineering*, pp. 648–56.
- Odstrcilik, J., R. Kolar, A. Budai, J. Hornegger, J. Jan, J. Gazarek, T. Kubena, P. Cernosek, O. Svoboda, & E. Angelopoulou (2013, Jun). Retinal vessel segmentation by improved matched filtering: evaluation on a new high resolution fundus image database. *IET Image Processing* 7(4), 373–383.
- Owen, C.G., A.R. Rudnicka, R. Mullen, S.A. Barman, D. Monekosso, P.H. Whincup, J. Ng, & C. Paterson (2009). Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (caiar) program. *Invest Ophthalmol Vis Sci* 50(5), 2004–10.
- Ricci, E. & R. Perfetti (2007). Retinal blood vessel segmentation using line operators and support vector classification. *IEEE Trans Med Imaging* 26(10), 1357–1365.
- Soares, J.V.B., J.J.G. Leandro, R.M. Cesar, Jr., H.F. Jelinek, & M.J. Cree (2006). Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. *IEEE Trans Med Imaging* 25(9), 1214–1222.
- Staal, J., M. Abramoff, M. Niemeijer, M. Viergever, & B. van Ginneken (2004). Ridge-based vessel segmentation in color images of the retina. *IEEE Trans Med Imaging* 23(4), 501–509.

This page intentionally left blank

# A registration algorithm for microscopy images of the capillary bed

H. Rieiro

*Universidade de Vigo, Vigo, Spain*

*Barrow Neurological Institute, Phoenix, AZ, USA*

J.L. Alba Castro

*Universidade de Vigo, Vigo, Spain*

S. Martinez-Conde & S.L. Macknik

*SUNY Downstate Medical Center, Brooklyn, NY, USA*

*Barrow Neurological Institute, Phoenix, AZ, USA*

**ABSTRACT:** We propose a new feature-based registration algorithm to corregister microscopy images of brain microvasculature. This algorithm is a three-dimensional adaptation of SIFT, modified here to further filter keypoint selection to highly distinctive regions of blood vessels, and to add three-dimensional rotational invariance. We show the algorithm is robust against rotations and additive noise in two-photon microscopy vasculature images. Future work will explore corregistration across different imaging modalities based on this method.

## 1 INTRODUCTION

Neuroimaging techniques are powerful tools in neuroscientific research. They are not only used to study neural disease and its models, but also as tools to research normal behavior of the brain, both from a structural point of view, such as studies of homeostatic regulation, and also from a functional perspective, since blood flow increases and decreases are directly correlated to neural activity (Kleinfeld et al. 1998; Chaigneau et al. 2007). The most well-known functional imaging technique, functional Magnetic Resonance Imaging (fMRI), uses blood oxygenation and volume as a proxy to study functional activity over large scale brain areas (Hyder et al. 2001).

There are a number of different imaging techniques that can be employed to study blood flow, each of them with a set of advantages and disadvantages. For example, fMRI can be used to study full brain activation with limited spatial and temporal resolution, on the order of 1 mm<sup>3</sup> and 1 second respectively (Logothetis 2003). Microscopy techniques, such as two-photon microscopy and confocal microscopy, work at much higher spatial and temporal resolutions (Denk et al. 1990). This makes them excellent tools to study blood flow regulation and microvascular structure at the microscopic level.

Several types of studies of microscopic blood flow require alignment of images, for example

those requiring the acquisition of data at distant points in time (Holtmaat et al. 2009). We propose a feature-based registration algorithm that will provide an automatic and precise alignment of images of the capillary bed acquired through microscopy techniques.

## 2 ALGORITHM DESCRIPTION

An image contrast registration algorithm functions by maximizing a similarity metric such as mutual information (Pluim et al. 2003). A full affine transformation model in three dimensions would thus require minimization in a twelve-dimensional space (rotations, scale, translation and shear along the three XYZ axes), which is computationally very expensive. The usual way to mitigate this problem is to use a multiresolution algorithm that carries out the maximization across increasing levels of detail, potentially using simplified models at low resolution stages (Jenkinson & Smith 2001). When we tried to apply this approach to images of the capillary bed, we found that, due to the small size of the capillaries and the high density of the microvascular network, it was not possible to downsample the images by a high enough factor to improve the method performance while keeping the results satisfactory.

Instead, we developed an feature-based algorithm for registration of capillary images, based

on a previous implementation of SIFT (Scale Invariant Feature Transform) (Lowe 2004). SIFT, described originally to work in two-dimensional images, defines points in the image (keypoints) that are invariant to image scaling, and feature descriptors that allow to obtain rotational invariance. We first tested a previous adaptation of SIFT for three-dimensional images (Cheung & Hamarneh 2009), but this implementation showed poor performance for our specific problem, specifically related to its rotational invariance and its capacity to extract robust features, so we enhanced it with application-specific knowledge.

### 2.1 Image preprocessing

Two-photon images of brain microvasculature present two sources of noise relevant to the registration algorithm. One is electrical noise in the sensors, with the main effect of causing background noise, which can have amplitude similar to that of very small vessels. The second noise source, of biological origin, is due to small changes in serum concentration over time, which can produce slight spatial mislocalizations of the features. These sources of noise, combined with the reduction of vessel brightness (signal) when imaging on deeper layers of tissue, can make small vessels hard to detect based on their contrast alone.

To diminish the effect of these noise sources we first filtered the images using a  $3 \times 3 \times 3$  voxel median filter, and then transformed the filtered image using a multilocal creaseness measure to detect ridges and valleys (López et al. 1999). Since the objects of interest were blood vessels, which presented as bright structures over dark background, we further removed negative creaseness values from the image, which corresponded to image valleys. The result was an image with reduced background noise and enhanced blood vessels contrast, in exchange for a small blurring (Fig. 1).

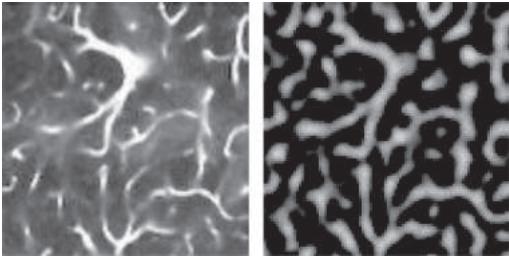


Figure 1. Input image and result of applying the creaseness transformation. The creaseness transformation has a two-fold effect, blurring the image and removing most of the background noise.

### 2.2 Keypoint localization

Keypoints over the preprocessed image were initially localized using a method like the one described by (Cheung & Hamarneh 2007), where maxima and minima of the multiscale image were found over 3 different Difference of Gaussians (DoG) scales (with  $\sigma = 1.5$ ), and three different image magnifications (0.5, 1, and 2). The keypoints found were further filtered using three application-specific criteria. We first removed those keypoints with negative amplitude in the DoG images, to remove keypoints that would fall outside the edge of a vessel, which therefore kept only keypoints falling within a vessel. Second, we set up a minimum intensity threshold to remove spurious maxima (0.1). Finally, we used a vesselness measure (Frangi et al. 1998) to keep only the keypoints that were localized in areas of the vessel that had high curvature, or, ideally, where the capillaries branched. The vesselness criteria used was:

$$V = 1 - e^{-\frac{R_a}{2\alpha}} \quad (1)$$

where  $R_a$  is the ratio between the two eigenvalues with the higher modulus of the Hessian matrix of the image at the keypoint, and  $\alpha$  is a constant set to 0.3. Figure 2 shows the result of the keypoint localization process.

### 2.3 Modified descriptor

After localizing the keypoints we built the feature descriptors as follows. The first step was to determine the keypoint orientation by building a histogram of orientations of the gradients (azimuth

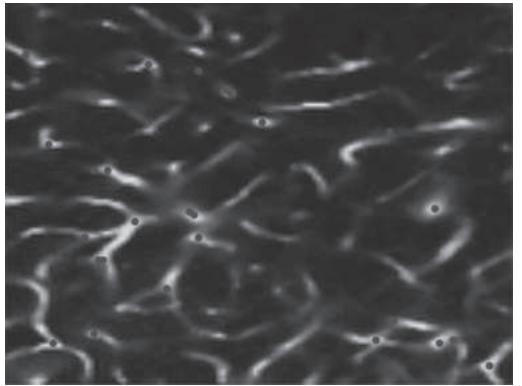


Figure 2. Example of keypoint localization in a two-photon microscopy section. Grey dots over imposed on blood vessels mark the detected keypoints, which tend to fall in the points of vessel branching.

and elevation) in a region of  $16 \times 16 \times 16$  voxels surrounding the keypoint. The histograms were weighed by the magnitude of their gradients and by the solid angle of the bin, to avoid giving too much weight to the low elevation bins (Scovanner et al. 2007). As in the original description of SIFT, we allow each keypoint to have several features if any orientation has an amplitude over 80% of the maximum value.

While the original description of SIFT included 2-D rotational invariance, the base implementation we used for our three dimensional images (Cheung & Hamarneh 2009) did not provide such a feature. Whereas rotational invariance within two dimensional images is relatively easy to implement due to the fact that there is only one possible rotation axis, when working with three dimensional images this rotation axis can be any three dimensional vector. Three-dimensional rotations lack other desirable properties such as commutability or the capability of being easily decomposed in orthogonal components. We opted to add rotational invariance to our features by adapting the original, two-dimensional mechanisms defined for the original SIFT descriptor to the 3-D case.

To achieve this, we rotated the feature gradients surrounding the keypoint so as the assigned feature orientation was always aligned with the positive direction of the X axis. The first step for this was to rotate by the opposite angle of azimuth (rotation around Z axis), and elevation (around Y axis) of the feature. Since this mechanism only corrects rotation around two angles, an uncompensated rotation around the X axis remains for all gradients that fail to precisely correspond to the main orientation, causing different rotated versions of the same feature, such as those found in two different images of the same object, to have different orientation histograms, reducing the registration robustness to rotations. To compensate this problem, we projected the realigned gradients on the YZ plane and calculated the principal components of the resulting cloud of points, to estimate the orientation of the main component with regards to the positive direction of the Y axis. We then used an additional rotation around X by the opposite angle to align the main component to the Y axis. We compensated for mismatched signs on the main component vector between different versions of the same feature by creating two different descriptors, one where the calculated main component is aligned with the positive direction of the Y axis, and another where it is aligned to the negative direction.

Finally, for each feature the  $16 \times 16 \times 16$  region surrounding the keypoint was divided into 64  $4 \times 4 \times 4$  subregions, and  $6 \times 6$  gradient orientation histograms, weighed by gradient magnitude, were built for each subregion, with the particularity

that the azimuth information was ignored in bins with maximal and minimal elevation ( $\pm\pi$ ). This is because at the extreme values of elevation the azimuth component becomes highly susceptible to noise, and therefore unreliable to use for matching.

## 2.4 Keypoint matching and model estimation

We compared the keypoints across two images ( $I_1$  and  $I_2$ ) using a histogram intersection metric (Sinha & Shukla 2013). The search for the best neighbor (the keypoint in the second image that maximizes the histogram intersection) was performed using a k-d tree based approximate method (Muja & Lowe 2009). To remove poor matches, we enforced two conditions:

1. We required the metric value for the best neighbor to be at least 25% higher than the value for the second best (Lowe 2004).
2. Due to the set of keypoints between the images being aligned is not identical, it is possible that a keypoint from  $I_1$  is being matched to a keypoint from  $I_2$  that does not exist in  $I_1$ . To reduce this problem, we removed cases in which the match was not reciprocal (i.e. the nearest neighbor for feature  $f_A$  is  $f_B$ , but the nearest neighbor for  $f_B$  is a different feature  $f_{A'}$ ).

From the matches of keypoints, we can estimate the transformation between images using a least-squares method, where an affine transformation  $A$  is estimated from the matched points  $P$  (from  $I_1$ ) and  $Q$  (from  $I_2$ ), corresponding to the first and second images, as:

$$\hat{A} = PQ(QQ')^{-1} \quad (2)$$

Since the results of the matching process are likely to produce a significant number of outliers, we used a RANSAC (Yaniv & Joskowicz 2005; Fischler & Bolles 1981) model estimation technique to minimize their effect in the estimated matrix.

## 3 EXPERIMENTAL RESULTS

We tested the developed algorithm using two-photon images acquired from the mouse olfactory bulb.

### 3.1 Animal preparation

C57B6 mice ( $n = 7$ ) were anesthetized using a mixture of ketamine and xylazine. Following surgical procedures similar to those described in (Kleinfeld et al. 1998), a small window ( $5 \times 5$  mm) was drilled over the left or right olfactory bulb. The blood serum was labeled through tail vein injection of

2000 kDa Tetramethylrhodamine, and two-photon imaging of the microvasculature was acquired using a laser wavelength of 857 nm, with resolution of  $2 \times 2 \times 2 \mu\text{m}$ .

### 3.2 Experimental procedure

To test the performance of our corregistration method when using biological images, and specifically to its robustness to changes in orientation, we compared each image to rotated versions of itself around an arbitrary axis. We used the normalized mutual information as a measure of registration quality, since it is a measure commonly used in other registration problems (Pluim et al. 2003).

We set an intensity threshold of 0.1 for keypoints in the DoG space. Keypoints where the vesselness value was below 0.5 were removed to avoid finding keypoints in straight vessel segments. The estimation of the model was performed using a RANSAC algorithm where the probability of outliers in the final estimation was set to be less than 0.05, and with matches classified as outliers when the Euclidean distance between the original and transformed keypoint was above 5  $\mu\text{m}$ .

### 3.3 Results

Figure 3 shows the average values of the mutual information metric between images when registering

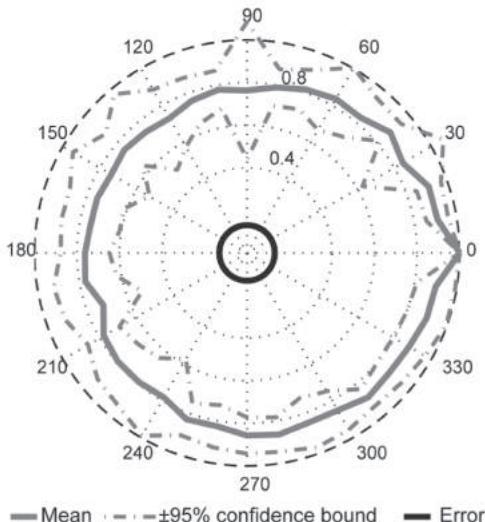


Figure 3. Results of registering images to rotated versions of themselves (gray) and unrelated images (black). Dashed lines show the 95% confidence interval for the registration metric values. Values of the mutual information metric are normalized to their maximum, which coincides with zero rotation (registration with itself).

images to rotated versions of themselves. The rotations were performed around an arbitrary rotation axis, with rotations around other axes yielding similar results. The values of the metric for this figure are normalized for each image such that a value of one (maximum radius in the figure) represents the maximum mutual information possible for each (i.e. the image entropy), to make the average and confidence intervals more meaningful when averaging over images with disparate contrasts. The method shows a uniform behavior over the range of three-dimensional rotations. In addition, registration of unrelated images provides much lower metric values, as represented by the darker line in Figure 3, which are used as the baseline for a bad registration.

To test the algorithm robustness against noise, we tested a 30 degree rotation while adding additive gaussian noise of varying energy (standard deviation between 0.02 and 0.1) to both the original and rotated images. The images themselves, without any added noise on top of that coming from the acquisition itself, have a standard deviation in the 0.04–0.2 range, so the noise levels used have significant magnitude. Figure 4 shows the values of the non-normalized mutual information metric when performing the registration under this conditions, with each line representing a single image used. This representation allows us to examine the robustness of the algorithm to noise as a function of the quality of the non-noisy registration. We can see that the algorithm is robust to the noise levels tested, with values of the registration metric

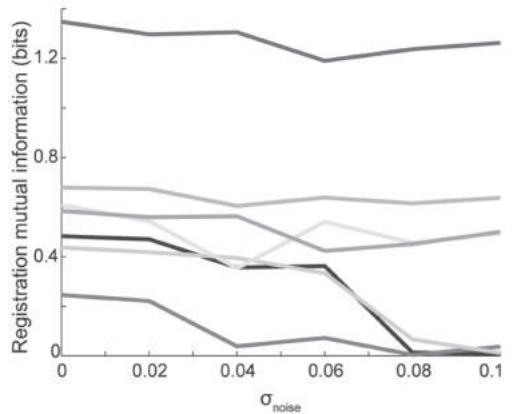


Figure 4. Registration results with additive gaussian noise over a 30 degree rotation. The value of the mutual information between the original and transformed image is shown for different standard deviations for the noise. Each line represents the result for a different image, each from a different animal. The algorithm is robust to noise in images having good contrast.

holding steady for noise in the lower and middle range except for the lowest contrast image (bottom line, Fig. 4).

## 4 CONCLUSIONS

We have developed an algorithm that provides the capability of automatically align images of blood vessels acquired through microscopy techniques. The algorithm is robust to changes in orientation over the whole 3-D range, and shows good behavior against the addition of significant levels of gaussian noise.

Future research lines will test the robustness of the method on images taken at disparate points in time, its behavior and changes in preprocessing needed to apply it to images acquired through different modalities such as confocal microscopy, and finally investigate the possibility of registering the images acquired across those different modalities.

## REFERENCES

- Chaigneau, E. et al., 2007. The relationship between blood flow and neuronal activity in the rodent olfactory bulb. *The Journal of Neuroscience: the official journal of the Society for Neuroscience*, 27(24), pp.6452–60.
- Cheung, W. & Hamarneh, G., 2007. n-SIFT: n-dimensional scale invariant feature transform for matching medical images. In *Proc. 4th IEEE Int. Symp. Biomedical Imaging: Biomedical Imaging: From Nano to Macro*. pp.720–723.
- Cheung, W. & Hamarneh, G., 2009. n-SIFT: n-dimensional scale invariant feature transform. *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society*, 18(9), pp.2012–21.
- Denk, W., Strickler, J.H. & Webb, W.W., 1990. Two-photon laser scanning fluorescence microscopy. *Science*, 248(4951), pp.73–76.
- Fischler, M. & Bolles, R., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), pp.381–395.
- Frangi, A.F. et al., 1998. Multiscale vessel enhancement filtering. *Medical Image Computing and Computer-Assisted Intervention—MICCAI'98. Lecture Notes in Computer Science*, vol 1496, 1496, pp.130–137.
- Holmaat, A. et al., 2009. Long-term, high-resolution imaging in the mouse neocortex through a chronic cranial window,
- Hyder, F. et al., 2001. Quantitative functional imaging of the brain: towards mapping neuronal activity by BOLD fMRI. *NMR in biomedicine*, 14(7–8), pp.413–31.
- Jenkinson, M. & Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Medical image analysis*, 5(2), pp.143–56.
- Kleinfeld, D. et al., 1998. Fluctuations and stimulus-induced changes in blood flow observed in individual capillaries in layers 2 through 4 of rat neocortex. *Proceedings of the National Academy of Sciences of the United States of America*, 95(26), pp.15741–15746.
- Logothetis, N.K., 2003. The underpinnings of the BOLD functional magnetic resonance imaging signal. *The Journal of neuroscience: the official journal of the Society for Neuroscience*, 23(10), pp.3963–71.
- López, A.M. et al., 1999. Evaluation of methods for ridge and valley detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4), pp.327–335.
- Lowe, D.G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), pp.91–110.
- Muja, M. & Lowe, D.G., 2009. Flann, fast library for approximate nearest neighbors. In *International Conference on Computer Vision Theory and Applications (VISAPP'09)*.
- Pluim, J.P.W., Maintz, J.B.A. & Viergever, M. a., 2003. Mutual-information-based registration of medical images: a survey. *IEEE transactions on medical imaging*, 22(8), pp.986–1004.
- Scovanner, P., Ali, S. & Shah, M., 2007. A 3-dimensional sift descriptor and its application to action recognition. *15th international conference on Multimedia*, (c), pp.357–360.
- Sinha, A.K. & Shukla, K.K., 2013. A Study of Distance Metrics in Histogram Based Image Retrieval. *International Journal of Computers & Technology*, 4(3), pp.821–830.
- Yaniv, Z. & Joskowicz, L., 2005. Precise robot-assisted guide positioning for distal locking of intramedullary nails. *IEEE transactions on medical imaging*, 24(5), pp.624–635.

This page intentionally left blank

# Bilateral filtering based biomedical image colorization

A. Popowicz & B. Smolka

*Institute of Automatic Control, Silesian University of Technology, Akademicka, Gliwice, Poland*

**ABSTRACT:** Most of biomedical images are monochromatic and thus their assessment may be significantly improved by means of the colorization techniques. Such colorized versions of images not only benefit from more pleasing appearance, but also some fine details are easily noticeable, what may substantially improve the diagnosis. In this article we present a novel method of colorization, which is based on a combination of distance transformation and modified bilateral filtering approach. Tunable parameters are introduced to allow for precise adjustment of final colorization outcomes. Our proposed algorithm enables a simple and efficient region highlighting and interactive segmentation of biomedical images. It may be also employed for the educational and presentation purposes. We show several examples of applications of our algorithm using various biomedical images obtained by ultrasonography, transmission electron microscopy, computer tomography, magnetic resonance and infrared imaging.

## 1 INTRODUCTION

Biomedical images obtained using the most widespread imaging techniques are usually presented in a gray scale form. As examples, we may enumerate the techniques such as the Computer Tomography (CT), the Ultrasonography (USG), the Magnetic Resonance (MRI), the Transmission Electron Microscopy (TEM) or the Infrared Imaging (IR) (Bankman 2000, Lin, Olivo, & Kung 2009, Walus M. 2015). Some new techniques emerge, like the retinal optical coherence tomography (Acharya U, Yun, Ng, Yu, & Suri 2008) or elastography (Ophir, Cspedes, Ponnekanti, Yazdi, & Li 1991), which also provide gray scale images (or video sequences) as the imaging output.

It is worth to notice, that the images presented in the gray scales are mostly less perceptible, than the ones with colors included. Therefore, the colorization techniques are important tools utilized during the image assessment phase. They may also improve the accuracy of the diagnosis.

We can divide the colorization techniques into two categories: automatic and semiautomatic (interactive). Although the methods from the first category allow for colorization without any user effort, they usually provide less appealing results. Moreover, the physicians prefer to supervise any image processing procedures, thus the semiautomatic methods are usually the solution of choice.

The pioneer colorization algorithm (Welsh, Ashikhmin, & Mueller 2002) was based on the color transfer idea presented in (Reinhard, Ashikhmin, Gooch, & Shirley 2001). In this original method, the colors were transferred from the

color reference to the gray scale target image utilizing the analysis of intensity statistics in distinct image patches. This approach was also adopted for the colorization of biomedical images in (Zhao, Wang, Jin, & Shi 2007).

Another automatic colorization algorithms can be found in (Zhen, Yan, & Lizhuang 2012) and in (Yan & Kankanhalli 2003). The authors of the first paper suggest the pattern continuity approach in contrast to traditional luminance-based schemes. The second publication presents the algorithm for infrared image colorization, wherein the reference image is obtained from the flash-assisted frame.

One of the most prominent semiautomatic colorization method was presented in (Levin, Lischinski, & Weiss 2004a, Levin, Lischinski, & Weiss 2004b). The authors utilize the minimization of color difference between adjacent pixels and assume that the neighboring pixels with similar intensities should also be given the same colors. Their proposed minimization formula may be easily solved using typical methods, like the Least Square algorithm.

In several papers (Lagodzinski & Smolka 2014, Yatziv & Sapiro 2006, Popowicz & Smolka 2014) the authors investigate the digital paths between adjacent image pixels taking into account the encountered intensity gradients. They obtain the degree of the membership of pixels to each of inserted scribbles and estimate the unknown color using the weighted averaged of all scribbled colors.

In this article we present a novel method of semiautomatic image colorization based on the algorithms presented in (Lagodzinski & Smolka

2014, Popowicz & Smolka 2014). We combine the distance transformation (Rosenfeld & Pfaltz 1968) and bilateral filtering (Tomasi & Manduchi 1998) to achieve high quality colorization outcomes. In the next Sections of this paper we show the basics steps of our concept, present the role of tunable parameters and exhibit exemplary results of application of the algorithm using various biomedical images.

## 2 ALGORITHM

The presented colorization method belongs to the family of semiautomatic approaches, in which the user indicates the color hints of so-called scribbles, which provide the information about required color in a given region of the image. Our aim is to assign an unknown chrominance value to each image pixel using the colors from the nearest scribbles. The weights utilized in such color blending are obtained by the magnitude of the intensity gradients encountered between the pixel and a given scribble. Therefore, the weights correspond with the membership measure of pixels to the inserted scribbles. The main steps of our algorithm are presented schematically in Figure 1.

In the first step, we calculate the topographic distances on the image domain between scribbles of the same color and any other pixel within the image. To do so, we employ the distance transformations (Borgefors 1986) and the double-scan algorithm (Rosenfeld & Pfaltz 1968). In this algorithm we utilize two processing masks, which are used in respectively first and second scan (see Fig. 2). The scans are performed from the left to right, up to bottom (first scan) and in the opposite direction—right to left, bottom to top (second scan). For details of scans directions see Figure 3.

In the double scan algorithm, initially all the pixels are given infinite distance value except for the pixels belonging to the scribble, which are given 0 (see Fig. 4b). On such initialized distance matrix, the double scan algorithm slides twice, performing following distance replacement in each step:

$$d_0 = \max\{d_1, d_2, d_3, d_4\} \quad (1)$$

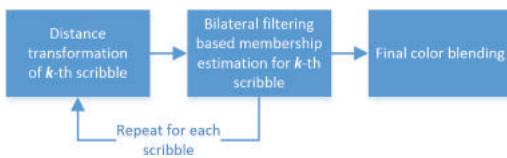


Figure 1. Three-step algorithm schema.

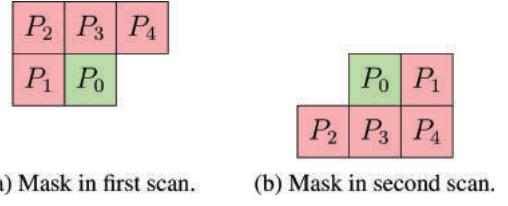


Figure 2. The masks utilized in double-scan algorithm.

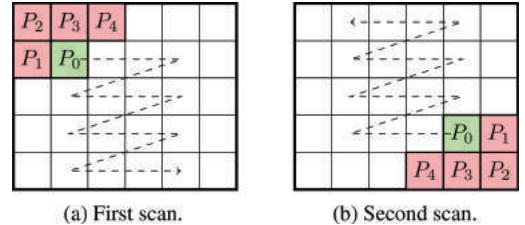


Figure 3. Direction of mask sliding in double-scan algorithm.

where  $d_i$  denotes the distance for pixel  $P_i$  in corresponding mask (see Fig. 2). Eventually, after both scans, the estimations of Euclidean distance are obtained (see the result in Fig. 4d). The procedure is repeated for each of inserted scribbles.

After the distance calculations, we initialize the membership arrays for each scribbled color. In such an array all the pixels are initially given 0 membership, except for the scribbled pixels, which receive 1 (full membership). We also make use of the previously calculated metric distances to obtain the order of pixels, which will be processed in the bilateral filtering membership estimation, explained below. The initial membership array for the scribble from Figure 4 with the processing order depicted by the filled gray levels, is presented in Figure 5.

In the next part of the algorithm we perform a modified version of classic bilateral filtering. The bilateral filtering is nonlinear image smoothing algorithm developed for suppression of Gaussian noise in both gray scale and color images (Tomasi & Manduchi 1998). The idea is based on the non-linear, adaptive filtering, in which the local window slides the image matrix, replacing the intensity of its central pixel with the weighted average of intensities of other pixels in window:

$$I_0 = \sum_{k=1}^{N^2} I_k w_k \quad (2)$$

where  $I_0$  is the filtered intensity of central pixel in  $N \times N$  local window,  $I_k$  and  $w_k$  denote respectively the

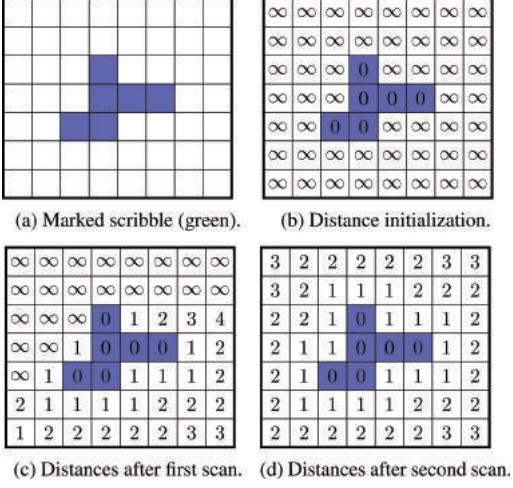


Figure 4. An exemplary application of distance transformations.

0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	0	0
0	0	0	1	1	1	0	0	0
0	0	1	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Figure 5. Initial membership array with order of pixel processing indicated by the gray scale levels.

intensity and the weight of corresponding  $k$ -th pixel in the window. In the bilateral filtering the weights are inversely proportional to the intensity differences and the metric distances to a central pixel:

$$I_0 = \frac{1}{Z} \sum_{k=1}^{N^2} I_k \exp\left(-\frac{(I_0 - I_k)^2}{2\sigma_I^2}\right) \exp\left(-\frac{d_k}{2\sigma_d^2}\right), \quad (3)$$

where  $Z$  is normalizing factor,  $d_k$  is the metric distance between  $k$ -th and central window pixel,  $\sigma_I$  and  $\sigma_d$  tune the influence of respectively, the intensity difference and the pixels distance on final smoothing. The bilateral filtering, in contrast to

linear filtering algorithms, is capable to preserve the sharp edges in filtered outcomes.

We utilize the presented filtering idea for the calculation of membership of a given pixel to scribbled color. However, instead of using the weighted average, we replace the membership by the maximum over the weighted membership values in a local filtering window. Similarly to the bilateral filtering approach, the weights are obtained from the intensity difference between the pixels. In each step of sliding window, the corresponding membership value replacement is performed:

$$M_0 = \max_{k=1, \dots, N^2} \left\{ M_k \exp\left(-\frac{(I_0 - I_k)^2}{2\sigma^2}\right) \right\}, \quad (4)$$

where  $M_0$  is the new estimation of membership of central pixel in a window,  $M_k (k > 0)$  is the membership of  $k$ -th pixel in a window,  $I_k$  stands for the intensity of  $k$ -th pixel and  $\sigma$  is the parameter responsible for the strength of membership decrease.

In contrast to classic image filtering methods, in our algorithm the window slides through the pixels in the order defined by previously calculated metric distance (as shown in Fig. 4d and visualized by shades of gray in Fig. 5). Therefore, our bilateral filtering window replaces the membership values starting from pixels neighboring with a scribble and moves gradually toward the image borders.

This bilateral filtering based procedure assigns higher membership values to pixels, which are connected with scribble by a digital, pixel-to-pixel path, within which the intensity changes smoothly. The image regions, which are separated from a scribble by high intensity gradients (like the boundaries of an image region), receive very low membership coefficients. Some membership arrays for four scribbled colors in an exemplary gray scale microscopy image of blood cells are presented in Figure 6.

The final step of the algorithm is the assignment of chrominance to each pixel. We perform the following color blending using weights taken directly from the membership arrays:

$$C_P = \frac{\sum_{s \in S} C_s M_s}{\sum_{s \in S} M_s}, \quad (5)$$

where  $C_P$  is an assigned chrominance of pixel  $P$ ,  $C_s$  is the chrominance of scribbled color  $s$  from the set  $S$  of all scribbled colors,  $M_s$  is the membership of pixel  $P$  to scribbled color  $s$ . After such a color blending procedure, each pixel receives the color in accordance to its membership value to the corresponding color scribble.

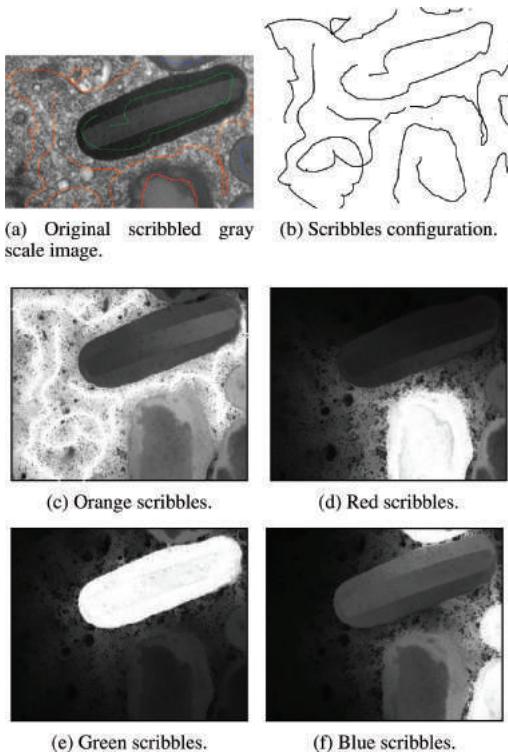


Figure 6. Exemplary membership arrays. The gray scale in presented arrays was chosen so that white and black level correspond respectively to full and no membership ( $M = 1$  and  $M = 0$ ).

### 3 TUNABLE PARAMETERS

Our method enables fine tuning of colorization outcomes by modifying the size of the window ( $N$ ) and the parameter  $\sigma$  in the bilateral filtering based membership replacement (4). While the first parameter influences the algorithm robustness against the noise, the second one modifies the strength of final color blending.

The size of the window ( $N$ ) allows to set the size of small details, which will be neglected during the memberships assignment. For large windows, the algorithm will be robust against the local noise and will be able to pass over thin lines that are associated with noise rather than with the scene objects. On the other hand, reducing the window size to the minimum (i.e.  $N = 3$ ) may be desirable if one wishes to colorize some tiny objects. The membership arrays and colorization results of ultrasound image of breast mass for different  $N$  parameters are depicted in Figure 7. As it can be seen, the larger mask ( $7 \times 7$ ) enabled the accurate colorization of whole region of breast mass starting from very simple scribbles.

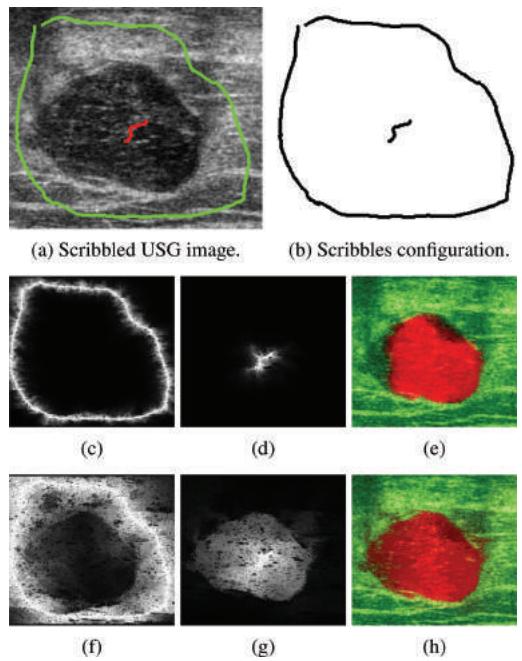


Figure 7. Examples of utilization of various mask sizes  $N$ : (a) is the original USG image with indicated color scribbles depicted also separately in (b); (c), (d) and (f), (g) are the membership maps, respectively for  $3 \times 3$  and  $7 \times 7$  masks; (e) and (h) are the corresponding colorization outcomes.

The second parameter  $\sigma$  is responsible for the degree of final color blending. The higher the  $\sigma$ , the more blurred is the final colorization outcome. For very small  $\sigma$  values, the edges between the regions of different colors become clearly visible. For higher  $\sigma$  the blending is stronger and the colorization result becomes in general more pleasing for human perception. Two extreme values of  $\sigma$  were used for the examples depicted in Figure 8.

We present an example of image colorization quality as a function of both tunable parameters. To evaluate the quality of the colorization, we utilized the known color image reference. Such test image, depicted in Figure 9, was converted to gray scale (by leaving only the luminance channel in YCbCr color space), scribbled and colorized by our algorithm. The outcomes were compared with the original image employing Peak-Signal-to-Noise (PSNR) measure:

$$MSE = \frac{1}{3XY} \sum_{c \in \{R,G,B\}} \sum_{i=1}^X \sum_{j=1}^Y [I_r^c(i,j) - I_s^c(i,j)]^2, \quad (6)$$

$$PSNR = 10 \log \left( \frac{255^2}{MSE} \right), \quad (7)$$

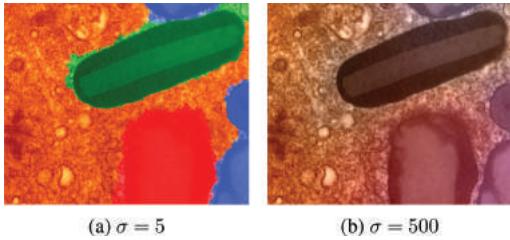


Figure 8. Colorization results for various  $\sigma$  parameter values. The scribbled original image can be seen in Figure 6a.

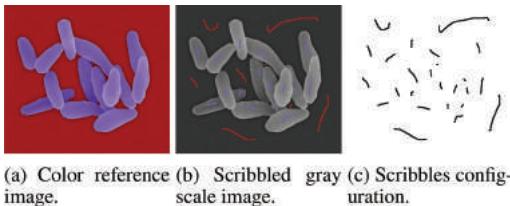


Figure 9. Original and scribbled gray scale test image samples utilized for colorization quality evaluation.

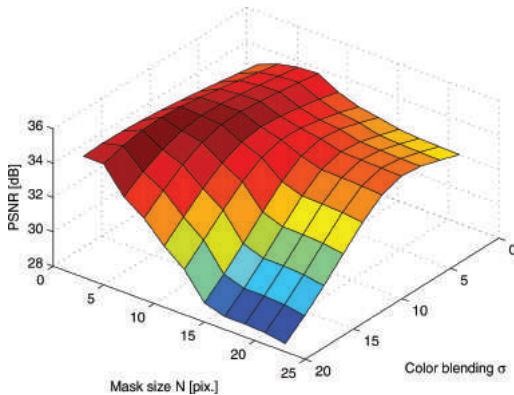


Figure 10. Colorization quality as a function of tunable parameters when colorizing the image Figure 9b.

where  $X, Y$  are width and height of the image,  $c$  is color channel (R, G or B),  $I_r^c(i,j)$  is the intensity of pixel at position  $(i,j)$  in reference color image (color channel  $c$ ),  $I_s^c(i,j)$  is the same as above but for the colorized version of the image.

In our tests, we used  $\sigma$  range from 2 to 20 and window sizes respectively 3, 5, ..., 23. In such parameters space we managed to find an optimal solution, which, as it can be observed in Figure 10, was obtained for  $N = 5$  and  $\sigma = 14$ . Extensive experiments revealed that for most of typical, properly sampled (i.e. no over- and under-sampling)

biomedical images, the optimal parameters should be found within the same range.

#### 4 COMPARISON WITH OTHER COLORIZATION TECHNIQUES

We compared the quality of colorization of our method with the results of the state-of-the-art algorithms. For this purpose we either implemented or utilized the available code for the following methods: *Colorization Using Optimization* (Levin, Lischinski, & Weiss 2004a, Levin, Lischinski, & Weiss 2004b) (CUO), *Chrominance Blending Colorization* (CBC) (Yatziv & Sapiro 2006), and *Distance Transformation based Colorization* (DTC) (Lagodzinski & Smolka 2014).

In our experiment the microscopic gray scale image was scribbled using 8 different colors (70 scribbles,  $680 \times 490$  pixels), as presented in Figure 11a. Then, for each method, we modified its parameters, so that the outcomes were visually most attractive. The final results are depicted in Figure 11. The computational time for each method is given in Table 1.

As it can be noticed, the colorized versions of gray scale image are of very similar quality. It confirms, that our method is capable to provide fine and comparable colorization outcomes. Some

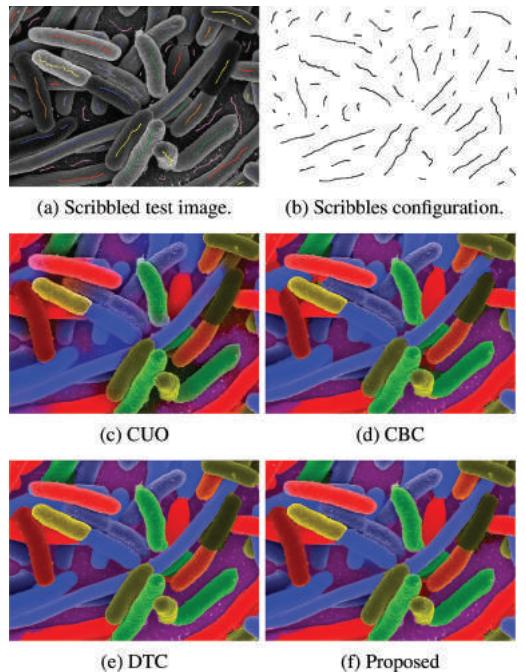


Figure 11. The outcomes of colorization of biomedical microscopy image provided by various algorithms.

Table 1. Computational time comparison.

Method	CUO	CBC	DTC	Proposed
Evaluation time	21 s	105 s	8 s	44 s

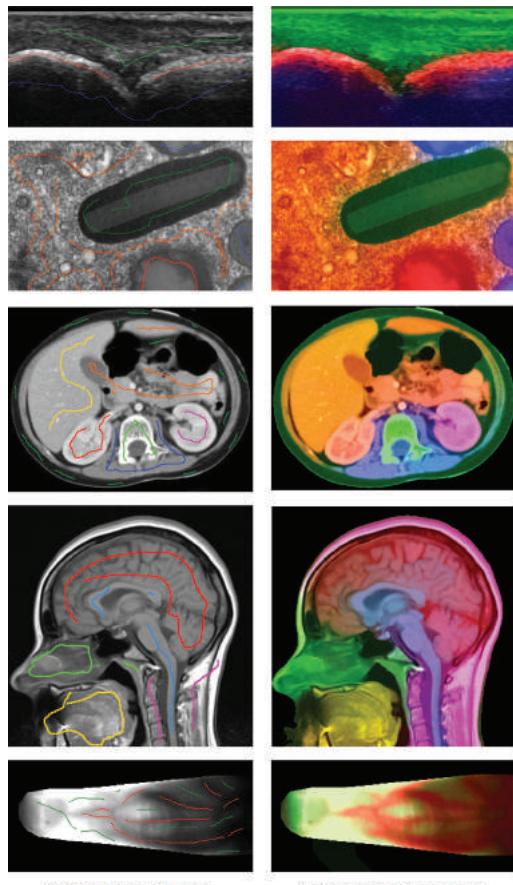


Figure 12. Examples of colorization of biomedical imaging by our algorithm. Corresponding images include (from up to bottom): USG image of human joint, the TEM image of a human leukocyte, a CT slice through the upper human abdomen, MRI image of human brain, infrared finger vascular system.

slight unwanted color leakage may be noticed for each of employed method, however it may be reduced by better scribbles placing. The evaluation time differs significantly and indicates, that our proposed algorithm (44 s) is much slower than the fastest method DTC (8 s), but still is more computationally efficient than CBC (105 s).

In Figure 12 we present some examples of applications of our method for the medical

imaging purposes. They include miscellaneous images obtained by various imaging techniques: USG, TEM, CT, MRI and IR. We inserted several scribbles for better discernibility of distinct regions, depicting tissues, cells, human organs etc.

## 5 SUMMARY

In the paper we have proposed a novel approach to colorization of biomedical images. Our method employs simple operations within a digital image, which involve the distance transformations and the replacement procedures based on the bilateral filtering scheme. The tunable parameters enable adjusting the final colorization outcomes, so that they meet the application requirements.

We presented some examples of applications of our method to images obtained by different techniques widely used in the medicine. The proposed algorithm allowed for achieving fine and medically valuable colorization results comparable with the results of other state-of-the-art algorithms. We plan to extend the capabilities of the algorithm after investigating the impact of non-square bilateral processing windows and different kernel functions.

## ACKNOWLEDGMENT

The research leading to these results has received funding from the Polish-Norwegian Research Programme operated by the National Center for Research and Development under the Norwegian Financial Mechanism 2009–2014 in the frame of Project Contract No. Pol-Nor/204256/16/2013.

This work was performed using the infrastructure supported by POIG.02.03.01-24-099/13 grant: GeCOniiI–Upper Silesian Center for Computational Science and Engineering.

Ultrasound image depicted in Figure 12, was created for MEDUSA project at Section for Rheumatology; Department for Neurology, Rheumatology and Physical Medicine, Central Hospital, Førde, Norway.

## REFERENCES

- Acharya U. R., W.L. Yun, E.Y. Ng, W. Yu, & J.S. Suri (2008, August). Imaging systems of human eye: A review. *J. Med. Syst.* 32(4), 301–315.
- Bankman, I.N. (Ed.) (2000). *Handbook of Medical Imaging*. Orlando, FL, USA: Academic Press, Inc.
- Borgefors, G. (1986, June). Distance transformations in digital images. *Comput. Vision Graph. Image Process.* 34(3), 344–371.

- Lagodzinski, P. & B. Smolka (2014, March). Application of the extended distance transformation in digital image colorization. *Multimedia Tools Appl.* 69(1), 111–137.
- Levin, A., D. Lischinski, & Y. Weiss (2004a, August). Colorization using optimization. *ACM Trans. Graph.* 23(3), 689–694.
- Levin, A., D. Lischinski, & Y. Weiss (2004b). Colorization using optimization. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, New York, NY, USA, pp. 689–694. ACM.
- Lin, F., M. Olivo, & S.-Y. Kung (2009). *Biomedical Imaging* (1st ed.). Springer Publishing Company, Inc.
- Ophir, J., I. Cspedes, H. Ponnekanti, Y. Yazdi, & X. Li (1991). Elastography: A quantitative method for imaging the elasticity of biological tissues. *Ultrasonic Imaging* 13(2), 111–134.
- Popowicz, A. & B. Smolka (2014, March). Isoline based image colorization. In *Computer Modelling and Simulation (UKSim), 2014 UKSim-AMSS 16th International Conference on*, pp. 280–285.
- Reinhard, E., M. Ashikhmin, B. Gooch, & P. Shirley (2001, September). Color transfer between images. *IEEE Comput. Graph. Appl.* 21(5), 34–41.
- Rosenfeld, A. & J. Pfaltz (1968). Distance functions on digital pictures. *Pattern Recognition* 1(1), 33–61.
- Tomasi, C. & R. Manduchi (1998, Jan). Bilateral filtering for gray and color images. In *Sixth International Conference on Computer Vision*, 1998, pp. 839–846.
- Walus M., Bernacki K., K.J. (2015, March). Nir finger vascular system imaging in angiology applications. In *Mixed Design of Integrated Circuits and Systems (MIXDES), 2015 22nd International Conference on*.
- Welsh, T., M. Ashikhmin, & K. Mueller (2002, July). Transferring color to greyscale images. *ACM Trans. Graph.* 21(3), 277–280.
- Yan, W.-Q. & M. Kankanhalli (2003, July). Colorizing infrared home videos. In *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on*, Volume 1, pp. I–97–100 vol.1.
- Yatziv, L. & G. Sapiro (2006, May). Fast image and video colorization using chrominance blending. *Image Processing, IEEE Transactions on* 15(5), 1120–1129.
- Zhao, Y., L. Wang, W. Jin, & S. Shi (2007, May). Colorizing biomedical images based on color transfer. In *Complex Medical Engineering, 2007. CME 2007. IEEE/ICME International Conference on*, pp. 820–823.
- Zhen, Z., G. Yan, & M. Lizhuang (2012, July). An automatic image and video colorization algorithm based on pattern continuity. In *Audio, Language and Image Processing (ICALIP), 2012 International Conference on*, pp. 531–536.

This page intentionally left blank

# Comparison of the internal structures of bones by microtomography

J.C.L. Stutz

*Universidade do Estado do Rio de Janeiro—Instituto Politécnico, Nova Friburgo, Rio de Janeiro, Brasil  
Universidade Estácio de Sá, Nova Friburgo, Rio de Janeiro, Brasil*

J.S. Domínguez & J.T. de Assis

*Universidade do Estado do Rio de Janeiro—Instituto Politécnico, Nova Friburgo, Rio de Janeiro, Brasil*

**ABSTRACT:** The use of computed microtomography ( $\mu$ CT) has revolutionized many areas of research, such as non-invasive, fast and high precision techniques, which allows immediate visualization of internal structures without any risk to the object of study. Its usefulness is undeniable for medicinal purposes, is also important for purposes of zoological research and paleontological. In this work, we used the  $\mu$ CT to investigate the internal structure of bones from mammalian and poultry. We studied the bones of rats and quail. Through of microtomography images, we observed that the bones of the poultry have a bony structure in the form of a trellis that is not present in bones of mammals. These trellises bony is an evolutionary adaptation that allowed the bones of the birds they become longer and lighter maintaining its strength. It was also observed that the percentage of the trabecular area in poultry is almost half of that observed in mammals. The results obtained validate the use of  $\mu$ CT as a technique that allows the study of bone structures in small samples. Enabling explore the morphological differences between the bones of those species.

**Keywords:** microtomography; bones; X-ray

## 1 INTRODUCTION

The computed microtomography ( $\mu$ CT) is a non-invasive method, which allows the study of the complex internal structures of an object. This technique has been used in various areas of knowledge. However, the biomedical applications are the most common (Morton et al., 1990). The  $\mu$ CT is being used in many medical studies as osteoporosis (Lima et al., 2006), osteoarthritis (Amanai et al., 2008), pseudoarthrosis (Qu and Schroeder, 2008), fragility and fracture risk (Hanson and Bagi, 2004) and the differences between healthy bones and unhealthy bones (Lima et al., 2011). These studies use  $\mu$ CT to investigate the internal structures of osseous tissue.

In microtomography samples are explored using some kind of radiation to produce detailed images of axial body.  $\mu$ CT can be analyzed as a two-step process: exploitation of samples and image reconstruction. During exploitation, the bodies of the samples under study are radiographed at different angles generating a set of radiographs. These radiographs are digitally processed in tomographic image reconstruction process to generate a three-dimensional image of the interior of the object under study from a series of two-dimensional images.

In this work, we use the  $\mu$ CT to compare the internal structure of mammalian bones with the

poultry bones, using rat bones (*RattusNorvegicus*) and quail bones (*CoturnixCoturnix*).

The main motivation of this work is to explore the internal morphological structures of pneumatic bones of birds and compare them with mammals. The pneumatization of bird bones is linked to its ability to fly, to provide strength and lightness of his bone structure.

## 2 MATERIALS AND METHODS

We analyzed four samples, two bones of rats and two bones of poultries. Figure 1 is a 3D rendering of the samples used.

To explore the samples of rat bones we use the scanner Skyscan model 1174, operated at 50 kV,



Figure 1. Scanner Skyscan model 1174.

Table 1. Dimensions of the samples.

Sample	Number of slices	Pixel size ( $\mu\text{m}$ )	Size in mm
Poultry1	2020	22.08	44.61
Poultry 2	2033	22.08	44.90
Rat 1	1323	33.29	44.04
Rat 2	1149	33.29	38.25

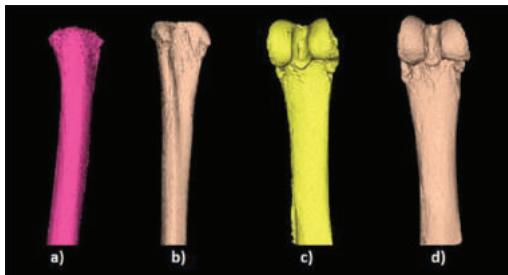


Figure 2. 3D rendering of samples: quail bones a and b, rat bones c and d.

with aluminum filter of 0.75 mm, and pixel size of 33.29  $\mu\text{m}$ . A scanner Skyscan model 1173, operated at 50 kV, with aluminum filter of 1.00 mm and pixel size of 22.08  $\mu\text{m}$ , produced the radiographs of poultry bones.

Table 1 summarizes the dimensions of the samples and the number of horizontal slices reconstructed from radiographs obtained in the exploration of samples.

### 3 RESULTS AND DISCUSSION

Figure 3 shows the poultry bones and their slices. In her is possible to observe the bone of the bird at its extremity, which is formed by a trabecular bony structure, followed by a trellis section followed by a cortical section. Figure 4 shows the rat bones and its slices, can see that the bone also begins with a trabecular bone structure, but followed by cortical section of bone, not presenting in its interior a trellis structure.

As we can see from Table 2, there is a significant decrease in the area of trabecular bone of the poultry, this area is replaced by bone trellises. Those trellises are displayed in 3D, from different angles for better viewing in Figure 5.

For a better view of the trellis region was chosen and separate a section of quail 1, composed of 40 slices, displayed in Figure 6. The selected area corresponds to a section of  $\sim 0.883$  mm or  $\sim 2\%$  of the sample area and  $\sim 6.41\%$  of the trellis region in the sample.

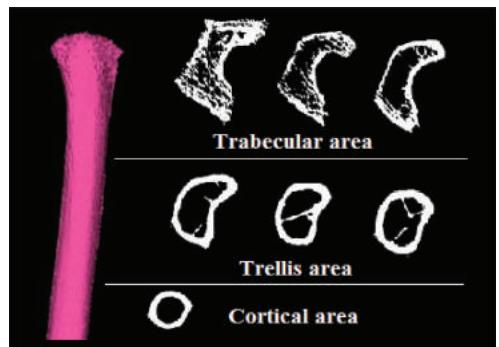


Figure 3. Poultry bone slices examples.

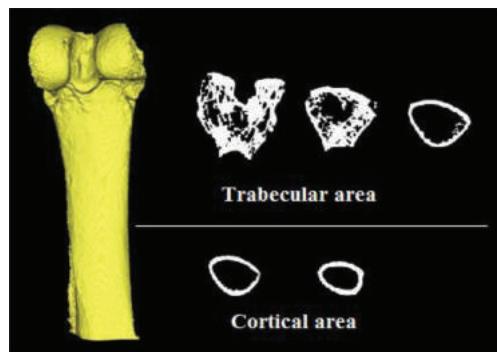


Figure 4. Examples of rat bone slices.

Table 2. Percentage of different types of internal bone structures.

Sample	% trabecular area	% trellis area	% cortical area
Poultry 1	27.48	31.63	40.89
Poultry 2	32.86	24.5	42.65
Rat 1	51.17	0	48.83
Rat 2	56.22	0	43.86

Based on this set of slices the 3D volume was rebuilt and is represented in Figure 7. In this image can be observed clearly the structural nature of the bone trellis.

Comparing the trellis region with the trabecular region of the quail bone, a huge difference between the two regions is observed. The trabecular region of the bone, by its definition, has a spongy structure without defining pattern filling the entire inside of the bone. The trellis region has a large empty area with several narrow and straight structures in the form of pillars which support between themselves and inner walls of the bone in a triangular arrangement.

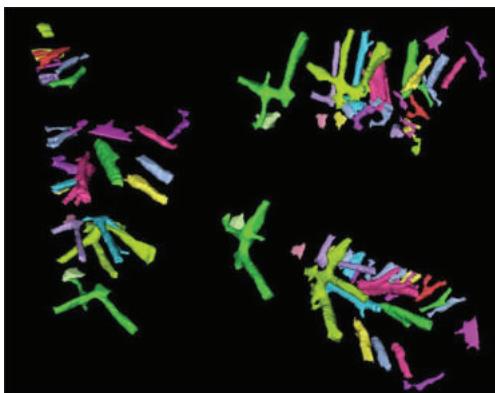


Figure 5. Rendering 3D bone trellis at different angles.

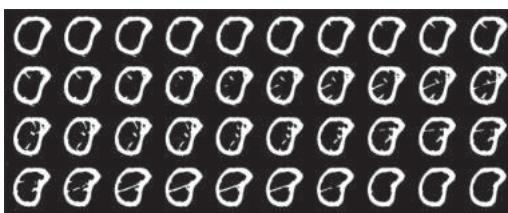


Figure 6. Set of slices of trellis area of Poultry1.



Figure 7. 3D volume based on a set of slices of trellis area.

This kind of structure reduces the amount of bone volume, making the bone bird lighter.

In Table 3 the values of five morphometric parameters used in bone assessment studies are presented: Total Volume (TV), Bone Volume (BV), Bone Volume to Tissue Volume (BV/TV), Bone Surface (BS) and Bone Surface to Bone Volume (BS/BV). These values were calculated by CTAn® software that composes the software package of Skyscan scanner used at work.

The TV is the total volume 3D Volume-of-Interest (VOI) it is measured based on the marching

Table 3. Morphometric parameters of the samples.

	Poultry 1	Poultry 2	Rat 1	Rat 2
TV ( $\mu\text{m}^3$ )	4.01E+12	4.11E+12	4.74E+12	3.56E+12
BV ( $\mu\text{m}^3$ )	1.68E+11	1.45E+11	2.36E+11	2.77E+11
BV/TV	4.1850	3.5274	4.9816	7.7837
BS ( $\mu\text{m}^2$ )	1.84E+09	2.09E+09	1.97E+09	2.07E+09
BS/BV	0.01099	0.0144	0.00833	0.007443 (1/ $\mu\text{m}$ )

cubes of the model. The BV is the total volume of binarised objects within the VOI. The BV/TV is the proportion of the VOI occupied by binarised solid objects. The BS is the surface area of all the solid objects within the VOI and the BS/BV is the surface to volume ratio or “specific surface” is a useful basic parameter for characterizing the thickness and complexity of structures.

#### 4 CONCLUSION

These results show that by using  $\mu$ CT images you can view and separate the existing bone trellises inside the long bones of poulties.

With the segmentation of these structures will be possible, using computational tools to study the load capacity and resistance they provide to the bone. This information may be used to compare with the bones of other birds and so describe the different adaptations in bone structures that contribute to their flight.

#### REFERENCES

- Amanai, T., Nakamura, Y., Aoki, S., Mataga, I., (2008). Micro-CT analysis of experimental *Candida* osteoarthritis in rats. *Mycopathologia* 166, 133–141.
- Hanson, N.A., Bagi, C.M., (2004). Alternative approach to assessment of bone quality using micro-computed tomography. *Bone* 35, 326–333.
- Lima, I.C.B., Oliveira, L.F., Lopes, R.T., (2006). Bone architecture analyses of rat femur with 3D microtomographic images. *J. Radio anal. Nucl. Chem.* 269, 639–642.
- Lima, I., Taam, P., da Costa, V., Fleiuss, M.F., Rosenthal, D., Lopes, R.T., (2011). Study of strontium ranelate bone issues by X-ray microtomography. *Nucl. Instrum. Methods Phys. Res. A* 652, 779–782.
- Morphometric parameters measured by Skyscan™ CT-analyser software. Disponível em: <[www.skyscan.be/next/CTAn03.pdf](http://www.skyscan.be/next/CTAn03.pdf)>. Acesso em: junho de 2015.
- Morton, E.J., Webb, S., Bateman, J.E., Clarke, L.J., Shelton, C.G., (1990). Threedimensional X-ray microtomography for medical and biological applications. *Phys. Med. Biol.* 35, 805–820.
- Qu, G., Schroeder, H.P.V., (2008). Trabecular microstructure at the human scaphoid Nonunion. *J. Hand Surg.* 33, 650–655.

This page intentionally left blank

# Model adaptation for mesh generation of biological structures

A. Ramos-de-Miguel, R. Montenegro & J.M. Escobar

*Division Discretization and Applications, University Institute on Intelligent System and Numerical Application in Engineering, Las Palmas of Gran Canaria, Spain*

**ABSTRACT:** In this work we propose a technique for the modellisation of biological structures for its applications in Finite Elements Methods (FEM). The key of this method consists on deforming the mesh of a generic structure in order to obtain a mesh of a specific one. The input data are both the mesh of the generic structure and the data collected by Computerised Tomography (CT) scan, Magnetic Resonance Imaging (MRI), etc... This technique involves genetic algorithms for adjusting the nodes of the input mesh to the collected points, Radial Basis Functions (RBF) to deform the input mesh, and optimisation of the final mesh.

## 1 INTRODUCTION

The generation of FEM meshes of biological structures for simulating pathologies, prothesis and behaviour of different organs (1) (2) (3), presents a great challenge when little information is available.

This problem can be found, for example, in CT for medical image acquisition and later 3D reconstruction. In this case, a compromise is required between emitted radiation, exposure time and desired image quality. Moreover if we use high radiation and long exposure time, there can be no guarantees that sufficient data will be collected.

What we propose in this work is a method to generate an FEM mesh by doing small deformations on a model mesh. Our inputs are a tetrahedral mesh that represents the organ of a generic patient and the acquired data,  $AD$ , from a particular patient, i.e., a CT scan, MRI, etc.

$M$  represents the tetrahedral mesh of a generic patient's organ,  $S_M$  the surface mesh of  $M$  and  $N_s$  its nodes.

The first stage is to superpose, searching for the best fit,  $N_s$  to  $AD$ , by doing rigid transformations of  $M$  in order to minimise the Hausdorff distance between  $N_s$  to  $AD$ . We define two sets of points,  $P_S \subset N_s$ , known as selected points, and  $H \subset AD$ , known as handle points. The pair  $(p, h)$  with  $p \in P_S$  and  $h \in H$  is selected among  $N_s$  and  $AD$  if the distance between them is below a certain threshold, given by the user. The surface mesh  $S_M$  is deformed in order to interpolate the points of  $H$ .

We use RBF to deform  $M$  due to the good behaviour of these functions and their low computational cost (4) (5) to get  $M'$ . The intrinsic problem of this method is that the resulting deformed mesh can be tangled, so an optimisation procedure

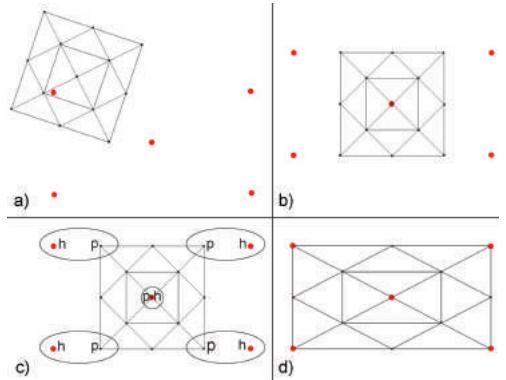


Figure 1. Summary of the main stages of the method in a simplified example. a) Input data: a generic mesh of a structure  $M$  (black) and the acquired data of the desired structure,  $AD$  (red); b) Affine transform of  $M$  after the point cloud superposition; c) Selection of pairs  $(p, h)$  to guide the deformation, where  $h$  are the handle points; d) The mesh after deformation.

must be applied. This optimisation consists on the application of an untangling and smoothing procedure: one for the surface  $S_{M'}$  and another for the tetrahedral mesh  $M'$ .

## 2 METHODS

### 2.1 Point cloud superposition

Our purpose is to minimise the Hausdorff distance between the points  $N_s$  to  $AD$  using rotations, translations and scaling. This distance indicates how far apart the two sets are.

A genetic algorithm was used for this task because there is not an analytic function that relates the affine transformation of a point cloud with the obtained Hausdorff distance.

### 2.1.1 Genetic algorithm

Each individual has 9 genes, 3 for translation, 3 for rotation and 3 for scale in each axis  $x$ ,  $y$  and  $z$ .

The genetic algorithm comprises 5 stages. In the first one, the initial population is created, where the genes of each individual were generated randomly. Then, an iterative process is initiated where, for each iteration, each individual is compared with the fitness function. After evaluating each individual, the leader is selected and a crossover and a mutation stage follows. The crossover and mutation are done gene by gene by using the gene information from the parents. The last step is the generation of the new individuals to keep the size of the population constant.

## 2.2 Deformation of the model mesh

Once the superposition of  $N_S$  and  $AD$  have been done,  $M$  is transformed into a deformed model  $M'$  by computing updated point locations  $\mathbf{x}' = \mathbf{x} + \mathbf{d}(\mathbf{x})$  for each original point  $\mathbf{x} \in M$ .

### 2.2.1 Deformation function

RBF are used for computing  $\mathbf{d}(\mathbf{x})$ ,  $\mathbf{x} = (x, y, z)$ . The deformation function  $\mathbf{d} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$  is given by:

$$\mathbf{d}(\mathbf{x}) = \sum_{j=1}^m \mathbf{w}_j \varphi_j(\mathbf{x}) + \sum_{k=1}^4 \mathbf{q}_k \pi_k(\mathbf{x}) \quad (1)$$

Equation (1) is defined as a linear combination of radially symmetrical nucleus functions  $\varphi_j(\mathbf{x}) = \varphi(\|\mathbf{x} - \mathbf{h}_j\|)$ , located in centres  $\mathbf{h}_j \in H$  and a linear polynomial to guarantee the affine transformations, where  $\{\pi_1, \pi_2, \pi_3, \pi_4\} = \{x, y, x, l\}$  is a base

of the linear polynomial space. The chosen kernel function to deform the model was  $\varphi(r) = r^3$  (7).

The weights  $\mathbf{w}_j$  and  $\mathbf{q}_j$  are determined by imposing the interpolation conditions,  $\mathbf{d}(\mathbf{p}_j) = \partial \mathbf{h}_j = \mathbf{p}_j - \mathbf{h}_j$ .

## 2.3 Mesh untangling and smoothing

As explained above, the deformation process on  $M$  can produce  $M'$  to be tangled, so mesh optimisation is required.

This procedure is divided in two stages: one for the surface mesh  $S_{M'}$  and another for  $M'$  inner nodes.

The optimisation process (untangling and smoothing) is an iterative process in which each node of the mesh is moved to a new position that minimises a certain objective function. The objective function is derived from some quality measure of the local mesh, that is, the set of elements (triangles for 2D or tetrahedrals for 3D) connected to the adjustable or free node. (6) (7) (8).

The optimisation process for  $S_{M'}$  takes place in a local parametric space. We search ideal triangles in this local parametric space that become optimal in the surface  $S_{M'}$ . After finding the optimal position in the parametric space, the nodes are projected to the local approximation using RBF. Finally,  $M'$  inner nodes are optimised using (9).

## 2.4 Skeleton deformation

When we work with geometries with very small differences, the superposition of  $M'$  in  $AD$  matches. But, some geometries are impossible to superpose well because they have structural differences. For example, a full body will be badly adapted to another because the arms and legs have different position on each mesh.

To solve this problem we propose to apply the deformation procedure to the skeletons of  $M$  and  $AD$ , which we will name as  $Sk_M$  and  $Sk_{AD}$ . First

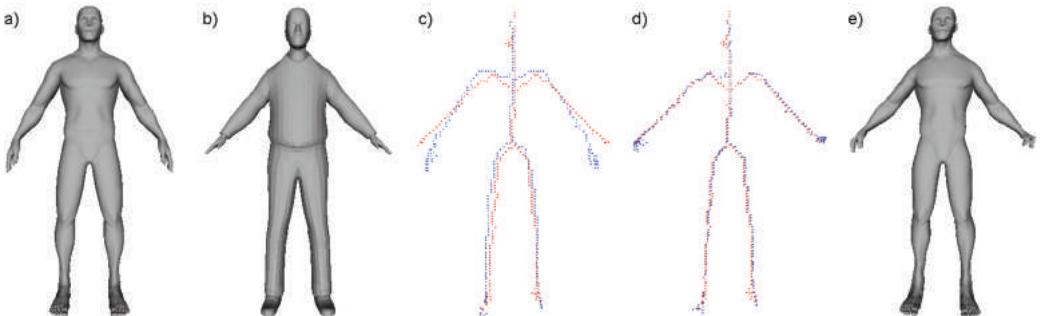


Figure 2. Skeleton deformation stages. a) Model input mesh; b) Patient input mesh; c) Superposed skeletons of model (blue) and patient (red) meshes; d) Deformed model (blue) skeleton to fit the patient (red) skeleton; e) Deformed model surface mesh.

we voxelize  $M$  and  $AD$  and then we calculate the skeleton 3D medial axis from a 3D binary volume, to get  $Sk_M$  and  $Sk_{AD}$  (as described in (10)). Then,  $Sk_M$  must be deformed to be adapted to  $Sk_{AD}$ , proposed in 2.1 and 2.2. Rigid transformation and deformations done to  $Sk_M$  are also applied to  $M$ .

Figure 2 shows the model (a) and patient (b) meshes and the corresponding skeletons (c). The skeleton of the model is deformed using RBF in order to fit with the skeleton of the patient (d). Finally, (e) shows the deformed mesh surface.

With this optional step, we reduce the fitting error between  $M'$  and  $AD$ .

### 3 RESULTS

#### 3.1 Superposition test

The superposition is a critical stage of this method. If a good adjustment between the input point clouds and the mesh is not reached, the deformation will select the wrong handle points and the final mesh will not fit the input data.

In this superposition test one mesh is adjusted with itself. In this case  $AD$  is a set of points extracted from the input mesh that simulates  $M$ 's CT scan.

Applying the Hausdorff distance minimisation procedure 2.1 to  $\{AD_1, AD_2, AD_3, AD_4\}$ , results in the affine transformations  $\{T_1, T_2, T_3, T_4\}$ . The Hausdorff distance is measured between  $T_i(N_s)$ ,  $i = 1, \dots, 4$  and  $N_s$  to evaluate the goodness of the fitting between patient data and input mesh. The data are shown in Table 1.

A genetic algorithm was used to calculate affine transformations with 1000 generations and a population size of 20 individuals.

#### 3.2 Deformation test

$AD$  will be taken as a set of points extracted from the mesh in Figure 4-b, that simulate a CT scan. Also, it will be assumed that  $AD$  and  $M$  are already superposed.

This test consists on deforming  $M$  to adapt it to  $AD$  following the procedure 2.2.

Table 1. Evolution of the Hausdorff distance as the number of points of the acquired data increases.

Mesh	Num. points CT scan	Hausdorff dist.
CT 1	101	0.1152
CT 2	200	0.1084
CT 3	413	0.0997
CT 4	825	0.0894

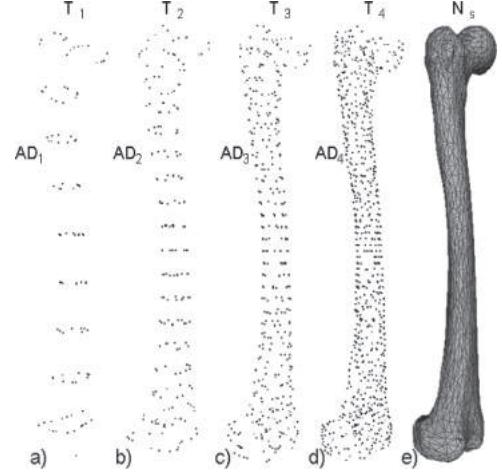


Figure 3. Different simulated CT scans extracted from the original mesh. a) CT 1 with 10 Femur slices; b) CT 2 with 20 Femur slices; c) CT 3 with 40 Femur slices; d) CT 4 with 80 Femur slices; e) Original Femur mesh.

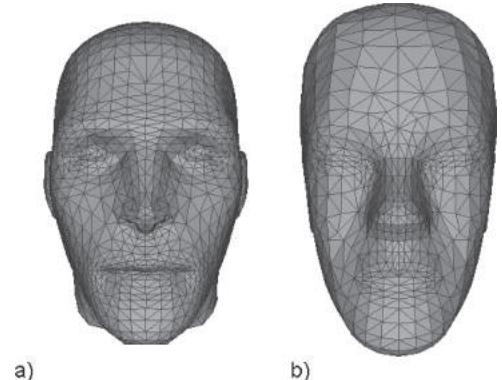


Figure 4. Left: 3D mesh model  $M$ ; right: mesh used to extract the input data (patient)  $AD$ .

In Figure 5 the different adjustments between  $AD$  and  $M$  are illustrated in terms of the number of handle points. It should be noted that the greater the number of handle points the better the adjustment between  $M$  and  $AD$ , Table 2.

The faces shown in Figure 5 have been optimised following the procedure 2.3. In Figure 6 a detail of the rightmost head in Figure 5 is shown, before a) and after applying the optimisation procedure b). Table 3 shows the number of tangled elements and the mean quality before and after optimisation of each head in Figure 5. Mean ratio quality metric was used to measure the mesh quality (11). The 0 value corresponds to a degenerated element and 1 to an equilateral triangle.

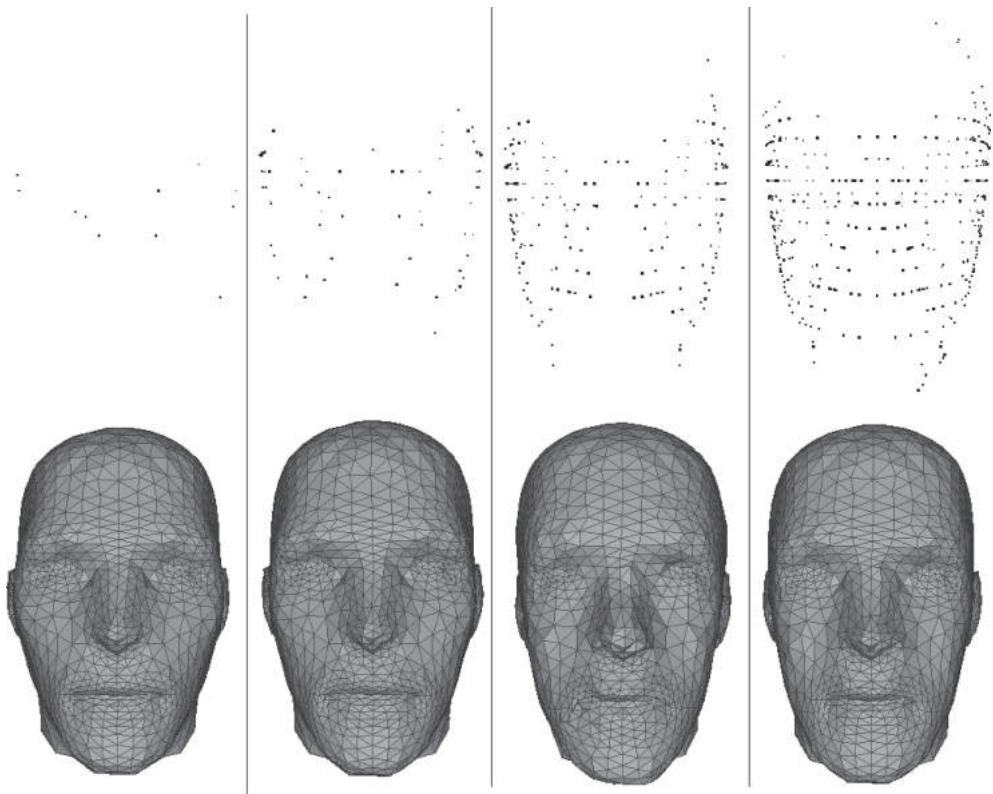


Figure 5. Deformation test after untangling stage, changing the number of handle points. From left to right, the number of handle points is increased. The first line shows the handle points used and the second line the deformation mesh.

Table 2. Evolution of the Hausdorff distance as the number of handle points increases.

Mesh	n. handle P.	Hausdorff dist.
Case 1	67	0.3833
Case 2	1094	0.3725
Case 3	1453	0.3179
Case 4	1557	0.3112

Table 3. Evolution of the final mesh quality after optimisation and untangling. All case finished with 0 tangled elements.

Mesh	n. tangled ele.	Mean Q.	
	Pre	Pre	Post
Case 1	571	0.568	0.756
Case 2	595	0.566	0.755
Case 3	637	0.557	0.755
Case 4	624	0.559	0.755

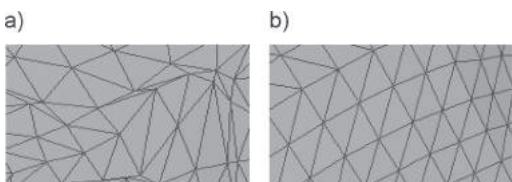


Figure 6. A detail of the rightmost head in Figure 5, before a), and after the smoothing and untangling stage b).

#### 4 DISCUSSION

We propose a method to create meshes suitable to be used in FEM and matched to the patient organ geometry.

The highest computational cost corresponds to the computation of the Hausdorff distance in the

present implementation. This computation can be improved by arranging the data in an octree structure and also through parallel computing.

The handle points guide the deformation. Hence, as illustrated in Table 2, the number of handle points is strongly related to the final mesh adjustment. Also, as shown on Figure 5, the dissemination of these points makes it possible to capture the singularities of the mesh.

In this method, the handle points are selected using the nearest point strategy, but, probably, it would be better to work with some landmarks on the mesh (13), to enforce a better adaptation of some parts of the mesh.

On the skeleton deformation stage, the procedure tends to flatten the surface mesh. This problem and a possible solution is described in (12).

Finally, it should be noted that the mesh deformation induced by RBF can sometimes cause tangled meshes, so the application of an untangling and smoothing procedure is mandatory in order to obtain a valid mesh for the application of FEM.

## 5 CONCLUSIONS

We propose a technique for the modellisation of biological structures for its application in FEM when a geometry is given with sparse data. This technique involves genetic algorithms for adjusting the nodes of the input mesh to the collected points, Radial Basis Functions (RBF) for deforming the input mesh, and mesh optimization.

A critical point is the selection of handle points. Future research will involve the improvement of the method adopted for this selection, e.g. using landmarks (13). A proper handle point choice has positive effect in the fitting of the acquired data to mesh.

## REFERENCES

- [1] I. Borazjani; “Fluid-structure interaction, immersed boundary-finite element method simulations of bio-prosthetic heart valves”; *Computer Methods in Applied Mechanics and Engineering*, Volume 257, 15 April 2013, Pages 103–116.
- [2] L. Hou-Guang, T. Na, R. Zhu-Shi; “Numerical modeling of human middle ear biomechanics using finite element method”; *Journal of Donghua University*, Volume 28, April 2011, Pages 115–118.
- [3] A. Idhammad, A. Abdali and N. Alaa; “Computational simulation of the bone remodeling using the finite element method: an elastic-damage theory for small displacements”; *Theoretical Biology and Medical Modelling*, Volume 10, April 2013.
- [4] D. Sieger, S. Menzel and M. Botsch; “On Shape Deformation Techniques for Simulation-based Design Optimization”; *New Challenges in Grid Generation and Adaptivity for Scientific Computing*; SEMA SIMAI Springer Series, 2014, Springer Milan.
- [5] Eloi Ruiz-Gironés, Xevi Rocab, Jose Sarratea; “Mesh deformation based on radial basis function interpolation”; *Procedia Engineering*; Volume 82 (2014) 101–113.
- [6] J.M. Escobar, G. Montero, R. Montenegro and E. Rodriguez; “An algebraic method for smoothing surface triangulations on a local parametric space”; *Int. J. Numer. Meth. Engng*; 2006; 66:740–760.
- [7] A. de Boer, M.S. van der Schoot, H. Bijl; “Optimizing mesh distortion by hierarchical iteration relocation of the nodes on the CAD entities”; *Computers and Structures*; Volume 85, Issues 11–14, June–July 2007, Pages 784–795.
- [8] José María Escobar, Eduardo Rodríguez, Rafael Montenegro, Gustavo Montero and José María González-Yuste; “Simultaneous untangling and smoothing of tetrahedral meshes”; *Computer Methods in Applied Mechanics and Engineering*; 2003; Vol. 192 25:2775–2787.
- [9] José María Escobar, Eduardo Rodríguez, Rafael Montenegro, Gustavo Montero and José María González-Yuste; SUS code: simultaneous mesh untangling and smoothing code”; <http://www.dca.iusiani.ulpgc.es/SUScode>; 2010.
- [10] Ta-Chih Lee, Rangasami L. Kashyap and Chong-Nam Chu; “Building skeleton models via 3-D medial surface/axis thinning algorithms.”; *Computer Vision, Graphics, and Image Processing*; 56(6):462478, 1994.
- [11] Patrick M. Knupp; “Algebraic Mesh Quality Metrics.”; *SIAM Journal on Scientific Computing*; 23(1), 193–218, 2001.
- [12] Zohar Levi and David Levin; “Shape Deformation via Interior RBF.”; *1062 IEEE Transactions on Visualization and Computer Graphics*; Vol. 20, No. 7, 2014.
- [13] Lorenzo Grassi, Najah Hraiech, Enrico Schileo, Mauro Ansaldi, Michel Rochette and Marco Viceconti; “Evaluation of the generality and accuracy of a new mesh morphing procedure for the human femur”; *Medical Engineering and Physics*, Volume 33, 2011, 112–120.

This page intentionally left blank

# Medical volume rendering based on gradient information

Thiago F. de Moraes, Paulo H.J. Amorim & Jorge V.L. da Silva

*Tridimensional Technology Division, Center for Information Technology Renato Archer, Campinas-SP, Brazil*

Helio Pedrini

*Institute of Computing, University of Campinas, Campinas-SP, Brazil*

Maria Inês Meurer

*Department of Pathology, Federal University of Santa Catarina, Florianópolis-SC, Brazil*

**ABSTRACT:** Several medical imaging techniques have allowed the acquisition of volumetric images for non-invasive diagnosis purposes, such as Computerized Tomography, Magnetic Resonance Imaging, X-Ray Microtomography, Positron Emission Tomography and Ultrasonography. A substantial increase in both data size and image resolution has been possible due to recent technological advances. Interactive exploration of volumetric data through visualization approaches can be used to enhance internal structures of interest. This work describes and evaluates a novel projection technique for volume rendering based on gradient information and viewing direction to generate values of color and opacity from the data values. Experiments are conducted on several data sets to demonstrate the effectiveness of the proposed technique compared to other projection approaches.

## 1 INTRODUCTION

Volume rendering can be defined as a set of techniques for visualizing bidimensional (2D) projections of tridimensional (3D) data sets. Volume rendering has applications in several knowledge domains, such as medicine, biology, geology, among others. Typically, 3D medical data sets are acquired through Computerized Tomography (CT), Magnetic Resonance Imaging (MRI), X-Ray Microtomography (microCT), Positron Emission Tomography (PET) and Ultrasonography scanners for diagnostic purposes.

The generation of 2D images from 3D data sets can be performed via rendering approaches, which are in general categorized into two groups: isosurface rendering and direct volume rendering.

The isosurface rendering techniques involve the visualization of intermediate structures extracted from the data. The algorithm of Marching Cubes (Lorensen and Cline 1987) is one of the most known techniques for extracting a polygonal mesh of an isosurface from a 3D scalar data set.

In direct volume rendering (Levoy 1990), attributes of color and opacity are based on voxel intensity, such that anatomical structures can be reconstructed by selecting proper threshold levels for each attribute. The mapping of each voxel value to color and opacity is done through a transfer function. Some widely used direct volume

rendering techniques are volume ray casting, splatting, shear warp, and texture-based volume rendering. Due to the intrinsically parallel nature of direct volume rendering, Graphics Processing Unit (GPU) volume rendering approaches have been proposed to speed up the visualization process (Rodríguez et al. 2014, Hadwiger et al. 2006, Okuyan and Gündükay 2014).

In contrast to the approaches that require every voxel value to be mapped to color and opacity values, projection techniques do not use transfer functions. A review of some relevant projection techniques is presented in Section 2.

This paper proposes and evaluates a novel projection technique for volume rendering based on gradient information and viewing direction, combined with Maximum Intensity Difference Accumulation (MIDA) projection, to map data values to color and opacity values. Experiments conducted on several data sets demonstrate the effectiveness of the proposed technique compared to other projection techniques for medical volume rendering. The evaluated techniques were implemented in InVesalius (2015), an open-source imaging framework.

## 2 BACKGROUND

The Maximum Intensity Projection (MIP) (Bruckner and Gröller 2009) selects and projects only the

voxels with maximum intensity along the viewing rays. Since a single value is displayed along each ray, visual ambiguities can be caused by the lack of depth information and shape perception of the structures.

Some variants of the MIP technique have been proposed to mitigate its main problems. The Local Maximum Intensity Projection (LMIP) (Sato et al. 1998) displays the first maximum value that is above a certain threshold, instead of the global maximum value, which adds spatial information. The Depth-Shaded Maximum Intensity Projection (DMIP) (Heidrich et al. 1995) weights the data value by a depth-dependent term, which reduces the chance of high data values to be projected in case they are located far way from the image plane. The Shape-Enhanced Maximum Intensity Projection (SEMIP) (Zhou et al. 2011) searches for a valid gradient for the maximum intensity of each viewing ray and applies a shading based on gradient to improve depth and shape perception of internal structures. Depth-Enhanced Maximum Intensity Projection (DEMIP) (Díaz and Vázquez 2010) uses depth and simple color blending to disambiguate the ordering of the structures. A related projection technique is the Minimum Intensity Projection (MinIP) (Mistelbauer et al. 2013), which displays the minimum value along the viewing ray and can be used to provide information about blood flow deviations, such as soft plaques. The Average Intensity Projection (AIP) (Wu et al. 2008) displays the mean of all values found in the viewing ray, which can lose image sharpness.

The Maximum Intensity Difference Accumulation (MIDA) (Bruckner and Gröller 2009) provides a method for controlling the volume compositing effect by introducing a weighting coefficient that takes into account new higher values along the viewing ray. Thus, when the maximum value changes from a low to a high value, the corresponding sample contributes a higher intensity in comparison to other samples. Occluded structures of interest can be displayed through MIDA, where the accumulated opacity is modulated according to the maximum intensity difference.

The Contour MIP (Csébfalvi et al. 2001) displays the data values based on the magnitude of local gradient information and on the angle between the viewing direction and gradient vector. The technique enhances the voxels with higher value of gradient information, emphasizing the surfaces of different objects.

### 3 METHODOLOGY

This work proposes a new projection technique for volumetric rendering, which is based on gradient information and viewing direction to produce values of color and opacity from the data values.

A comparative analysis with other projection techniques (described in Section 2) is conducted on a number of medical data sets.

The 3D volume data sets are represented by a finite number of 2D cross-sectional slices. A volume element (voxel) corresponds to a position in the data set and stores a data value. In a direct volume rendering method (Levoy 1998), a ray is shot through an image plane into the 3D volume, such that color and opacity values along the ray are computed for each pixel.

The accumulated color and opacity values can be computed following the emission-absorption model (Fernando 2004) as

$$C = \sum_{i=1}^n C_i \prod_{j=1}^{i-1} (1 - \alpha_j) \quad (1)$$

$$\alpha = 1 - \prod_{j=1}^n (1 - \alpha_j)$$

where  $C$  and  $\alpha$  are the accumulated color and opacity values computed iteratively, whereas  $C_i$  and  $\alpha_i$  are the color and opacity obtained from the segment  $i$  along the viewing ray.

Equation 1 can be efficiently computed if samples are sorted in a certain order. In back-to-front order, the accumulated color and opacity values are computed as

$$\hat{C}_i = C_i + (1 - \alpha_i) \hat{C}_{i+1} \quad (2)$$

$$\hat{\alpha}_i = \alpha_i + (1 - \alpha_i) \hat{\alpha}_{i+1}$$

where  $\hat{C}_i$  and  $\hat{\alpha}_i$  are the accumulated color and opacity, respectively, from the back of the volume.

In front-to-back order, the accumulated color and opacity values are computed as

$$\hat{C}_i = (1 - \hat{\alpha}_{i-1}) C_i + \hat{C}_{i-1} \quad (3)$$

$$\hat{\alpha}_i = (1 - \hat{\alpha}_{i-1}) \alpha_i + \hat{\alpha}_{i-1}$$

where  $\hat{C}_i$  and  $\hat{\alpha}_i$  are the accumulated color and opacity, respectively, from the front of the volume.

As described in Section 2, there are some variations of the composition strategies, such as MIP, MinP, LMIP, DEMIP, SEMIP, AIP, MIDA, among others.

The accumulated color  $\hat{C}_i$  and opacity  $\hat{\alpha}_i$  for the  $i$ -th sample along viewing ray in the MIDA technique are expressed as

$$\hat{C}_i = \beta_i \hat{C}_{i-1} + (1 - \beta_i \hat{\alpha}_{i-1}) \alpha_i C_i \quad (4)$$

$$\hat{\alpha}_i = \beta_i \hat{\alpha}_{i-1} + (1 - \beta_i \hat{\alpha}_{i-1}) \alpha_i$$

where  $\beta_i = 1 - \delta_i$  and  $\delta_i = \max(I_i - I_{\max,i}, 0)$ , such that  $I_{\max,i}$  is the current maximum intensity value along the viewing ray and  $I_i$  is the intensity from the segment  $i$  along the viewing ray.

The gradient information can also be used to determine the optical properties (color and opacity) of voxels. The gradient vector  $\nabla f$  points in the direction of largest variation

$$\nabla f = \left( \frac{\delta f}{\delta x}, \frac{\delta f}{\delta y}, \frac{\delta f}{\delta z} \right) \quad (5)$$

In the Contour MIP (Csébfalvi et al. 2001), the data values are mapped to color and opacity values based on both the magnitude of local gradient information and the angle between the gradient vector and viewing direction. The MIP technique is used as the projection technique for rendering the image.

In our work, the gradient information is also employed to emphasize the surface of object. However, the MIDA technique is used to generate the color and opacity values from the data values and gradient information.

Initially, the method calculates the pixel intensity based on the magnitude of the slices taking the viewing direction into account, similarly as the work proposed by Csébfalvi et al. (2001)

$$s(P, V) = (1 - |\nabla(P) \cdot V|)^n \quad (6)$$

where  $\nabla(P)$  is the gradient magnitude of pixel  $P$ ,  $V$  is the viewing direction and  $n$  is a sharpness coefficient for the contour. Viewing direction is  $(0, 0, 1)$

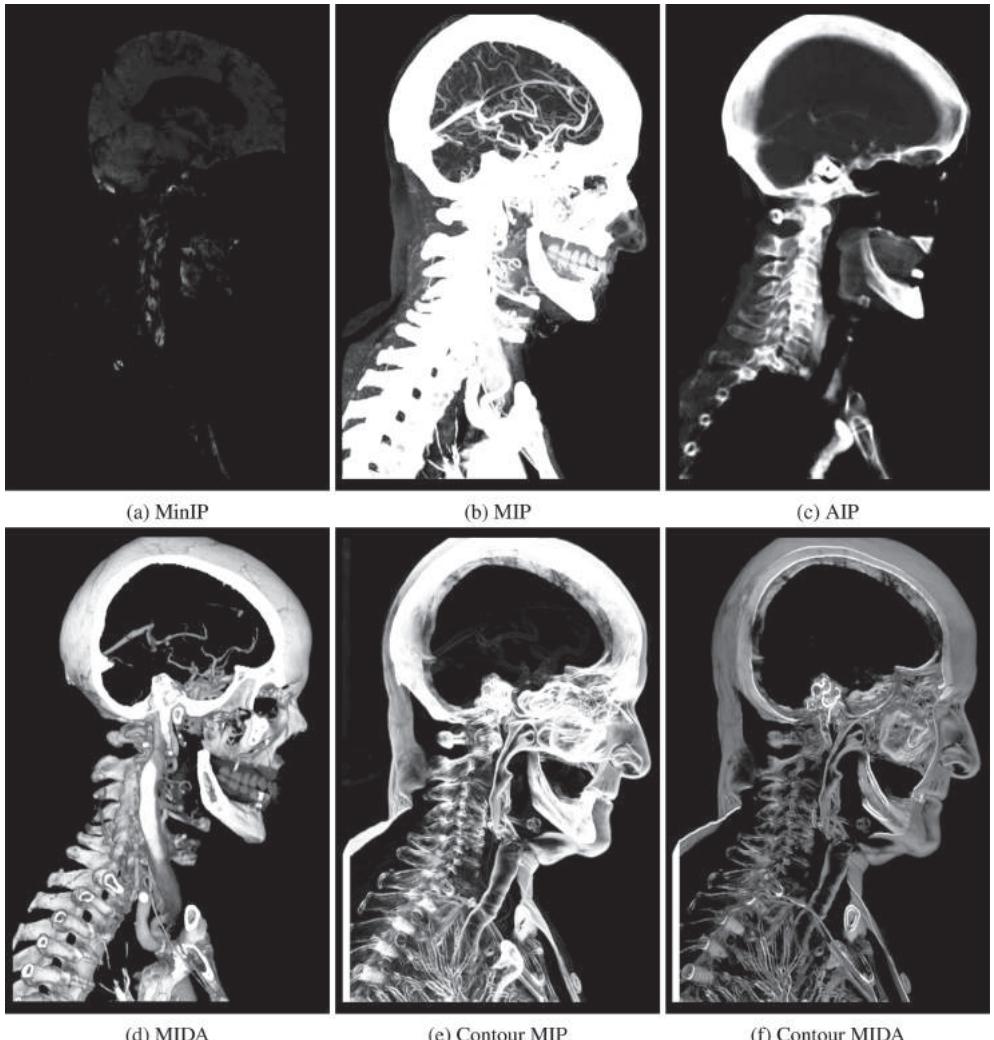


Figure 1. Results for Manix data set for different projection techniques.

for axial, (0,1,0) for coronal and (1,0,0) for sagittal plane.

Equation 6 is computed for all pixels of the input slices. Then, the MIDA technique is applied to the generated set of slices and produces a 2D image as final result. One advantage of the proposed method is to reduce effects caused by lack of depth perception.

#### 4 EXPERIMENTAL RESULTS

We conducted tests on several medical volumetric data sets available in InVesalius (2015) and OsiriX (2015). Due to space limitation, only three data sets will be used to illustrate the results. All evaluated projection techniques were implemented in the open source imaging framework InVesalius.

Figures 1, 2 and 3 show the results for Manix, Chest and Colonix data sets, respectively, obtained through previously described projection techniques: MIP (Maximum Intensity Projection), MinIP (Minimum Intensity Projection), AIP (Average Intensity Projection), MIDA (Maximum Intensity Difference Accumulation), Contour MIP and Contour MIDA (proposed method).

MIP is useful for rendering high contrast regions, for instance, visualization of bone regions and nodules. MinIP allows visualization of regions with low data values, for instance, detection of excessive fat in the liver. AIP offers an overall view of the slices, since it corresponds to a mean of the data values. MIDA provides results with depth information, which helps the comprehension of lesions or other interest regions. Contour MIP and Contour MIDA highlight regions with high contrast, allowing radiologists to focus on such regions.

In Figure 1, it is interesting to observe some differences between MIP (Fig. 1b) and MIDA (Fig. 1d). Both images show bone regions, however, the depth information is not present in the MIP image, such that it is very difficult to distinguish the bone structures from the aorta and other structures in the neck region. On the other hand, it is possible to observe bone orientation and other volumetric information from the MIDA image. A similar comparison is possible between Contour MIP (Fig. 1e) and Contour MIDA (Fig. 1f), where both show bone structures, soft tissues and the pulmonary bronchioles (which were not visible in the MIP and MIDA images). However, in the Contour MIDA image, it is possible to observe the soft

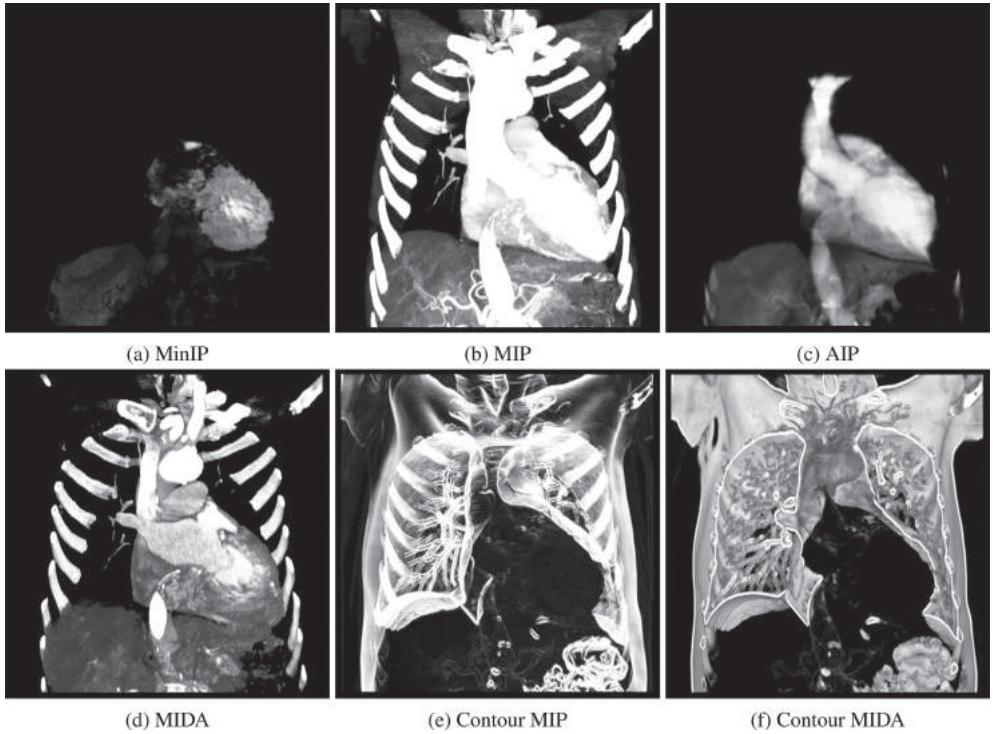


Figure 2. Results for Chest data set for different projection techniques.

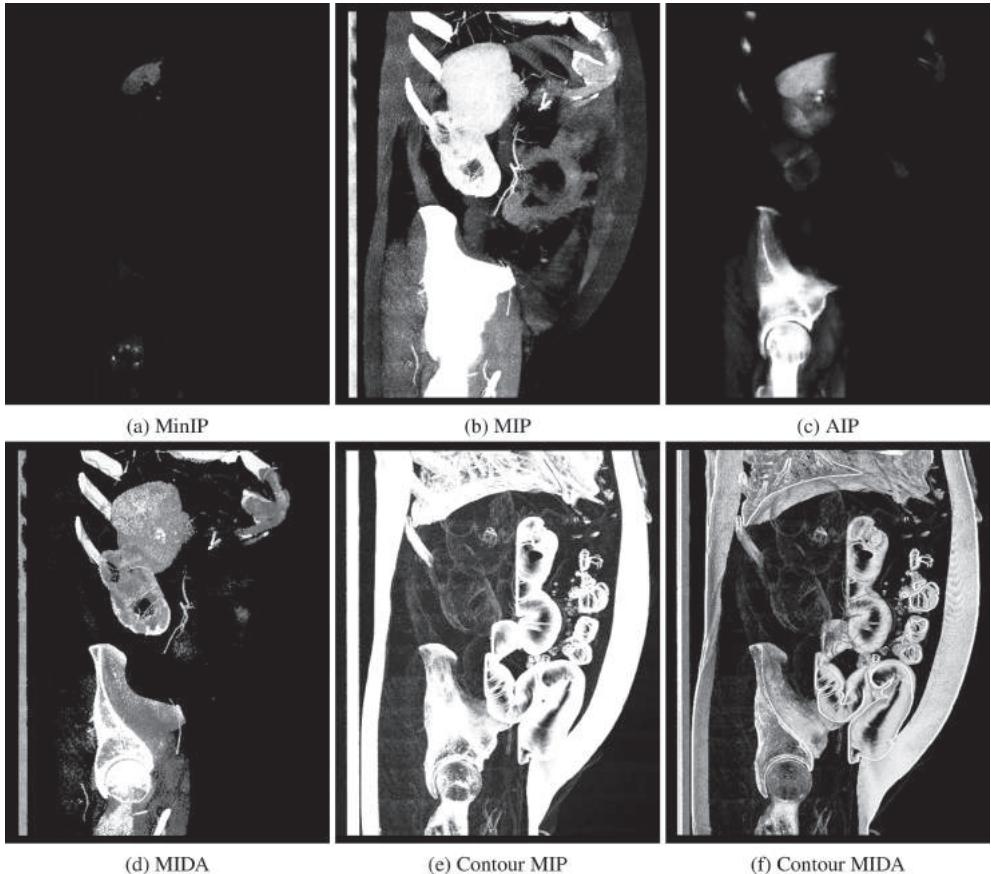


Figure 3. Results for Colonix data set for different projection techniques.

tissue covering the jaw bone, which is not visible in the image generated by Contour MIP.

In Figure 2, it is possible to observe that only Contour MIP (Fig. 2e) and Contour MIDA (Fig. 2f) are able to show the pulmonary bronchioles and the grooves on the base of the left lung. The main difference between both is the depth and volumetric information. The depth information makes it possible to perceive that the bones of the ribs are behind the lungs, not in front of them.

In Figure 3, only Contour MIP (Fig. 2e) and Contour MIDA (Fig. 2f) are able to show the colon in the images. However, the Contour MIDA allowed to differentiate the bends of the colon, as well as to show the internal parts of the pelvis bone.

## 5 CONCLUSIONS

This work proposed and evaluated a novel projection technique for medical volume rendering based

on gradient information and viewing direction to generate a final image.

Experiments are conducted on several data sets and the results obtained through the proposed method are compared with other projection approaches.

## ACKNOWLEDGEMENTS

The authors are grateful to FAPESP, CNPq and CAPES for the financial support.

## REFERENCES

- Bruckner, S. & M.E. Gröller (2009, June). Instant Volume Visualization using Maximum Intensity Difference Accumulation. *Computer Graphics Forum* 28(3), 775–782.
- Csébfalvi, B., L. Mroz, H. Hauser, A. König, & M.E. Gröller (2001). Fast Visualization of Object Contours by Non-Photorealistic Volume Rendering. In *Computer Graphics Forum*, Volume 20, pp. 452–460.

- Díaz, J. & P. Vázquez (2010). Depth-Enhanced Maximum Intensity Projection. In *8th IEEE/EG International Symposium on Volume Graphics*, pp. 93–100.
- Fernando, R. (2004). *GPU Gems: Programming Techniques, Tips and Tricks for Real-Time Graphics*. Pearson Higher Education.
- Hadwiger, M., A. Kratz, C. Sigg, & K. Bühler (2006). GPU-Accelerated Deep Shadow Maps for Direct Volume Rendering. In *SIGGRAPH/EUROGRAPHICS Conference On Graphics Hardware: 21st ACM SIGGRAPH/Eurographics Symposium on Graphics Hardware*, Volume 3, Vienna, Austria, pp. 49–52.
- Heidrich, W., M. McCool, & J. Stevens (1995, October). Interactive Maximum Projection Volume Rendering. In *IEEE Conference on Visualization*, pp. 11–18.
- InVesalius (2015). Open Source Software for Reconstruction of Computed Tomography and Magnetic Resonance Images. <http://www.cti.gov.br/invesalius/>.
- Levoy, M. (1988). Display of Surfaces From Volume Data. *IEEE Computer Graphics and Applications* 8(3), 29–37.
- Levoy, M. (1990). Efficient Ray Tracing of Volume Data. *ACM Transactions on Graphics* 9(3), 245–261.
- Lorensen, W.E. & H.E. Cline (1987, July). Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *Computer Graphics* 21(4), 163–169.
- Mistelbauer, G., A. Morar, A. Varchola, R. Schernthaner, I. Baclija, A. Köchl, A. Kanitsar, S. Bruckner, & E. Gröller (2013, June). Vessel Visualization using Curvicircular Feature Aggregation. *Computer Graphics Forum* 32(3), 231–240.
- Okuyan, E. & U. Güdükbay (2014). Direct Volume Rendering of Unstructured Tetrahedral Meshes using CUDA and OpenMP. *The Journal of Supercomputing* 67(2), 324–344.
- OsiriX (2015, January). DICOM files. <http://www.osirixviewer.com/datasets/>.
- Rodríguez, M.B., E. Gobbetti, J.I. Gutián, M. Makhinya, F. Marton, R. Pajarola, & S. Suter (2014). State-of-the-art in Compressed GPU-Based Direct Volume Rendering. In *Computer Graphics Forum*, Volume 33, pp. 77–100.
- Sato, Y., N. Shiraga, S. Nakajima, S. Tamura, & R. Kikinis (1998). Local Maximum Intensity Projection (LMIP): A New Rendering Method for Vascular Visualization. *Journal of Computer Assisted Tomography* 22(6), 912–917.
- Wu, Q., F. Merchant, & K.R. Castleman (2008, April). *Microscope Image Processing* (1st ed.). Academic Press.
- Zhou, Z., Y. Tao, H. Lin, F. Dong, & G. Clapworthy (2011). Shape-Enhanced Maximum Intensity Projection. *The Visual Computer* 27(6–8), 677–686.

# Chaos theory-based quantification of ROIs for mammogram classification

Jarosław Kurek & Bartosz Świderski

*The Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences, Warsaw, Poland*

Sami Dhahbi

*Research Team on Intelligent Systems in Imaging and Artificial Vision (SIIVA)—RIADI Laboratory, Ariana, Tunisia*

Michał Kruk

*The Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences, Warsaw, Poland*

Walid Barhoumi

*Research Team on Intelligent Systems in Imaging and Artificial Vision (SIIVA)—RIADI Laboratory, Ariana, Tunisia*

Grzegorz Wieczorek

*The Faculty of Applied Informatics and Mathematics, Warsaw University of Life Sciences, Warsaw, Poland*

Ezzeddine Zagrouba

*Research Team on Intelligent Systems in Imaging and Artificial Vision (SIIVA)—RIADI Laboratory, Ariana, Tunisia*

**ABSTRACT:** The paper presents improved method for breast cancer diagnosis. The previous method has been presented in Dhahbi et al. 2015. The suggested improvements regards applying more diagnostic features such as wavelet packet decomposition, Hilbert matrix, fractal texture features, etc. Moreover we investigated several classifiers such as Random Forest, Support Vector Machine and Decision Tree. In this paper larger database (number of images/trials) has been used and reached better accuracy.

**Keywords:** mammography; diagnostic feature; breast cancer diagnosis; wavelet transform

## 1 INTRODUCTION

Breast cancer is one of the common women-cancer worldwide. It does not matter which region we take into consideration: the developed or less developed world, statistic states that over 508 000 women died in 2011 due to breast cancer [1] and 522 000 in 2012 [2].

Breast cancer is arising from cells of the breast that develops locally in the breast and metastasizes to the lymph nodes and internal organs (e.g. lungs, liver, bones and brain). Breast cancer is the most common malignancy in women [1]. This represents approximately 23% of all cases of cancers in women and approximately 14% of deaths because of this [2]. It is estimated that each year breast cancer is diagnosed in 1.5 million women worldwide, and about 522,000 die because of

this [2]. It is the most common cancer among people of highly civilized countries, such as USA, Canada, Australia and Western European countries. Least of breast cancer recorded in southern Asia and Africa. Breast cancer, which is the most common cancer in women, in men is fortunately rare.

To cope with this issue, screening program is applied. Usually for women this screening program made up of palpable breast examination by a doctor (like self-examination performed regularly carried out by the woman herself) and mammography (ionizing radiation to create mammography images). Such research is usually done in women over 50 years of age every three years, although many experts believe that it is worth it carried out every two years. In the USA, women in first-degree relatives of individuals who develop breast cancer

are advised to use mammography every 2 years starting at age test 10 years younger than that in which the person became ill relative.

Screening mammograms can detect lumps or other breast abnormality at an early stage when they are not detectable by a woman or doctor, which increases very significantly the chance of cure [3]. A significant negative effect of the research is a phenomenon observed that the negative result of mammography (i.e. not confirming the presence of the disease) in many women creates a false sense of security and reluctance and negligence in systematic self-examination of the breast. This procedure is wrong, because in about 20–25% cases, mammography cannot detect a cancer developing (these are called false negative results).

The Mammography screening is currently the best tool for nationwide mammogram screening programs reducing the mortality rate of breast cancer [4, 5]. Hence the improved methods of the breast cancer are usually based on the analysis of mammography images.

## 2 DATABASES

Total number of mammography images used in this paper is 10168 and it is a part of Digital Database for Screening Mammography (DDSM University of South Florida) which is currently the largest public mammogram database [6]. The dataset includes 2604 cases. Screening mammography typically involves taking two views of the breast, from above (Cranial-Caudal view, CC) and from an oblique or angled view (Medio-Lateral-Oblique, MLO). Hence when we take into consideration left and right breast, we have 4 images for every case in the used challenging dataset.

### 2.1 Region of Interest (ROI) Cropping

This is an important step to extract and focus on appropriate part of mammography image. Since the whole mammographic image comprises the pectoral muscle and the background with a lot of noise, features were computed on a limited ROI that contains the prospective abnormality [2]. Thus, cropping a region of interest removes the unwanted parts. Image cropping was performed manually based on the physician annotation provided in the dataset. We can describe ROI as rectangular space where in its center we can observe the lesion (Fig. 1).

### 2.2 Trial subsets used in experiments

Based on the above-described database, 10168 ROIs have been used in this work among the 10416 available ones [7]. Some number of all available

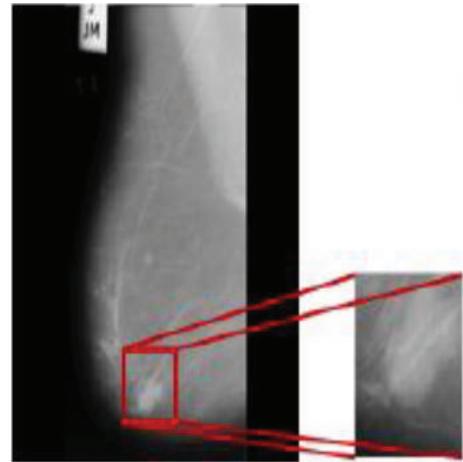


Figure 1. Example of Region of Interest (ROI) extraction.

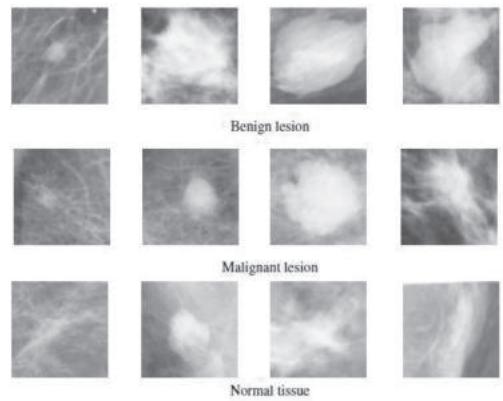


Figure 2. Sample of used ROIs from DDSM database.

ROIs has been rejected due to small quality of image. The dataset consists of three group of mammography result (Fig. 2):

1. Normal tissue—8254 ROIs
2. Benign lesion—862 ROIs
3. Malignant lesion—1052 ROIs

Examples of ROIs for every of three results are depicted on Figure 2.

Due to the complexity of separating the three classes, we decided to merge ROI's belong to benign lesions and malignant ones. It means the issue is reduced to binary classification of normal tissue and abnormal tissue trials. Finally, we obtained:

1. Normal tissue—8254 ROIs
2. Abnormal tissue—1914 ROIs

### 3 FEATURES GENERATION

Based on the all extracted ROI's, we generated 122 potential diagnostic features for binary classification. Dataset of potential diagnostic features consist of the following group:

- features generated based on Hilbert's image representation (23 features).
- features generated based on segmentation-based fractal texture analysis (36 features).
- Kolmogorov-Smirnov descriptors (38 features).
- statistics from the gray-level co-occurrence matrix and other statistic (15 features).
- Maximum sub-region descriptors (10 features) using K-S statistic and Minkowsky approach.

#### 3.1 Hilbert's image representation

Every image, which has regular representation in form of matrix of size  $n \times m$ , has been transformed to one vector using our previously proposed Hilbert curve and recursive modeling. Building Hilbert curve of order 4 is depicted on Figure 3.

#### 3.2 Trial subsets used in experiments

By means of Hilbert transformation of every image, we can obtain image vector representation for each image, which can be further treated as "time-series". Based on every image vector  $X$  representation, (where  $n$  denoted the vector length), we can calculate the following potential features:

- Mean:

$$\mu = \frac{1}{n} \sum_{i=1}^n X_i$$

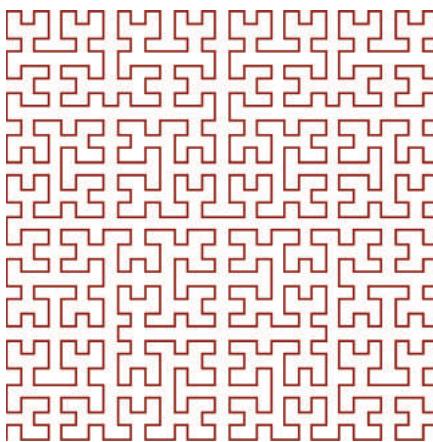


Figure 3. Example of Hilbert curve, which is a space filling curve that visits every pixel in a matrix (image).

- Standard deviation:

$$\sigma = \frac{1}{n} \sqrt{\sum_{i=1}^n (x_i - \mu)^2}$$

- Variance:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

- Skewness:

$$S = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \mu}{\delta} \right)^3$$

- Kurtosis:

$$K = \frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \mu}{\delta} \right)^4 - 3$$

- Root Mean Squared (RMS):

$$RMS = \sqrt{\frac{\sum_{i=1}^n |x_i|}{n}}$$

- Crest factor (peak-to-rms ratio):

$$C = \frac{X_{peak}}{X_{RMS}}$$

- Wavelet energy of 4-level wavelet packet decomposition of db10 wavelet family:

After wavelet packet decomposition, the portion of energy for every terminal node has been calculated based on the following equation:

$$E = \sum_{k=1}^N |S_{jk}|^2$$

where,  $S_{jk}$  is appropriate coefficient discrete wavelet transformation.

#### 3.3 Segmentation-based fractal texture analysis

This approach let us generate 36 texture features based on SFTA algorithm (Segmentation-based Fractal Texture Analysis) and returns  $1 \times 6$  vector  $D$  extracted from the input grayscale image. When we apply 6th fractal order, we can obtain 36 texture features.

The SFTA algorithm decomposes images into various thresholded images using several sets of lower and upper threshold values. Implementation has been based on Costa approach [9]. Thresholded images are used to extract the fractal dimension. Since images with more jagged edges and prominent color differences tend to have higher fractal dimension, images with more

uniform texture properties will have closer fractal dimension.

### 3.4 Kolmogorov-Smirnov descriptors

Kolmogorov-Smirnov descriptors reflect the change distribution of intensity of pixels placed in the rings of the increasing geometrical distances from the central point. The division of the images into coaxial rings is illustrated in Figure 4. It represents 4 concentric rings of equal number (56) pixels in each ring (equal number of pixels leads to more stable distribution of KS statistics) [10].

The central point is travelling along pixels uniformly distributed in the image. The results of the statistical analyses of these coaxial rings will be combined together by concatenating the pixel intensities corresponding to the same rings placed in equal distances, at different positions of the central pixel. Then cumulative Kolmogorov-Smirnov (KS) distances between the intensity of pixels  $x_i$  and  $x_j$  belonging to two different rings using KS test is then estimated. Based on above approach we can generate the following 7 features:

- dKS\_12 (mean KS statistics between ring no 1 and ring no 2).
- dKS\_13 (mean KS statistics between ring no 1 and ring no 3).
- dKS\_14 (mean KS statistics between ring no 1 and ring no 4).
- the ratio dKS\_13/dKS\_12.
- the ratio dKS\_14/dKS\_12.
- the coefficient  $\alpha_0$  of the approximation line  $d_{KS} = \alpha_0 + \alpha_1 l + \varepsilon$ .
- the slope coefficient  $\alpha_1$  of the approximation line  $d_{KS} = \alpha_0 + \alpha_1 l + \varepsilon$ .

Generally, we obtained 14 K-S features: 7 for mean and 7 for median approach. Rest of 24 similar

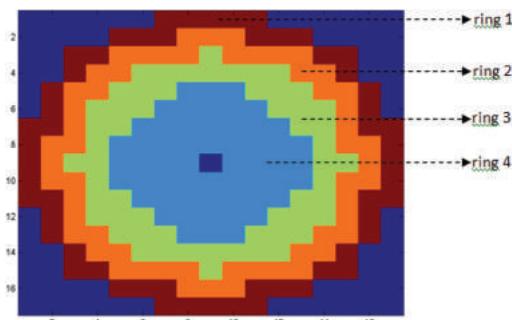


Figure 4. Illustration of 4 coaxial rings around the central pixel.

features generated based on K-S approach has been described in [10].

### 3.5 Statistics from the gray-level co-occurrence matrix and other statistic

Four features have been generated using Gray-Level Co-occurrence Matrix (GLCM) from a given image. GLCM calculates how often a pixel with gray-level (grayscale intensity) value  $i$  occurs horizontally adjacent to a pixel with the value  $j$ . Based on GLCM approach, the following statistic features were generated:

- Contrast: measure of the intensity contrast between a pixel and its neighbour over the whole image.
- Correlation: measure of how correlated a pixel is to its neighbour over the whole image.
- Energy: sum of squared elements in the GLCM.
- Homogeneity: value that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

Further 11 features are generated from general statistic measures of the treated image as one-dimensional vector. Hence, the following features were generated:

- median
- mean
- standard deviation
- kurtosis
- min
- max
- high order cumulates and moments

### 3.6 Maximum sub-region descriptors using K-S statistic and Minkowsy approach

The box-counting fractal dimension is measure characterizing the fractal complexity. It is the particular case of the Mandelbrot fractal dimension and is based on the notion of self-similarity of the structure at different scales. It measures how the length of the complex curve is changing when the measurement is performed with the increased accuracy [11].

To characterize any curve, this curve is covered with the set of regular squared areas of the size  $\varepsilon$ . Thus, we have to calculate the number of squared areas containing any part of the given curve. Number of founded areas is denoted as  $N(\varepsilon)$ .

$$d = \lim_{\varepsilon \rightarrow \infty} \frac{\log(N(\varepsilon))}{\log(\varepsilon)}$$

In our approach, we have used 8 boxes. We suggested 122 various features based on which we can choose the best subset for binary classification.

Table 1. Result of binary classification (normal vs. abnormal tissues) of mammography images.

Setting	Accuracy [%]	Standard deviation [%]
SVM	80%	5.1%
Decision tree	79%	5.5%
Random forest	81%	4.5%

#### 4 FEATURES SELECTION

To choose appropriate diagnostic features, from all 122 potentially available, the best separating two class of tissue (normal and abnormal) sequential feature selection has been performed. This approach selects a subset of features from the data matrix  $X$  that best predicts the data in  $y$  by sequentially selecting features until there is no improvement in prediction [8]. There is possibility to apply chosen classifier (e.g. no-linear) for this selection method e.g. SVM, random forest, decision tree. For every candidate feature subset, sequential feature selection performs 10-fold cross-validation by repeatedly calling function with different training subsets of  $X$  and  $y$ , XTRAIN and ytrain, and test subsets of  $X$  and  $y$ , XTEST and ytest. The result is logical vector indicating which features are finally chosen. After applying sequential features selection, we obtained 48 diagnostic features from the 122 available ones.

#### 5 NUMERICAL EXPERIMENTS

The authors have applied three type of classifier: SVM, decision tree and random forest. For numerical experiments, the equal subset of normal and abnormal tissues trials have been chosen in randomize approach. The 10-fold stratified cross validation has been performed to compute the misclassification error for each classifier. Results are given in Table 1.

#### 6 CONCLUSION

The obtained accuracy result ( $= 81\%$ ) is more reliable than the one presented ( $= 86.46\%$ ) in [2],

because numerical tests are performed on 3828 total trials: 1914 trials of abnormal tissue and 1914 trials of normal tissue randomly chosen from 8254 available trials. In [2], accuracy is 86.46% but numerical experiments have been performed based on only 200 trials (ROIs): 100 trials normal and 100 trials abnormal. So, current method presented in this paper is more reliable based on the chosen selected features. The best result obtained for random forest classifier applying 200 trees in this algorithm.

#### REFERENCES

- [1] "Breast cancer: prevention and control", WHO 2015.
- [2] S. Dhahbi, W. Barhoumia, E. Zagroubaa "Breast cancer diagnosis in digitized mammograms using curvelet moments", Comp. Biol. Med. 64 (2015) 79–90.
- [3] A. Jotwani, J. Gralow, Early detection of breast cancer, Mol. Diagnosis and Ther. 13(6) (2009)349–357.
- [4] H. Nelson, K. Tyne, A. Naik, C. Bougatsos, B. Chan, P. Nygren, L. Humphrey, Screening for breast cancer: Systematic evidence review update for the U. S. preventive services task force, Ann. Intern. Med. 151(10) (2009) 727–W242.
- [5] S. Hofvind, G. Ursin, S. Tretli, S. Sebuodegård, B. Moller, Breast cancer mortality in participants of the norwegian breast cancer screening program, Cancer 119 (2013) 3106–3112.
- [6] M. Heath, K. Bowyer, D. Kopans, W. Kegelmeyer, R. Moore, K. Chang, S. Munishkumar, Current status of the digital database for screening mammography, in Digital Mammography, Springer Netherlands (1998) 457–460.
- [7] M. Jiang, S. Zhang, H. Li, N. Metaxas, Computer-aided diagnosis of mammographic masses using scalable image retrieval, IEEE Transactions on Biomedical Engineering 62(2) (2015) 783–792.
- [8] Matlab user manual, MathWorks, Natick, 2014.
- [9] Costa, Alceu Ferraz, Humpire-Mamani, Gabriel, and Traina, Agma Juci Machado. An efficient algorithm for fractal analysis of textures. 2012 25th SIBGRAPI Conference on Graphics, Patterns and Images 39–46, 2012.
- [10] Swiderski, S. Osowski, M. Kruk, J. Kurek, Texture characterization based on the Kolmogorov-Smirnov distance, Expert Systems with Applications 42(1) (2015) 503–509.
- [11] Schroeder M., Fractals, Chaos, Power Laws, 6 Ed. New York: W.H. Freeman and Company, 2006.

This page intentionally left blank

# M5L: A web-based Computer Aided Detection system for automated search of lung nodules in thoracic Computed Tomography scans

A. Traverso & M. Agnello

*Dipartimento di Scienza Applicata e Tecnologia, Politecnico di Torino, Italy  
Istituto Nazionale di Fisica Nucleare Sezione di Torino, Italy*

P. Cerello, M. Saletta & S. Bagnasco

*Istituto Nazionale di Fisica Nucleare Sezione di Torino, Italy*

C. Peroni & E. Fiorina

*Dipartimento di Fisica, Universita' degli Studi di Torino, Italy  
Istituto Nazionale di Fisica Nucleare Sezione di Torino, Italy*

M.E. Fantacci

*Dipartimento di Fisica, Universita' degli Studi di Pisa, Italy  
Istituto Nazionale di Fisica Nucleare Sezione di Pisa, Italy*

A. Retico

*Istituto Nazionale di Fisica Nucleare Sezione di Pisa, Italy*

E. Lopez Torres

*CEADEN, Havana, Cuba  
Istituto Nazionale di Fisica Nucleare Sezione di Torino, Italy*

**ABSTRACT:** M5L, a Web-based fully automated Computer-Aided Detection (CAD) system for the automated detection of lung nodules in thoracic Computed Tomography (CT), is based on a multi-thread analysis with two independent CAD subsystems, the lung Channeled Ant Model (lungCAM) and the Voxel-Based Neural Analysis (VBNA), and on the combination of their results. The M5L performance, extensively validated on 1043 CT scans from 3 independent data-sets, including the full LIDC/IDRI database, is homogeneous across the databases: the sensitivity is about 0.8 at 6–8 False Positive findings per scan, despite the different annotation criteria and acquisition and reconstruction conditions. In order to make CAD algorithms and their results available to users without requiring the installation of CAD software or new hardware for CAD computations, the project has proposed a Cloud SaaS (Software as a Service) approach composed by three main building blocks: a front-end web which handles the workflow, image upload, CAD results notification and direct on-line annotation of the exam by the radiologist; the OpenNebula-based cloud IaaS (Infrastructure as a Service) batch farm allocates virtual computing and storage resources; the M5L CAD provides the nodule detection functionality.

## 1 INTRODUCTION

Lung Cancer is one of the main public health issues in developed countries, accounting for about 19% and 28% of cancer-related deaths in Europe (Ferlay et al. 2010) and the United States of America (Jemal et al. 2009), respectively, with a five-year survival rate of only 10–16% (Jemal et al. 2010). Lung cancer is commonly manifesting itself as non-calcified pulmonary nodules. Computed Tomography (CT) has been shown to be the most

sensitive image way for the detection of small pulmonary nodules: low dose high resolution CT-based screening trials are regarded as a promising technique for detecting early-stage lung cancers (Team et al. 2011). The identification of early stage pathological Region of Interests (ROIs) in low dose high resolution CT scans is a very difficult task for radiologists, considering as first main issue the high number (300/400) of noisy slices to be analyzed. In order to support radiologist, researchers have started developing CAD methods

to be applied to CT examinations. Several studies (Das et al. 2006) (Brochu et al. 2008) (Matsumoto et al. 2008) reported an improvement in the sensitivity of radiologists when assisted by CAD systems. In addition, CAD systems act as detection rates equalizers between observers of different level of experience (Brown et al. 2005). The most common approach in hospitals, in order to make CAD algorithms available in clinical routine, is the deployment of standalone workstations, usually equipped with a provider dependent Graphic User Interfaces (GUI). This approach presents several drawbacks, such as the high fixed cost of the software licenses or the difficulty in upgrading the software release. Furthermore, the computational power required by CAD computations is often very high (increasing with the complexity of the algorithms) and unpredictable (for example during large screening campaigns when a big amount of data need to be analyzed as faster as possible), often requiring new powerful and more expensive hardware to be bought with impacts on cost for the hospitals. The emerging of Cloud computing, accessible via secure Web protocols, solves almost at all the previous issues. A further important reason favors the SaaS (Software as a Service) approach: the possibility of combining different CADs with no particular effort. It was shown that the combination of multiple CAD algorithms increases the overall sensitivity (van Ginneken et al. 2010). This paper is aiming at presenting the M5L on-demand CAD system developed within the MAGIC-5 collaboration. The MAGIC-5 project, funded by the Istituto Nazionale di Fisica Nucleare (INFN, Italy) and Ministero della Universita' e della Ricerca (MIUR, Italy) is aiming at:

- Developing and combining CAD algorithms for lung nodules identification.
- Making CAD algorithms available to users without requiring the installation of dedicated software or high cost hardware.
- Validating the system inside the clinical routine in hospitals, studying the impact of CAD algorithms in the overall performance of the radiologists.

## 2 THE ALGORITHMS

M5L is the combination of two independent CAD subsystems: the Channeler Ant Model (CAM) and the Voxel-Based Neural Approach (VBNA). These two algorithms have a common start line, which is the parenchymal volume segmentation using a 3D region growing algorithm, which produces the separation of trachea, bronchi and lungs. The results are saved as mask of bits and used to perform further steps.

### 2.1 THE *lungCAM CAD*

The lungCAM structure follows a standard approach: the pre-processing stage (equalisation and lung volume segmentation) is followed by a search for Regions Of Interest (ROIs), an analytic filter and a neural classifier. Before starting the current analysis, CT scans in DICOM standard format are pre-processed to reduce the noise contribution: each 2D slice is analysed with a Savitzky-Golay filter (Rajagolopan, 2003) that provides noise reduction without loss of resolution. From then on, every step of the lungCAM algorithm is intrinsically 3-dimensional.

#### 2.1.1 Lung segmentation

The lung segmentation (De Nunzio et al. 2011) proceeds according to four main steps:

- Analysis of the CT Hounsfield Unit level distribution and evaluation of the intensity threshold to be applied in the following stages.
- 3D region growing of the lung volume with the detected threshold.
- Wave-front algorithm for the definition of the lung surface on the inner side and the removal of the trachea and the main bronchi.
- Morphological closing with a cylinder from the out-side in order to include pleural nodules and close the holes left by vessels.

#### 2.1.2 ROI hunting

The segmentation algorithm is performed with the Channeler Ant Model (CAM) (Cerello et al. 2010) based on Virtual Ant Colonies. The ant life cycle is a sequence of atomic time steps, during which the behaviour is determined by a set of rules that control the pheromone release, the movements and the variations of the ant energy, a parameter related to breeding and death. The lung internal structures are segmented by iteratively deploying ant colonies in voxels with intensity above a pre-defined threshold (anthills). Ants live according to the model rules until the colony extinction: the pheromone deposition generates pheromone maps. By an iterative thresholding of pheromone maps a list of ROI candidates is obtained. ROIs with a radius larger than 10 mm are post-processed in order to disentangle nodules attached to internal lung structures like vessels and bronchi. When no more voxels meet the condition to become an anthill, the information provided by the global pheromone map is analysed.

#### 2.1.3 Filtering and classification

Ideally, any computable quantity which is expected to show a different pattern for true nodules and false candidates would be a useful feature for the neural network. However, the choice to select a

small number of features for the neural classifier training aims at optimising the generality and keeping the performance stable as the validation data-set size increases.

Selected features are shown in Figure 1. The average number of ROIs after the nodule hunting, depending on the number of slices, ranges between several hundreds to few thousands per CT scan, a number far too large to be used as input for a neural network classifier. The vast majority of findings is easily rejected with an analytic filter based on correlations between the radius, the sphericity and the fraction of voxels connected to the lung mask. A Feed Forward Neural Network (FFNN) was selected as nodule candidate classification method. The training sample was made of 5 and 69 CTs from the ANODE09 and LIDC/IDRI databases, respectively. The training was carried on in cross-validation mode. The FFNN configuration was defined as follows: 13 input neurones, 1 hidden layer with 25 neurones and 1 neurone in the output layer, representing the probability of the finding to be relevant.

## 2.2 The VBNA CAD

The VBNA CAD system deals differently with internal and juxtapleural nodules, by means of two dedicated procedures: CADI for internal and CADJP for juxtapleural nodules (Retico et al. 2008) (Retico et al. 2009) (Camarlinghi et al. 2011). Both are three-step procedures. The first step consists in the lung segmentation; the second step consists in the ROI (Region Of Interest) hunter and performs the candidate nodule selection; the third step consists in the FP reduction.

### 2.2.1 Segmentation

The 3-dimensional segmentation algorithm is based on four main steps. Once the scans have been isotropically resampled, to separate the low-intensity lung parenchyma from the high-intensity surrounding tissue (fat tissue and bones), the voxel

intensities are thresholded at a fixed value. Vessels and airways are not included in the segmented lung at this stage since their volume is outside the segmented lung volume. To include them without modifying the pleura surface morphology a combination of morphological operators (dilation and erosion) is applied. Finally, the logical OR operation between the so obtained mask and the original lung mask provides the final mask. The identified lung mask is used for CADI, whereas its boundary is used for CADJP.

### 2.2.2 ROI hunting for internal nodules

In the CADI, the internal nodules are modelled as spherical objects with a Gaussian profile, following the approach proposed in (Li et al. 2003). To detect this kind of objects (Retico et al. 2008), a dedicated Dot-Enhancement (DE) filter, determining local characteristics of each voxel by the eigenvalues of the Hessian matrix, is implemented. To enhance the sensitivity of this filter to nodules of different sizes, a multi-scale approach combining the DE function with Gaussian smoothing is performed following prescriptions by (Li et al. 2003).

### 2.2.3 ROI hunting for juxtapleural nodules

In the CADJP (Retico et al. 2009) pleura surface normals (evaluated using the triangular mesh representing the pleura surface) are constructed and each voxel is assigned a score proportional to the number of normals intersecting in it. Since the evaluation of the normal intersections in the real 3D space is a complex and computationally intensive operation, it is implemented in the voxel space. To deal with noise, cylinders with Gaussian profile are considered instead of segments (Paik et al. 2004). This information is collected in the score matrix S (x, y, z). The local maxima of the 3D matrix S (x, y, z) are the juxtapleural candidate nodule locations.

### 2.2.4 Classification

In order to classify the candidate nodule findings obtained in the previous step, an original procedure, the Voxel-Based Neural Approach (Gori et al. 2007), performs the reduction of the number of FPs in the lists of internal and juxtapleural candidate nodules. First, a ROI including voxels belonging to the candidate nodule is defined from each location provided by the previous step. The basic idea of the VBNA is to associate with each voxel of a ROI a feature vector defined by the intensity values of its 3D neighbours (in this case  $5 \times 5 \times 5$  intensity values) and the eigenvalues of the gradient matrix and of the Hessian matrix. Support Vector Machines (SVM), by which have been obtained better results, are implemented for the classification procedure. The training sample

Geometrical features	Intensity-related features
Center of gravity $Xi=x,y,z(*)$	Average
Radius (mm)	Average outside mask
Sphericity	Std. Deviation
Skewness of distance from $Xi$	Std. Deviation outside mask
Kurtosis of distance from $Xi$	Maximum
Volume (mm <sup>3</sup> ) (*)	Entropy
Fraction of voxels connected to the pleura	Entropy outside mask

Figure 1. List of features extracted from the nodule output mask. Features labeled with the asterisk were not used in the classification stage.

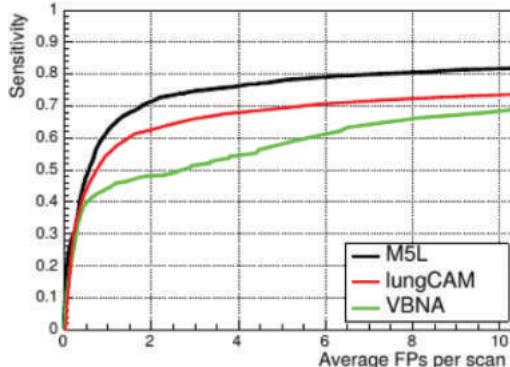


Figure 2. FROC curves of the lungCAM and VBNA subsystems and of their M5L combination on the LIDC test validation dataset (949 CT scans).

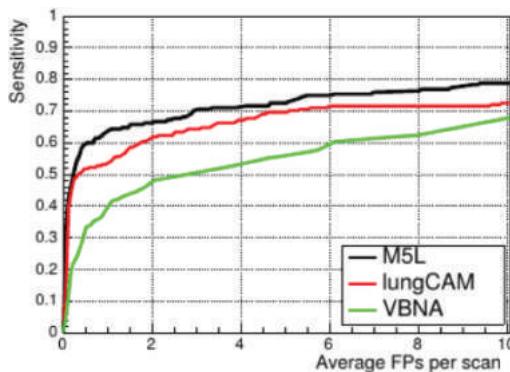


Figure 3. FROC curves of the lungCAM and VBNA subsystems and of their M5L combination on the ANODE09 test validation dataset (50 CT scans).

was made of 69 CTs from the LIDC/IDRI database. At the end of this step, each ROI is assigned a degree of suspicion averaging the score of all the voxels belonging to it.

### 2.3 Combination and results

To obtain the combined M5L result the following matching criterion has been used: a CAD finding is considered a true positive if its Euclidean distance from the center of the lesion annotated by the radiologists is less than 1.5 times the radius of the annotated lesion. The results have been evaluated in terms of FROC (Free-response Receiver Operating Characteristic) curves.

Figure 2 shows the results obtained for the lungCAM and VBNA separate subsystems and for the combined M5L on the 949 scans of the LIDC test dataset (949 scans). The M5L sensitivity at 8 FP/

scan reaches 80% which, given the size and heterogeneity of the dataset, is quite remarkable (Torres et al. 2015).

In the case of ANODE09 the FROC curves are shown in Figure 3.

## 3 THE M5L WEB-BASED ON-DEMAND SYSTEM

In order to make CAD algorithms and their results available to radiologists through a user-friendly web interface, without limiting the users to install standalone code on their workstations and in order to guarantee an adaptable computational power to run CAD algorithms, the project has built a system which is basically composed by the following subsystems:

- The Web front-end: The Web front-end is thought as a service to help doctors and reviewers in the world to exchange imaging studies and compare diagnoses on the same studies. The front-end handles the part of CT submission for analysis, the possibility to insert on-line the medical review of the study case and to access to CAD results.
- The Cloud computing back-end: the cloud computing backend is though as a service to guarantee flexibility in the available computing resources. It allows to allocate computing resources (i.e. virtual machines) according to the user request. The back-end handles the part of algorithms computation.

### 3.1 The WEB front-end

The service is available as a web application accessible with every browser from desktop and mobile devices. Having proper credentials, DICOM images can be uploaded to the remote repository and reviewers can insert their medical diagnosis and see other reviewers ones. The notification system is based on emails which basically advertises the submitters when CAD results are ready and advertise the reviewer that a new case has been associated to him for review. The entire front-end web has been developed using the free and open source tool DRUPAL, an open source content management platform powering millions of websites and applications (Coombs 2009). Drupal core is modular, defining a system of hooks and callbacks, which are accessed internally via an API (Application Programming Interface), allowing third-party contributed modules. We have essentially built two different modules, and the corresponding user profiles which are briefly described in the following sections.

### 3.1.1 Submission module

This module is conceived to be used by a technical radiologist (which acts as submitter role), uploads in zip format the CT study to be submitted and selects a radiologist which will review the study inserting the annotation. All CAD computations will be performed in a separated cloud-back end, this feature allows the user to submitted in sequence the cases without having to wait for CAD results. Before being submitted for computation, all cases are subject to further checks, such as re-anonymise the case if not yet.

### 3.1.2 Review module

This module is conceived to be used by a radiologist. Accessing with proper credentials (as reviewer role) he/she can insert the medical annotation of cases assigned to him/her for review during the submission process. A notification system, based on e-mail, tells the radiologist when CAD results will be ready. He will be able to access the results in different formats such as DICOM Structured Report, HTML, PDF and review his annotation according to these results. An additional plugin for the medical imaging viewer Osirix has been developed to see CAD marks directly on the exams.

## 3.2 The cloud back-end

One of the issues when running CAD algorithms and combining their results is that the required computing power is large and often unpredictable. Furthermore, if the tool is to be used in potentially massive screening campaigns, which can be concentrated in a given period of the year, some flexibility is required in the available computing resources. This can be referred to as the *scientific elastic computing paradigm*, in which two requirements need to be satisfied:

- An application must be able to easily and quickly expand its resources during peak times.
- Resources should never get wasted (i.e. at any given time the amount of idle resources should be minimal).

In the past, there have been some attempts to use an approach similar to the GRID infrastructure used in high energy physics to overcome those problems (Bellotti et al. 2007) (Lamanna 2004). The main issue with GRID computing is the rigidity, complexity of the structure and the man power required to manage the system. Furthermore, this solution does not fit the majority of Medical Physics projects, that require a custom environment. For previous reasons, the use of Cloud Computing solutions is progressively growing (Mell and Grance 2011).

### 3.2.1 M5L elastic cluster at INFN torino cloud

The M5L CAD web service is hosted by the INFN Torino Computing Center, which is a Tier-2 of LHC (Large Hadron Collider) Computing Grid. In order to reduce the manpower effort to manage the Tier-2, support an increasing number of applications and offer computational power to small research projects with specific needs, a Private Cloud infrastructure has been created. The facility is managed by OpenNebula, a free and open-source Cloud Management Platform which allows hardware and virtual infrastructure control and monitoring, adding the possibility of virtual machine lifecycle management (Milojić et al. 2011). OpenNebula orchestrates storage, network, virtualisation, monitoring, and security technologies to deploy computing services as virtual machines on distributed infrastructures. The resources used by the M5L are one physical host for the web-server and several virtual machines for the computational power. The VMs are deployed in a specific sandbox (IaaS) with an internal private virtual network and a virtual router with a public address for the external access and the communication with the web server (that has another public IP). At this moment we are allowed to deploy up to 18 VMs, with a total of 48 cores. In order to achieve the capability to scale up or down the resources according to the workload, we decided to create an elastic cluster, composed by one head node and many workers, using CernVM Online (Buncic et al. 2010). With these tools, it is possible to contextualise the VMs, define the use of resources and enable (between other options) HTCondor and Elastiq. HTCondor is a specialised workload management system (batch system) for compute-intensive jobs. Users submit their serial or parallel jobs to HTCondor, HTCondor places them into a queue, chooses when and where to execute the jobs based upon a policy, carefully monitors their progress, and ultimately informs the user upon completion. Elastiq, on the other hand, is a lightweight Python daemon that allows a cluster of virtual machines running a batch system to scale up and down automatically. Each CT uploaded on the web service is copied on the head node of the cluster and the M5L CAD execution is controlled by HTCondor. If some CTs to be examined waits in the HTCondor queue over a defined period of time, new workers will be deployed up to the limits of resources. Conversely if a worker is idle over a defined period of time, it will be deleted. In this way when no exam is uploaded we occupy only a small part of the resources available, freeing the remaining for others users of the cloud. At the moment we are using VMs with 6 cores and 15 GB of RAM. Only the head node is always active. Stress tests are in progress to define the best configuration and the

prove the solidity of the system. With this configuration we are able to complete the analysis of a single CT in nearly 30 minutes, and process up to 42 exams simultaneously.

#### 4 CONCLUSIONS

The project has developed two CAD algorithms, and their combination, for automatic lung nodules identification. The validation has been performed over the biggest public available database with an overall sensitivity of about 80% at 6–8 FP per scan. This represents a very remarkable result, not only for the high sensitivity reached, but also because this represents the only work with a validation performed over such a big database. Furthermore, the project has proposed a new approach to make CAD algorithms and CAD results available to users without requiring the installation of software or hardware. In addition, the computational solution based on a cloud computing back-end allows to solve all issues described in Chapter 4 regarding the need of powerful resources for CAD computations. The prototype system has already been tested by some Italian hospitals with good results. One of the final goals of the project is to investigate the usage of CAD as *second reader* inside clinical routine, meaning that the radiologist will see CAD results only after having completed the annotation of the case. He/She will be then able to review his annotation after having seeing these results. The comparison of the annotations (before and after CAD) will allow to evaluate the impact of CAD into the performance of the radiologist. In order to achieve these goals the project has started a collaboration with the radiology department of the IRCCS (Institute for Cancer Research) in Candiolo (Turin, Italy). The project has decided to adopt a perspectival approach, acquiring exams of patients undergoing chest CT during the daily activity of the hospital (with a rate of about 50 CT scans per week). We have decided to adopt this approach because it will allow to have a validation with a data-set up to 1–2 thousands CTs per year, one order greater than data-set found in literature (Das et al. 2006) (Brochu et al. 2008).

#### REFERENCES

- Bellotti, R., P. Cerello, S. Tangaro, V. Bevilacqua, M. Castellano, G. Mastronardi, F. De Carlo, S. Bagnasco, U. Bottigli, R. Cataldo, et al. (2007). Distributed medical images analysis on a grid infrastructure. *Future Generation Computer Systems* 23(3), 475–484.
- Brochu, B., C. Beigelman-Aubry, J. Goldmard, P. Raffy, P. Grenier, & O. Lucidarme (2008). Thorax-évaluation de l'impact d'un système cad sur la performance des radiologues pour la détection des nodules pulmonaires sur des examens scanographiques multicoupes du thorax.
- Brown, M.S., J.G. Goldin, S. Rogers, H.J. Kim, R.D. Suh, M.F. McNitt-Gray, S.K. Shah, D. Truong, K. Brown, J.W. Sayre, et al. (2005). Computer-aided lung nodule detection in ct: Results of large-scale observer test 1. *Academic radiology* 12(6), 681–686.
- Buncic, P., C.A. Sanchez, J. Blomer, L. Franco, A. Harutyunian, P. Mato, & Y. Yao (2010). Cernvm—a virtual software appliance for lhc applications. In *Journal of Physics: Conference Series*, Volume 219, pp. 042003. IOP Publishing.
- Camarlinghi, N., M. Fantacci, I. Gori, & A. Retico (2011). A computer-aided detection system for lung nodules in ct images.
- Cerello, P., S.C. Cheran, S. Bagnasco, R. Bellotti, L. Bolanos, E. Catanzariti, G. De Nunzio, M.E. Fantacci, E. Fiorina, G. Gargano, et al. (2010). 3-d object segmentation using ant colonies. *Pattern Recognition* 43(4), 1476–1490.
- Coombs, K. (2009). Drupal done right. *Library journal* 134(19), 30–32.
- Das, M., G. Mhlenbruch, A.H. Mahnken, T.G. Flohr, L. Gndel, S. Stanzel, T. Kraus, R.W. Gnther, & J.E. Wildberger (2006). Small pulmonary nodules: Effect of two computeraided detection systems on radiologist performance 1. *Radiology* 241(2), 564–571.
- De Nunzio, G., E. Tommasi, A. Agrusti, R. Cataldo, I. De Mitri, M. Favetta, S. Maglio, A. Massafra, M. Quarta, M. Torsello, et al. (2011). Automatic lung segmentation in ct images with accurate handling of the hilar region. *Journal of digital imaging* 24(1), 11–27.
- Forlay, J., H.-R. Shin, F. Bray, D. Forman, C. Mathers, & D.M. Parkin (2010). Estimates of worldwide burden of cancer in 2008: Globocan 2008. *International journal of cancer* 127(12), 2893–2917.
- Gori, I., F. Bagagli, M.E. Fantacci, A.P. Martinez, A. Retico, I. De Mitri, S. Donadio, C. Fulcheri, G. Gargano, R. Magro, et al. (2007). Multi-scale analysis of lung computed tomography images. *Journal of Instrumentation* 2(09), P09007.
- Jemal, A., R. Siegel, E. Ward, Y. Hao, J. Xu, & M.J. Thun (2009). Cancer statistics, 2009. *CA: a cancer journal for clinicians* 59(4), 225–249.
- Jemal, A., R. Siegel, J. Xu, & E. Ward (2010). Cancer statistics, 2010. *CA: a cancer journal for clinicians* 60(5), 277–300.
- Lamanna, M. (2004). The lhc computing grid project at cern. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 534(1), 1–6.
- Li, Q., S. Sone, & K. Doi (2003). Selective enhancement filters for nodules, vessels, and airway walls in two- and threedimensional ct scans. *Medical physics* 30(8), 2040–2051.
- Matsumoto, S., Y. Ohno, H. Yamagata, D. Takenaka, & K. Sugimura (2008). Computer-aided detection of lung nodules on multidetector row computed tomography using three dimensional analysis of nodule candidates and their surroundings. *Radiation medicine* 26(9), 562–569.

- Mell, P. & T. Grance (2011). The nist definition of cloud computing. Milojičić, D., I.M. Llorente, & R.S. Montero (2011). Open nebula: A cloud management tool. *IEEE Internet Computing* (2), 11–14.
- Paik, D.S., C.F. Beaulieu, G.D. Rubin, B. Acar, R.B. Jeffrey Jr, J. Yee, J. Dey, & S. Napel (2004). Surface normal overlap: a computer-aided detection algorithm with application to colonic polyps and lung nodules in helical ct. *Medical Imaging, IEEE Transactions on* 23(6), 661–675.
- Retico, A., P. Delogu, M.E. Fantacci, I. Gori, & A.P. Martinez (2008). Lung nodule detection in low-dose and thinslice computed tomography. *Computers in biology and medicine* 38(4), 525–534.
- Retico, A., M.E. Fantacci, I. Gori, P. Kasaé, B. Golosio, A. Piccioli, P. Cerello, G. De Nunzio, & S. Tangaro (2009). Pleural nodule identification in low-dose and thinslice lung computed tomography. *Computers in biology and medicine* 39(12), 1137–1144.
- Team, N.L.S.T.R. et al. (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. *The New England journal of medicine* 365(5), 395.
- Torres, E.L., E. Fiorina, F. Pennazio, C. Peroni, M. Saletta, N. Camarlinghi, M. Fantacci, & P. Cerello (2015). Large scale validation of the m5l lung cad on heterogeneous ct datasets. *Medical physics* 42(4), 1477–1489.
- van Ginneken, B., S.G. Armato, B. de Hoop, S. van Amelsvoortvan de Vorst, T. Duindam, M. Niemeijer, K. Murphy, A. Schilham, A. Retico, M.E. Fantacci, et al. (2010). Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the anode09 study. *Medical image analysis* 14(6), 707–722.

This page intentionally left blank

# Non-intrusive and calibration free visual exploration analysis in children with Autism Spectrum Disorder

Dario Cazzato & Francesco Adamo

*Faculty of Engineering, University of Salento, Lecce, Italy*

Giuseppe Palestra

*Department of Computer Science, University of Bari, Italy*

Giulia Crifaci, Paola Pennisi & Giovanni Pioggia

*Institute of Clinical Physiology, National Research Council of Italy, Messina, Italy*

Liliana Ruta

*Department of Developmental Neuroscience, Stella Maris Scientific Institute, Calambrone, Pisa, Italy*

Marco Leo & Cosimo Distante

*National Institute of Optics, National Research Council of Italy, Arnesano (LE), Italy*

**ABSTRACT:** Assistive technology is a generic system that is used to increase, help or improve the functional capabilities of people with disability. Recently, its employment has generated innovative solutions also in the field of Autism Spectrum Disorder (ASD), where it is extremely challenging to obtain feedback or to extract meaningful data. In this work, a study about the possibility to understand the visual exploration in children with ASD is presented. In order to obtain an automatic evaluation, an algorithm for free gaze estimation is employed. The proposed gaze estimation method can work without constraints nor using additional hardware, IR light sources or other intrusive methods. Furthermore, no initial calibration is required. These relaxations of the constraints makes the technique particularly suitable to be used in the critical context of autism, where the child is certainly not inclined to employ invasive devices. In particular, the technique is used in a scenario where a closet containing specific toys, that are neatly disposed from the therapist, is opened to the child. After a brief environment exploration, the child will freely choose the desired toy that will be subsequently used during therapy. The video acquisition have been accomplished by a Microsoft Kinect sensor hidden into the closet in order to obtain both RGB and depth images, that can be processed by the estimation algorithm, therefore computing gaze tracking by intersection with data coming from the well-known initial disposition of toys. The system has been tested with children with ASD, allowing to understand their choices and preferences, letting to optimize the toy disposition for cognitive-behavioural therapy.

## 1 INTRODUCTION

Autism is a neurodevelopmental disorder strongly characterized by deficits in social interaction and impaired understanding of the mental states of others (Siegel, Vukicevic, & Spitzer, 1990), a dysfunction that persists even in people with autism who have IQs in the normal range. Autistic Spectrum Disorders (ASD) is no longer a rare impairment with an extremely poor prognosis, but a multi-factorial disorder whose qualitative impairments can be mitigated with early intensive intervention. Worldwide population prevalence is about 1%. Early diagnosis of ASD is a central topic of

research in developmental psychology. Signs of autism are not reliably present at birth, but emerge through a process of diminishing, delayed, or atypical development of social-communication behaviours, starting between the ages of 6 and 12 months. Intervention and support should be individualized and, if appropriate, multidimensional and multidisciplinary (Lai MC et al., *The Lancet*, 2014).

Aesthetic preferences of subjects with ASD are very different (Bogdashina, 2003). The preferred toy of a child can be intolerable for another child. Often, to understand whether an object is attracting or not the attention of an autistic

child can be impossible because autistic children can use gestures or vocalization to express their needs by protoimperative gestures, but usually they don't communicate objects of shared interest by protodeclarative gestures (Baron-Cohen, S., 1989; Curcio, F., 1978). Also in pre-verbal children, absence of protodeclarative gestures is considered a discriminating items between infants who had later been diagnosed with autism and typically developing infants (Clifford et al. 2007). Baron-Cohen et al. (1996) showed that absence of protodeclarative pointing, gaze-monitoring and pretend play in children 18 months olds, in 83.3% of cases predicts a future diagnosis of autism. In Modified Checklist for Autism in Toddlers, protodeclarative gestures are considered a key behaviour (Robins et al. 2009).

Autistic subjects show atypically gaze behaviours: in real social situations, subjects with ASD show reduced salience of eyes and increased salience of mouths, bodies and objects (Klin et al. 2002; Shic et al. 2011) and look significantly less at the partner in dyadic interaction rather than typically development children (Noris et al. 2012). In artificial social situations, (for example when they see a cartoon or a movie with social actors, or in static photos, or in virtual reality, etc.) they attend to characters' faces for less time than typically development subjects (Riby et Hancock 2009; Riby et Hancock 2008) and show less the centre of subjects' faces rather than control group (Trepagnier et al. 2002).

Visual atypical behaviours also affect the way in which subjects with ASD observe the scene; Shic et al. (2011) showed that children with ASD, in comparison to typically development subjects, are more attracted by background object. Many studies show an abnormal exploration of object stimuli (Ozonoff et al. 2008); the work of Sasson et al. (2008), for example, showed that, compared with typically development subjects, subject with ASD are more perseverative in fixations of details of images and that their fixations are more detail oriented, while Hutman et al. (2012) showed that infants later diagnosed with ASD are more attentive to non-social stimuli during interactions with an unfamiliar experimenter and that they shift attention among visual targets less frequently than high and low-risk typically developing children. By a simple preferential looking task, Pierce et al. (2011) showed that 40% of the autistic cross sample spent greater than 50% of viewing time fixating dynamic geometric images rather than dynamic social images. Several experimental protocols for the diagnosis and understanding of developmental disorders make use of video footage analysis to measure such elements as response time, attention changes and social interaction (Noris et al.,

2007). Many approaches to technology-enhanced intervention rely on educational methods shown to result in good outcomes and can be used to specify design principles needed for engineering successful technology-enhanced intervention tools (Tentori M, Hayes G, 2010).

Considering the scarcity of such specialized medical practitioners compared to the volume of patients, along with considering the high-cost involved, there is a strong motivation for investigating technological alternatives to solve the problem (Sivalingam et al., 2012). From the other side, computer vision technology has a unique opportunity to impact that study of children's behaviour, by providing a means to automatically capture behavioural data in a non-invasive manner and analyse behavioural interactions. Computational sensing and modelling techniques can play an important role in the capture, measurement, analysis, and understanding of human behaviour; this research area is called Behaviour Imaging.

In clinical settings, assessments of behaviour are typically based on the direct observation of a child by an experienced clinician. The ability to automatically measure behavioural variables using computer vision-based sensing could be valuable in enabling the collection of behavioural data on a large scale without requiring substantial human effort. Behaviour imaging technology can play several roles in support of a screening instrument. It can provide cost-effective tools for managing large collections of video and other data sources recorded during screening sessions. In particular, it can enable summarization, content-based retrieval, visualization, and comparison of observational data across populations and over time, to an extent that is not feasible using conventional manual methods (Rehg, 2011).

In this work, a study about a possibility to understand the visual exploration in children with ASD is presented. In order to obtain an automatic evaluation, an algorithm for free (i.e. without constraints nor using additional hardware, IR light sources or other intrusive methods) gaze estimation is employed. Furthermore, no initial calibration is required. In particular, the technique is used in a scenario where a closet containing specific toys, that are neatly disposed from the therapist, is opened to the child. After a brief environment exploration, the child will freely choose the desired toy that will be subsequently used during therapy. The video acquisition have been accomplished by a Microsoft Kinect sensor hidden into the closet in order to obtain a depth image that can be processed by the estimation algorithm, therefore computing gaze tracking by intersection with data coming from the well-known initial disposition of toys. The rest of the manuscript is organized as follows.

Section 2 introduces the proposed method in terms of employed system and selected participants. Section 3 explores the experimental setup and the sensitive data acquisition phase, necessary to evaluate results off-line, and results are presented, while Section 4 discusses them. Finally, Section 5 concludes the paper.

## 2 METHODS AND PARTICIPANTS

The employed solution is based on the work of (Cazzato et al. 2012). A block diagram of the proposed method is showed in Fig. 1.

Method's workflow is the following: first of all, images are acquired from a depth sensor, taking into account both RGB and depth channels. A search for a face over the input image is then carried out on the RGB image. A detected face is then tracked over the time, and a parameterized and people independent 3D face mask is subsequently overlapped on the human face. The face tracker is based on the Active Appearance Model (AAM), while the alignment is performed using Iterative Closest Point (ICP). The depth information allows to reconstruct the 3D positions of all of the tracked points and, therefore, to estimate the head pose in terms of yaw, pitch and roll angles. Starting from the average point between the two detected three-dimensional positions of the eyes and using the head pose information, a gaze track is finally estimated as well as its intersection with a vertical plane with regard to the ground and passing from the centre of the sensor, representing the observed point. Our gaze estimation method works as follows. First of all, the 2D position of the detected eye centre points are taken from the face mask. Note that small occlusions are handled, and the eye centre point is always estimated in the used model when the overlapping with the face successes. After that, the average value is taken, in order to take a point corresponding more or less with the nose septum and to use it as the origin of the gaze track. This value is converted into 3D coordinates with regard to a Cartesian coordinate system centred inside the sensor. Starting from the head pose information, two angles are taken, i.e.  $\omega_x$  and  $\omega_y$ , corresponding respectively to pitch and yaw. Then, the gaze track is computed and its intersection with a plane vertical with regard to the ground and passing from the centre of the sensor is calculated. The intersection point is computed separately for each angle and using the same method, and the Euclidean distance from the sensor can hereafter be computed. The procedure is showed for one angle in Fig. 2.

In case of translations on the  $x$  and  $y$  axes, the vector can be algebraically summed up with the

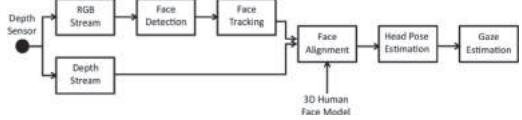


Figure 1. A block diagram of the proposed method.

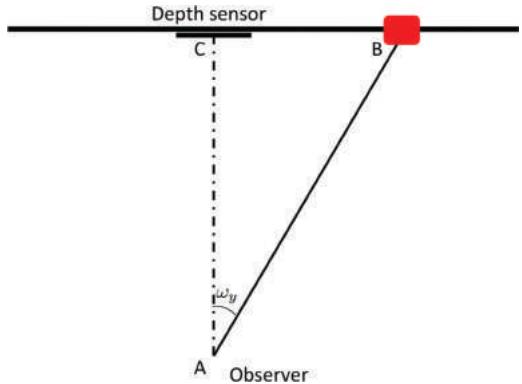


Figure 2. A scheme of the gaze estimation solution.

computed value, in order to translate the gaze vector to the right position. This way, the proposed method can be easily extended to compute the (possible) intersection with an object of interest in the plane. In this case, the virtual plane with lying objects is modelled with all the possible bounding boxes of the objects of interest. Therefore, the algorithm is able to determine if the gaze track ray has an intersection with a particular bounding box or not, determining which object in which frame have been observed.

About participants, the sample consisted of three boys (mean age: 50.3 months, range: 42–55 months), two with Autism Spectrum Disorder and one with Autistic Disorder, recruited at the Pervasive Healthcare Center of the Institute of Clinical Physiology of the National Research Council of Italy in Messina. The diagnostic assessment was based on the Autism Diagnostic Observation Schedule, Griffith's Developmental Scale, Vineland Adaptive Scale, McArtur Language Test and Child Behaviour Checklist (CBCL).

## 3 DATA ACQUISITION AND EXPERIMENTS

Our method has been employed in a scenario where a closet containing specific toys, that are neatly disposed from the therapist, is opened to the child. Fig. 3 shows the disposition of the toys into the closet. The closet has size of 85 × 195 cm, but

only three sectors (in the bottom) have been used, considering the age of the involved children. Thus, the useful part has a size of  $85 \times 117$  cm.

In Fig. 3, 9 Areas Of Interest (AOI) are designed. They have been manually computed, since a static disposition has been employed. Each AOI has a size of  $28.3 \times 39$  cm. They are labelled from 1 to 9, and to each one of them correspond a specific toy. Exception, the cell number 2 will represent the Kinect area. Gaze data was extracted for each of these AOI. Fixations can be aggregated at many different levels, resulting in a wealth of psychophysical measures. In particular, the following parameters have been taken into account:

- Fixation count: the number of fixation in a specific AOI. For our system, a fixation of a toy is considered to be consistent if at least 15 consecutive frames present an hit on the square in exam;
- First fixation: the first consistent AOI looked by the patient;
- Sequence: the order of gaze hits into the AOIs;
- Selected Toy: the final choice of the patient;
- Most viewed toy: the cell with the maximum number of hits.



Figure 3. An example of toys disposition into the closet.

After a brief environment exploration, the child freely chooses the desired toy that will be subsequently used during therapy. Concerning the face of the child, video acquisition has been accomplished by a Microsoft Kinect sensor hidden into the closet in order to obtain a depth image that will be processed by the estimation algorithm, that will compute gaze tracking by intersection with data coming from the well-known initial disposition of toys. Depth image is necessary to estimate the head pose and to know the X, Y, Z world coordinate of the detected facial points. The green LED of the sensor has been darkened, to avoid to be appealing to the child and to invalidate our experiment. The depth sensor was connected on an Ultrabook Intel i3 CPU @ 1.8GHz with 4 GB of RAM, work at 30fps. The usage of our technique on a common Ultrabook was made in order to facilitate for raw installations like in the topmost shelves of the closet. This is a very encouraging result, since it allows to provide the useful data in real-time.

All recordings were suitable for off line analysis with our algorithm. For this experiment, one well-known disposition of the toys is used, and a patient gaze track is firstly computed and projected on the image plane; in a further step, its intersection with the toys' AOIs have been computed.

For each patient, these values have been recorded. Furthermore, these values have been elaborated, and results are summed up in Table 1.

2D visual exploration on a plane has been tracked over time, and this datum had been related with the chosen toy by means of the belonging bounding box. Fig. 4 shows the x and the y visual in an ordered sequence of frames where detection occurred, while Fig. 5 shows the computed hit map for the same patient during the interaction. The system was tested with three children, showing a completely different behaviour; in particular, the system tracked the first child only few frames, due to his very fast and out of field movements. The child selected a toy after a very short exploration, the first fixation was in the AOI 5 and the most viewed toy was in the AOI 2. There are no relations between what he watched most and the selected toy. The second child has a lower number of fixations than the other children. The selection went toward the last toy observed, in the opposite direction in

Table 1. Gaze measures during tasks.

	Child 1	Child 2	Child 3
Fixation Count	4	3	6
Sequence	5-2-5-7	2-5-6	5-2-3-5-2-1
First Fixation Cell	5	2	5
Selected Toy	7	4	1
Most Viewed toy	2	6	1

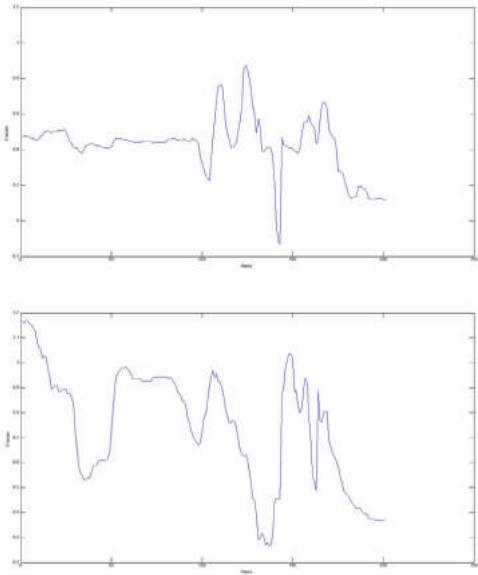


Figure 4. Recorded screen coordinates (X on top, Y on the bottom) for person #1.

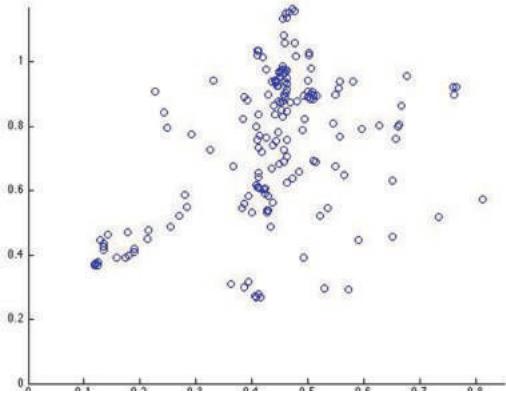


Figure 5. The computed hit map for person #1.

respect to the toys seen previously. In this case, only with the usage of our system is possible to capture details and, in this case, to understand the real will of the child. Finally, in the third video, the system track different interval of the visual exploration, observing a large exploration of toys before concentrating on a point and taking the corresponding toy. This child has a higher number of fixations than the other children. The selected toy was the most viewed, He spent more time to see the chosen toy than the other. Figures 6 and 7 show the hit maps for the person #2 and #3, respectively.

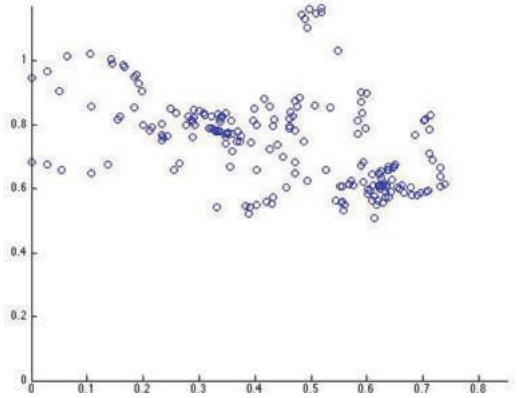


Figure 6. The computed hit map for person #2.

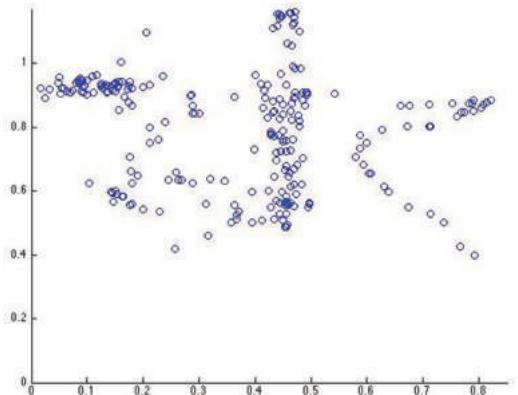


Figure 7. The computed hit map for person #3.

#### 4 DISCUSSION

In 2008, a case study reported by Vismara et Rogers showed a reduction in autistic symptoms thanks to a novel paradigm for treatment of autism based on a early intervention: the Early Start Denver Model (ESDM). Dawson et al. (2010) showed the efficacy of the ESDM for children with autism in improving cognitive and adaptive behaviours and reducing severity of ASD diagnosis. Rogers and Dawson (2010) emphasize the role of motivation to treat children with the ESDM. To motivate children they recommend to put subjects in a situation where there are many age-appropriate toys and then observe what children do. In this situation, children will probably watch toys more interesting for them, however the absence of protodeclarative gestures in children with autism (Baron-Cohen, S., 1989; Curcio, F., 1978) can be a problem for the therapist. In this contest, gaze estimator can be

a good instrument to understand where children have focused its attention. Eye tracking is a useful methodology for investigating spontaneous priorities and patterns of attention. Relative priorities in attention can be inferred by analysing which regions of a stimulus a participant looks at, for how long the regions are looked at and when during the course of viewing the various regions are looked at.

This non-invasive method can facilitate our understanding of underlying cognitive processes involved in object perception and the more complex challenges related to social functioning in ASD (Boraston and Blakemore 2007). Knowing where a subject is looking provides a wealth of information regarding his motivations, expectations, and innate preferences. For this reason, eye tracking has become a standard tool for cognitive and psychological investigation. However, though the direct examination of the scan path obtained by eye-tracking systems tells us where individuals are looking, it does not tell us why they choose to look at different locations, and, of course, it does not tell us what these differences mean. In order to answer the more difficult question, that of how scan patterns should be interpreted, more sophisticated methods for distilling information from raw visual trajectories are required.

Before concluding, a discussion that considers the differences between a commercial eye tracker and the proposed system is reported. First of all, all the parameters of interest during a visual exploration of the patient have been taken into account. About them, it was possible to obtain a precise estimation of the number of fixations, the sequence of points of interest and where the attention goes at first. Two parameters that can also be estimated, but further development would be necessary, are the time of permanence and the average fixation length. Although it could be easily extracted adding a temporization on the detected frames, there were still a number of misdetection that could falsify the estimation, since we cannot automatically detect what happens in those frames (or it could be rebuilt by interpolation, but its value would not be reliable for the particular application context in exam). For this reason, we decided to do not report these values. From the other hand, advantages of the proposed systems are that the patient was able to participate to the session in a completely ecological way, i.e. without requiring any initial calibration, nor constraints about its position. Moreover, the proposed system is considerably inexpensive compared to commercial eye tracker solutions. This was a very encouraging results, that could provide a new inexpensive technique for evaluating focus-of-attention in autistic children.

## 5 CONCLUSIONS

Computer vision technology has a unique opportunity to impact that study of children's behaviour, by providing a means to automatically capture behavioural data in a non-invasive manner and analyse behavioural interactions between children and toys. Eye tracking behavioural assessments may provide spatial information than face-to-face assessments. These tools aid the clinician in the behavioural assessment task by providing accurate and objective measurements.

The early detection of developmental disorders is key to child outcome, allowing interventions to be initiated that promote development and improve prognosis. Research on ASD suggests behavioural markers can be observed late in the first year of life. Many of studies involved extensive frame-by-frame video observation and analysis of a child's natural behaviour. Although non-intrusive, these methods are extremely time-intensive and require a high level of observer training. With the goal of aiding and augmenting the visual analysis capabilities in evaluation and developmental monitoring of ASD, we proposed computer vision tools to observe specific behaviours related to ASD elicited during toy selection tasks, providing both new challenges and opportunities in video analysis.

## ACKNOWLEDGEMENTS

This work has been supported by the "2007–2013 NOP for Research and Competitiveness for the Convergence Regions (Calabria, Campania, Puglia and Sicilia)" with code PON04a3\_00201.

## BIBLIOGRAPHY

- Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., Nightingale, N.A.T.A.S.H.A., Morgan, K.A.T.E. & Charman, T. (1996). Psychological markers in the detection of autism in infancy in a large population. *The British Journal of Psychiatry*, 168(2), 158–163.
- Baron-Cohen, S. (1989). Perceptual role taking and protodeclarative pointing in autism. *British Journal of Developmental Psychology*, 7(2), 113–127.
- Bogdashina, O. (2003). *Sensory perceptual issues in autism and Asperger Syndrome: different sensory experiences, different perceptual worlds*. Jessica Kingsley Publishers.
- Boraston, Z. & Blakemore, S.J. (2007). The application of eye-tracking technology in the study of autism. *The Journal of physiology*, 581(3), 893–898.
- Cazzato, D., Leo, M. & Distante, C. (2014). An Investigation on the Feasibility of Uncalibrated and Unconstrained Gaze Tracking for Human Assistive Applications by Using Head Pose Estimation. *Sensors*, 14(5), 8363–8379.

- Clifford, S., Young, R. & Williamson, P. (2007). Assessing the early characteristics of autistic disorder using video analysis. *Journal of autism and developmental disorders*, 37(2), 301–313.
- Curcio, F. (1978). Sensorimotor functioning and communication in mute autistic children. *Journal of autism and childhood schizophrenia*, 8(3), 281–292.
- Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J. & Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics*, 125(1), e17-e23.
- Hutman, T., Chela, M.K., Gillespie-Lynch, K. & Sigman, M. (2012). Selective visual attention at twelve months: Signs of autism in early social interactions. *Journal of autism and developmental disorders*, 42(4), 487–498.
- Klin, A., Jones, W., Schultz, R., Volkmar, F. & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of general psychiatry*, 59(9), 809–816.
- Lai, M.C., Lombardo, M.V. & Baron-Cohen, S. (2014) Autism. *Lancet* 383(9920):896–910.
- Noris, B., Benmachiche, K., Meynet, J., Thiran, J.P. & Billard, A.G. (2007). Analysis of Head-Mounted Wireless Camera Videos for Early Diagnosis of Autism. In *Computer Recognition Systems 2* (pp. 663–670). Springer Berlin Heidelberg.
- Noris, B., Nadel, J., Barker, M., Hadjikhani, N. & Billard, A. (2012). Investigating gaze of children with ASD in naturalistic settings. *PloS one*, 7(9), e44144.
- Ozonoff, S., Macari, S., Young, G.S., Goldring, S., Thompson, M. & Rogers, S.J. (2008). Atypical object exploration at 12 months of age is associated with autism in a prospective sample. *Autism*, 12(5), 457–472.
- Rehg, J.M. (2011, June). Behavior Imaging: Using Computer Vision to Study Autism. In *MVA* (pp. 14–21).
- Riby, D. & Hancock, P.J. (2009). Looking at movies and cartoons: eye-tracking evidence from Williams syndrome and autism. *Journal of Intellectual Disability Research*, 53(2), 169–181.
- Riby, D.M. & Hancock, P.J. (2008). Viewing it differently: Social scene perception in Williams syndrome and autism. *Neuropsychologia*, 46(11), 2855–2860.
- Robins, D.L., Fein, D. & Barton, M. (2009). Modified Checklist for Autism in Toddlers, Revised, with Follow-Up (M-CHAT-R/F) TM.
- Rogers, S.J. & Dawson, G. (2010). Early Start Denver Model for young children with autism: Promoting language, learning, and engagement. New York: Guilford Press.
- Sasson, N.J., Turner-Brown, L.M., Holtzclaw, T.N., Lam, K.S. & Bodfish, J.W. (2008). Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. *Autism Research*, 1(1), 31–42.
- Siegel, B., Vukicevic, J. & Spitzer, R.L. (1990). Using signal detection methodology to revise DSM-III-R: Re-analysis of the DSM-III-R national field trials for autistic disorder. *Journal of psychiatric research*, 24(4), 293–311.
- Shic, F., Bradshaw, J., Klin, A., Scassellati, B. & Chawarska, K. (2011). Limited activity monitoring in toddlers with autism spectrum disorder. *Brain research*, 1380, 246–254.
- Tentori, M. & Hayes, G.R. (2010, September). Designing for interaction immediacy to enhance social skills of children with autism. In *Proceedings of the 12th ACM international conference on Ubiquitous computing* (pp. 51–60). ACM.
- Trepagnier, C., Sebrechts, M.M. & Peterson, R. (2002). Atypical face gaze in autism. *Cyberpsychology & Behavior*, 5(3), 213–217.
- Vismara, L.A. & Rogers, S.J. (2008). The Early Start Denver Model: A case study of an innovative practice. *Journal of Early Intervention*.

This page intentionally left blank

# Semi-automatic tumor contouring method using PET and MRI medical images

Szabolcs Urbán

*Institute of Informatics, University of Szeged, Szeged, Hungary*

László Ruskó

*GE Healthcare, Szeged, Hungary*

Antal Nagy

*Institute of Informatics, University of Szeged, Szeged, Hungary*

**ABSTRACT:** Tumor contouring is a challenging task for physicians especially when functional as well as anatomical images are incorporated. This paper presents a semi-automated technique to solve this problem. The proposed method is a variant of the level-set segmentation where the initial region is based on the functional image and the speed function combines both images. The presented approach allows the user to balance between the information of two input images. The method was evaluated on registered head and neck PET-MRI image pairs using manually defined tumor contours as reference. The algorithm was tested for various types of tumors using different weights to combine the functional and the anatomical information. The best results showed good correlation with the reference (3% volume difference and 80% DICE similarity in average).

## 1 INTRODUCTION

The number of patients living with cancer is increasing from year to year. 3-dimensional (3D) medical imaging plays central role in the management of patients with cancer during diagnosis, therapy planning, and monitoring. There are various 3D imaging techniques available in clinical routine. Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) imaging supplies complementary information about the tumor being examined (AR. 2014). CT and MRI provides anatomical information at high spatial and contrast resolution, while PET gives functional information at lower spatial resolution.

Tumor contouring is an important task of oncology because the location and the change of the volume of the lesion plays crucial role in therapeutic decisions. The more accurate the tumor quantification is the more reliable is the evaluation of the therapy's effectiveness. Manual tumor contouring is very time consuming and subject to operator variability, especially when multiple images shall be incorporated to define the lesion boundary. It is also a very challenging task (even for experienced physicians) due to the large variation in localization, shape, size, and intensity heterogeneity of the

lesions. That is why there is significant need for computer assisted tools which make this process more accurate, efficient, and reproducible.

Large number of tumor contouring methods have been published which rely whether on anatomical or functional images, but only a few of them can incorporate both of them, especially PET and MRI. El Naqa et al. (El Naqa, Yang, Apté, Khullar, Mutic, Zheng, Bradley, Grigsby, & Deasy 2007) presented a level-set based method which uses the mean intensity values from the inside and outside region of the tumor contour. They tested their method on two tumor cases (using PET-CT). On (artificially generated) phantom data they achieved 90% DSC in average (using PET, CT and MRI), however the shown input images contained almost the same, homogeneous sphere-like object. Bagci et al. first published a tumor segmentation approach based on PET/CT images (Bagci, Udupa, Yao, & Mollura 2012). They represent the images in a product lattice and perform random walk image segmentation. In this approach the foreground and the background is selected based on the PET image such that background points are only taken from the low PET intensity regions. Tumors can be significantly larger in the MRI image compared to the PET, therefore the background points selected this way can be inaccurate. Later

they generalized this method to use more than two images (Bagci, Udupa, Mendhiratta, Foster, Xu, Yao, Chen, & Mollura 2013). According to their results this approach performed well for various clinical cases including MRI-PET scans (93.1% DSC in average). However these images contained almost no registration errors and the shown MRI-PET case contained similar information about the tumor on each of the input images.

## 2 THE PROPOSED METHOD

The goal of this work was to develop an interactive tool that can combine the information of registered PET and MRI images to facilitate tumor contouring in the head and neck region which contains various-shaped tumors and the mentioned images can contain complementary/contradictoire information. The main idea was to extend a standard segmentation technique to incorporate functional and anatomical images in a user-friendly way. Our primary goal was to develop a MR-PET tumor contouring tool the result of which is consistent with both input images and the user's expectations. This goal was primarily motivated by discussions with physicians working at the field of radiotherapy planning. They are challenged with defining the target volume (i.e. the tumor) when both functional and anatomical images are given because they have to switch back and forth between the input images when they manually contour the different parts of the tumor. This way they include all regions which may belong to the tumor based on any of the input images. According to our experiments with head-and-neck cases fully automated approach is hardly achievable, while the presented semi-automated technique can facilitate this complex task. The proposed method consists of two main steps. First, an initial region is extracted from the user-defined Region Of Interest (ROI) based on the functional image. Then, 3D fastmarching approach is applied to generate a set of contours, from which the user can select one (referred as *level*) that fits both images as well as the user's expectations. The contour propagation is driven by a hybrid speed function that incorporates both PET and the MRI image intensities such that the weight of the PET (referred as  $\lambda$ ) and MRI (referred as  $1 - \lambda$ ) images can be specified by the user. This way the user can balance between the PET and MRI information according to the characteristics of the tumor to be contoured. Figure 1 demonstrates the main steps of the algorithm.

The segmentation requires a *starting region*. This region is the basis of the contour propagation, so it must be fully located inside the tumor. In order to get such a region, only the functional image is used.

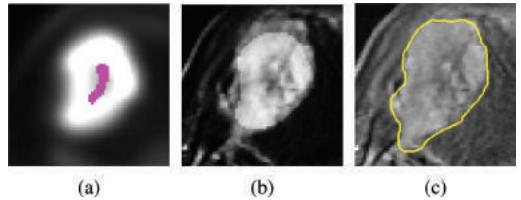


Figure 1. Main steps of the method: PET with initial region in magenta (a), hybrid speed function using  $\lambda = 30\%$  (b), and result contour for  $level = 23$  in MRI (c).

In FDG-PET images high uptake values show high glucose metabolism (high intensities), which can be normal inside some regions (e.g. brain, muscles), but it usually reflects pathology (tumor, inflammation) in the area of the neck. In the input PET images the intensities are represented by Standardized Uptake Value (SUV). Inside the user-defined ROI the maximal ( $SUV_{max}$ ) value is calculated and its 40% is used for thresholding the high-uptake region. The largest 3D-connected component of the detected region, which can be considered as the active part of the tumor, is the basis of the starting region. Due to the lower spatial resolution of the functional image, the border of this region cannot be always considered as part of the tumor in the anatomical image (that has much higher spatial resolution). Furthermore the input images were globally registered, which cannot guarantee the local accuracy of the registration at each tumor. To take these observation into account, the initial region is defined by the skeleton of the connected region. The applied 3D skeletonization algorithm (Lee, Kashyap, & Chu 1994) guarantees the connectedness of the result. The generated skeleton is a thin structure, therefore dilation is applied to that. The dilation incorporates the thresholded uptake region, so that the final skeleton is always located inside that. The proposed initial region represents the shape of the functionally active region and it is not sensitive for small local registration inaccuracy. An example for initial region is shown in Figure 1a.

In order to combine the functional and the anatomical information of the input images a *hybrid speed function* is defined in the following way. First, the intensity range of both PET and MRI images is normalized to the [0,1] interval, using the min-max normalization technique. Since the intensity maximum of an MRI images can be associated with some voxels affected by local imaging artifacts, we use the largest intensity that is higher than 99% of the voxels of the MRI image. This step is substantial because the input images involve different information, generated by different acquisition techniques. The functional (PET)

and the anatomical (T2w water MRI) images used in this study have similar characteristics. In case of both image types the higher intensities represent the tumorous tissue and lower intensities represent the background. After the normalization the intensity range is the same for both images, which makes it easier to combine the two different image information.

The hybrid speed function  $F$  is defined at each voxel in the following way:

$$F = \lambda \cdot \frac{1}{1 + e^{-\frac{I_{fn} - \beta_{fn}}{\alpha_{fn}}}} + (1 - \lambda) \cdot \frac{1}{1 + e^{-\frac{I_{an} - \beta_{an}}{\alpha_{an}}}} \quad (1)$$

where  $\lambda \in [0,1]$  is the weighting factor;  $I_{an}$  and  $I_{fn}$  are the normalized voxel intensity value in the anatomical and the functional images, respectively;  $\beta$  and  $\alpha$  are the parameters of the sigmoid function: its centre and the steepness factor, respectively. The speed function is defined as the weighted sum of two sigmoid filters applied to the input images voxel by voxel. This approach exploits that the meaning of intensity values are similar in both input images (i.e. high value represents tumor). Therefore, the combined intensities represent the same features which characterize the input images. The higher the value is in the speed function the faster the contour propagation is at the given voxel. This way, the contour propagation slows down when it reaches voxels which are located outside the tumor according to any of the input images. Note that the speed function can be defined based on other image features like edges if the intensity does not characterize the tumor in the anatomical input image. In such a case the speed is high in the homogeneous regions and low at the region boundaries.

Sigmoid functions are commonly used for better separation of higher and lower values in the speed function. The center of the function ( $\beta$ ) is defined as 0.5 and the steepness factor ( $\alpha$ ) is set to 0.1 for both input images ( $fn, an$ ). Therefore the intensity values of the input images below or above the range  $[\beta - \alpha/2; \beta + \alpha/2]$  become significantly lower or higher than the original values, respectively.

The definition of the weighting factor ( $\lambda$ ) depends on the input images and the tumor to be contoured. There are cases when the functional image can give more reliable information about the tumor (e.g. when the tumorous region and its surroundings are very heterogeneous in the MRI image). In other cases, the anatomical image allows more precise definition of the tumor contour. Consequently, the weighting factor shall be set according to the given tumor contouring problem.

The 3D fast marching method (Sethian 1999) is a special case of the level-set segmentation approach

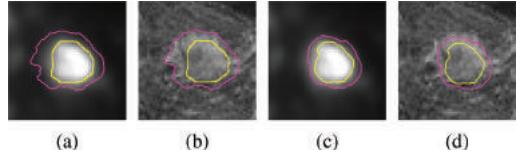


Figure 2. Tumor contour using different segmentation parameters:  $\lambda = 30\%$  (a and b) and  $\lambda = 70\%$  (c and d);  $level = 10$  (yellow) and  $level = 30$  (magenta).

with the restriction that the speed function can never change sign. It can be defined as follows:

$$F(x, y, z) |\nabla T(x)| = 1, x \in \Omega \quad (2)$$

where  $F$ : speed function ( $F > 0$ ),  $T$ : arrival function,  $\Omega$ : a domain in  $R^n$ .

The initial curve contains those  $x$ -es which belong to the zero level ( $T(x) = 0$ ). The main advantage of the fast-marching technique is that it can be efficiently computed as a narrow-band level set. In our case *alive points* are defined based on the initial region and *narrowband points* as its neighbors. The point of this definition is that the seed region is always inside the tumor and represents the shape of the tumor based on the functional image. Therefore, starting from this region the whole tumor volume can be reached. The above-defined hybrid speed function allows incorporating the information from both input images. Consequently, the contour propagation purely relies on the joint information. The method generates an image, which represents the shortest distances from the seed region at each voxel. The result contour can be extracted by thresholding the distance map with a particular value ( $level$ ). Similar to parameter  $\lambda$  there is no single value for  $level$  that provides the best tumor contour in every cases. Therefore, the user shall adjust manually the  $level$  according to the given tumor. Different  $(\lambda, level)$  combinations result in different tumor contours as shown in Figure 2. In this case the tumor is larger in the anatomical image, therefore using  $\lambda = 30\%$  and  $level = 30$  produced the best contour.

### 3 RESULTS AND DISCUSSION

The method was evaluated using 8 registered PET/MR image pairs. The test images originated from 3-modality examinations involving a (FDG) PET-CT and a subsequent MRI scan at the University Hospital Zurich. The MRI images were acquired with a 3 Tesla MRI scanner (GE Healthcare) using multiecho IDEAL gradient-echo imaging. The test exams covered the head and neck

region and involved a total number of 15 tumors. Each MRI case included multiple sequences, among which the T2 weighted IDEAL water image was used because the tumor contour was the most visible in this type of image.

For each case the CT image was registered globally to the MRI image using rigid registration technique that was based on mutual-information. The registration transform was applied to the PET image that was also resampled using the resolution and voxel spacing of the MRI image. This way the corresponding PET information was available at each voxel of the MRI image. In order to reduce noise, anisotropic diffusion filter (Perona & Malik 1990) was applied to the MRI images.

In each case all tumors were manually segmented by three physicians, separately. First, they delineated the tumor contour in the functional image, then they adjusted the contour based on the anatomical image taking also the PET image into account. After all three contours were available they selected the best one and (if needed) made some manual adjustments in consensus. The contour that was confirmed by all operators was used as reference. For the evaluation the Dice Similarity Coefficient (Dice 1945) and Absolute volume difference were used. The first metric is used to compare the overlap of two segmentation, while the second one shows the difference of two volumes, which is very important in the case of tumor assessment. The test cases involved typical head-and-neck tumors characterized by different size, shape, and internal and external heterogeneity, which made them very challenging to contour (in contrast to sphere-shaped tumors presented in some papers). In some cases the MR and the PET provided very different information about the tumor, which introduced significant inter-observer variability in manual contouring (DSC 55-90, AVD 2-39) based on the modality preference of the physician. The average time of manual contouring was 15 minutes per tumor.

The method has two main parameters which have significant effect on the final contour: the weight factor ( $\lambda$ ) that represents the weight of the PET image and the level (*level*) that is used to generate the contour. To demonstrate the effect of the parameters the result contour was created using several ( $\lambda$ , *level*) combinations. The  $\lambda$  was set from 0% to 100% (with 10% stepping) and the value of *level* ranged from 5 to 100. The method was applied to each lesion using all parameter combinations and after each test run the result contour was compared with the reference using standard segmentation error metrics.

Figure 3 shows the segmentation accuracy belonging to the different parameter settings and the best result for three typical tumors. In the first

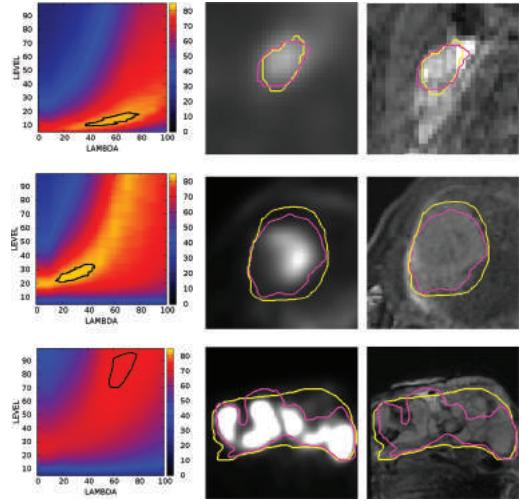


Figure 3. Effect of the parameters  $\lambda$  and *level* to the DSC accuracy (left), the reference (yellow) and the best result (magenta) in PET (center) and MRI (right) images.

column the Dice Similarity Coefficient (DSC) is plot for all parameter combinations (yellow means high accuracy). The second and third columns show the best contour (magenta) and the reference (yellow) in a characteristic axial slice of the tumor in the PET and the MRI images, respectively.

In the top row of Figure 3 a small tumor is shown with similar anatomical and functional information. The best result (DSC: 82%) was obtained using  $\lambda = 70$  and *level* = 16. In this case using larger  $\lambda$  eliminates the over-segmentation due to the ambiguous MRI information. The segmentation of a medium sized tumor is displayed in the second row. The best result (DSC: 83%) was achieved using  $\lambda = 30$  and *level* = 23. In this case the MRI shows more precise boundary information about the tumor because some part of it is only visible in the anatomical image. The third row demonstrates a difficult case that involves a large heterogeneous tumor. In this case the best result (DSC: 74%) was generated using  $\lambda = 60$  and *level* = 86. Using larger weight for the PET image prevents large over-segmentation which occur when the method relies on the very ambiguous MRI information. The large value of the parameter *level* was due to the large size of the tumor.

According to the accuracy plots of Figure 3 various combinations of  $\lambda$  and *level* can be used to generate similar contour (see dark contour representing 95% of the maximum). Although, there is always a pair that maximizes the accuracy, very similar result can be achieved using a different  $\lambda$  with a corresponding level. Thanks to the low

computation time the result contour can be automatically generated for several combinations of the two parameters so that the corresponding contour shall be only displayed when the user modifies a parameter. The user can find a good contour in an intuitive way by adjusting the two parameters (the balance between the two image information and the size of the contoured region) and having real time visual feedback on the change of that parameter. The method guarantees that a larger *level* generates larger contour. This observation together with the smoothness of the plots guarantees that the user can find the best contour in an intuitive way: the user can adjust the  $\lambda$  according to the importance of the input images then adjusting the *level* parameter to generate a more precise tumor contour.

The proposed method was integrated into the Medical Imaging Interactive Toolkit (MITK) framework (Nolden, Zelzer, Seitel, Wald, Mller, Franz, Maleike, Fangerau, Baumhauer, Maier-Hein, Maier-Hein, Meinzer, & Wolf 2013) as a plug-in (Fig. 4). The implemented workflow consists of the following steps: the two input images

are set, the volume of interest (i.e. bounding box of the tumor) is defined, the segmentation is executed (for all parameter combinations automatically), the result is displayed for the default parameter combination, after adjusting a parameter the contour is immediately updated.

The prototype was applied to each tumor in the test set. Table 1 shows the accuracy of the best result for each tumor involved in the evaluation. The table shows the DICE Similarity Coefficient (DSC) and the Absolute Volume Difference (AVD) in addition to the reference lesion volume and the parameter combination used to generate the result. According to the error metrics DSC was 80.6% and AVD was 3% in average, which shows the results correlate well with the reference. Note that average AVD was significantly affected by one tumor (case 1). In most cases the best result was associated with lower PET image weight (average  $\lambda$  was 35%). This fact was in line with the observation that water T2w MRI images give more accurate information about the tumor boundaries than the PET images.

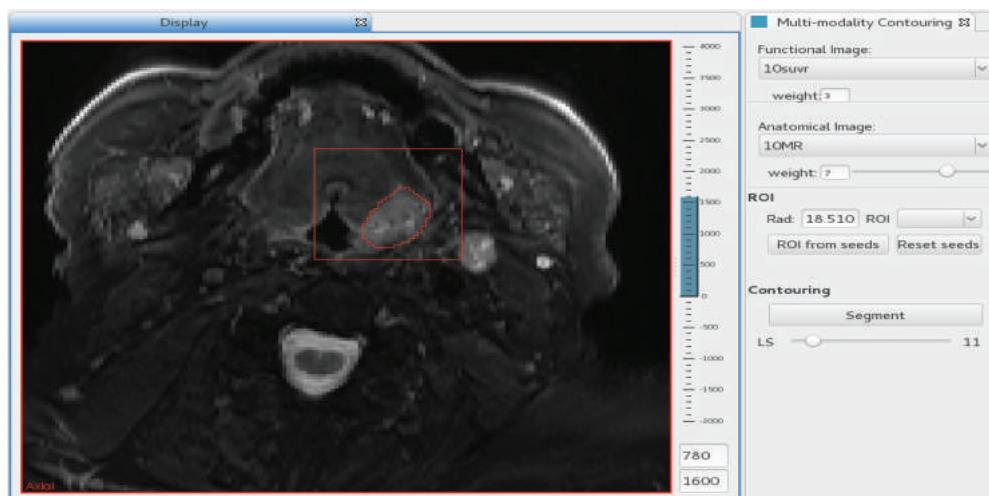


Figure 4. The method integrated into the MITK framework: the generated red contour inside the defined ROI using  $\lambda = 3$  and *level* = 11.

Table 1. Segmentation results.

Cases	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
DSC (%)	74	71	80	83	81	75	83	86	79	78	82	81	81	87	88
$\lambda$ (%)	60	30	20	30	30	70	30	20	20	50	70	20	30	30	20
Level	96	16	6	16	30	46	23	11	14	26	16	11	13	11	21
Volume (ml)	149.0	8.0	1.9	3.4	23.9	55.0	41.6	3.4	2.2	17.0	1.8	11.7	8.5	8.9	2.8
AVD (%)	25.5	3.5	1.4	0.1	1.4	1.5	1.3	1.3	1.0	1.0	1.7	1.9	1.0	2.9	0.3

The evaluation demonstrated that the value of the parameter *level* primarily depends on the size of the tumor. For large tumors higher value was used because the initial region was always the skeleton of the high intensity PET region. For some results the DSC was below 80%. In case 1 and 6 the reference segmentation involved regions which were not fully consistent with the image data: in some parts the contour involved high-uptake regions in the PET image, but in other parts such regions were excluded. Furthermore, in case 2 the tumor was weakly differentiable from its surrounding in the MRI image. These issues can be justified with the fact that inter-observer variability was high (mentioned before) in the above-mentioned cases. It is important to note that volume difference was very low in all cases (except for case 1).

The presented method was implemented in C++ using the Insight Toolkit (Johnson, McCormick, Ibáñez, & Consortium 2013) library. The evaluation was executed on a workstation with 3.4 GHz processor and 16 GB RAM. The running time of generating a contour was below 0.4 second in all cases using any parameter configuration (the average running time was 0.13 second). The registration (including automated method) took 1–5 minutes/case. The segmentation (including the parameter tuning) took about 1–3 minutes/tumor.

## 4 CONCLUSIONS

According to the evaluation the presented segmentation approach allows the user to combine MRI and PET information to create a tumor contour that is consistent with both input images. Experiments with complex head and neck tumors showed the parameters of the method can be adjusted in a very intuitive way, which helps the user to find the most accurate tumor contour in efficient way. The presented approach is a straightforward extension of the level-set segmentation to multi-modality images. The speed function can be generalized by incorporating edge-based, textural, or other features of morphological images so that other image combinations can be also used with this approach. In the future, the usability of the method shall be confirmed by more extensive clinical study involving other image combinations.

## ACKNOWLEDGEMENTS

We would like to thank Prof. Dr. Patrick Veit-Haibach and Gaspar Delso from the University Hospital Zurich for providing the input images.

We are also grateful for Dr. Zsuzsanna Besenyi, Dr. Zsuzsanna Fejes and Prof. Dr. László Pávics from the University of Szeged for creating the reference tumor contours.

This work was supported by Analitic Health-care Quality User Information Program of the National Research, Development and Innovation Fund, Hungarian Government, Grant VKSZ\_12-1-2013-0012.

## REFERENCES

- AR., P. (2014). Recommendations for cross-sectional imaging in cancer management. *The Royal College of Radiologists*.
- Bagci, U., J. Udupa, J. Yao, & D. Mollura (2012). Cosegmentation of functional and anatomical images. In N. Ayache, H. Delingette, P. Golland, and K. Mori (Eds.), *Medical Image Computing and Computer-Assisted Intervention MICCAI 2012*, Volume 7512 of *Lecture Notes in Computer Science*, pp. 459–467. Springer Berlin Heidelberg.
- Bagci, U., J.K. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, X. Chen, & D.J. Mollura (2013). Joint segmentation of anatomical and functional images: Applications in quantification of lesions from pet, pet-ct, mri-pet, and mri-pet-ct images. *Medical Image Analysis* 17(8), 929–945.
- Dice, L.R. (1945). Measures of the amount of ecological association between species. *Ecology* 26(3), pp. 297–302.
- El Naqa, I., D. Yang, A. Apte, D. Khullar, S. Mutic, J. Zheng, J.D. Bradley, P. Grigsby, & J.O. Deasy (2007). Concurrent multimodality image segmentation by active contours for radiotherapy treatment planninga). *Medical Physics* 34(12), 4738–4749.
- Johnson, H.J., M. McCormick, L. Ibáñez, & T.I.S. Consortium (2013). *The ITK Software Guide* (Third ed.). Kitware, Inc. *In press*.
- Lee, T., R. Kashyap, & C. Chu (1994). Building skeleton models via 3-d medial surface axis thinning algorithms. *CVGIP: Graphical Models and Image Processing* 56(6), 462–478.
- Nolden, M., S. Zelzer, A. Seitel, D. Wald, M. Müller, A. Franz, D. Maleike, M. Fangerau, M. Baumhauer, L. Maier-Hein, K. Maier-Hein, H.-P. Meinzer, & I. Wolf (2013). The medical imaging interaction toolkit: challenges and advances. *International Journal of Computer Assisted Radiology and Surgery* 8(4), 607–620.
- Perona, P. & J. Malik (1990). Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 12(7), 629–639.
- Sethian, J.A. (1999). *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*. Cambridge University Press.

# Automatic detection of bones based on the confidence map for Rheumatoid Arthritis analysis

K. Radlak, N. Radlak & B. Smolka

*Institute of Automatic Control, Silesian University of Technology, Gliwice, Poland*

**ABSTRACT:** In this paper an application of confidence map for bones surface detection in ultrasound images for Rheumatoid Arthritis is presented. The bone feature descriptor is based on confidence maps, which are directly related with the likelihood of ultrasound transmission and allows for highly accurate localization of the bone surface. The quality of the confidence map has been significantly improved by applying the Trimmed Non-Local Means filtering. The results were evaluated on the publicly available MEDUSA database that contains manually annotated markers of bones surface. This research is a preliminary step in the development of fully automatic system for the assessment of degree of synovitis inflammation.

## 1 INTRODUCTION

Ultrasound (US) imaging is a powerful method of visualization of human body structures, which is often combined with the Power Doppler (PD) technique used to measure the blood circulation in tissues. Ultrasonography with PD signal has indicated its efficiency, when diagnosing joint inflammation in Rheumatoid Arthritis (RA) (ØStergaard and Szkludlarek 2005), but it involves some degree of subjectivity. The disease level is scored from 0 to 3 according to the OMERACT definitions (McQueen et al. 2003), by human examiners based on their experience or standardized atlases (Terslev et al. 2003, Kulbacki et al. 2014). However, such an evaluation system can cause some discrepancies in the final diagnosis and treatment. Therefore, the most challenging task is the elaboration of an automatic method of assessment of joint inflammation level in RA using ultrasound images acquired with PD.

The degree of synovitis can be assessed by estimation of the size of the dark gray/black area above the finger bones or the joint. Consequently, the task of automatic determination of the level of synovitis can be considered as a segmentation problem of dark image regions in the vicinity of bones and the joint. To determine the proper orientation in an image, we propose to detect the bones regions in an image as a reference for the further analysis to determine the synovitis region. Another possibility is to estimate the joint localization (Wereszczyński et al. 2014), but this approach is not considered in this work.

To evaluate the novel methodology for automated assessment of joint synovitis activity, a large set of ultrasound images with manually prepared annotations was collected within the MEDUSA project

(Kulbacki et al. 2014) from the patients with chronic arthritis during routine visits at the Rheumatology Department of Helse Førde in Norway. Exemplary ultrasound images with manually marked regions of interest are presented in Figure 1.

In this work, we focus on applying an ultrasound confidence map (Karamalis et al. 2012) for bones region detection. The confidence map is directly related with the likelihood of ultrasound transmission. The idea of an ultrasound confidence map is based on the concept of Random Walk introduced for image segmentation in (Grady 2006) and has been addressed in the literature with different

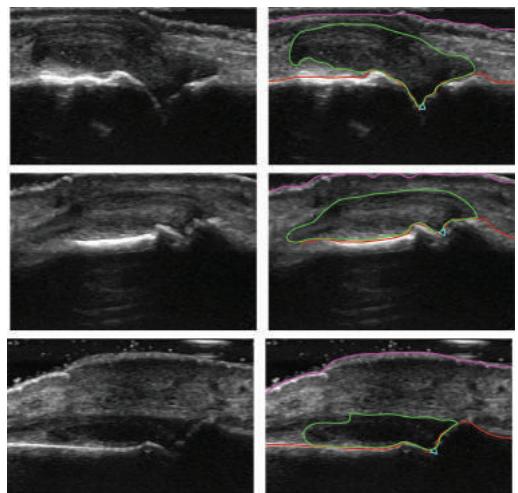


Figure 1. Ultrasound images on the left and manually delineated synovitis region (green), bones (red), skin (magenta) and joint (cyan) on the right.

approaches and in different contexts (Karamalis et al. 2012). Methods that allow for estimation of ultrasound attenuation can be used to detect shadowed and attenuated areas, thus, providing information about unreliable image regions. The estimation of ultrasound transmission allows algorithms to detect unreliable regions, like shadows, in order to improve the robustness and accuracy of their underlying methods.

The bone feature descriptor (Karamalis 2013) based on confidence maps allows for highly accurate localization of the bones surface in continuous acquisition of US images of a human leg, what encouraged us to apply the method for the ultrasound images of fingers. However, the main difference is the fact, that in the US image of a finger there are two bones, which need to be recognized.

The results obtained when applying the proposed algorithm were evaluated on the MEDUSA project database that contains manually annotated markers of bones regions. This research is a preliminary step in the development of fully automatic system that will support the diagnosis of RA.

This paper is organized as follows. The next Section provides a brief description of the proposed method, Section 3 presents the numerical results of bone detection accuracy evaluated on automatically found and manually marked bone surfaces utilizing the MEDUSA database and Section 4 concludes the paper.

## 2 BRIEF DESCRIPTION OF THE METHOD

The automatic bones detection is an important problem in Computer-Aided Orthopedic Surgery, in which precise information about the bone localization is very useful in the pre-operative planning phase and navigation during surgery (Beitzel et al. 2012), but to the best of our knowledge, no studies on bones detection for Rheumatoid Arthritis are available. The previously published papers are related with femur bone detection. The bone in ultrasound imaging appears as a bright and connected ridge with a shadow region and is much more distinguishable from other parts of the US image. However, an automatic detection is still a challenging task due to the presence of speckle noise, image low contrast, various artifacts and shadowing in ultrasound images and also due to the fact that most ultrasound segmentation methods are usually limited to a specific target (Noble and Boukerroui 2006).

There are several works related with bone detection in surgery. Jain et al. (Jain and Taylor 2004) calculate the probability that a pixel belongs to the bone region using Bayesian Probabilistic Framework based on the features which characterize the bones as structures with high intensity and high gradient accompanied by multiple reflections.

In (Foroughi et al. 2007) the authors also calculate the probability map based on properties of the bones in US images, but finally they use dynamic programming to determine where the bone starts and ends. In (Kowal et al. 2007) the authors propose to use thresholding, morphological operations and the detected pixels are merged in an iterative manner. In the papers (Karamalis 2013, Quader et al. 2014), the authors propose to use the confidence maps to determine the bone surface position. The values of the confidence map can be interpreted as relative signal strengths at different image locations and we expect to observe the stronger reduction of signal strength, when the signal passes through the bone.

In our approach, we applied the confidence map concept proposed in (Karamalis et al. 2012, Karamalis 2013), but the original US images were enhanced by our Trimmed Non-Local Means (TNLM) filter designed for the reduction of both mixed Gaussian and impulsive and speckle noise (Radlak and Smolka 2013, Radlak and Smolka 2014). The comparison of the quality of the confidence map with and without the proposed filtering method is presented in Figure 2. Another difference lies in the fact, that the authors used the series of US images along the femur bone and they can combine the results of detection coming from many images.

In the applied framework, the bones regions are detected similarly to the work of Karmalis (Karamalis 2013) using bone confidence localizer, but without the last step in which the author combines the bone localizer responses from series of US images. In the MEDUSA project, we have only single images of a finger of each patient.

The first step in the proposed method is the calculation of the bones confidence localizer map  $L$  defined as

$$L_{ij}^* = \sum_{a=i-s}^{i+1} \sum_{b=j-s}^{j+s} F(a,b) - \sum_{a=i-1}^{i+s} \sum_{b=j-s}^{j+s} M(a,b), \quad (1)$$

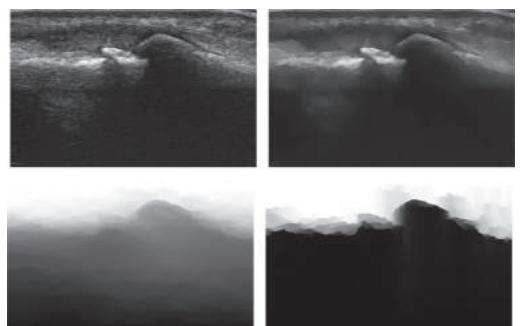


Figure 2. Ultrasound confidence map for noisy US image (on the left) and using the Trimmed Non-Local-Means filtering (on the right).

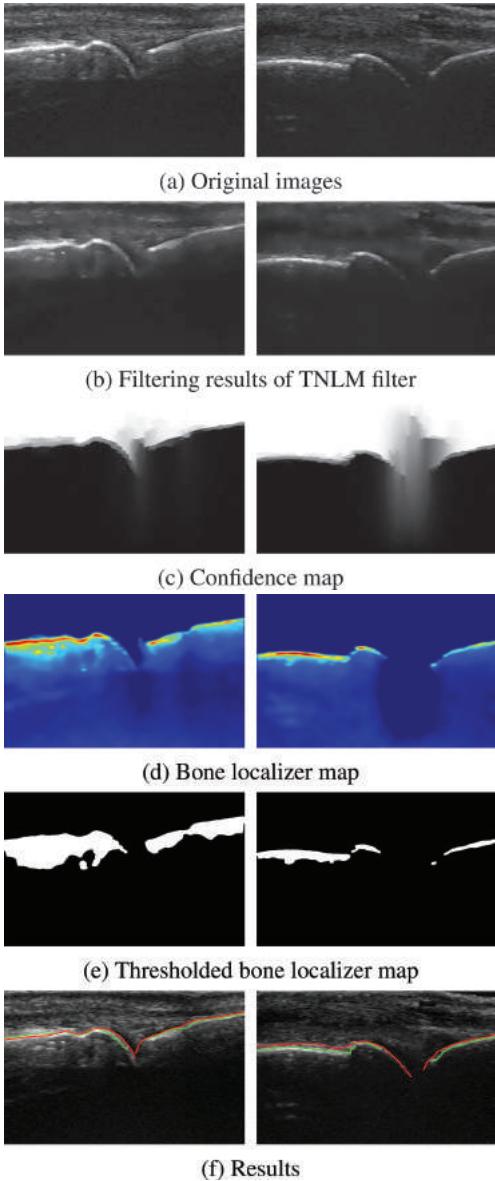


Figure 3. Exemplary results of the following steps of the algorithm performed on the images from MEDUSA project. The rows from (a)–(f) show the temporary results of the most significant algorithm steps. In the last row the detected bones are marked as green and manually annotated contours as red.

$$L_{ij}^* = \begin{cases} L_{ij}^* & \text{if } L_{ij}^* > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where  $M$  is the confidence map,  $F$  is the image filtered using TNLM algorithm,  $s$  is the filtering window size and  $i, j$  are row and column indices.

The confidence map  $M$  is calculated according to the method described in (Karamalis et al. 2012). Then, the obtained map  $L$  is thresholded using the Otsu method (Otsu 1979) and the obtained blobs are filtered using morphological operations and labeled. To overcome the difficulties caused by the discontinuities between the blobs, those for which the distance between contours is less than the predefined threshold are merged into one group. Finally, the two largest groups of blobs are considered as possible bones. To detect the bones region, we took the upper boundaries and the final bones contours were interpolated pixel by pixel using a piecewise cubic Hermite interpolating polynomial. The exemplary results of each step of the algorithm are presented in Figure 3.

### 3 BONE DETECTION ACCURACY

For the quantitative and objective evaluation the following procedure is proposed. A subset of images without Power Doppler coming from MEDUSA project has been chosen. These images were taken around the joint and there are two bones depicted in each image. However, due to the fact that some noise and the shadowing artifacts may occur, the bones can be less distinguishable. Therefore, all images were divided into three groups according to the subjective assessment of visibility of the bones in the image. These groups are: (a) low quality images with low contrast between bones region (110 images), (b) images with single bone that is visible (404 images), (c) images in which two bones are visible (429 images). Exemplary images are presented in Figure 4.

For each image in this dataset the bones contours were annotated by an expert to provide the necessary ground-truth and allow for automatic evaluation of the bones localization. Due to the fact that it is our preliminary work and to give the opportunity to compare the results, this database is available for researchers at <http://dip.aei.polsl.pl/medusa> upon request.

The contours were represented as sets of points  $\mathcal{A} = \{a_1, \dots, a_m\}$  and  $\mathcal{B} = \{b_1, \dots, b_m\}$  for manual and automatic bones localization respectively. Direct one-to-one comparison of the points sets is not possible, as they do not have a direct mapping and usually have different cardinality. Therefore, the Hausdorff distance (Huttenlocher et al. 1993) is a common measure for comparing two non-empty sets of feature points. The Hausdorff metric is defined as follows:

$$H(\mathcal{A}, \mathcal{B}) = \max(h(\mathcal{A}, \mathcal{B}), h(\mathcal{B}, \mathcal{A})), \quad (3)$$

$$h(\mathcal{A}, \mathcal{B}) = \max \min_{a \in \mathcal{A}} \min_{b \in \mathcal{B}} \|a - b\|, \quad (4)$$

where  $\|\cdot\|$  denotes the Euclidean distance.

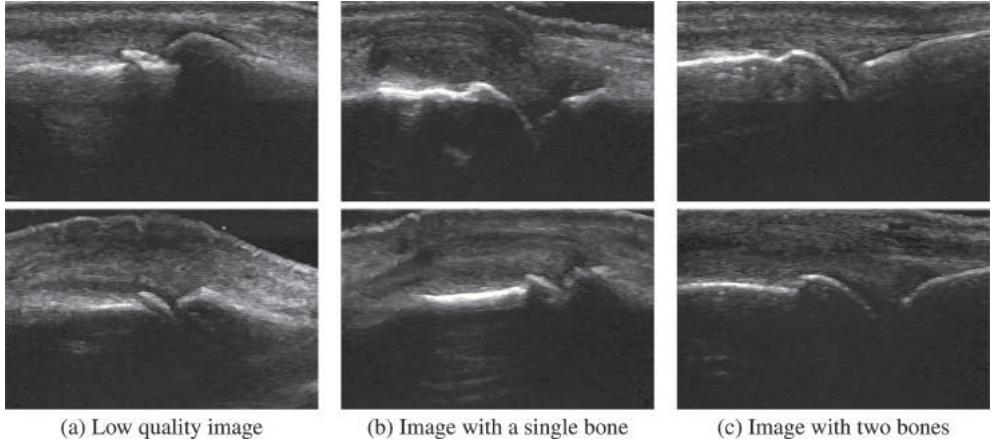


Figure 4. Test images.

For image processing applications and pattern recognition tasks the Modified Hausdorff Distance (MHD) introduced in (Dubuisson and Jain 1994) is an appropriate choice to diminish the impact of noise in the sets (Jesorsky et al. 2001). The MHD is defined as:

$$h_{mod}(\mathcal{A}, \mathcal{B}) = \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} \min_{b \in \mathcal{B}} \|a - b\|. \quad (5)$$

In this work, we propose to present the detection accuracy according to the methodology used in facial landmarks detection (Jesorsky et al. 2001, Cristinacce and Cootes 2006). A mean error between detected and manually annotated curves is not too much informative, therefore the distribution function of the relative error is used providing a good overview of detection variability within the datasets. Data proportion on the plot, shows for how many images from the evaluation dataset the modified Hausdorff distance is less than a given threshold. The error of the bones detection using the modified Hausdorff distance in pixels is presented in Figure 5. The curves show the accuracy separately for each of the two bones (the green color for the first and the red color for the second bone) in the image for our three datasets with two bones visible (DB1, DB2), single bones visible (SB1, SB2) and low quality images (LB1, LB2). As can be observed, the detection results are much better for the subset with two visible bones. We also propose to normalize the modified Hausdorff distance, by dividing the result by the length of manually annotated bone contour. These results are presented in Figure 6. The confidence map were created for default parameters suggested in the source code provided by the authors in (Karamalis et al. 2012). Exemplary results of correctly detected bones are shown in Figure 7.

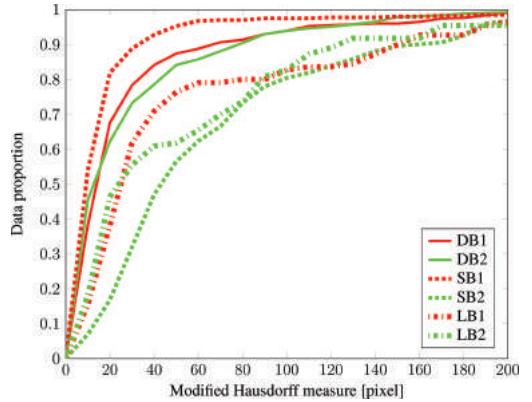


Figure 5. Cumulative error plots of modified Hausdorff measure in pixels for our dataset.

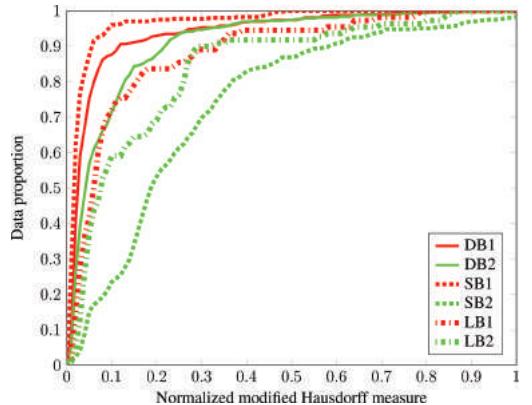


Figure 6. Cumulative error plots of normalized modified Hausdorff measure for our dataset.

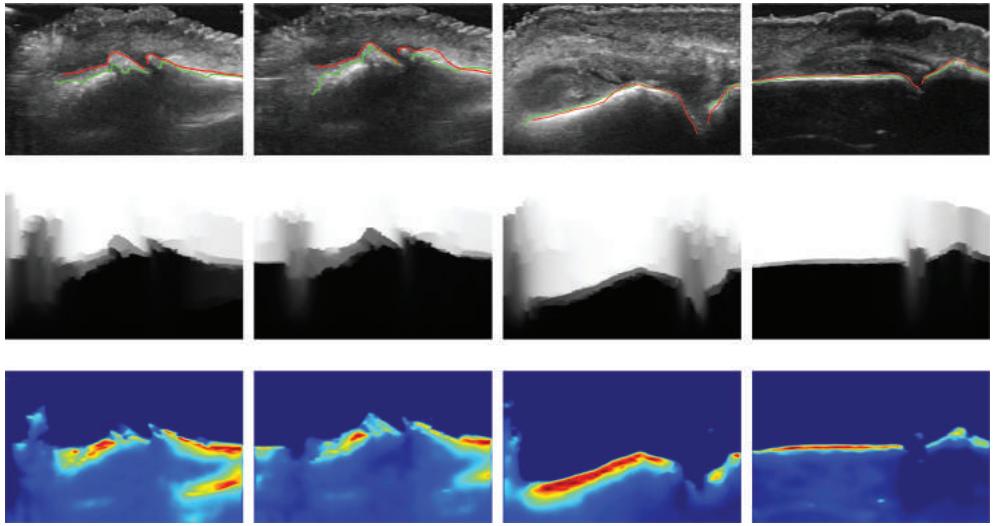


Figure 7. Exemplary results of correctly detected bones in ultrasound images and corresponding confidence and bone localizer maps.

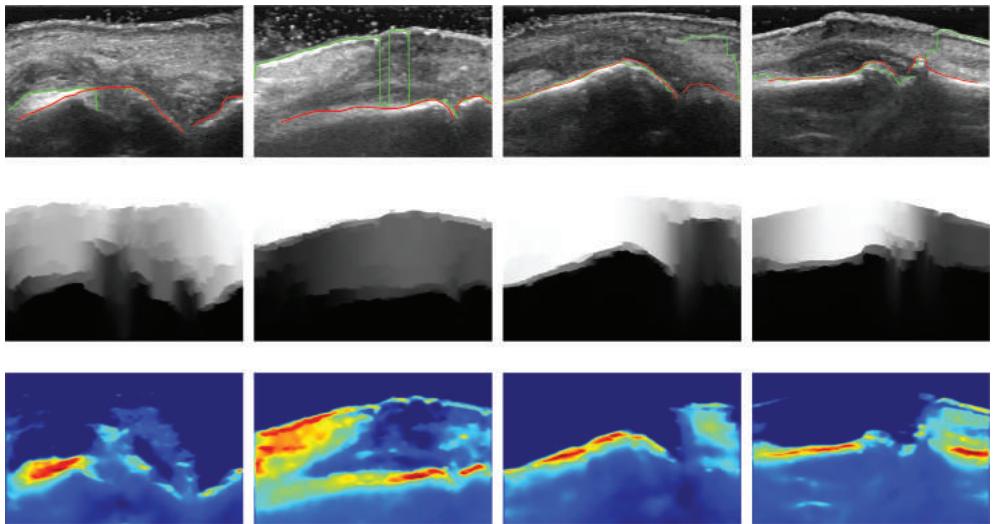


Figure 8. Some failure examples of the bone detection and corresponding confidence and bone localizer maps.

Our algorithm produces also some errors and some examples are presented in Figure 8. These cases are extremely difficult due to fact that in some images the skin surface is represented by strong edges or the edges produced by bones are not so distinguishable. This algorithm can be improved by analyzing where the values of the confidence map are significantly reduced, but this issue will be considered in the future work.

#### 4 CONCLUSION

In this paper an application of confidence map for bones region detection in ultrasound images has been presented. The bone feature descriptor is based on the confidence map that is directly related with the likelihood of ultrasound transmission and allows for accurate localization of the bones in US images. The quality of the confidence map has been significantly

improved by the application of Trimmed Non-Local Means filtering. The results were evaluated on the MEDUSA project database that contains manually annotated markers of bones surface.

This research is an initial step in the development of fully automatic system that will support the assessment of the degree of synovitis inflammation in Rheumatoid Arthritis.

## ACKNOWLEDGMENT

The research leading to these results has received funding from the Norwegian Financial Mechanism 2009–2014 under Project Contract No. Pol-Nor/204256/16/2013. Ultrasound images were created for the MEDUSA project at Section for Rheumatology; Department for Neurology, Rheumatology and Physical Medicine, Central Hospital, Førde, Norway.

## REFERENCES

- Beitzel, J., S.-A. Ahmadi, A. Karamalis, W. Wein, & N. Navab (2012). Ultrasound bone detection using patient-specific CT prior. In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2664–2667.
- Crustenacce, D. & T.F. Cootes (2006). Feature detection and tracking with constrained local models. In *Proc. BMVC*, pp. 95.1–95.10.
- Dubuisson, M.-P. & A. Jain (1994). A modified Hausdorff distance for object matching. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, Volume 1, pp. 566–568.
- Foroughi, P., E. Boctor, M. Swartz, R. Taylor, & G. Fichtinger (2007). P6d-2 ultrasound bone segmentation using dynamic programming. In *IEEE Ultrasonics Symposium*, pp. 2523–2526.
- Grady, L. (2006). Random walks for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(11), 1768–1783.
- Huttenlocher, D., G. Klanderman, & W. Rucklidge (1993). Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(9), 850–863.
- Jain, A.K. & R.H. Taylor (2004). Understanding bone responses in B-mode ultrasound images and automatic bone surface extraction using a Bayesian probabilistic framework. In *Proc. SPIE 5373*, Medical Imaging 2004: Ultrasonic Imaging and Signal Processing, pp. 131–142.
- Jesorsky, O., K.J. Kirchberg, & R. Frischholz (2001). Robust face detection using the Hausdorff distance. In *Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication*, AVBPA ’01, pp. 90–95.
- Karamalis, A. (2013). *Ultrasound Confidence Maps and Applications in Medical Image Processing*. Ph.D. thesis.
- Karamalis, A., W. Wein, T. Klein, & N. Navab (2012). Ultrasound confidence maps using random walks. *Medical Image Analysis* 16(6), 1101–1112.
- Kowal, J., C. Amstutz, F. Langlotz, H. Talib, & M.G. Ballester (2007). Automated bone contour detection in ultrasound B-mode images for minimally invasive registration in computer-assisted surgery in vitro evaluation. *The International Journal of Medical Robotics and Computer Assisted Surgery* 3(4), 341–348.
- Kulbacki, M., J. Segen, P. Habela, M. Janiak, W. Knieć, M. Fojcik, P. Mielnik, & K. Wojciechowski (2014). Collaborative tool for annotation of synovitis and assessment in ultrasound images. In *Computer Vision and Graphics*, Volume 8671 of *Lecture Notes in Computer Science*, pp. 364–373. Springer International Publishing.
- McQueen, F., M. Lassere, J. Edmonds, P. Conaghan, C. Peterfy, P. Bird, P.O’Connor, B. Ejbjerg, M. Klarlund, N. Stewart, P. Emery, R. Shnier, H. Genant, & M. stergaard (2003). OMERACT rheumatoid arthritis magnetic resonance imaging studies. Summary of OMERACT 6 MR imaging module. *The Journal of Rheumatology* 30(6), 1387–1392.
- Noble, J. & D. Boukerroui (2006). Ultrasound image segmentation: a survey. *IEEE Transactions on Medical Imaging* 25(8), 987–1010.
- ØStergaard, M. & M. Szkudlarek (2005). Ultrasonography: A valid method for assessing rheumatoid arthritis? *Arthritis & Rheumatism* 52(3), 681–686.
- Otsu, N. (1979). A threshold selection method from graylevel histogram. *IEEE Trans. Syst. Man Cybern* 9(1), 62–66.
- Quader, N., A. Hodgson, & R. Abugharbieh (2014). Confidence weighted local phase features for robust bone surface segmentation in ultrasound. In *Clinical Image-Based Procedures. Translational Research in Medical Imaging*, Volume 8680 of *Lecture Notes in Computer Science*, pp. 76–83. Springer International Publishing.
- Radlak, K. & B. Smolka (2013). Trimmed non-local means technique for mixed noise removal in color images. In *IEEE International Symposium on Multimedia (ISM)*, pp. 405–406.
- Radlak, K. & B. Smolka (2014). Adaptive non-local means filtering for speckle noise reduction. In *Computer Vision and Graphics*, Volume 8671 of *Lecture Notes in Computer Science*, pp. 518–525. Springer International Publishing.
- Terslev, L., S. Torp-Pedersen, A. Savnik, P. von der Recke, E. Qvistgaard, B. Danneskiold-Samse, & H. Bliddal (2003). Doppler ultrasound and magnetic resonance imaging of synovial inflammation of the hand in rheumatoid arthritis: A comparative study. *Arthritis & Rheumatism* 48(9), 2434–2441.
- Wereszczyński, K., J. Segen, M. Kulbacki, P. Mielnik, M. Fojcik, & K. Wojciechowski (2014). Identifying a joint in medical ultrasound images using trained classifiers. In L. Chmielewski, R. Kożera, B.-S. Shin, and K. Wojciechowski (Eds.), *Computer Vision and Graphics*, Volume 8671 of *Lecture Notes in Computer Science*, pp. 626–635. Springer International Publishing.

# Improved computer recognition of Fuhrman grading system in analysis of Clear-Cell Renal Carcinoma

M. Kruk & J. Kurek

*Warsaw University of Life Sciences, Warsaw, Poland*

S. Osowski

*Warsaw University of Technology, Military University of Technology, Warsaw, Poland*

R. Koktysz

*Military Institute of Medicine, Warsaw, Poland*

**ABSTRACT:** The paper presents the improved system of recognition of the Fuhrman grading in the analysis of clear-cell renal carcinoma. The previous system was described in (Kruk et al. 2014). The proposed improvement is based on the new segmentation method which gives better results. A redundant wavelet transformation is used for denoising the image and enhancing the edges in multiple levels of resolution. The image gradient is estimated using the wavelet transformation. The watershed transform is then applied to the obtained gradient image, thanks to this the segmented regions that do not satisfy specific criteria are removed.

## 1 INTRODUCTION

A Clear-Cell Renal Carcinoma (CC-RCC) is a very common renal disease in adults aged  $\geq 50$  years. The incidence of renal cell carcinoma has gradually increased each year over the past 30 years. The tumors show a solid, nesting, and tubular pattern of growth with tumor cells arranged in nests and separated from each other by an extensive network of delicate sinusoidal vascular channels. The most popular and widely used system for grading CC-RCC is a nuclear grading system described in 1982 by Fuhrman (Fuhrman et al. 1982).

The Fuhrman grading scheme describes 4 nuclear features for assigning a specific grade: nuclear size, nuclear shape, chromatin pattern, and size of the nucleoli. It defines a scale of 1–4, where grade 1 carries the best prognosis and grade 4 the worst. The grade is strictly correlated with the stage of development of illness. Nuclear grade has been shown to be independent on the tumor type as a prognostic factor, but its value in specific histological subtypes of renal cell carcinoma is still in question.

The grading schema of CC-RCC is based on the microscopic image of the neoplasm cells with Hematoxylin and Eosin (H&E) staining. Grade 1 tumors have the round, uniform nuclei with inconspicuous or absent nucleoli. Nuclear contours at grade 2 are more irregular than in grade 1; the

nuclei are about 15 microns in diameter. They may be visible at high magnification. At grade 3 the nuclear contours are even more irregular in size and shape. The nuclear diameters can approach 20 microns, and the nucleoli are readily seen. The size of cells at grade 4 exceeds 20 micrometers, and the pleomorphic and hyperchromatic nuclei, and prominent nucleoli in a minority of cells, are usually observed.

The main task of this paper is to improve an automatic computer system which was described in (Kruk et al. 2014) that would evaluate the Fuhrman grade of renal cells with the accuracy comparable to that of human experts. To solve the problem, we apply the segmentation algorithm described in (Jung et al. 2002) which gives better results than standard watershed algorithm. To perform classification we use the Support Vector Machine (SVM) classifier based on the data representing over 100 patients. To counteract the problem of biasing the score, we use 3 medical experts. Each expert performed the manual estimation of the Fuhrman grade of cells for the same set of images. In the learning phase of the system we have used only the cells for which class recognition was the same in the assessment of all experts. In this way we incorporate the average knowledge and experience of different experts in the learning process of the classifier, thereby reducing the interobserver and intraobserver variabilities.

## 2 MATERIALS

All medical materials used in experiments were gathered from Warsaw Military Institute of Health Services. 70 microscopic images representing all 4 grades were prepared and then used to learn the system. In the verification of the established system we used the additional 62 patients of different Fuhrman grade of the illness and analyzed 94 randomly selected images representing these patients.

The microscopic digital images of Hematoxylin and Eosin (H&E) stained kidney neoplasm cells of the appropriate tumor area were collected. They were acquired in a magnification of  $400 \times$  through an Olympus BX-61 microscope and registered with an Olympus DP-72 camera in RGB format at a resolution of  $2070 \times 1548$ . The example of such images is presented in Figure 1.

## 3 METHODS

The proposed system contains five main steps: (1) segmentation of nuclei, (2) generation of numerical descriptors, (3) evaluation and selection of descriptors, (4) application of the classifier determining the Fuhrman grade, (5) saving results to the database.

### 3.1 The segmentation algorithm

The previous segmentation algorithm was based on the gradient method of nuclei detection, instead of the often-used automatic thresholding such as Otsu method (Gonzales et al. 2008). However, its main disadvantage is too high sensitivity to the local extrema, hence sometimes the over segmentation has been observed. The proposed new approach to segmentation was described

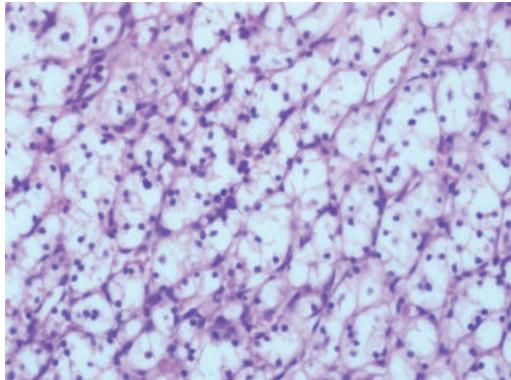


Figure 1. Example of the microscopic image of kidney tissue.

in (Jung et al. 2002) and is based on the wavelet method of image denoising and edge enhancement in multiple resolutions. In this approach, the gradients of the denoised and enhanced image are estimated using the wavelet transform, and then the watershed transform is applied to the obtained gradient image. In these experiments we have applied Daubechies wavelet function db4, and 4 levels of decomposition. Wavelet transformation improves the robustness of the watershed transformation. The watershed implementation uses the immersion simulation approach. A post-processing stage is finally applied to remove over-segmented regions with small areas, and to merge erroneously segmented regions, that are separated by weak borders after denoising and enhancement.

The example of an original image is shown in Figure 2a and the result of its segmentation is depicted in Figure 2b.

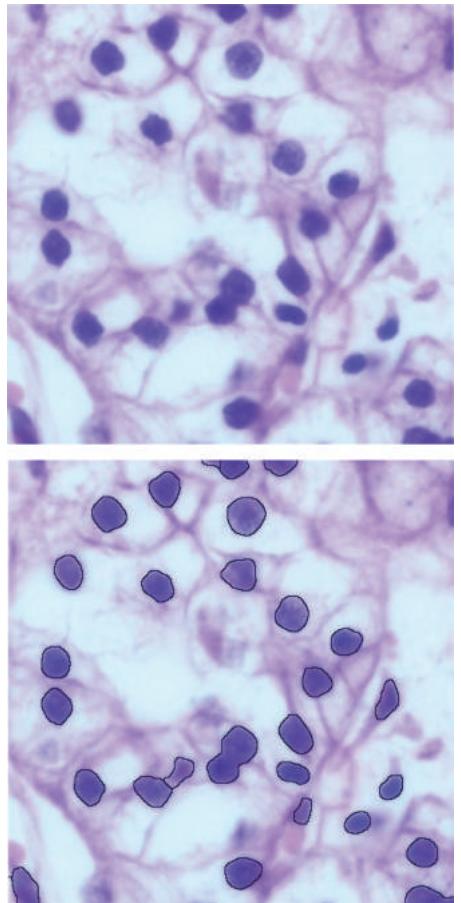


Figure 2. Original image (a) and segmented nuclei (b).

### 3.2 The set of numerical descriptors

All segmented nuclei are described by the numerical features used as the input matrix for classifier. The large set of features has been defined by authors and then feature selection process was applied to evaluate each feature and avoid these features without any value in classification task. The evaluation method was described in next subsection. The numerical features used in our experiment are divided into the following groups:

- **Texture Haralick features:** The following 13 texture features described in (Huang 2010) have been generated.

1. Energy
2. Contrast
3. Correlation
4. Sum of variances
5. Inverse difference moment
6. Entropy
7. Information measures of correlation
8. Sum average
9. Sum entropy
10. Sum variance
11. Difference variance
12. Contrast
13. Difference entropy

- **Morphometric features:**

14. Area
15. Major axis length
16. Minor axis length
17. Eccentricity
18. Convex area
19. Equivalent diameter:  $ed = \sqrt{4 * \text{Area}/\pi}$
20. Solidity:  $\text{solidity} = \text{Area}/\text{Convex area}$
21. Perimeter

- **Features describing colors:**

22. Relative sum of pixels of red components:  $sr = R/(R+G+B)$
23. Relative sum of pixels of green components:  $sg = G/(R+G+B)$
24. Relative sum of pixels of blue components:  $sb = B/(R+G+B)$

- **Features from histogram:**

25. Mean of histogram
26. Standard deviation of histogram
27. Kurtosis of histogram

- **Modified features:** Four texture features described in (Gianazza et al. 2010; Young IT et al. 1986) have been defined:

28. Heterogeneity
29. Homogeneity
30. Clump
31. Condensation

In this way 31 numerical features have been generated.

### 3.3 Features selection

Feature selection has been performed by applying Fisher measure (Duda et al. 2003):

$$Sf_{cl1c2} = |\text{mean}(c1) - \text{mean}(c2)| / (\sigma_{c1} + \sigma_{c2})$$

where  $\text{mean}(c1)$ ,  $\text{mean}(c2)$  are the mean of feature  $f$  in the first and second class, and  $\sigma_{c1}, \sigma_{c2}$  are the standard deviations of the feature in class 1 and 2, respectively.

Fisher measure arranges the features from the most important (the highest value of measure) to the least important. To find their proper population we have to set appropriate value of threshold. On the basis of some introductory experiments we have found the best results corresponding to the value of threshold equal 1.

In the numerical experiments of classification we have applied the strategy one-versus-all (Scholkopf & Smola, 2002). In this approach we have to build 4 classifiers, each responsible for recognition of one Fuhrman scale (Fuhrman scale associated with a class).

Table 1 shows the sets of features selected for each classification system arranged in this way. The numbers in the table indicate the numbering of features described earlier. As we can see the population of selected features depends on the class and extends from 14 (class 1 versus all) to 30 (class 2 versus all).

### 3.4 Classification system

As a classifier we have used Support Vector Machine of the Gaussian kernel (Scholkopf & Smola, 2002). To deal with a problem of many classes, we applied the one versus all approach. In this approach we train as many local two-class recognition classifiers as is the number of classes. On the basis of the results of all classifiers the final winner for each observation is found by applying the majority competition (Scholkopf & Smola, 2002). In learning the classifier systems we have used 3537 images of nuclei which were earlier

Table 1. Selected features.

Class 1 vs all	2 5 10 11 12 14 15 16 18 19 21 24 25 26
Class 2 vs all	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
Class 3 vs all	1 2 3 4 5 6 8 9 10 11 12 13 14 15 16 17 18 19 21 22 23 24 25 26 27 28 29 30
Class 4 vs all	2 4 5 6 7 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

Table 2. The population of cells of different classes taking part in numerical experiments.

Class	1	2	3	4
Number of cells	1193	1167	808	369

separated by the segmentation algorithm and then annotated by human experts. The population of all classes is presented in Table 2.

The number of cells forming classes differ significantly, from 1193 (class 1) to 369 (class 4). This is due to the availability of cells representing these classes in the available data base of the Military Institute of Health Services.

#### 4 RESULTS

The results of experiments will be given in the numerical and graphical forms. Figure 3 presents the annotated cells corresponding to the image of Figure 2. The recognized cells have been described by numbers, where number 1 refers to Fuhrman grade 1, number 2 to grade 2 and number 3 to grade 3.

The numerical results correspond to the accuracy, sensitivity and precision and compare the new values to the previous results presented in (Kruk et al. 2014). The confusion matrices of recognition of all Fuhrman grades of cells in the previous and our improved system is shown in Table 3 and 4. It is easy to observe that most mistakes are made between two adjacent classes. This problem is very hard to omit because the border between two classes is rather fuzzy.

It is easy to observe that the improved system has generated better results (higher values of diagonal terms and smaller values of off-diagonal elements). Similar improvement of sensitivity and precision is shown in Table 5.

Slides of images representing all investigated cases were reevaluated independently by 3 experts blind to the previous scores. Each slide was assigned with respect to Fuhrman grade according to the opinion of the particular expert. The results of expert assessments were compared to the outcome of our computerized Automatic System (AS). The comparison was done on the basis of Cohen's kappa methodology (Carletta, 1996). Kappa is widely accepted and used in the field of content analysis. It is interpretable, allows different results to be compared, and suggests a set of diagnostics in cases where the reliability results are not good enough for the required purpose. The results are presented in Table 6.

In all cases the Kappa values were larger for our improved system than for the inter experts results.

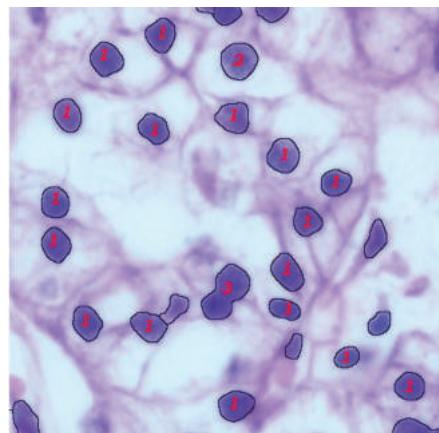


Figure 3. The graphical results of classification system: a).

Table 3. Confusion matrix of the Fuhrman grade recognition of the previous system.

	Grade 1	Grade 2	Grade 3	Grade 4
Grade 1	0.967	0.025	0.008	0
Grade 2	0.035	0.942	0.015	0.008
Grade 3	0.006	0.048	0.916	0.030
Grade 4	0.010	0.048	0.099	0.843

Table 4. Confusion matrix of the Fuhrman grade recognition of the improved system.

	Grade 1	Grade 2	Grade 3	Grade 4
Grade 1	0.978	0.015	0.007	0
Grade 2	0.030	0.951	0.011	0.008
Grade 3	0.005	0.041	0.922	0.032
Grade 4	0.010	0.038	0.091	0.861

Table 5. Comparison of class sensitivity and precision of the previous and improved versions of the Fuhrman grade recognition system.

Fuhrman grade	Grade 1	Grade 2	Grade 3	Grade 4
Previous sensitivity	96.7%	94.2%	91.6%	84.3%
Actual sensitivity	97.8%	95.1%	92.2%	86.1%
Previous precision	94.99%	88.01%	88.25%	95.69%
Actual precision	95.4%	89.9%	90.01%	96.21%

Table 6. Cohen Kappa values for the results of recognition of 4 Fuhrman grades.

	AS	Expert 1	Expert 2	Expert 3
AS	—	0.4341	0.5242	0.4730
Expert 1	0.4341	—	0.3660	0.2499
Expert 2	0.5242	0.3660	—	0.2476
Expert 3	0.4730	0.2499	0.2476	—

## 5 CONCLUSION

The paper shows that improvement of segmentation algorithm significant has influence on results of classification system.

The important advantage of the improved system is its repeatability of scores, which is in great contrast with the human expert results, significantly dependent on the particular choice of the expert and his/her mental and physical condition in the time of assessment. Moreover, the our system allows a reduction of time required for the image analysis in comparison to the human expert. Thanks to this the acceleration of the research in this area is possible.

The main task for the future is to find the method to recognize the most important area in whole specimen. Now the main region of interest must be indicated by human expert to further processing.

## REFERENCES

- Carletta J. (1996). Assessing agreement on classification tasks: The kappa statistic. Computational Linguistics 22:249–254.
- Huang P.W., Lai Y.H. (2010). Effective segmentation classification for HCC biopsy images. J Pattern Recognition 43:1550–156.
- Duda R.O., Hart P.E., Stork P. (2003). Pattern classification and scene analysis. Wiley, New York.
- Fuhrman S.A., Lasky L.C., Limas C. (1982). Prognostic significance of morphologic parameters in renal cell carcinoma. Am. J. Surg. Pathol. 6:655–663.
- Gianazza E., Chinello C., Mainini V., Cazzaniga M., Squeo V., Albo G., Signorini S., Di Pierro S.S., Ferrero S., Nicolardi S., van der Burgt Y.E., Deelder A.M., Magni F. (2012). Alternations of the serum peptidome in renal cell carcinoma discriminating Benin malignant Sidney tumors, J Proteom 76:125–140.
- Gonzalez R., Woods R., 2008: Digital Image Processing. New Jersey, Prentice Hall.
- Huang, M., & Haralick, R.M. (2010). Recognizing Patterns in Texts. Pattern Recognition and Machine Vision, 6, 19.
- Jung, C.R., & Scharcanski, J. (2002). Robust watershed segmentation using the wavelet transform. In Computer Graphics and Image Processing. Proceedings. XV Brazilian Symposium on (pp. 131–137). IEEE.
- Kruk M., Osowski S., Markiewicz T., Ślądkowska J., Koktysz R., Kozłowski W., Świderski B., (2014). Computer approach to recognition of Fuhrman grade of cells in clear-cell renal cell carcinoma. Analytical and Quantitative Cytology and Histolgy 36(3):147–160.
- Schölkopf, B., & Smola, A.J. (2002). Learning with kernels: Support vector machines, regularization, optimization, and beyond. MIT press.
- Young I.T., Verbeek P.W., Mayal B.H. (1986). Characterization of chromatin distribution in cell nuclei. Cytometry 7:467–474.

This page intentionally left blank

## A proof of concept of an augmented reality system for Nuss surgery

A. Ferreira

*ICVS/3B's—PT Government Associate Laboratory, Braga/Guimarães, Portugal*

P. Morais

*ICVS/3B's—PT Government Associate Laboratory, Braga/Guimarães, Portugal*

*INEGI, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal*

S. Queirós

*ICVS/3B's—PT Government Associate Laboratory, Braga/Guimarães, Portugal*

*Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal*

F. Veloso

*ICVS/3B's—PT Government Associate Laboratory, Braga/Guimarães, Portugal*

N.F. Rodrigues

*ICVS/3B's—PT Government Associate Laboratory, Braga/Guimarães, Portugal*

*DIGARC, Polytechnic Institute of Cávado and Ave, Barcelos, Portugal*

J. Correira-Pinto

*ICVS/3B's—PT Government Associate Laboratory, Braga/Guimarães, Portugal*

J.L. Vilaça

*ICVS/3B's—PT Government Associate Laboratory, Braga/Guimarães, Portugal*

*DIGARC, Polytechnic Institute of Cávado and Ave, Barcelos, Portugal*

**ABSTRACT:** Pectus Excavatum (PE) is the most common congenital chest wall deformity, affecting 1 in 400 live births. This deformity is commonly corrected using the minimally invasive Nuss procedure, where a bar is positioned under the sternum. Although recent procedure advances based on patient-specific prosthesis were proposed, correct bar placement is still challenging.

In this work, we propose a novel augmented reality system to guide the surgeon during PE bar placement. This system combines a 3D sensor with a projector to superimpose the thoracic ribs cage on the chest wall of the patient, thus indicating the optimal insertion and bar placement points.

This system was validated in three different scenarios: 1) simulated chest surface models; 2) 3D printed phantom; and 3) 3D commercial thoracic phantom. An error of  $3.93 \pm 3.44$  mm, and  $3.08 \pm 1.57$  mm were obtained in the first and second experiments, respectively. In the final experiment, visual assessment of the result proved that a high similarity was obtained between the projected model and the real ribs cage position.

Overall, the proposed system showed high feasibility with low error, proving that 3D projection of the ribs on the patient's chest wall may facilitate PE bar insertion and ultimately provide useful information to guide Nuss procedure.

**Keywords:** augmented reality; pectus excavatum; Nuss surgery

### 1 INTRODUCTION

Pectus Excavatum (PE) is characterized by a depression of the anterior chest wall. It is the most common congenital chest wall deformity, accounting for 90% of all congenital chest wall abnormalities and affecting around 1 in 400 live

births (Koumbourlis 2009). The most common treatment of this condition is the Minimally Invasive Surgery (MIS) introduced in 1998 by Donald Nuss, which consists in inducing remodelling of the chest wall to a normal shape by inserting a bar behind the sternum. This procedure has become well accepted by pediatric and thoracic surgeons

due to its technical simplicity, reproducibility and long-term functional and cosmetic results. Despite allowing the successful PE correction in 93%–96% of the cases, the Nuss procedure has an overall complication rate of 33%, namely associated to persistent pain after operation, pneumothorax, seromas, infection and bar migration (Fallon et al. 2013). Such complications are mainly related to the conventional bar modelling process, which consists in the intra-operative manual modelling of the bar, and the lack of surgery planning.

Vilaça et al. (2013) proposed a system to automatically bend a personalized bar for PE correction based on the patient's anatomy and tools for virtual simulations of the postsurgical outcome (Moreira et al. 2012, Vilaça et al. 2011). Similarly, Zhao et al. (2014) proposed a framework for chest modelling and personalized surgical planning for PE. Although the abovementioned frameworks present several advantages for the patient through an optimal deformity correction with superior final cosmetic result, the bar placement task is still dependent of the surgeon, which could be difficult due to: 1) loss of visual drive; 2) reduced depth perception; 3) limited angle of view; and 4) limited touch sensation (Pessaux et al. 2014).

Augmented Reality (AR) is a computer-based technology that can display virtual information in the real environments, such that virtual objects appear to coexist in the same space as the real world. This technology has been used in medical procedures such as neurosurgery, laparoscopy, otolaryngology, orthopaedics and maxillofacial surgery. Since these systems can create virtual transparency of the patient anatomy and consequently project it at the body surface, an increase of the surgeon vision is obtained, which facilitates the surgical procedure and, consequently, reduces the rate of procedural complications (Okamoto et al. 2015).

In this work, we propose a system to solve the aforementioned problems in minimally invasive PE procedures based on an augmented reality system to superimpose the patient-specific ribs cage on the patient's chest wall. Such system will increase the safety and accuracy of the Nuss surgery by providing important visual information directly on the patient and, therefore, allowing the identification of the correct sites for insertion of the bar, while following the steps established during the pre-operative planning. In this way, a reduction of the possible procedure complications, due to the incorrect bar placement, is expected.

## 2 SYSTEM OVERVIEW

The projector-based AR system for Nuss surgery assistance is composed by a Microsoft Kinect

version 2 sensor and a projector Epson EH-TW550. The Kinect is a physical device with depth sensing technology, a built-in color camera and infrared (IR) emitter, enabling 3D real-time scanning of the patient's chest wall. To guarantee that the patient's ribs cage is projected in the right place, both Kinect and projector are calibrated using a gray code projection. Then, the 3D model obtained from pre-operative Computed Tomography (CT) images segmentation is registered with the patient's chest wall and ultimately projected over him. Figure 1 presents the flowchart of the current system.

### 2.1 Kinect/projector calibration

Since two independent modules (Kinect sensor and projector) are used to create an accurate projection of the chest wall, a calibration step should be performed to define a common world between them. The proposed system uses gray code patterns to obtain the dense point correspondences between the projector and Kinect images (Fig. 1.a). Each pixel in the projected patterns encodes the corresponding projector image coordinates. The patterns are observed by the Kinect and decoded to obtain the correspondences (Schmalz et al. 2011). These correspondences are used to transform the patient's CT information from the Kinect's coordinate system to the projector's coordinate system.

Note that gray code patterns recognition is an active calibration method that makes usage of self-identifying patterns that can be shown by the projector. As such, the tedious manual identification of marks required by a passive calibration method become unnecessary (Schmalz et al. 2011).

### 2.2 Virtual information registration

To perform a markerless registration of the virtual information with the patient, the proposed system uses the Kinect sensor to acquire the 3D model of the patient's chest wall (Khoshelham and Elberink 2012). Then, the chest wall surface, segmented from the pre-operative CT images using a simple threshold, is registered with the Kinect's information (Fig. 1.b, c, d). This registration is based on the Iterative Closest Point algorithm (ICP). However, a first alignment between the two point sets is performed using the Sample Consensus Initial Alignment (SAC-IA) algorithm proposed by Rusu et al. (Rusu et al. 2009).

The SAC-IA performs fast searches in an exhaustive Fast Point Feature Histograms (FPFH) correspondence space to find a good alignment solution, which can be further refined using the ICP algorithm (Rusu et al. 2009). The FPFH are the optimization of the Point Feature Histograms,

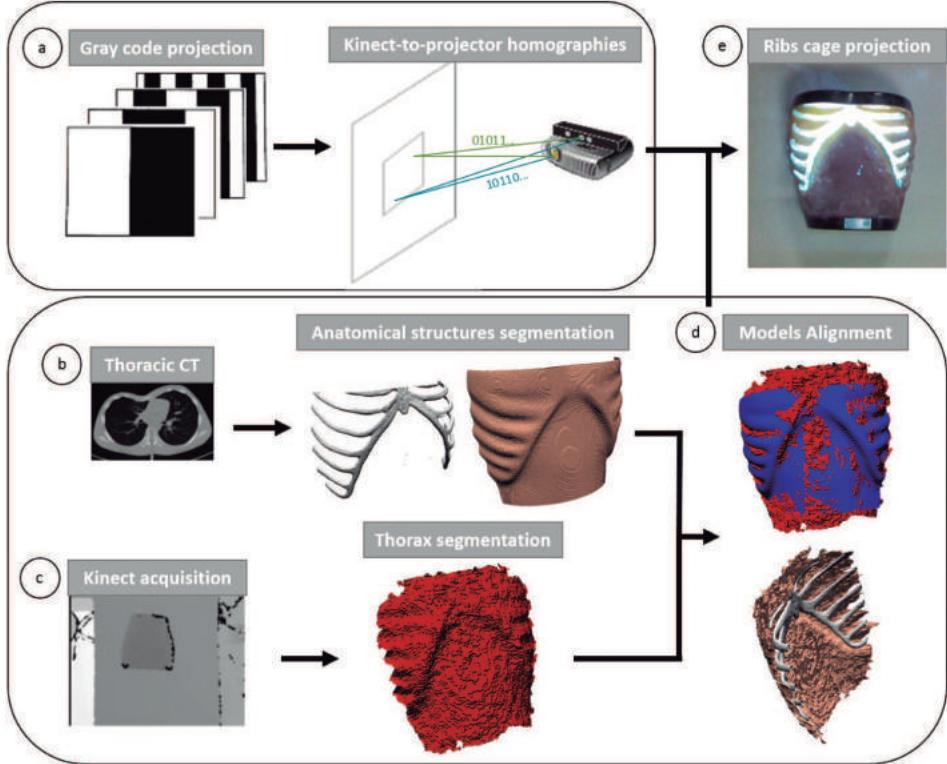


Figure 1. System overview: (a) Microsoft Kinect Sensor/Projector calibration with gray code patterns; (b) Segmentation of the chest wall, rib bones, cartilages and sternum from pre-operative Computed Tomography (CT) data; (c) Patient's chest wall 3D scanning obtained with Kinect; (d) Alignment of the 3D CT models with the Kinect model; (e) Projection of the rib bones, cartilages and sternum on top of the patient's chest wall.

a robust multi-dimensional features that describe the local geometry around a point for 3D point clouds datasets. After estimation of the FPFH for the Kinect and CT point clouds, the SAC-IA proceed as follow:

- Select sample points from one 3D model while making sure that their pairwise distances are greater than a user-defined minimum distance;
- For each of the sample points, find a list of points whose histograms are similar to the sample points' histogram. From these, select one randomly which will be considered as sample points' correspondence;
- Compute the rigid transformation defined by the sample points and their correspondences, and compute an error metric for the point cloud that computes the quality of the transformation (Rusu et al. 2009).

These three steps are repeated, and the transformation that presents the best error metric is stored and used to roughly align the two point clouds.

Consequently, the ICP algorithm is used to refine the alignment.

The ICP is a popular algorithm employed to register two sets of point clouds  $X, P$ , by iteratively matching points in one set to the closest points in another set, consequently refining the transformation between the two sets in order to minimize the distance between both point clouds. Then, given two corresponding point sets:  $X = \{x_1, \dots, x_n\}, P = \{p_1, \dots, p_n\}$ , the ICP algorithm computes the translation  $t$  and rotation  $R$  that minimizes the sum of the squared error:

$$E(R, t) = \frac{1}{N_p} \sum_{i=1}^{N_p} \|x_i - Rp_i - t\|^2, \quad (1)$$

where  $x_i$  and  $p_i$  are the corresponding points (Besl and McKay 1992). To compute the transformation between  $X$  and  $P$ , the algorithm iterate the following steps:

- i. Application of the optimal transformation estimated in the previous iteration;
- ii. Find the closest point in the second set, for each point in the first transformed point set;
- iii. Update the point matches by discarding outliers;
- iv. Compute the transformation using the updated point matches, until convergence of the estimated transformation (Zhang 2014).

Finally, after alignment between the CT information with the Kinect 3D data, the other CT segmentation structures, namely bones, cartilages and sternum, can be transformed to the projector's coordinates and overlaid in the patient's chest wall (Fig. 1.e).

### 3 EXPERIMENTS AND RESULTS

To validate the proposed system, three different scenarios are used. First, a set of translated and rotated 3D chest wall models were registered with a set of chest walls. Second, a 3D printed object is used as a phantom model to validate the proposed system. Finally, preliminary results with a commercial thoracic phantom are also presented.

#### 3.1 3D registration

A 3D ideal model of a chest wall was used to assess the accuracy of the proposed 3D registration method. As such, it was created a set of novel models through the addition of white Gaussian noise on the ideal model. Note that the white noise varied from 25 to 35 dB of Signal-to-Noise Ratio (SNR). With these distorted models, we intend to mimic the 3D models obtained with the Kinect, which is susceptible to have some noise due to the technology limitations and lighting conditions.

Regarding the ideal model, a set of translations and rotations ( $30^\circ$  to  $270^\circ$ ) were performed to simulate misalignment between the ideal model and simulated model. It is important to note that the translation performed in the second set of 3D models shifted them in a manner that no overlap occurred with the original model on the original position.

Then, the proposed 3D registration method was applied between the two model sets, where the set with white Gaussian noise was used as target while the second set was used as the registration model.

Figure 2 shows the mean distance error between the registered models and the original CT model at different SNRs.

#### 3.2 Phantom test

In order to validate the calibration between projector and Kinect, a 3D printed phantom (Fig. 3.a) was used. The current phantom consists on a set of cylinders with different radii and heights. Note that this phantom was initially created in a Computer-Aided Design (CAD) tool and then printed through a 3D-printing process.

Since the Kinect is fixed, only the surface directly in sight of it are acquired (see Fig. 3.b). As such, one needs to perform the registration of the phantom's complete virtual model with the acquired phantom's partial 3D model. After registration, the top surfaces of each cylinder were projected on the phantom (Fig. 3.b), allowing to measure the displacement of the centers from the phantom's cylinders and the reprojected surfaces.

As a final remark, one should note that different projection distances were used (1000 and 1500 mm) to prove that an accurate result can be obtained independently of the focal distance.

Regarding the results, a mean reprojection error of  $3.08 \pm 1.57$  and  $6.87 \pm 1.35$  mm at 100 and 150 cm distance, respectively, were obtained.

#### 3.3 Thoracic model test

The proposed system was also tested with an ABDFAN thoracic model (commercial phantom). Initially, a 3D CT acquisition of the phantom model was performed. Using this dataset, a 3D ribs surface and a 3D chest surface were extracted by applying a simple threshold on the CT image. Similarly, the proposed AR system was applied and a 3D point set of the chest surface from the commercial phantom was acquired by the Kinect sensor. This point set is aligned with the segmented

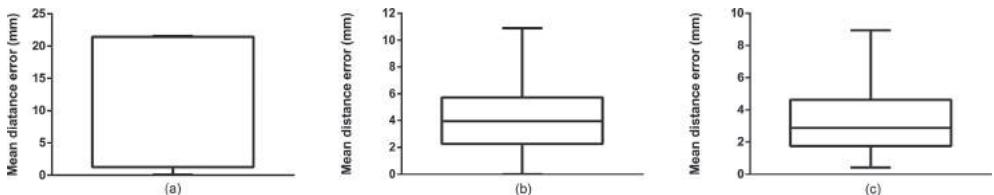


Figure 2. Mean distance error of the 3D registration using targets with different SNR of white Gaussian noise: (a) 25 dB, 30 dB and 35 dB.



Figure 3. 3D printed phantom for system accuracy test: (a) 3D virtual model, (b) test result at 1500 mm distant from the phantom's base.



Figure 4. Projection of the rib cage on top of a Kyoto Kagaku ABDFAN thoracic model.

3D chest wall surface using the ICP strategy, defining an optimal transformation between them. As a final step, this optimal transformation is used to project the CT 3D ribs surface onto the phantom model. Note that a previous calibration between the Kinect and projector was performed, as described in Section 2.1.

Figure 4 shows the projection result, where the 3D ribs model is superimposed on the commercial phantom.

#### 4 DISCUSSION

Figure 2 presents the registration error with different SNR of white noise. As expected, lowest SNR increase the registration errors. Moreover, a minimum SNR of 30 dB should be used to obtain an acceptable result. Once the target presents a SNR

equal to or higher than 30 dB, the registration method presents good results with a mean error distance around 5 mm, being in accordance with previous studies available in literature (Almhdi et al. 2007). Since the 3D models obtained with Kinect does not present a SNR with such low magnitude, the developed registration method is expected to present a good performance for the intended application.

Regarding the second experiment, the results obtained with the 3D printed phantom indicates that the system have a low reprojection error, inferior to 7 mm, in the different positions. This fact is important to ensure that the system will perform as projected, even in different surgical setups. As the common distance between ribs is approximately 15 mm (Yilmaz et al. 2014), the reported low error magnitude also shows that the present system can represent the intercostal space with the necessary accuracy. Thus, the results corroborate that the described AR system can project useful information to assist surgeons placing the pectus bar. It is also important to refer that the system has better performances when is placed closer to the projection target. This is explained by the operation of the Kinect, which can acquire a 3D surface with more details and vertices when is closer to the target.

The commercial thoracic model test allowed to evaluate the performance of the proposed system in a setup closer to the real intended application. By observing Figure 4, it is possible to support the utility of the system during the Nuss surgery, since the projection of the ribs cage on top of the model provides valuable visual information of the ribs location to the surgeons. The access to this information during Nuss surgery will assist the correct placement of the PE bar, which presents potential to overcome some complications like persistent pain after operation, pneumothorax, seromas, infection and bar migration.

#### 5 CONCLUSION

The proposed AR system promises to be useful on surgical procedures like the Nuss surgery, providing valuable information throughout PE bar placement. In this work, the proposed AR system was validated in three different experiments, proving the feasibility and accuracy of the presented system.

In the near future, the described augmented reality system will be tested in clinical practice. Given the experimental results obtained so far, it is expected to have a good performance, helping the surgeons to identify the location of the ribs and perform the surgical procedure as pre-operatively planned.

## REFERENCES

- Almhdié, A., C. Léger, M. Deriche, & R. Lédée (2007). 3d registration using a new implementation of the icp algorithm based on a comprehensive lookup matrix: Application to medical imaging. *Pattern Recognition Letters* 28(12), 1523–1533.
- Besl, P.J. & N.D. McKay (1992). A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14(2), 239–256.
- Fallon, S.C., B.J. Slater, J.G. Nuchtern, D.L. Cass, E.S. Kim, M.E. Lopez, & M.V. Mazzotti (2013). Complications related to the nuss procedure: Minimizing risk with operative technique. *Journal of Pediatric Surgery* 48(5), 1044–1048.
- Khoshelham, K. & S.O. Elberink (2012). Accuracy and resolution of kinect depth data for indoor mapping applications. *Sensors* 12(2), 1437–1454.
- Koumbourlis, A.C. (2009). Pectus excavatum: Pathophysiology and clinical characteristics. *Paediatric Respiratory Reviews* 10(1), 3–6.
- Moreira, A.H., P.L. Rodrigues, J. Fonseca, A. Pinho, N. F. Rodrigues, J. Correia-Pinto, & J.L. Vilaca (2012). Pectus excavatum postsurgical outcome based on preoperative soft body dynamics simulation. In *SPIE Medical Imaging*, pp. 83160 K–83160 K. International Society for Optics and Photonics.
- Okamoto, T., S. Onda, K. Yanaga, N. Suzuki, & A. Hattori (2015). Clinical application of navigation surgery using augmented reality in the abdominal field. *Surgery Today* 45(4), 397–406.
- Pessaux, P., M. Diana, L. Soler, T. Piardi, D. Mutter, & J. Marescaux (2014). Robotic duodenopancreatectomy assisted with augmented reality and real-time fluorescence guidance. *Surgical Endoscopy* 28(8), 2493–2498.
- Rusu, R.B., N. Blodow, & M. Beetz (2009). Fast Point Feature Histograms (FPFH) for 3D registration. In *IEEE International Conference on Robotics and Automation, 2009. ICRA'09.*, pp. 3212–3217. IEEE.
- Schmalz, C., F. Forster, & E. Angelopoulou (2011). Camera calibration: active versus passive targets. *Optical Engineering* 50(11), 113601–113601–10.
- Vilaça, J.L., A.H. Moreira, L. Pedro, N. Rodrigues, J.C. Fonseca, A. Pinho, & J. Correia-Pinto (2011). Virtual simulation of the postsurgical cosmetic outcome in patients with pectus excavatum. In *SPIE Medical Imaging*, pp. 79642 L–79642 L. International Society for Optics and Photonics.
- Vilaça, J.L., P.L. Rodrigues, T.R. Soares, J.C. Fonseca, A.C. Pinho, T. Henriques-Coelho, & J. Correia-Pinto (2013). Automatic prebent customized prosthesis for pectus excavatum minimally invasive surgery correction. *Surg Innov* 21(3), 290–296.
- Yilmaz, M., C.A. Dickman, & N.R. Crawford (2014). Morphometric analysis of the thoracic cage in adults: Anatomical considerations in relation to neurosurgical thoracoscopic procedures. *Journal of Neurological Sciences (Turkish)* 31(1), 107–114.
- Zhang, Z. (2014). Iterative Closest Point (ICP). *Computer Vision: A Reference Guide*, 433–434.
- Zhao, Q., N. Safdar, C. Duan, A. Sandler, & M. Linguraru (2014). *Chest Modeling and Personalized Surgical Planning for Pectus Excavatum*, Volume 8673 of *Lecture Notes in Computer Science*, Book section 64, pp. 512–519. Springer International Publishing.

# Usage of mobile devices in a bone fracture reduction process

J.R. Jiménez, F. Paulano, J.M. Noguera & J.J. Jiménez

*Graphics and Geomatics Group of Jaén, Universidad de Jaén, Jaén, Spain*

**ABSTRACT:** The market share of mobile devices has spectacularly increased in a few years. The novel characteristics of this kind of devices: intuitive interfaces and everywhere/everytime access, has promoted its widespread use in many different scenarios. In addition, medical trainees and specialist are familiarized with them. On the other hand, wearable devices are a very promising new line of products even for medical applications. In recent years, many computational techniques have been proposed to help the specialist in the fracture reduction process. However, due to its intrinsic complexity and its high technological requirements, this research line is still open. The increasing computational power and the birth of new technology favor the chances of automating more tasks reducing risks and surgery time exposition. Nevertheless, the use of mobile devices is not widespread in this specific medical context. In this paper, we briefly introduce the possibilities of mobile and wearable devices in a fracture reduction process from a global perspective. Different areas of research have to be taken into account to get a practical overview of the problem and the connections among them. Specifically the focus is placed on the visualization of medical data, the interaction, collaborative aspects and augmented reality possibilities.

**Keywords:** mobile devices; bone fracture reduction; 3D medical images; visualization; computer supported cooperative work; user interaction; augmented reality

## 1 INTRODUCTION

Their growing capabilities, their everywhere/everytime access and their friendly user interface have favored the proliferation of mobile devices in the last years. In fact, market sales data confirm that this new line of products has surpassed the sales of traditional desktop PCs and laptops. This is a long-term tendency based on current users behavior and preferences (Gartner 2013).

Mobile devices have become ubiquitous in medical-care professional settings (Kho, Henderson, Dressler, & Kripalani 2006, Whipple, Allgood, & Larue 2012, Berkowitz, Kung, Eisenberg, Donohoe, Tsai, & Slanetz 2014) and therefore medical trainees and physicians have already been using their mobile devices for a decent amount of time. For this reason they are already familiarized with the natural multi-touch interfaces and interaction metaphors that these platforms provide (Botha, Preim, Kaufman, Takahashi, & Ynnerman 2012), and it seems natural to use them for as many tasks as possible (Butson, Tamm, Jain, Fogal, & Kruger 2013).

In addition, this new trend has not passed unnoticed to the research community that is increasingly taking advantage of the mobile devices GPU-enabled capacities in different fields such as 3D visualization techniques, Augmented Reality (AR), and interaction methods (Noguera & Torres 2013, Noguera & Jimenez 2015).

Furthermore, a related new promising line of products is consolidating its technology and features: wearable devices. This new kind of gadget keeps one of the main features of handheld devices its everywhere/everytime access.

Yet, to the best of our knowledge we have not found proposals that cover totally or partially a bone fracture reduction process in a mobile device. This paper reviews the possibilities and challenges of using mobile devices and the like in the processes and activities related with a bone fracture reduction. We place together the bone fracture reduction issues and those novel characteristics of this kind of devices that somehow can be of help. This pursuing leads us to different areas of research that we briefly introduce. Consequently, the set of topics associated with the scope of this study conveys us to devise human-computer interaction, augmented reality, visualization and Computer Supported Cooperative Work (CSCW) challenges, see Figure 1.

The paper is structured as follows: Section 2 briefly introduces the common stages and activities related with a bone fracture reduction. Based on these stages, next sections emphasize the techniques based on mobile devices that can be applied to some of them, their requirements and challenges. Specifically, Section 3 reviews the visualization of 3D medical images, Section 4 details the possibilities of utilizing augmented reality in a bone fracture

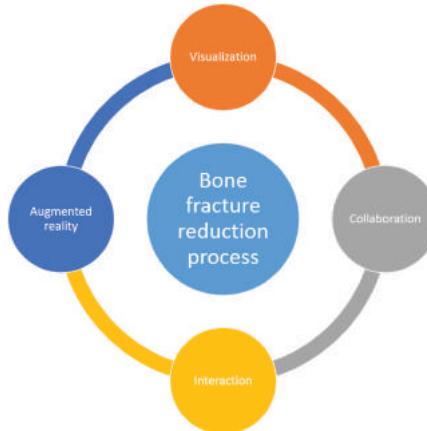


Figure 1. Research areas of interest related with the topics of this study: bone fracture reduction activities and mobile devices features.

reduction process and Section 5 highlights the corresponding human computer interaction requirements when part of the activities are developed with a mobile device and Section 6 focuses on the conditions and difficulties to design adequate collaborative environments. Finally, Section 7 discusses among the different issues and concludes the paper.

## 2 THE BONE FRACTURE REDUCTION PROCESS

The main stages of a bone fragment reduction process are: diagnosis, planning and intervention. The activities related with this process can also be classified according to their position with respect to the surgery as: pre-surgery, intra-surgery, post-surgery (Jaramaz, Hafez, & DiGioia 2006), see Figure 2. Next, we briefly describe the most common activities and tasks related with a bone fragment reduction.

Computer Tomography (CT) is the most common method to acquire 3D images in order to distinguish bone tissue. Collective or individually, the diagnosis of a bone fracture is usually accomplished by directly using this kind of images.

One of the most important activities in order to increase the chances of a successful surgery is planning. An adequate planning can reduce surgery time and minimize risks, favoring a faster recovery (Sikorski & Chauhan 2003, Sugano 2003). Firstly, the planning of a bone fracture reduction requires generating 3D models. This task is achieved by segmenting the bone tissue and label all the fragments (Paulano, Jiménez, & Pulido 2014). The special features that present the fractured bone in a CT

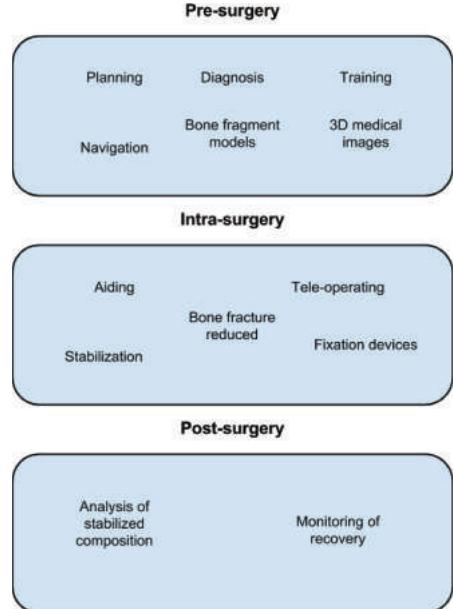


Figure 2. Activities and resources classified according to their position with respect to surgery.

scan make these tasks harder (Paulano, Jiménez, & Pulido 2013). The resulting 3D model can be a cloud of points, a surface or a soup of triangles depending on the problem to be solved. Secondly, the next step consists of virtually reducing the fracture. In this second task, bone fragments are placed in their correct position. Some methods have been proposed in order to deal with this process (Willis, Anderson, Thomas, Brown, & Marsh 2007) (Chowdhury, Bhandarkar, Robinson, & Yu 2009). In order to avoid that fragment loses their correct position, the fracture need to be stabilized by including fixation devices. Several works propose to perform this process in a virtual way (Hu, Li, Qiao, Liu, Ji, & Ye 2011). Finally the reduced fracture can be analyzed to check the goodness of the stabilization. For this purpose, finite element analysis is used most of the times (Blessy, Anburajan, & Ramkumar 2011). An analysis can be performed to determine the goodness of a virtual reduction or as a post-surgery activity.

Furthermore, a continuous training program for medical specialists and practitioners improves the common knowledge about new methodologies and technology at hand. A decent variety of cases are required in order to provide a satisfactory training program. This necessity is a relevant issue especially regarding bone fractures. As a result, a virtual generator of bone fractures is an important challenge.

Regarding intra-surgery activities it is not a novelty the use of systems related with Computer

Assisted Surgery (CAS) or tele-operating. The main goal are to shorten the surgery time, to reduce the radiation exposure of the patient, to increment the precision and to increase the knowledge of what the surgeon is doing. Real time navigation systems for intramedullary nails or robot-guided surgery are good examples of these systems.

### 3 VISUALIZATION OF 3D MEDICAL IMAGES

As it has been previously commented 3D CT images are the most recommended to distinguish bone tissue. The CT scan technique allows acquiring a set of 2D images representing a 3D model of a part of a body. The value stored in a cell of one of these image represents an intensity. A specific range of these intensities represents bone tissue. Consequently, by using a transfer function it is possible to remove some components (tissue) and obtain an image where almost nothing but bones are visible, see Figure 3. In fact, a bone has different components that are stored with different intensities. Two components can be distinguished in a bone: cortical (outer part) and trabecular (inner part) tissue. As a consequence, in the case of bone fragments, sometimes it is not straightforward to define a valid transfer function.

This section is focused on introducing the main strategies to visualize a 3D medical image in order to obtain a diagnostic. There are two main approaches to resolve this visualization problem: a 2D and a 3D approach. A 2D visualization demands few computational requirements and, as a consequence, it is more applicable to mobile devices. A 2D visualization is based on the displacement of three planes along their corresponding axes. The result, is a composite image that is valuable to specialists, see Figure 4.

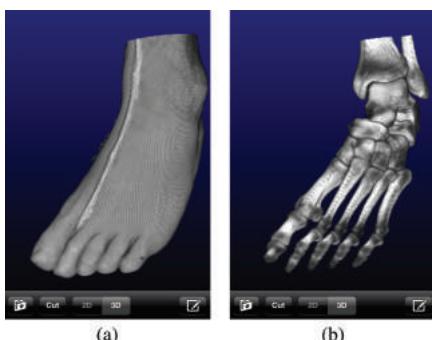


Figure 3. The same 3D medical images visualized with different transfer functions on a mobile device. Figure (b) uses an adequate transfer function to distinguish bone tissue.

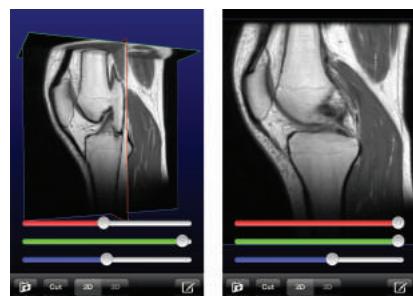


Figure 4. An example of a 2D visualization.

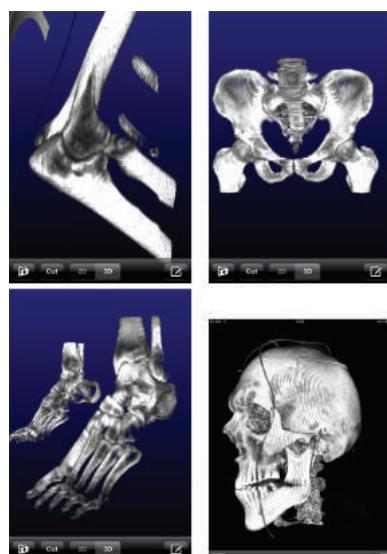


Figure 5. Examples of CT 3D medical images on a mobile device.

On the other hand, direct volume rendering techniques have traditionally been preferred to get a 3D visualization of 3D medical images (Hadwiger, Kniss, Rezk-Salama, Weiskopf, & Engel 2006). The first attempts to visualize 3D medical images in a mobile device date back to 2005 (Lamberti & Sanna 2005, Zhou, Qu, Wu, & Chan 2006). In these initial studies, the low graphics capacity of handheld devices impeded an independent rendering of 3D images. As a consequence, a server was mandatory to fulfill this task and submit the resulting images to the device.

Since the arrival of GPU-enabled mobile devices the whole rendering process can be carried out within their own graphics processors (Campolalegre, Brunet, & Navazo 2012, Movania & Feng 2012, Noguera, Jiménez, Ogáyar, & Segura 2012, Noguera & Jiménez 2012). A few examples are depicted in Figure 5.

However, the power limitation of these devices in comparison with their desktop PC counterpart severely limits the size of the volume and the required Frames Per Second (FPS) for a real interactive user experience. (Vázquez Alcocer & Balsa Rodríguez 2012, Noguera & Jimenez 2015). In spite of depending on a reliable network bandwidth and a powerful server in order to provide a good service to the client interaction demands, the referred limitations of mobile devices are one of the reasons that justify the current necessity of client-server architectures in real environments.

In fact, the mobile volume rendering research area is still open and it is focused on improving the performance and the compression techniques in order to be able to increase the dimension of the volume in a real interactive and useful application (Noguera & Jimenez 2015).

#### 4 AUGMENTED REALITY SYSTEMS

Augmented Reality (AR) consists of a mix visualization of live real world and artificially generated objects or media. In the context of fracture reduction, it would be the mixture of real patient images and extra images and information generated during the planning or even in real time during surgery.

In the last decade, many developments have taken place in the application of AR techniques in surgical procedures. Bockholt et al. (Bockholt, Bisler, Becker, Muller-Wittig, & Voss 2003) use AR techniques to overlap preoperative scanned image data and results of the planning to the operation field. Pandya and Auner (Pandya & Auner 2005) superimpose live video with the synchronized CT-derived segmented objects of interest. Fischer et al. (Fischer, Eichler, Bartz, & Straß er 2007) present a hybrid tracking method for medical AR. The final pose estimation takes the initial infrared tracking data as well as salient features in the camera image into account. Soler et al. (Soler, Nicolau, Fasquel, Agnus, Charnoz, Hostettler, Forest, Mutter, & Marescaux 2008) utilize an AR system that superimpose preoperative data onto the real view of the patient during a laparoscopic procedure. The system allow tracking instruments and improve pathology targeting. Liao et al. (Liao, Inomata, Sakuma, & Dohi 2010) develop an image overlay navigation system for open MRI-guided surgery. For that purpose, they present image calibration and markers-based registration methods for patient-to-image registration.

Regarding bone fracture reduction, Yao et al. (Yao, Pang, Liu, & Hu 2009) presents an aided planning and training interactive system frame for defective bone fracture reduction surgery. Shen et al. (Shen, Chen, Guo, Qi, & Shen 2013) use AR

techniques to develop a preoperative reconstruction plate design system for unilateral pelvic and acetabular fracture reduction and internal fixation surgery.

Regarding mobile devices, generally speaking, the use of tablets or smartphones are not advisable during surgery. However, wearable devices, such as smart glasses, can fit in this activity and offer interesting novelties and opportunities as a computer aid system during surgery. Nevertheless, special tests have to be passed in order to meet the required policies and regulations.

#### 5 HUMAN COMPUTER INTERACTION ISSUES

In order to enrich the visualization of a bone fracture model, medical specialists usually demand navigational features to mark the model with some useful measures such as angles, distances, etc., and to set correspondences between fragments. In this sense, Park et al. (Park, Kim, & Ihm 2008) proposal allows the user to draw vector data by using the touch-screen for annotation purposes. On the other hand, Jeong et al. (Jeong & Kaufman 2007) presented an interactive virtual colonoscopy and let the user to enter the volume.

Additionally, the reduction process requires manipulating the fragments: selection, rotations and translations. We refer the reader to the work of Telkenaroglu et al. (Telkenaroglu & Capin 2013) for an extended review of touch-screen user interaction techniques. The selection of the tools to handle bone fragments in the mobile device requires considering the material of the surrounding tissue, blood vessels, etc. In addition, it is also relevant the response to the user regarding the goodness of the union between two or more fragments or an adequate metaphor to indicate the proximity among them. Special care has to be taken with these aspects as, by using touch-screen devices, a finger partially hides the neighbor area. Similar considerations apply to the manipulation of fixation devices. Observe that this is still an open field of research due to the intrinsic difficulties of a user interaction with a touchscreen device and the continuous increment of the number of applications. In fact, this task of manipulating bone fragments is complex even in case of using haptic devices within a full virtual reality equipment.

#### 6 COLLABORATIVE SYSTEMS

The interface of mobile devices facilitates the interactive visualization of medical images. Thus these devices could help the specialist to visualize and discuss the results of a planning, the medical images

of the patient, the previously identified bone fragments or even their expected final position.

The ubiquity of mobile devices is also a key factor for their use in the planning/training of a fracture reduction. The special features of this type of devices make them useful for cooperative visualization and interaction, the so-called Computer Supported Cooperative Work (CSCW). A mobile CSCW solution (Johnson 2013) would allow the clinician to access medical imaging at the patient bed-site while discussing and sharing images with colleagues in different locations, or even consult the images directly in the operating theater. In the context of a bone fracture reduction process a mobile CSCW should provide a comprehensive user interface to facilitate a truly cooperation and the usefulness of the shared information. This approach would involve the synchronization with a remote server, where all the patient data would be stored.

Nonetheless, at this point it is important to remember the essential aim of the CSCW research area (Fitzpatrick & Ellingsen 2013): to design enhanced collaborative environments by technological means. The experience in this field of research tells us that the technology by itself does not guarantee an improvement on the collaborative experience (Tang & Carpendale 2008). A detailed knowledge of the final socio-technical system and the impact over all the stakeholders is always required to success.

## 7 DISCUSSION AND CONCLUSIONS

In this paper, we have briefly reviewed the opportunities of using mobile devices as a technological help in a bone fracture reduction process. The versatility and ubiquity properties of mobile devices make them very attractive. We have described this process and its main activities and tasks. Afterwards, we have mainly disclosed the chances of mobile and wearable devices for diagnosis, training, composition of bone fragments and computer aid during surgery. Additionally, we have summarized the requirements regarding human computer interaction issues and for designing effective collaborative environments. We note that medical systems demand especial legal regulation assessments in order to be certified for clinical usage.

The bone fracture reduction is a complex process composed of different activities. This fact becomes especially true in case of complex fractures with several fragments. This is the reason why in spite of the number of medical apps in the market (iMedicalApps) none of them cover the entire reduction process. Those activities that require everytime/everywhere user interaction are more alike to be

entirely managed in a handheld device. However, some automatic tasks that demand extra computational power can be deployed in a server by using a distributed architecture model. For instance, the analysis of a composition, the segmentation of a bone fragment, the identification of a fracture zone, etc. It is important to remember that the computational power of a mobile device will continuously be behind its desktop PC counterpart as it is battery dependent and must focus on energy saving instead of developing pure computational power (Noguera & Jimenez 2015).

It is also expected than in a close future the use of wearable devices (such as watches, smart glasses and the like) will not be uncommon. These new pervasive devices will likely open up a new range of possibilities still to be developed. For instance, in this paper, we have introduced the possibility of using smart glasses during surgery.

Concluding, we have presented the reduction process problem and the related areas of research in order to propose adequate solutions based on mobile devices. Besides, the increasing number of specific journals and magazines devoted to handheld and wearable devices issues supports the relevance of this new research area.

## ACKNOWLEDGEMENTS

This work has been partially supported by the Spanish Ministerio de Ciencia e Innovacións and the European Union (via ERDF funds) through the research project TIN2011-25259.

## REFERENCES

- Berkowitz, S.J., J.W. Kung, R.L. Eisenberg, K. Donohoe, L.L. Tsai, & P.J. Slanetz (2014). Resident ipad use: Has it really changed the game? *Journal of the American College of Radiology* 11(2), 180–184.
- Blessy, C., M. Anburajan, & S. Ramkumar (2011, April). Evaluation of mandibular reconstruction plate using Finite Element Analysis. In *2011 3rd International Conference on Electronics Computer Technology*, Volume 3, pp. 344–349. IEEE.
- Bockholt, U., A. Bisler, M. Becker, W. Muller-Wittig, & G. Voss (2003). Augmented reality for enhancement of endoscopic interventions. In *Virtual Reality, 2003. Proceedings. IEEE*, pp. 97–101. IEEE Comput. Soc.
- Botha, C.P., B. Preim, A. Kaufman, S. Takahashi, & A. Ynnerman (2012). From individual to population: Challenges in Medical Visualization. *ArXiv e-prints*.
- Butson, C.R., G. Tamm, S. Jain, T. Fogal, & J. Kruger (2013, January). Evaluation of interactive visualization on mobile computing platforms for selection of deep brain stimulation parameters. *IEEE Transactions on Visualization and Computer Graphics* 19(1), 108–117.

- Campoalegre, L., P. Brunet, & I. Navazo (2012). Interactive visualization of medical volume models in mobile devices. *Personal and Ubiquitous Computing*, 1–12.
- Chowdhury, A.S., S.M. Bhandarkar, R.W. Robinson, & J.C. Yu (2009, July). Virtual multi-fracture craniofacial reconstruction using computer vision and graph matching. *Computerized medical imaging and graphics: the official journal of the Computerized Medical Imaging Society* 33(5), 333–42.
- Fischer, J., M. Eichler, D. Bartz, & W. Straß er (2007, January). A hybrid tracking method for surgical augmented reality. *Computers & Graphics* 31(1), 39–52.
- Fitzpatrick, G. & G. Ellingsen (2013). A review of 25 years of cscw research in healthcare: Contributions, challenges and future agendas. *Computer Supported Cooperative Work (CSCW)* 22(4–6), 609–665.
- Gartner, I. (2013). Gartner says worldwide pc, tablet and mobile phone combined shipments to reach 2.4 billion units in 2013. <http://www.gartner.com/newsroom/id/2408515>.
- Hadwiger, M., J.M. Kniss, C. Rezk-Salama, D. Weiskopf, & K. Engel (2006). *Real-time Volume Graphics*. Natick, MA, USA: A.K. Peters, Ltd.
- Hu, Y., H. Li, G. Qiao, H. Liu, A. Ji, & F. Ye (2011). Computer-assisted virtual surgical procedure for acetabular fractures based on real CT data. *Injury* 42(10), 1121–1124.
- iMedicalApps. <http://www.imedicalapps.com> [accessed 17 July 2015].
- Jaramaz, B., M. Hafez, & A. DiGioia (2006, Sept). Computer-assisted orthopaedic surgery. *Proceedings of the IEEE* 94(9), 1689–1695.
- Jeong, S.-J. & A.E. Kaufman (2007, July). Interactive wireless virtual colonoscopy. *Visual Computer* 23(8), 545–557.
- Johnson, D. (2013). Mobile support in cscw applications and groupware development frameworks. *International Journal of Interactive Mobile Technologies (iJIM)* 7.
- Kho, A., L. Henderson, D. Dressler, & S. Kripalani (2006). Use of handheld computers in medical education. *Journal of General Internal Medicine* 21(5), 531–537.
- Lamberti, F. & A. Sanna (2005). A solution for displaying medical data models on mobile devices. In *SEPADS'05*, Stevens Point, Wisconsin, USA, pp. 1–7. World Scientific and Engineering Academy and Society (WSEAS).
- Liao, H., T. Inomata, I. Sakuma, & T. Dohi (2010, June). 3-D augmented reality for MRI-guided surgery using integral videography autostereoscopic image overlay. *IEEE transactions on bio-medical engineering* 57(6), 1476–86.
- Movania, M.M. & L. Feng (2012). Ubiquitous medical volume rendering on mobile devices. In *Information Society (i-Society), 2012 International Conference on*, pp. 93–98.
- Noguera, J. & J. Jiménez (2012). Visualization of very large 3d volumes on mobile devices and WebGL. In *WSCG Communication Proceedings*, pp. 105–112.
- Noguera, J., J. Jiménez, C. Ogáyar, & R. Segura (2012). Volume rendering strategies on mobile devices. In *International Conference on Computer Graphics Theory and Applications (GRAPP 2012)*. Rome (Italy), pp. 447–452.
- Noguera, J. & J. Torres (2013). Interaction and visualization of 3d virtual environments on mobile devices. *Personal and Ubiquitous Computing* 17(7), 1485–1486.
- Noguera, J.-M. & J.-R. Jimenez (2015). Mobile volume rendering: Past, present and future. *Visualization and Computer Graphics, IEEE Transactions on PP(99)*, 1–1.
- Pandya, A. & G. Auner (2005). Simultaneous augmented and virtual reality for surgical navigation. In *Fuzzy Information Processing Society, 2005. NAFIPS 2005. Annual Meeting of the North American*, pp. 429–435.
- Park, S., W. Kim, & I. Ihm (2008). Mobile collaborative medical display system. *Computer Methods and Programs in Biomedicine* 89(3), 248–260.
- Paulano, F., J. Jiménez, & R. Pulido (2013). *Trends on identification of fractured bone tissue from CT images*, pp. 263–269. CRC Press.
- Paulano, F., J.J. Jiménez, & R. Pulido (2014, May). 3D segmentation and labeling of fractured bone from CT images. *The Visual Computer* 30(6–8), 939–948.
- Shen, F., B. Chen, Q. Guo, Y. Qi, & Y. Shen (2013). Augmented reality patient-specific reconstruction plate design for pelvic and acetabular fracture surgery. *International Journal of Computer Assisted Radiology and Surgery* 8(2), 169–179.
- Sikorski, J.M. & S. Chauhan (2003). Computer-assisted orthopaedic surgery: Do we need caos? *Journal of Bone & Joint Surgery, British Volume* 85-B(3), 319–323.
- Soler, L., S. Nicolau, J.-B. Fasquel, V. Agnus, A. Charnoz, J.M. Hostettler, C. Forest, D. Mutter, & J. Marescaux (2008). Virtual reality and augmented reality applied to laparoscopic and notes procedures. In *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*, pp. 1399–1402.
- Sugano, N. (2003). Computer-assisted orthopedic surgery. *Journal of Orthopaedic Science* 8(3), 442–448.
- Tang, C. & S. Carpendale (2008). Evaluating the deployment of a mobile technology in a hospital ward. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, CSCW '08*, New York, NY, USA, pp. 205–214. ACM.
- Telkenaroglu, C. & T. Capin (2013, October). Dual finger 3d interaction techniques for mobile devices. *Personal Ubiquitous Comput.* 17(7), 1551–1572.
- Vázquez Alcocer, P. & M. Balsa Rodríguez (2012). Practical volume rendering in mobile devices. In *Proc. International Symposium on Visual Computing*, Volume 7431 of *Lecture Notes in Computer Science (LNCS)*, pp. 708–718. Springer Verlag.
- Whipple, E.C., K.L. Allgood, & E.M. Larue (2012). Third-year medical students' knowledge of privacy and security issues concerning mobile devices. *Medical Teacher* 34(8), e532–e548.
- Willis, A., D. Anderson, T. Thomas, T. Brown, & J.L. Marsh (2007). 3D reconstruction of highly fragmented bone fractures. In *Medical Imaging 2007: Image Processing, Proceedings of the SPIE*, pp. 65121.
- Yao, Y., X.-x. Pang, T. Liu, & Q.-x. Hu (2009). Design of a prototype for augmented reality defective bone repair simulation system. In *2009 First International Workshop on Education Technology and Computer Science*, pp. 66–70. IEEE.
- Zhou, H., H. Qu, Y. Wu, & M. Chan (2006). Volume visualization on mobile devices. In *14th Pacific Conference on Computer Graphics and Applications*, pp. 76–84.

# Automated peritumoral edema segmentation in preoperative brain MRI scans

E. Binaghi

*Dipartimento di Scienze Teoriche e Applicate, Università degli Studi dell'Insubria, Varese, Italy*

P. Melchiorre & L. Romitelli

*Università degli Studi dell'Insubria, Varese, Italy*

S. Balbi

*Dipartimento di Biotecnologie e Scienze della Vita, Università degli Studi dell'Insubria, Varese, Italy*

D. Lattanzi

*Scuola di Specializzazione in Neurochirurgia, Ospedale di Circolo e Fondazione Macchi, Varese, Italy*

**ABSTRACT:** Peritumoral brain edema is a frequently encountered phenomenon that strongly contributes to neurological signs and symptoms. This paper presents a fully automated procedure based on Support Vector Machine for the accurate and efficient edema delineation in brain MRI imagery. The segmentation procedure is conceived fully automatic and specifically oriented to non-volumetric data to facilitate the insertion in clinical practice. The major efforts in building a robust classifier were concentrated in the extraction and selection of a set of significant features from multispectral brain MR images. Preliminary experimental results obtained by processing in-house collected data, are encouraging creating the premises for an effective use of the automated segmentation procedure in brain tumor clinical studies. To see if the proposed supervised method can be considered an alternative to other contemporary approaches, the results obtained were compared with those obtained by an unsupervised region-based segmentation procedure based on Graph Cut.

## 1 INTRODUCTION

Magnetic Resonance (MR) imaging plays an important role in the diagnosis of brain pathologies. In brain tumor clinical studies it represents the first adopted exam to forward a diagnostic hypothesis, typically based on tumor imaging features including active tumor area, necrosis and edema, usually in MRI sequences with no or slight enhancement after gadolinium injection (Duffau 2005). Accurate and efficient segmentation of tumor and surrounding tissues of the brain in the MRI imagery is fundamental in order to select the most appropriate treatment, to plan the best surgical approach and to envisage possible alternative solutions supporting surgery, when the resection appears not at once indicated (Clark 1995, Kaus et al. 2001). Segmentation accomplished through a complete manual tracing is a difficult, time consuming task usually affected by strong intra- and inter-variation (Warfield et al. 2004, Binaghi et al. 2014a).

The problem of automatically segmenting MRI brain tumor images continues to be investigated

giving rise to a variety of approaches. Most of the work has been focused on the automated delineation of the whole pathological area or the active tumor area basing on semi-automatic or fully automated methods (Gordillo et al. 2013).

In recent years we assist to a wide spread utilization of Machine Learning procedures successful for multimodal MRI segmentation. From a viewpoint of reproducibility, unsupervised methods are clearly desirable, but most of them still present low accuracy (Gordillo 2013). Supervised approaches circumvent the problem of explicitly and analytically describing the parts to be segmented lying to a learning stage the charge of inducing the classifier from supervised data available. Parametric and nonparametric methods based on statistical, neural and fuzzy frameworks are proposed showing in general a competitive behavior in capturing complex multivariate relationships in the data (Bezdek et al. 1993, Gordillo et al. 2013, Bauer et al. 2013). Support Vector Machine (SVM) methods (Vapnik, 1995) have shown excellent performances in MRI segmentation studies aimed at identifying a variety

of neurological conditions (Bauer et al. 2013, Binaghi et al. 2014b).

In the present work we focus the attention on peritumoral brain edema segmentation. Peritumoral brain edema is a frequently encountered phenomenon that strongly contributes to neurological signs and symptoms. The role of peritumoral edema as a prognostic factor has been investigated in several studies that have led to different conclusions depending on the pathology under study. A quantitative evaluation of the edema from MR imagery is of basilar relevance to formulate surgical indications in elderly patients harboring intracranial meningiomas (Greenberg et al. 1999, Caroli et al. 2005). In such a critical context the availability of automated segmentation procedure specifically oriented to the accurate and efficient edema detection and delineation is of great importance to easily conduct retrospective, large scale studies and to insert the segmentation task in the clinical practice.

We address the segmentation task by proposing a supervised strategy based on SVM. The supervised classification makes use of multichannel intensities of Post Gadolinium T1-weighted (T1c) and T2-weighted FLAIR image (T2f). Additional textural and contextual features are considered and increasingly added during the experimentation in order to select the best configuration. The segmentation procedure is fully automatic and oriented to non-volumetric data characterized by poor inter-slice spacing to facilitate the insertion in current clinical practice.

## 2 THEORETICAL BACKGROUND

SVM is a classification algorithm based on kernel methods (Vapnik 1995, Schoelkopf & Smola 2002) able to map the original parameter vectors into a higher dimensional feature space through a kernel function. Classes which are non-linearly separable in the original space can be linearly separated in the higher dimensional feature space. Thus SVM is capable of solving complex nonlinear classification problems. The learning algorithm is based on the choice of current worst-classified patterns. These training examples are the most difficult to classify and during most of the training period are on the wrong side of the current decision boundary. At the end of the training stage, these patterns will be *support vectors*.

Suykens (2002) proposed a new formulation of SVM by adding a Least Squares (LS) term in the original formulation of the cost function. This modified version of the original Vapnik's SVM formulation avoid solving a more difficult quadratic programming problem and only requires the solution of a set of linear equations thus reducing significantly the computational complexity.

Let vector  $x \in R^n$  denote a pattern to be classified, and let scalar  $y \in \{-1,+1\}$  denote corresponding class label. Let  $\{(x_i, y_i), i = 1, 2 \dots l\}$  a supervised training set of  $l$  elements for a two-class classification problem, with  $x_i \in X \subseteq R^n$  and  $y_i \in \{-1,+1\}$ . Considering the case of linearly separable data, the solution consists in the construction of the decision function  $f_{w,b}(x) = \text{sgn}(g_{w,b}(x))$  with  $g_{w,b}(x) = w^T x + b$  that separates training examples and correctly classifies an input pattern  $x$  that is not necessarily from the training set. SVM classifier defines the *Maximum Margin Separating Hyperplane* that causes the largest separation between "borderline" examples from the two classes. Mathematically, this hyperplane can be found by minimizing the cost function:

$$J(w) = \frac{1}{2} \|w\|^2 \quad (1)$$

subject to

$$\begin{aligned} w^T x_i + b &\geq +1 \text{ for } y_i = +1 \\ \text{or} \\ w^T x_i + b &\leq -1 \text{ for } y_i = -1 \end{aligned} \quad (2)$$

For nonlinear classification the function.

$$g' = w^T \varphi(x) + b \quad (3)$$

is specified in which the non liner operator  $\varphi(\cdot)$  is introduced to map the input pattern  $x$  into a higher dimensional space. In case of non linear classification a non-negative slack variable  $\xi$  is introduced, in order to have a tolerance to errors. The cost function to be minimized become:

$$J(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \xi_i \quad (4)$$

subject to

$$y_i(w^T \varphi(x_i) + b) \geq +1 - \xi_i \text{ with } \xi_i \geq 0 \quad i = 1, \dots, l \quad (5)$$

$C$  is a constant value that expresses the balance between margin and error and must be provided a priori. Linearly non separable data are analyzed with kernel functions such as higher order polynomials and Gaussian RBF.

## 3 FULLY AUTOMATED EDEMA SEGMENTATION

The SVM model learns from a set of labeled image elements the invariant common properties of the edema and not-edema classes. The trained SVM classifier acts as a dichotomizer; it receives in input

a multidimensional pattern extracted from T1c and T2f MR elements and assigns labels to elements never seen during the training phase.

Features composing the multidimensional patterns are extracted from the co-registered imaging data. The extraction of an appropriate set of local features is a crucial task strongly affecting the performances of the overall segmentation strategy.

In addition to the image intensities in the two MR modalities, we consider other significant features to describe quantitatively neighbor relationships and texture of the images. Texture analysis covers a wide range of techniques based on first- and second order image texture parameters (Tucelian & Jain 1998).

In more detail, the following features are considered: multimodal intensities from T1c and T2f scans ( $I_{t1c}$ ,  $I_{t2f}$ ), first order texture features: mean (M), variance (Var), skewness (S), kurtosis (K) and entropy (E) computed on T1c and T2f scans, intensities in neighborhoods of multimodal voxels (I1-I26). All the proposed features have been analyzed systematically in order to determine the combination that is most appropriate for the classification task (see section 4). The features have been normalized to have zero mean and unit variance.

## 4 EXPERIMENTS

The segmentation method was experimented on a in-house collected multispectral datasets of 10 patients. Each dataset is composed of T1c and T2f scans. The T1c is acquired using a 3D sequence characterized by 1 mm isotropic voxels, the inter-slice spacing of 1 mm and the slice thickness of 1 mm; the T2f sequence includes an in-plane resolution between 0.75 and 0.81 mm and slice thickness of 5 mm. The spacing between slice is 6 mm for cases 1,3,6,7,8,10 and of 6.5 mm for cases 2,4,5,9.

Figure 1 shows an axial slice of one patient with the two MRI modalities considered.

The masks obtained by the automated segmentation were compared with the masks obtained through a manual segmentation of the T2f images. Manual segmentations were performed by a team of experts with the support of an interactive software tool for slice-by-slice labeling. Manual segmentations provided by the experts were fused by applying a Majority Voting procedure.

### 4.1 Evaluation metrics

Performances were evaluated by comparing spatial distribution of the masks obtained by the automated segmentation with spatial distribution of reference masks manually segmented by experts. The adopted similarity measure is the Dice Coefficient (DC) defined as follows (Bioux et al. 2000):

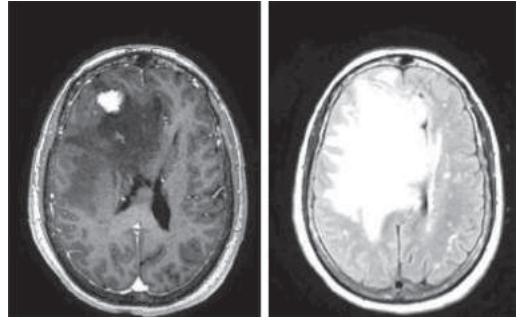


Figure 1. Axial slice of one patient. From left to right: the T1c and T2f scans.

$$DC = \frac{2T_p}{2T_p + F_p + F_n} \quad (6)$$

with  $T_p$  = True Positive,  $T_n$  = True Negative,  $F_p$  = False Positive,  $F_n$  = False Negative.

### 4.2 Results

Several experiments were conducted in which different configurations of the segmentation procedure were considered varying:

- parameters involved in the SVM model
- training and test data
- features composing multidimensional patterns.

After a trial and error procedure the value of the standard deviation ( $\sigma$ ) for Gaussian RBF kernel was chosen as 0.5 in the SVM classifier configured as soft-margin LS model.

Different configurations of the classification procedure were also evaluated varying the number and the source of training and test examples randomly selected from the ground truth masks. We considered patient-specific (intra-patient) analysis and inter-patient analysis. In patient-specific analysis the training and test data were obtained from the ground truth masks of the image to be segmented. Inter-patient analysis was performed training on several patients, and testing in one never seen during the training phase.

Different configurations of the multispectral pattern presented in input to the SVM model have been also considered in the experiments basing on the following selected feature sets:

- intensity values from T1c and T2f scans
- intensity values from T1c and T2f scans + texture measures
- intensity values from T1c and T2f + gray scale values of neighborhoods voxels of each scan including incrementally 6, 18, and 26 neighbors (I1-I26).

The configuration that optimizes the balance between performance accuracy, computational efficiency, robustness/generalisation capabilities, was *intensity values from T1c and T2f scans + texture measures computed on 6 neighborhoods voxels*. The multidimensional pattern is then structured as follows:

$$x_i = \begin{bmatrix} x_{t1c}, x_{t2f} \end{bmatrix} \text{ with} \\ x_{t1c} = \begin{bmatrix} I^i_{t1c}, M^i_{t1c}, Var^i_{t1c}, S^i_{t1c}, K^i_{t1c}, E^i_{t1c} \end{bmatrix} \quad (7) \\ x_{t2f} = \begin{bmatrix} I^i_{t2f}, M^i_{t2f}, Var^i_{t2f}, S^i_{t2f}, K^i_{t2f}, E^i_{t2f} \end{bmatrix}$$

and used a training set of 10.000 labeled samples randomly chosen within the ground truth masks in the percentage of 70% in edema region, 30% in the remaining brain area. Table 1 shows performance values obtained by averaging Dice coefficients obtained by the SVM model trained on patient specific data (intra-analysis). Performances values

Table 1. Dice similarity coefficients (mean and standard deviation) for the datasets under study.

	Dice coefficient
Intra-patient analysis	$0.84 \pm 0.05$
Inter-patient analysis	$0.83 \pm 0.07$

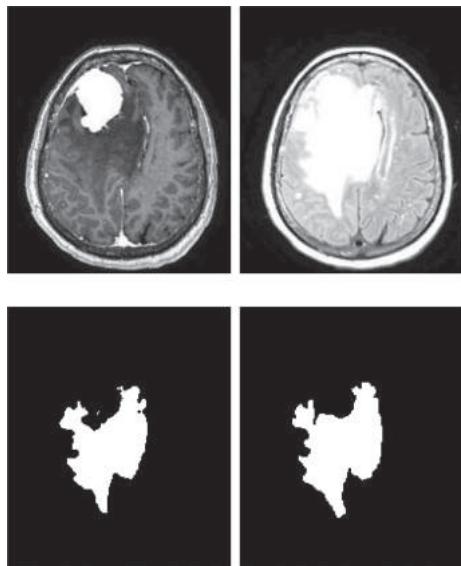


Figure 2. Segmentation results of one axial slice obtained by the interpatient analysis. First row, from left to right: original T1c image, original T2f image; second row, from left to right: edema segmentation mask obtained by the automated procedure, truth mask.

obtained by performing inter-analysis on the available data sets are also reported. In this case Dice coefficients are obtained by averaging the individual leave-one-out cross-validation results.

With the intra-patient analysis, the Dice coefficient is 0.84. Performances obtained with the inter-patient analysis slightly decrease obtaining a Dice coefficient equal to 0.82. Results obtained confirm the good and robust behaviour of the segmentation procedure that maintains acceptable performances without requiring patient specific training. Figure 2 presents the segmentation result for one slice in case of inter-patient analysis. The edema region is well delineated although few omission and commission errors are visible in some areas.

## 5 COMPARISON ANALYSIS

To see if the proposed supervised method can be considered an alternative to other contemporary approaches, the results obtained were compared with those obtained by the unsupervised region-based segmentation procedure based on graph cuts (Boykov & Funka-Lea 2006). The s/t Graph Cuts framework offers a globally optimal object extraction method for N-dimensional images. Graph cuts optimize a segmentation energy function that combines boundary regularization with regularization of regional properties.

The application of Graph Cut requires the identification of object and background prototypes with which to initialize the overall segmentation process.

Usually this task is accomplished through an interactive session in which users manually select seeds on the image. Recently there have been some proposals exploiting solutions for automatically initializing Graph Cut-based segmentations of biomedical images (Santle et al., 2012).

The present study uses the max-flow/min-cut algorithm (Boykov and Kolmogorov 2004) as optimization framework and adopts an automated initialization procedure based on k-means clustering algorithm.

Figure 2 presents the segmentation results for one axial slice obtained by the SVM and Graph Cut. The SVM segmentation produces a mask with a higher level of underestimation compared to the mask produced by the Graph Cut segmentation.

In Table 2 the means of the DC scores obtained by the CEM-Graph Cut and SVM when segmenting Edema part for the datasets under study are reported. SVM performances are obtained by averaging results of the inter- and intra-analysis. The Dice coefficient obtained by the Graph Cut model over the available data set is equal to 0.87. The two methods show comparable performances. Graph-Cut slightly prevails, however the comparison must

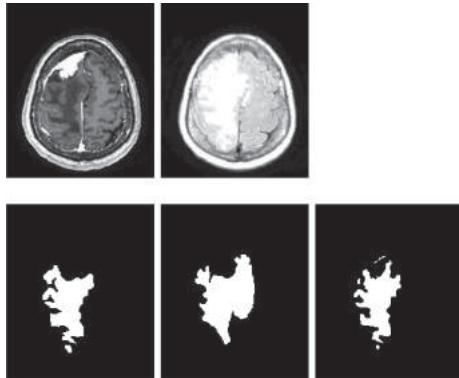


Figure 3. Segmentation results of one axial slice. First row, from left to right: original T1c image, original T2f image; second row, from left to right: edema truth mask, segmentation mask obtained by SVM and by Graph Cut.

Table 2. Mean Dice Scores (*DS*) obtained by the CEM-Graph Cut and SVM (averaging results of the inter- and intra-analysis), when segmenting Edema part for the datasets under study.

	Dice coefficient
Graph Cut	$0.87 \pm 0.09$
SVM	$0.84 \pm 0.06$

take into account that, differently from the supervised method, Graph Cut is semi-automated and information provided by the users for each image to be processed, may improve performances thus reducing stability and reproducibility of the results.

## 6 CONCLUSIONS

The objective of the present study was to develop a supervised segmentation procedure based on SVM specifically oriented to peritumoral edema delineation in non-volumetric MRI data. As seen in our experimental context the procedure is competitive with other approaches. It shows a robust behaviour and it is oriented to the application in clinical practice. Future plans contemplate the acquisition of new data with which we will be able to perform a more significant inter-patient analysis and to better analyse the generalisation capabilities of the trained segmentation model.

## REFERENCES

Bauer, S., Wiest, R., Nolte, L.-P., Reyes, M. 2013. A survey of MRI-based medical image analysis for brain tumor studies. *Phys. Med. Biol.* 58:R97–R129.

- Bezdek JC, Hall LO, Clarke LP. 1993. Review of MR image segmentation techniques using pattern recognition. *Medical Physics* 20(4):1033–1048.
- Binaghi, E., Pedoia, V., Balbi S. 2014a. Collection and fuzzy estimation of truth labels in glial tumour segmentation studies. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualisation* 2:1–15.
- Binaghi E., Omodei M., Pedoia V., Balbi S., Lattanzi D. and Monti E. 2014b. Automatic Segmentation of MR Brain Tumor Images using Support Vector Machine in Combination with Graph Cut. In Proc. of the International Conference on Neural Computation Theory and Applications (NCTA-2014), Rome, 22–24 Ottobre 2014, SciTePress.
- Bouix, S., Martin-Fernandez, M., Ungar, L., Koo, M.N.M.S., McCarley, R.W., Shenton, M.E., 2007. On evaluating brain tissue classifiers without a ground truth. *NeuroImage*, 36:1207–24.
- Boykov, Y., & Kolmogorov, V. 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision., *PAMI*, 26:1124–37.
- Boykov, & Funka-Lea, G. 2006. Graph cuts and efficient n-d image segmentation. *Int. J. Comput. Vision*, 70:109–131.
- Caroli, M., Locatelli M., Prada F., Beretta F., Martinelli-Boneschi F., Campanella R., Arienta C. 2005. Surgery for intracranial meningiomas in the elderly: a clinical-radiological grading system as a predictor of. *J Neurosurg*, 102:290–294.
- Clarke, L.P., Velthuizen, R., Camacho, M., Heine, J., Vaidyanathan, M., Hall, L., Thatcher, R., Silbiger, M.S. 1995. MRI segmentation: methods and applications. *Magn Reson Imaging* 13(3):343–36.
- Duffau, H. 2005. Lessons from brain mapping in surgery for low-grade glioma: insights into associations between tumour and brain plasticity. *Lancet Neurol* 4(8):467–487.
- Gordillo, N., Montseny E., Sobrevilla, P. 2013. State of the art survey on MRI brain tumor segmentation. *Magnetic Resonance Imaging* 31:1426–1438.
- Greenberg, H., Chandler, W., Sandler, H. 1999. *Brain Tumors*, Oxford: Oxford University Press.
- Kaus, M.R., Warfield, S.K., Nabavi, A., Black, P.M., Jolesz, F.A., Kikinis R. 2001. Automated segmentation of MRI of brain tumors. *Radiology* 218:586–591.
- Santle, K., Camilus, Govindan, V.K., 2012. ‘A Review on Graph Based Segmentation.’ *IJIGSP* 4(5):1–13.
- Schoelkopf, B., Smola, A. 2002. *Learning with kernels: support vector machines, regularization, optimization, and beyond*, Cambridge: MIT Press.
- Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B., Vandewalle, J., 2002. *Least Squares Support Vector Machines*. Singapore: World Scientific Publishing Co.
- Tuceryan, M. & Jain, A. 1998. *Texture analysis*. Inc. River Edge, NJ: World Scientific Publishing Co.
- Vapnik, V.N. 1995. *The Nature of Statistical Learning Theory*, New York: Springer-Verlag.
- Warfield, S.K., Zou, K.H., Wells, W.M. 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans Med Imaging* 23(7): 903–921.

This page intentionally left blank

## Evaluation of segmentation techniques for wound area identification

Guanghua Zhang

*Chongqing University of China, Chongqing, China*

*Graphics and Imaging Laboratory, University of Girona, Girona, Spain*

Pau Xiberta, Anton Bardera & Imma Boada

*Graphics and Imaging Laboratory, University of Girona, Girona, Spain*

Angel Romero

*Faculty of Infirmary, University of Girona, Girona, Spain*

**ABSTRACT:** One of the key parameters that has to be considered in wound treatment is the size of the wound. Usually, this is monitored by an expert at regular intervals. To carry out this process, image-based approaches can be applied. Generally, these approaches start applying a segmentation technique to isolate the wound from the rest of the image. Due to the importance of segmentation, our aim is to evaluate the performance of four segmentation techniques (grabcut, snakes, region growing and Support Vector Machine (SVM)) on pressure ulcer images. In our study, we have considered twenty images captured following normal clinical practice and under non-controlled conditions which makes the processing more difficult. The images have been manually segmented by two experts and the obtained results have been compared with the automatic segmentation obtained with the different methods. From the comparison, we can conclude that both snakes and grabcut methods clearly behave better than region growing and SVM.

### 1 INTRODUCTION

A pressure ulcer, a type of wound, is a localized injury to the skin or underlying tissue usually over a bony prominence, as a result of pressure, or pressure in combination with shear. Once a pressure ulcer has developed, immediate treatment is required and the treatment strategies vary according to the stage of the wound and the purpose of the treatment. Generally, an expert periodically examines the ulcer progress to monitor the administered treatment. This monitoring is based on visual inspection and different wound parameters are analyzed, amongst them, shape, size, depth and tissue type (Gethin 2006).

Different methods are used to measure pressure ulcer shape and size. The simplest one assumes that the wound is rectangular in shape and, using a ruler, the clinician measures wound surface by multiplying the greatest length by the perpendicular greatest width (Langemo, J. D. Hanson, S. Hunter, & P. Thompson 2008). This method overestimates wound area and accuracy tends to diminish as the wound size increases. A more accurate method traces the wound perimeter on a graduated clear plastic or acetate paper placed over the wound and calculates the wound area by determining the number of squares contained in the wound

region (Thawer, Houghton, & Woodbury 2002, G. Gethin & S. Cowman 2006).

Beside these manual techniques, there are also more sophisticated methods based on electronic devices such as the planimeter that uses a transparent film placed over the wound while the margins are traced, and computer calculates the area after digitally scanning the tracing (Mayrovitz & Soontupe 2009). There are also more complex and expensive vision-based technologies that use stereophotogrammetry or structured lighting to obtain wound images (Krouskop, Baker, & Wilson 2002, Plassman & Jones 1998). These images are reconstructed using a computer to produce a 3-D model of the wound. The wound border is then traced on the computer image, and the computer software determines the area and volume of the wound. More recently, Sprigle et al. (Sprigle, Nemeth, & Gajjala 2011) proposed a new device to photo-document and measure the wound area that uses four laser diodes arranged in a square with the camera lens lying in its center. These lasers are used to calculate wound area and skew. For a review and comparison of wound measurement techniques see (M. Bilgin & U.Y. Gunes 2014, Ahmad-Fauzia, Khansab, Catignanic, Gordillo, Send, & Gurcan 2015).

In this context, it is also common the use of digital cameras to record wound progress. Several techniques

have been proposed to measure wound parameters from taken images and also to automatically classify them. These methods definition is challenging due to the variability of image parameters such as illumination or patient position that makes the method reproducibility difficult. Focusing on image-based methods, a key step of all them is the application of a segmentation method that after image capture identifies the wound area on the image. Once it has been identified other techniques such as color treatment, feature extraction and classification techniques are applied to determine wound parameters (B. Song & A. Sacan 2012, Veredas, Mesa, & Morente 2010, Kolenik & Fexa 2004, M. Burns, J. Enderle, Rosow, & Zhu 2002, Weber, N. Watermann, J. Jossinet, J. Byrne, J. Chantrey, S. Alam, K. So, Bush, S. OKane, & E.T. McAdams 2010, Sundeep-Kumar & Eswara-Reddy 2014). Due to the importance of segmentation techniques in these wound processing pipelines, in this paper, we evaluated the performance of four different segmentation methods: grabcut, snakes, region growing and SVM on pressure ulcer images. The methods are tested on twenty images and the obtained results are compared with wound area segmentation performed by experts.

## 2 WOUND SEGMENTATION

There is a wide variety of image segmentation methods in the literature (Gonzalez & Woods 2006). To carry out our study, we have selected four segmentation methods to cover edge-based, region-based, graph-based, and clustering-based segmentation techniques, respectively. Next, we describe these segmentation techniques.

### 2.1 Snakes

Snakes method (Kass, Witkin, & Terzopoulos 1988) is an edge-based segmentation technique. From an initial user-defined contour, also called snake, the method iteratively updates this contour in order to minimize its total energy. The total energy is composed of two parts: the internal energy that depends on contour shape (smoother the contour, smaller the internal energy) and the external energy that depends on the energy field, usually defined from the image gradient. The contour reaches minimum at the local energy extreme that correspond to the image edges in case of image gradient is used. In our framework, we have used the cvSnakeImage() implementation provided by the opencv library (Bradski 2000). This method requires as input parameters:  $\alpha$  that represents the weight of continuity energy,  $\beta$  that represents the weight of curvature energy, and  $\gamma$  the weight of image energy. For the tests we set  $\alpha = 0.1$ ,  $\beta = 0.4$ , and  $\gamma = 0.5$ . To apply the method the user interaction has to draw an initial contour in the image. In our case, the initial contour is defined considering 20 points to ensure that the wound is included in this initial contour. For the test (see section 3) we set the number of iterations at 1000. The first columns of Figure 1 and 2 illustrate two examples of the initial contour and the obtained results.

### 2.2 Region growing

Region growing is a region-based technique. From an initial seed point, neighbour pixels are added into the region depending on a similarity criterion, typically based on a given threshold. The process is iterated on until there are not new inclusions.

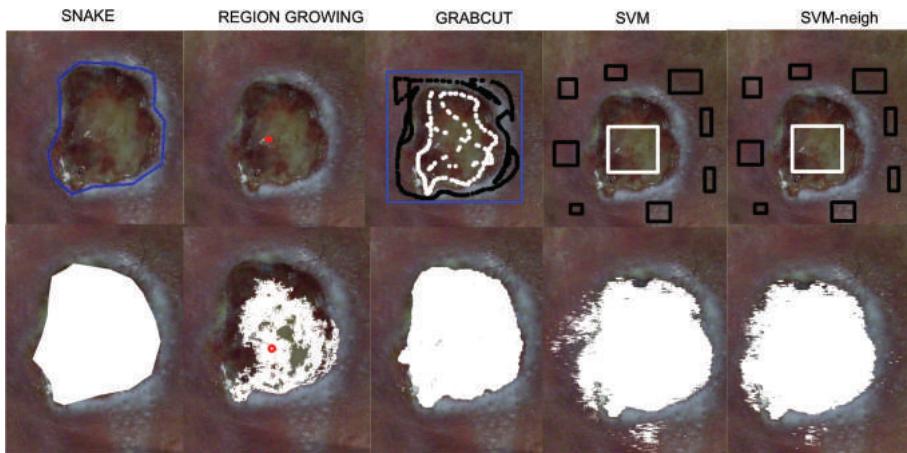


Figure 1. From left to right the user interaction and the obtained segmentation with snakes, region growing, grab cut, SVM and SVM-neigh techniques.

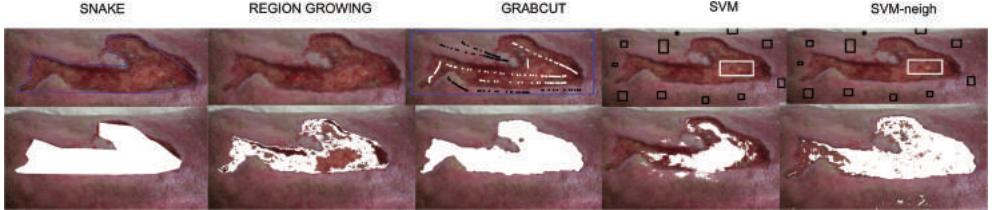


Figure 2. From left to right the user interaction and the obtained segmentation with snakes, region growing, grab cut, SVM and SVM-neigh techniques.

In our implementation, the method computes the standard deviation of the image intensity on a window centered in the seed point. Then, the neighbour pixels are included if their intensity is in the range  $[I - k\sigma, I + k\sigma]$ , where  $I$  is the intensity of the seed point,  $k$  is a parameter of the method, and  $\sigma$  is the standard deviation computed in the initial step. Thus, the method has two parameters: the size of the window where the standard deviation is computed and  $k$ . We have used the `floodfill()` implementation provided by the opencv library (Bradski 2000). In this method, the user interaction with the image simply consists in the selection of the seed point and the adjustment of  $k$  and the window size to find out the best result. e set  $\sigma$  between 10 and 50. The second columns of Figure 1 and 2 show two examples of the seed placement and the obtained results.

### 2.3 Grabcut

The grabcut method (Rother, Kolmogorov, & Blake 2004) is a graph-based technique and, in particular, based on graph cuts. To start the process, the user defines a bounding box around the object to be segmented. Then, the method estimates the color distributions of the target object and the background using a Gaussian mixture model. This model is used to construct a Markov random field over the pixel labels, with an energy function that prefers connected regions with the same label, and running a graph cut based optimization to infer their values. Since this estimate is more accurate than the one taken from the bounding box, this two-step procedure is repeated until convergence. To correct the estimates the user can point out misclassified regions and rerun the optimization. The method also corrects the results to preserve edges. In our framework, we have used the `grabCut` method from the opencv library (Bradski 2000). This implementation does not require any parameter. The user interaction for this method consists in drawing a square contour containing the wound to get an initial segmentation and also selecting the background and foreground with some points.

The third columns of Figure 1 and 2 show two examples of selected regions by the user and the obtained results.

### 2.4 Support Vector Machine (SVM)

SVM method is a clustering-based technique that classifies each pixel using the Support Vector Machine algorithm (Cortes & Vapnik 1995) from a training set, which is given by the user. For each pixel, a vector of features is obtained and used by the SVM algorithm to perform the classification. In our implementation, the RGB color values and the distance to the lesion center (given by the user) are used as features. In order to do not have features with different range of values, the distance to the lesion center has been divided by the image diagonal and multiplied by 255. In this way, every feature takes values in the range [0,255]. In our framework, we have used the `SVM()` implementation provided by the opencv library (Bradski 2000). The parameters of this implementation are: the four selected features corresponding to R, G, and B values and the pixel to the center of selected foreground distance. We have also implemented the SVM method adding local information. For each pixel, we have considered the R, G, and B values from the top, bottom, left and right neighbour pixels obtaining a feature vector with 12 values. We denote this method SVM-neigh. The user interaction for SVM method consists in the definition of some rectangular regions corresponding to either background or foreground wound pixels. The fourth and fifth columns of Figure 1 and 2 show two examples of selected regions by the user and the obtained results with the SVM and SVM-neigh, respectively.

## 3 EXPERIMENTAL SETUP

This study was done in collaboration with clinicians from the Faculty of Infirmary from the University of Girona. For the experiments we used 20 images of  $2304 \times 3072$  resolution stored

in JPEG format (see Fig. 3). Clinicians captured these images following normal clinical practice and under non-controlled conditions which makes the processing more difficult. We developed a tool to capture the ground truth. Clinicians use this tool to trace the wound boundary on the image. From this trace the tool computes a mask where black and white pixels represent background and wound area, respectively.

All testing images were manually segmented by two experts and also using the implemented methods. We compared the masks obtained from manual segmentation with the masks obtained by each method and classified pixels in four groups: False Positives (FP), True Positives (TP), False Negatives (FN), and True Negatives (TN). An example is presented in Figure 4. From the quantity of pixels at each group, the measures precision ( $P$ ) and recall ( $R$ ) can be computed. Precision is given by the proportion of the segmented pixels that are correctly classified as lesion, that is,



Figure 3. The 20 images of the wound database used in this study.

$$P = \frac{TP}{TP + FP}.$$

Recall indicates the proportion of the lesion that has been segmented, that is

$$R = \frac{TP}{TP + FN}.$$

Since both measures are complementary (typically when one grows, the other decreases), the harmonic mean of both, also known as  $F$ -score, is also used to evaluate the results, since it is a combination of both.  $F$ -score is given by

$$F\text{-score} = 2 \cdot \frac{P \cdot R}{P + R}.$$

## 4 RESULTS

The obtained results are presented in Table 1, where the mean and standard deviation of precision ( $P$ ), recall ( $R$ ), and  $F$ -score are shown. As it can be seen, the best results are achieved by the snakes and grabcut methods, while region growing, SVM, and SVM-neigh achieve much poorer results. We have also shown the inter-operator variability, that has been obtained by considering the operator 1 as the ground truth and analyzing the manual segmentations of operator 2 in terms of precision, recall and  $F$ -score. As it can be seen, all the methods still behave poorer than manual edition.

With respect to the method interactivity, snakes requires the definition of an initial contour that, in our implementation, is done by manually defining a polygon of 20 points. After this, the user

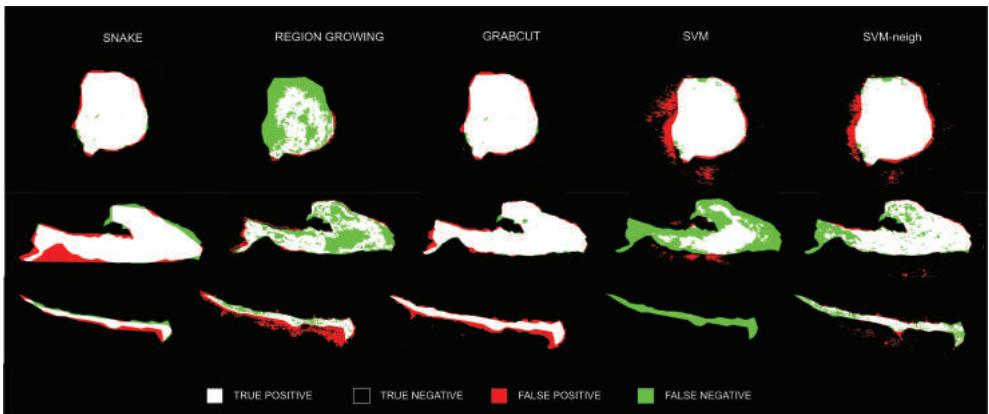


Figure 4. Masks of manual segmentation and the automatic ones are compared and points are labelled as true and false positive or true and false negative.

Table 1. Segmentation results for the four methods in terms of precision ( $P$ ), recall ( $R$ ), and  $F$ -score.

Method	$P$	$R$	$F$ -score
Snakes	$0.922 \pm 0.065$	$0.915 \pm 0.083$	$0.915 \pm 0.058$
Region growing	$0.866 \pm 0.132$	$0.644 \pm 0.186$	$0.713 \pm 0.139$
Grabcut	$0.896 \pm 0.100$	$0.947 \pm 0.076$	$0.914 \pm 0.063$
SVM	$0.838 \pm 0.157$	$0.690 \pm 0.271$	$0.704 \pm 0.217$
SVM-neigh	$0.869 \pm 0.081$	$0.841 \pm 0.080$	$0.851 \pm 0.579$
Inter-operator	$0.966 \pm 0.045$	$0.937 \pm 0.056$	$0.950 \pm 0.033$

can define some parameters that can be optimally set, for most of cases, after about 10 iterations. The total time to obtain the final result is about 50 seconds. The grabcut algorithm also achieve very good results, but the interaction takes more time than snakes, since the method is much more computationally demanding. The final result is achieved in this case in about 2 minutes. Despite the poor results of the region growing, its interaction is very fast and a final result can be obtained in 30 seconds. SVM and SVM-neigh are the most computationally demanding methods, but the manual interaction is not very high. In this case, the total time is about 2 minutes.

In Figure 4, we show three different cases that represent different wound shapes: a regular one, an irregular one with concavities, and a lengthened one, respectively. From these results we can observe that the snakes method performs better with regular shapes. Snakes method assumes a smooth contour and, when the wound has an irregular or lengthened shape, the method does not achieve correct results. Region growing method has a bad performance in all the cases which is due to irregularities on the lesion colour but not on the shape. Grabcut method has a good performance in all the cases and irregular shapes do not affect the results. Moreover, the interaction is almost the same, independently of the wound shape. If we compare SVM and SVM-neigh, we can observe that in all the cases considering neighbour information leads to better results. Although SVM and SVM-neigh are not affected on the contour shape themselves, in the case of lengthened shapes, the pixels of wound extremes are misclassified. This is due to the fact that the distance to the central point of manual interaction is considered as a SVM feature.

## 5 CONCLUSIONS

The standard practice for the treatment of wounds includes monitoring the size of the wound at regular intervals. Different methods based on image-processing techniques have been proposed to

automate wound measurement from images. Wound segmentation is a key step of these methods. Due to the importance of segmentation in these approaches we have evaluated four segmentation techniques and compared with manual segmentation. From our tests we observed that snakes and grabcut methods obtain the best results, which are slightly less accurate than manual edition. As future work, we will improve the user interface to easily interact with the method parameters, we will integrate the methods into an e-learning tool and we will investigate new image features to improve the SVM method results.

## ACKNOWLEDGMENTS

This work was supported by the Spanish Government (Grant No. TIN2013-47276-C6-1-R) and by the Catalan Government (Grant No. 2014-SGR-1232). Guanghua Zhang received financial support from the European Commission (Erasmus Mundus project Techno, ref. 204323-1-2011-1-FR-EMA21).

## REFERENCES

- Ahmad-Fauzia, M., I. Khansab, K. Catignanic, G. Gordillo, K. Send, & M. Gurcan (2015). Computerized segmentation and measurement of chronic wound images. *Computers in Biology and Medicine*, 74–85.
- Bilgin, M. & U.Y. Gunes (2014). A comparison of 3 wound measurement techniques. *J Wound Ostomy Continence Nurs* 40(6), 590–593.
- Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Burns, M. J. Enderle, E. Rosow, & Q. Zhu (2002). Development of a wound assessment system for quantitative chronic wound monitoring. In *Proc. IEEE Annual North-east Bioeng. Conference*.
- Cortes, C. & V. Vapnik (1995). Support-vector networks. *Machine Learning* 20(3), 273–297.
- Gethin, G. (2006). The importance of continuous wound measuring. *Wounds* 2(2), 60–68.
- Gethin, G. & S. Cowman (2006). Wound measurement comparing the use of acetate tracings and visitrak digital planimetry. *J Clinical Nursing* 15, 422–27.

- Gonzalez, R. & R. Woods (2006). *Digital Image Processing* (3rd Edition). Prentice-Hall.
- Kass, M., A. Witkin, & D. Terzopoulos (1988). Snakes: Active contour models. *International Journal of Computer Vision* 1(4), 321–331.
- Kolenik, M. & A. Fexa (2004). Segmentation of wounds in the combined color-texture feature space. *SPIE Med. Imaging* 5370, 549–556.
- Krouskop, T., R. Baker, & M. Wilson (2002). A non contact wound measurement system. *Journal of Rehabilitation Research and Development* 39(3), 337–46.
- Langemo, D., J.A. J., D. Hanson, S. Hunter, & P. Thompson (2008). Measuring wound length, width, and area; which technique? *Adv Skin Wound Care* 21(1), 42–47.
- Mayrovitz, H. & L. Soontupe (2009). Wound areas by computerized planimetry of digital images: accuracy and reliability. *Advances in skin and wound care* 22(5), 222–229.
- Plassman, P. & T. Jones (1998). Mavis: a non-invasive instrument to measure area and volume of wounds. *Medical Engineering and Physics* 20(5), 332–38.
- Rother, C., V. Kolmogorov, & A. Blake (2004). Grab-cut-interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (SIGGRAPH)* 23(3), 309–314.
- Sprigle, S., M. Nemeth, & A. Gajjala (2011). Iterative design and testing of a hand-held, non-contact wound measurement device. *Journal of Tissue Viability*.
- Song, B. & A. Sacan (2012). Automated wound identification system based on image segmentation and artificial neural networks.
- Sundeep-Kumar, K. & B. Eswara-Reddy (2014). Wound image analysis classifier for efficient tracking of wound healing status. *Signal and Image Processing* 5(2), 15–27.
- Thawer, H., P. Houghton, & G. Woodbury (2002). A comparison of computer-assisted and manual wound size measurement. *Ostomy Wound Manage* (48), 46–53.
- Veredas, F., H. Mesa, & L. Morente (2010). Binary tissue classification on-wound images with neural networks and bayesian classifiers. *IEEE Trans. Medical Imaging* 29(2), 410–27.
- Weber, S., N. Watermann, J. Jossinet, J. Byrne, J. Chantrey, S. Alam, K. So, J. Bush, S. OKane, & E.T. McAdams (2010). Remote wound monitoring of chronic ulcers. *IEEE Trans. Inf. Technol. Biomed* 14(2), 371–77.

# Automated image segmentation based on multiobjective optimization and machine learning

S.M. Shontz

*The University of Kansas, Kansas, USA*

J.S. Tahara

*BSMP Student, The University of Kansas, Kansas, USA*

*BSMP Student, The University of Sorocaba, Sorocaba, Brazil*

D.O. McLaurin

*CD-adapco*

D.J.L. Colbry

*Michigan State University, Michigan, USA*

B. Parikh

*The University of Kansas, Kansas, USA*

**ABSTRACT:** Image segmentation is a challenging problem due to sensor-based noise, artifacts caused by motion, and partial obstruction of objects. Although numerous algorithms for image segmentation have been invented, there is no single algorithm which can be used to successfully segment all images. Similarly, there is no fully automated image segmentation tool despite recent progress in this direction. In this talk, we present a machine learning tool for automated segmentation of a set of images. Our technique first uses an existing image segmentation algorithm to semi-automatically segment the first image in the data set. It then proceeds by using our multiobjective optimization technique to match the segmented shapes in the initial image to their most similar shapes in the second image in the data set. Optimal shapes are determined based upon factors such as color, geometry, motion, and texture. This process is then repeated in order to segment the remaining images in the data set. Machine learning is used throughout the optimization procedure in order to determine the most appropriate values for the weights of the relevant terms in the objective function for a set of images. We present results obtained from testing our algorithms on various sets of biological and medical images. In addition, we present results on the scalability of our method.

**Keywords:** automated image segmentation; multiobjective optimization; machine learning; unsupervised algorithm

## 1 INTRODUCTION

Image segmentation is the process of dividing pixels in an image into categories based on shared visual characteristics. Such characteristics may include the color, geometry, shape, texture, or other key features. Image segmentation is often an important first step for applications in that the process partitions the image into subimages that are easier to analyze. Image segmentation is a challenging problem due to sensor-based noise, artifacts caused by motion, and objects being only partially visible due to obstructions. Numerous algorithms (e.g., level set methods, model-based methods, graph partition-

ing methods, and algorithms based on compression) have been invented for image segmentation; however, there is no single algorithm which can be used to successfully segment all images. Similarly, there is no fully automated image segmentation tool despite recent progress in this direction (e.g., Chuang 2014, Walsh & Skala 2014).

## 2 METHODOLOGY

In this talk, we present a machine learning tool for automated segmentation of a set of images. Our technique first uses an existing image segmentation

algorithm to semi-automatically segment the first image in the data set. The method then proceeds by using our multiobjective optimization technique to match the segmented shapes in the initial image to their most similar shapes in the second image in the data set. Optimal shapes are determined based upon factors such as color, geometry, motion, and texture. A segmentation for the image is achieved by constructing a weighted combination of these segmentations.

For example, in order to segment the image according to color, the images are converted from RGB to HSV. Three separate segmentations of the image are obtained according to the hue, saturation, and value. The final image which is segmented according to color is given according to the following formula

$$f(w_1, w_2, w_3) = w_1(hue_{new} - hue_{segmented})^2 + w_2(saturation_{new} - saturation_{segmented})^2 + w_3(value_{new} - value_{segmented})^2.$$

Here a least-squares formulation is used to determine the segmentation according to hue, saturation, and value, and  $w_1$ ,  $w_2$ , and  $w_3$  are weights which yield a final segmentation of the image based on a convex combination of the individual segmentations. This least-squares formulation can be extended in a straightforward manner in order to obtain segmentations of the images according to other factors such as geometry, motion, and texture.

This process is then repeated in order to segment the remaining images in the data set. Machine learning is used throughout the optimization procedure in order to determine the most appropriate values for the weights of the relevant terms in the objective function for a set of images.

### 3 EXPERIMENTAL RESULTS

We present some of the results of our automated image segmentation based upon the color of the image. Figure 1 shows an input flower image which is to be segmented according to color. Figure 2 and its manual segmentation are used as input to the automated image segmentation algorithm. The segmentations of the flower image in Figure 1 that are generated according to the hue, saturation, and value are shown in Figures 4, 5, and 6. The final segmentation of the flower image which results from setting  $w_1 = 0.3$ ,  $w_2 = 0.3$ , and  $w_3 = 0.4$  in the multiobjective optimization formula is shown in Figure 7.



Figure 1. Input flower image to be segmented.



Figure 2. Input flower image which is the basis for the multiobjective optimization.



Figure 3. Manual segmentation for the flower image shown in Figure 2. The manual segmentation is also input to the multiobjective optimization.

Note that while the structure of the flower in Figure 7 is recognizable, its segmentation can be improved using other factors, such as geometry, motion, and texture, as the basis for the segmentation. Incorporation of machine learning can also be used to improve upon the results.



Figure 4. Segmentation of flower image in Figure 1 according to hue.



Figure 7. Final segmentation of flower image in Figure 1.

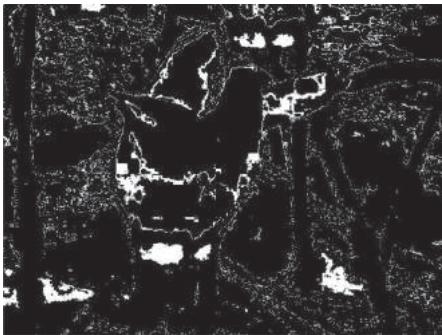


Figure 5. Segmentation of flower image in Figure 1 according to saturation.



Figure 6. Segmentation of flower image in Figure 1 according to value.

In our talk, we will present additional results obtained from testing our automated image segmentation tool on various sets of biological (Charbonneau 2013, Dyer 2013, Ingle et al. 2012) and medical images (Cocosco et al. 1997, West 2004). In addition, we present results on the scalability of our method.

## REFERENCES

- Charbonneau, A. 2013. Image provided as part of a study to quantify phenotypic differences between weedy and native radish populations, Michigan State University, Work in progress.
- Chuang, H.-C. 2014. Automated Image Segmentation for Analysis of Materials Microstructures, Ph.D. Thesis, Purdue University).
- Cocosco, C.A. et al. 1997. BrainWeb: Online interface to a 3D MRI simulated brain database, Neuroimage 5(4): 2/4, S425, Proc. of the 3rd International Conference on Functional Mapping of the Human Brain, Copenhagen. See <http://www.bic.mni.mcgill.ca/brainweb/>.
- Dyer F. 2013. Image provided as part of a behavior study of eye movements. The most recent publication is Ingle et al. 2012.
- Ingle N. et al. 2012. Coordination of Vision and Action in Chameleons, Proc. of the 49th Animal Behavior Society Annual Meeting, Albuquerque, USA.
- Walsh, A.J. and Skala M.C. 2014. An automated image processing routine for segmentation of cell cytoplasms in high-resolution autofluorescence images, Proc. SPIE 8948, Multiphoton Microscopy in the Biomedical Sciences XIV, 89481M.
- West J. 2004. Application of the Level Set Method to Hydrocephalus: Simulating the Motion of the Ventricle, M.S. Thesis, University of Waterloo.

This page intentionally left blank

## Fractal descriptor on holographic images of cervical cells

Mona Mihailescu, Eugen I. Scarlat & Irina Alexandra Paun

*Applied Sciences-Physics Department, University Politehnica Bucharest, Bucharest, Romania*

Irina Grigorescu

*Computer Science and Information Technology Department, University Politehnica Bucharest, Bucharest, Romania*

Roxana Radu

*LOTUS Hospital SRL, Romania*

Oana Tatiana Nedelcu

*Institute of Microtechnology, Bucharest, Romania*

**ABSTRACT:** This paper proposes a method to statistically characterize the proprieties of cervical cells extracted from samples labeled as normal and abnormal such as the ones affected by pathological modifications or subjected to various processes like electroporation. The characterization is based on fractal descriptors applied on 3D images obtained employing digital holographic microscopy technique. Fractal descriptions are using the measures of the dilated volume and surface of the holographic phase images of the cells. The method is a promising tool to discriminate among the normal and abnormal cervical cells by exploiting the whole image of the cell not only limited zones or random cross sections, with potential to be automatically implemented.

### 1 INTRODUCTION

The scope of this study is to characterize the 3D phase images of the cervical cells using fractal descriptors. These images are obtained after reconstruction of the holograms recorded using Digital Holographic Microscopy (DHM) technique.

Various methods are currently in place to discriminate distinct species of cells, or to characterize slightly different features inside the same cell species. The classical techniques are based on chromatographic markers able to identify the key differences between cells such as to separately counting the subspecies of interest and to establish, for example, the differential leukocyte formulae, the weight of immature red blood cells, or the ratio of normal to abnormal cells that help the clinical diagnostic in the Papanicolaou (PAP) test, etc.

Recently, the chromatographic techniques are going to be replaced with the more sophisticated methods based on phase imaging, with the advantage of using no markers; this aspect is preferred because the influence of the markers on the cells characteristics is partially unknown. These images achieved enough spatial resolution to reveal sub-cellular features in order to help the wholeblood analysis (Marquet et al. 2005), red blood cell identification like mature-immature red blood cells

(Mihailescu et al. 2011), leukocytes classification (Scarlat et al. 2013), analysis of red blood cells stored in standard conditions for different periods (Moon et al. 2013, Bhadury et al. 2014), of stressed and unstressed red blood cells (Park et al. 2011, Mihailescu et al. 2012), of anisocytosis stages in a red blood cells population (Mir et al. 2013).

The difficulties appear when optical changes from different sources are mixing in a single phase map thus shadowing the cause under investigation. For example, there are cells with no well-defined nucleus shape or with no nucleus at all so that such silhouette is not sufficient for in-depth analysis. In these cases more complex instruments are needed from both the perspective of image taking devices and the processing software. For example, cervical cells subjected to electroporation process are suffering optical changes at spatial scales that are differing from the ones caused by pathological reasons.

Fractals are powerful tools dedicated to finding the intricate rules between each pixels of the cell image and its relevant optical and morphological features. Fractal descriptor (Bruno et al. 2008) is the instrument which uses not a single (mean) value of the fractal dimension over the whole scale but a set of values over distinct portions of the scale. For any spatial surface expressing the values of the

fractal dimension computed under different scales of analysis, these descriptors provide a complete mapping of the complexity of the image, from the local details until the global pattern distribution.

Here, the descriptors use algorithms to quantitatively describe the properties embedded in the content of 3D reconstructed digital image of the cervical cell such as to ensure the detection of the relevant optical changes that allow for grouping them in normal and abnormal cells.

## 2 CERVICAL CELLS

The cervical cells are harvested in hospital conditions (LOTUS Hospital), respecting ethical and legal regulations, and hygienic rules as well. Cervical cells harvested through the standard protocol are subjected for liquid cytology PAP testing performed using the PAP Cyto fast kit. Typical normal samples consist of cervical cells which include strongly-individualized nuclei with small dimensions (4–10  $\mu\text{m}$ ) as compared to the whole cell (30–80  $\mu\text{m}$ ) and with significant coloration with respect to the embedding cytoplasm. In Figure 1 are the images of a normal and an abnormal cell marked with chromatographic agents, recorded in bright field microscopy. It is clearly observed that the abnormal cells have larger nucleus than the normal ones, the coloration intensity is weaker, and they exhibit a halo around the nucleus (as is depicted in standard procedure (Solomon et al. 2002)).

According to the PAP criteria the result would be negative or positive depending on the number and the weight of the abnormal cells identified in the smear.

## 3 HOLOGRAM AQUISITION AND RECONSTRUCTION

Digital Holographic Microscopy (DHM) technique allows for 3D imaging of cells in their natural, marker free environment. The procedure starts

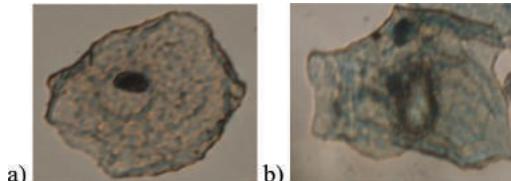


Figure 1. Typical normal cervical cell (a) and abnormal cervical cell (b) in a marked smear on glass plate (Nikon Ti-U microscope, 40 $\times$  objective).

with holograms acquisition in a home-made experimental setup based on the Mach-Zehnder interferometer with additional identical objectives in both arms. The angle between the reference and the object beams could be trimmed to improve the lateral resolution  $\Delta r$  while maintaining a convenient depth range (Wang et al. 2008)

$$\Delta r = \frac{2\lambda}{\pi A}, \quad (1)$$

where  $A$  is the numeric aperture and  $\lambda$  is the laser wavelength. A CCD camera (Pike F421, 7.4  $\mu\text{m}$  pixel pitch, 2048  $\times$  2048 pixels) grabs the hologram and send it to the computer unit.

In the case of our system, we used 40 $\times$  objectives with numerical aperture of 0.6 at a wavelength of  $\lambda = 632.8$  nm. The transverse resolution is, as in classical optical microscopy, diffraction-limited (Cucu et al. 1999). For the present setting, a lateral resolution of  $\Delta r = 0.7 \mu\text{m}$  was achieved in good agreement with the Rayleigh's criterion for coherent illumination (Born et al. 1999). It is demonstrated that the lateral resolution is independent on the recording distance (Doblas et al. 2015).

The reconstruction of the 3D object image is performed digitally using an algorithm based on the scalar diffraction theory, Fresnel approximation (Koala®). In Figure 2 left are presented the holograms at the level of single cell for a normal (a) and an abnormal cell (b) where the fringe displacement is proportional to the phase shift introduced

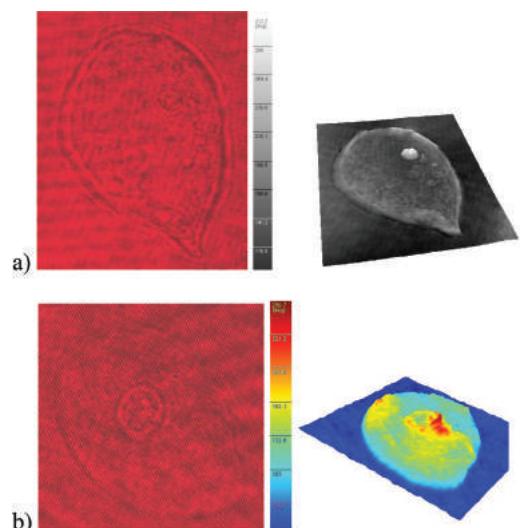


Figure 2. Hologram of a cervical cell (left) and the corresponding phase image of the reconstructed object (right) for a) normal cell, b) abnormal cell.

by the cell in the optical path relative to the substrate. The 3D reconstructed phase maps of the same cells are in the Figure 2 right. The axial resolution is given by the reconstruction numerical procedure which can simulate the Fresnel transform at distances in nanometers range. Consequently, in the reconstruction process, DHM provides multi-focus quantitative phase contrast imaging with an axial resolution up to approximate 5 nm (Kemper et al. 2007).

The histograms are processed taking into account the whole area of the cell according to a standard procedure consisting in the following steps: 1) numerical filtering of the diffraction orders in the Fourier domain to drop off the twin image, 2) numerical filtering inside the first diffraction order to eliminate the noise due to parasitic fringes from the experimental setup, 3) tilt and aberration corrections, followed by three specific steps devoted to this kind of cells, i.e. 4) profile extraction and measurements (phase shift and diameters), 5) implementation and running of the dilation and packing algorithm, and 6) statistical characterization of the cells. The last three steps are implemented in MATLAB and are applied for all cells following the same rules.

An enough large number of holographic images were grabbed and reconstructed from both normal and abnormal samples, followed by the extraction of 1D cross sectional profiles passing through the corresponding nucleus (Fig. 3). One can observe that for

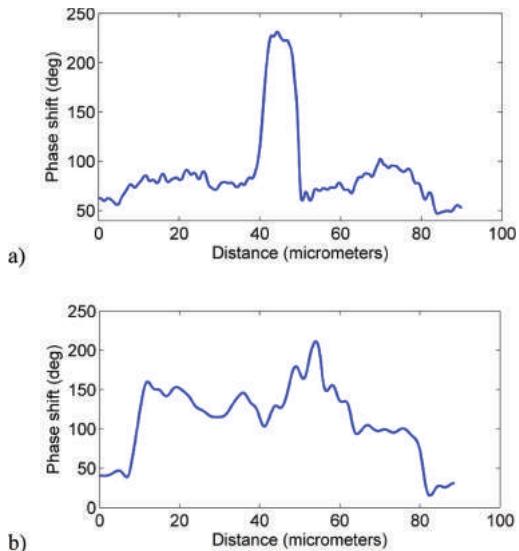


Figure 3. Cross sections through the phase map of a normal (a) and an abnormal (b) cervical cell. The section plane is passing through the nucleus surrounded by the cytoplasm.

normal cells the cross section profile exhibits a sharp peak corresponding to the refractive index of the compact nucleus (Fig. 3a), differing from the cases of abnormal cells (Fig. 3b) with diffused nucleus inside the cytoplasm or modified membranes.

While the diameters of the nuclei of the normal cells are between 4–10  $\mu\text{m}$ , for abnormal cells they are larger than 15  $\mu\text{m}$ . The range of the phase shift introduced by the cells along the optical path is also different for normal and abnormal cells: it is of the order of 100 degrees in the first case but no more than tenths of degrees in the second case. The phase shift distribution over the cell profile according to Figure 3 is assumed to be bimodal; in the present work the normal cells are considered when the Ashmann's coefficient  $D_A$  is larger than two

$$D_A > 2, \quad (2)$$

such as the distributions are clearly separated (Ashmann et al. 1994).

Each holographic image  $H \in [1:p] \times [1:q] \rightarrow I$  has different values for  $p$ , and  $q$  pixels number in transverse directions, but the same  $I_{\max} = 256$  phase levels. The phase image is converted in a three-dimensional surface  $S$  in such a way that each pixel with coordinate  $(x, y)$  in the image is mapped onto a point in the 3D space with coordinates  $(x, y, z(x,y))$ , where  $z$  is the pixel phase shift, that is:

$$S = \{(i, j, k(i, j)) | (i, j) \in [1:p] \times [1:q]\}, \quad (3)$$

and

$$k(i, j) = \{1, 2, \dots, I_{\max}\} | k = H(i, j)\}, \quad (4)$$

where  $I_{\max}$  corresponds to the maximum phase shift in the image. The dimensionless quantities result via the normalization of  $p$ ,  $q$  and  $2\pi$  respectively.

#### 4 FRACTAL DESCRIPTORS

A fractal object is a geometrical set of points in a multidimensional topological space exhibiting a structure characterized by a theoretically infinite degree of self-similarity, each part of the object corresponding to a copy of the whole. The fractal dimension is the most used measure to describe such a fractal object. Since in the real-world the set of points  $S$  is a digital images represented in a discrete and finite space, several approximation methods have been proposed in the literature (Falconer 1986). Such methods aim to compute a measure of self-similarity and complexity of the object by generalizing the definition of the Euclidean dimension. Thus the object is measured by a rule unit with the same topological dimension of the object. The length  $r$  of this unit is ranged along an interval to compute the

number of units  $N(r)$  necessary to cover the object. In this context, the fractal dimension is given by

$$d \approx -\lim_{r \rightarrow 0} \frac{\log N(r)}{\log r}. \quad (5)$$

Fractal descriptors extend the fractal dimension concept by computing  $d$  under different scales of analysis. Here the descriptors exploit the fractal properties of the volume  $V_s(i,j,k)$  and the corresponding external frontier surface  $\Sigma_s(i,j,k)$  of the digital holographic images  $S(i,j,k)$  when covering it with structural elements as suggested by Falconer (Falconer 1986).

Our analysis applies these descriptors on the 3D reconstructed phase images containing single cell structure. Simultaneous use of volume and surface is necessary because distinct cells can exhibit the same volume but different surfaces.

In the Bouligand–Minkowski method, any element of  $S$  is dilated by a spheric ball  $s(r)$  with variable radius  $r$ . For each  $r > 0$ , the set  $s(r)$  is defined to be the  $r$ -neighborhood of  $S$ , i.e. the set of all points in  $\mathbf{R}^3$  which are at distance less than  $r$  from  $S$  (or equivalently,  $s(r)$  is the union of all the spheres of radius  $r$  which are centered at a point in  $S$ ). The dilation of  $S$  by  $s_r$  is  $S_r = S \oplus s(r)$  with the dilated volume  $V_r = V(S \oplus s(r))$ , and the dilated frontier surface  $\Sigma_r = \Sigma(S \oplus s(r))$ .

Computationally, the dilating spheres  $s(r)$  are centered on the pixels of  $S$  with increasing radii following a scaling rule

$$\rho = \{r_{\min}, r_1, r_2, \dots, r_{\max}\}, \quad (6)$$

where the radii are computed according to the Euclidean metric

$$r_{ijk} = \sqrt{i^2 + j^2 + k^2}, \quad (7)$$

with  $i, j, k$  given by Eqs. (3)–(4).

For the frontier case the dynamic of the fractal descriptor  $D_\Sigma$  is computed using Eq. (5) according to the number of disjoint balls of radii given by Eqs. (6)–(7) that are covering the surface  $\Sigma_r$  (see Fig. 4):

$$D_\Sigma = \left\{ \frac{\log N_\Sigma(s(r_{ijk}))}{\log r_{ijk}} \right\}. \quad (8)$$

Similarly, the descriptor  $D_V$  for volume case is the number of disjoint balls filling the volume  $V_r$  (see Fig. 5):

$$D_V = \left\{ \frac{\log N_V(s(r_{ijk}))}{\log r_{ijk}} \right\}. \quad (8')$$

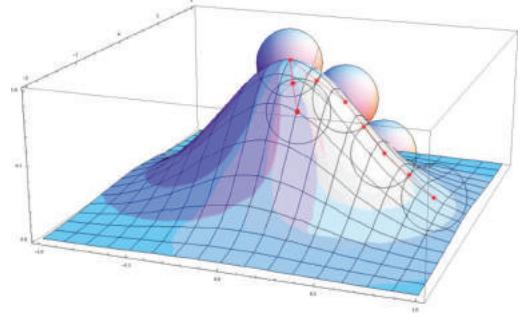


Figure 4. The principle of covering the surface with disjoint balls.

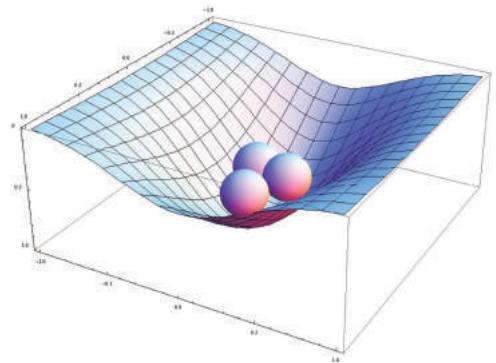


Figure 5. The principle of filling the volume with disjoint balls.

In the present work the descriptors (8) and (8') are used together following a version of the method used by Da Silva Oliveira et al. (Da Silva Oliveira et al. 2014):

$$D_{\Sigma V} = \left\{ \frac{\log N_\Sigma(s(r_{ijk}))}{\log r_{ijk}}, \frac{\log N_V(s(r_{ijk}))}{\log r_{ijk}} \right\}. \quad (9)$$

The algorithm for the packing of spheres onto the surface or inside the volume initiates with randomly positioning of the first sphere followed by a neighborhood strategy. For the following positions concentric spheres around this first position are considered at increasing distances such as no overlapping are allowed between the candidate element and the all other already placed spheres. If the candidate does not overlap with any of the previously placed spheres, then it will be kept and the procedure moves forward with one step. The algorithm stops when all possible candidates from a given distance are outside the cell's boundary. For every radius of the structural element the

number of covering spheres is averaged over all position of the first element. The averaged values of the numbers of spheres at the given radius are provided automatically at the end of the algorithm.

The procedure was equally applied for all investigated cells. MATLAB codes were written separately for surface covering and volume filling and run for all values of the radii of spheres given by Eq. (6). In this case, the background surrounded the cell is set to zero values.

## 5 RESULTS

For the present study the sources of the normal cells were the samples qualified as negative by the result of the PAP smear analysis, while the sources of the abnormalities were the samples qualified as positive. The study further focuses on single cell analysis performed on cells extracted from negative samples and subsequently selected to fulfill Eq. (2); if so, they were labeled as normal. Similarly the abnormal label was assigned to the cells extracted from the positive samples that did not fulfill Eq. (2). 100 normal and 100 abnormal cells were selected in this way.

For every cell the local fractal dimension is computed by estimating the linear slope  $\hat{\mu}$  with the variance table for the fitted model (ANOVA) under the constraints of statistical significance of  $p$ -value and  $R^2$  coefficient. By imposing  $p$ -value  $< 5\%$  and  $R^2 > 95\%$ , the fractal descriptor  $D_\Sigma$  (Eq. (8)) for abnormal cells exhibits a critical point that splits the abscissa interval in two different regions with distinct slopes corresponding to small scale  $2\sqrt{3} \leq r_{ijk} \leq 15\sqrt{3}$ , and big scale  $15\sqrt{3} < r_{ijk} \leq 45\sqrt{3}$  (see Fig. 6). A pyramidal shape is inserted as benchmark (the blue curves in Fig. 6 and Fig. 7). These regions remain valid for the volume descriptor  $D_V$  (Eq. (8')), see Fig. 7).

Differing from the pyramid shape which exhibits a diminishing slope (in absolute values) at big scale

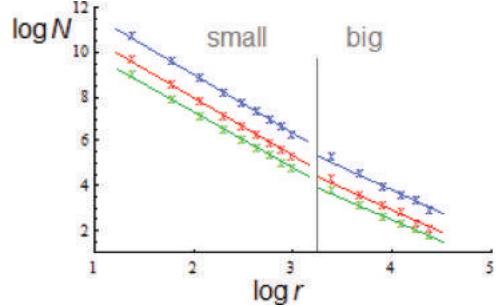


Figure 7. Typical descriptor (volume case  $D_V$ ): normal cells (green), abnormal cells (red), and pyramid shape (blue).

Table 1. Estimated local slopes.

Descriptor/ cell type	Local slope ( $-\hat{\mu} \pm \sigma_{\hat{\mu}}$ )	
	Small	Big
$D_\Sigma/\text{normal}$	$2.105 \pm 0.004$	$2.308 \pm 0.003$
$D_\Sigma/\text{abnormal}$	$2.139 \pm 0.007$	$2.561 \pm 0.012$
$D_V/\text{normal}$	$2.569 \pm 0.006$	$2.037 \pm 0.007$
$D_V/\text{abnormal}$	$2.703 \pm 0.009$	$2.392 \pm 0.020$

irrespective the descriptor, the cells behave contradictory: the local slope increases at big scale in the surface descriptor and decreases in the volume descriptor. However, the quadratic approximation over the whole scale (small and big) is negatively shaped in surface description and positively shaped in the case of volume description.

The estimated local slopes for the investigated regions are given in Table 1. All the data are presented as mean values and standard errors of the means  $\sigma_{\hat{\mu}}$ .

The trends of the mean values of the local fractal dimensions are opposite in the descriptors when moving from small to big scale, therefore the measures could be concatenated to improve the discriminating power of the method.

Since we are interested in evaluating the differences among normal and abnormal cells, a popular measure to address the issue is the Bhattacharyya distance (Wallin et al. 2015). Assuming Gaussian distributions of slopes for both normal and abnormal cells  $G(\hat{\mu}, \hat{\sigma})$  with  $\hat{\sigma} = 10\sigma_{\hat{\mu}}$  the distances increase from small scale to big scale; the largest distance is in the case of the surface descriptor, at big scale (see Table 2). Conversely, since the smallest distance is found at small scale in the surface description, it would be an additional argument to concatenate the descriptors as suggested by Eq. (9).

The small distances at small scales are partially the effect of noise that affects the images below the

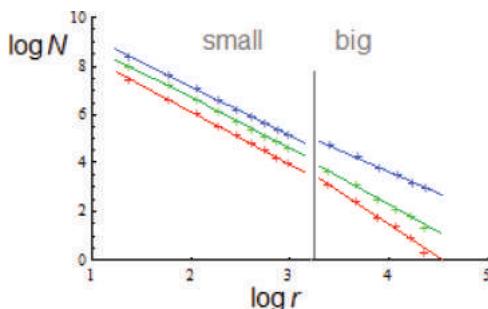


Figure 6. Typical descriptor (surface case  $D_\Sigma$ ): normal cells (green), abnormal cells (red), and pyramid shape (blue).

Table 2. Bhattacharyya distances between the distributions of slopes of normal and abnormal cells.

Descriptor	Scale	
	Small	Big
$D_\Sigma$	0.118	1.392
$D_V$	0.416	0.928

diffraction limit where operate the coverings at the lower values of the radii in the vicinity of  $r_{\min}$  (see Eq. (6)).

The distributions of the slope and consequently the distributions of the local fractal dimension are more compact in the case of normal cells in the sense of smaller kurtosis whatever the volume or the surface description.

## 6 CONCLUSIONS

The cervical cells were analyzed using an image processing method based on fractal descriptors in order to reveal the intricate elements able to classify the cervical cells in normal and abnormal categories. An algorithm based on fractal descriptors was developed and applied on the reconstructed holographic 3D images which contain the phase shift introduced by the cell in the optical path (unmarked cell).

Due to the sharp delimitation of the normal nuclei evidenced in a bimodal distribution of the cross sectional phase profile, the dynamics of the local slope—equivalently, the local fractal dimension—in surface and volume descriptions of the cell images was presented. The surface descriptor at big scale and the volume descriptor at small scale have greater sensitivity to phase changes exhibited by the abnormal cells, therefore is recommended to use them together. However the small-big scale partition and the establishing a threshold for Bhattacharyya distance are subjects to more research.

The characterization of holographic images with fractal descriptors is a promising tool to detecting information about the phase changes of the cervical cells to the advantage of taking into account the entire cell not only a random cross section. Moreover the method is marker free and suitable for automate analysis.

## ACKNOWLEDGEMENT

This research was financed by the Romanian Authority for Scientific Research and Innovation contract UEFIS-CDI, PN-II-PT-PCCA no. 30/2014.

## REFERENCES

- Ashman K.M. et al., 1994, *Detecting bimodality in astronomical datasets*, Astron. J. 108 2348–2361.
- Bhaduri B. et al., 2014, *Optical Assay of Erythrocyte Function in Banked Blood*, Scientific Reports, 4, (6211).
- Born M., Wolf E., eds., 1999, “*Principles of Optics*”, (Cambridge University Press, 467–472).
- Bruno O.M. et al., 2008, *Fractal dimension applied to plant identification*, Information Sciences 178 (12), 2722–2733.
- Cuche E. et al., 1999, *Simultaneous amplitude-contrast and quantitative phase-contrast microscopy by numerical reconstruction of Fresnel off-axis holograms*, Appl. Opt. 38, 6994–7001.
- Da Silva Oliveira et al., 2014, *Enhancing Fractal Descriptors on Images by Combining Boundary and Interior of Minkowski Dilation*, Physica A, 393, 2623–2635.
- Doblas A. et al., 2015, *Study of spatial lateral resolution in off-axis digital holographic microscopy*, Opt. Comm. 352, 63–69.
- Falconer K.J., 1986, “*The Geometry of Fractal Sets*”, Cambridge Univ. Press, N.Y.
- Kemper B. et al., 2007, *Digital Holographic Microscopy A New Method for Surface Analysis and Marker-Free Dynamic Life Cell Imaging*, Biophotonics 2, 41–47.
- KOALA <http://www.lynceetec.com/content/view/497/193/>.
- Marquet P. et al., 2005, *Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy*, Opt. Lett. 30, 468–470.
- Mihaiilescu M. et al., 2011, *Automated Imaging Identification and Counting of Similar Cells From Digital Hologram Reconstruction*, Applied Optics, 50 (20), 3589–3597.
- Mihaiilescu M. et al., 2012, *Diffraction pattern study for cell type identification*, Optics Express, 20, (2), 1465–1474.
- Mihaiilescu M. et al., 2010, *Wavelet filters for programmable object investigation in digital holographic microscopy*, Journ. Optoelect. Adv. Mat., 12 (1), 94–99.
- Mir M. et al., 2010, *Blood screening using diffraction phase cytometry*, J. Biomed. Opt. 15, 2, 027016, 2010.
- Moon I. et al., 2013, *Automated quantitative analysis of 3D morphology and mean corpuscular hemoglobin in human red blood cells stored in different periods*, Opt. Expr. (21), 25, 30947–30957.
- Park YK. et al., 2011, *Measurement of the nonlinear elasticity of red blood cell membranes*, Phys. Rev. E, 83, 051925.
- Scarlat E.I. et al., 2013, *Discriminating the main representatives of the white blood cell species on the basis of the fractal properties of the DHM phase profile*, UPB. Sci. Bull., A, 75 (2), 147–154.
- Solomon D. et al., 2002, *The 2001 Bethesda system. terminology for reporting results of cervical cytology*, Jour. Americ. Medical Assoc., 287, 2114–2119.
- Wallin, J. et al., 2015, *Latent modeling of flow cytometry cell populations*. arXiv preprint arXiv:1502.04058.
- Wang H. et al., 2008, *Simple and robust digital holography for high-resolution imaging*, Chin. Opt. Lett., 6, (3), 165.

# An integrated two Time-of-Flight camera system to determine knee flexion movement: Comparison with standard motion analysis system

Emeric Veron-Tocquet

*Laboratoire de Traitement de l'Information Médical, LaTIM-INSERM U1101, France  
Université de Bretagne Occidentale, France*

Valérie Burdin

*Laboratoire de Traitement de l'Information Médical, LaTIM-INSERM U1101, France  
Institut Mines Telecom-Telecom Bretagne, France*

Joel Savéan

*Laboratoire de Traitement de l'Information Médical, LaTIM-INSERM U1101, France  
Université de Bretagne Occidentale, France*

Julien Leboucher & Olivier Rémy-Nériss

*Laboratoire de Traitement de l'Information Médical, LaTIM-INSERM U1101, France  
Université de Bretagne Occidentale, France  
CHU Brest, France*

**ABSTRACT:** Medical rehabilitation therapies depend on the accurate motion of bone structures, which is often impossible to acquire without using invasive markers. Current clinical studies that evaluate measurements of motion use skin markers even though its drawbacks due to skin motion are well-known. Recent developments in motion analysis use markerless Time-of-Flight (ToF) cameras that can acquire three dimensional data. This study was focused on integrating two synchronized ToF cameras to evaluate knee flexion movement and compare the findings with standard Vicon system data. A rigid structure was developed and the best relative location of ToF cameras on this structure relative to the patient was analyzed. In the experimental process, the rigid structure was placed in the motion analysis lab and standard skin markers were placed on the lower leg of a volunteer. The volunteer performed knee flexion movements while lying on a bed and knee flexion data was simultaneously acquired by both the Vicon system as well as the integrated ToF cameras during nine camera placement combinations. The RMS error between the two measurement methods ranged from 5.61° to 9.64° depending on camera locations.

## 1 INTRODUCTION

Motion analysis is used in many cases of rehabilitation: comprehension of bone deformities, diagnosis of a deficient function, post-surgical follow-up evaluation. Methods that provide motion information are mainly based on marker position determined by an acquisition system. One of the most popular, in clinical diagnosis, is the optoelectronic system (e.g. Vicon motion analysis system). This kind of system acquires 3D positions of each marker placed, on the patient's skin representing anatomical landmarks. These methods provide an accuracy within 1 mm [1] for marker position but skin artifacts appear due to the relative displacement between bones and skin [2], [3]. Furthermore, patient instrumentation and preliminary calibration

of the optoelectronic cameras system is time concerning for clinicians and patients.

Recently, new technologies have appeared for acquiring scene in 3 dimensions and without markers. This include structured light cameras, multi-view RGB cameras, and Time-of-Flight (ToF) cameras. The ToF cameras system is an interesting way to acquire dynamic movement with acceptable accuracy compared with RGB cameras (low accuracy) [4] and structured light (no dynamic acquisition) [5]. A preliminary study for determining knee flexion movement with a single ToF camera provided encouraging results [6]. In this study, angle differences were compared between flexion angle measured with single ToF camera and flexion angle calculated with actual bones positions on a cadaveric lower limb. The maximum difference

calculated was  $9^\circ$  and the RMS, for three trials, was  $8.03^\circ$ ,  $5.90^\circ$  and  $5.60^\circ$ . Reinschmidt and colleagues [7] compared bone locations with the segment location by markers on the skin to estimate flexion angle. They found RMS error equal to  $5.3^\circ$  with a maximal difference of  $7.9^\circ$ .

In the current study, the surface coverage of the ToF was increased by using a synchronized two-ToF camera system. The aim of this study was to determine the accuracy of knee flexion movement using two-ToF system as well as to bring out the best position of cameras in reference with each other and to the patient.

## 2 MATERIALS AND METHODS

A structure was build to mount the ToF cameras in order to acquire the knee flexion movement in the transvers plane. The structure was made of metal ladders to which a one-meter-radius semi circle bar was attached (Fig. 1-a). This arc was marked with angle graduations in order to mark the position of the two ToF-cameras (Fig. 1-b, where the angle  $\alpha$  represented angle between the two cameras). Each camera was fixed on a 3-dimensions articulated pliers, which permitted to orient them in the same plane (plane formed by the arc). Figure 1-c represents the operating structure with the angle between cameras ( $\alpha$ ) and the angle between the two arcs ( $\theta$ , that represents the position of the two-cameras system relative to the patient).

### 2.1 Laser calibration method

We developed a system based on plane lasers to calibrate the cameras together. We created

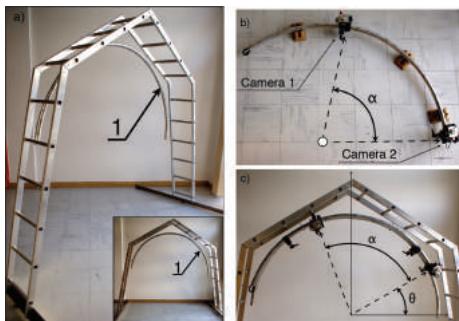


Figure 1. a) Ladder and the static semi-circle fixed on. b) Definition of  $\alpha$  angle between the two cameras. c) Represents the operating structure with the angle between cameras ( $\alpha$ ) and the angle between the two arcs ( $\theta$ , that represents the position of the two-cameras system relative to the patient).

aluminiboxes to receive the plane lasers (Fig. 2). We installed them to obtain the vertical and horizontal axis of cameras ( $\bar{x}$  and  $\bar{y}$ ) when laser were projected on a wall. Lasers had two degrees of freedom in rotation, by this way we were able to fix them in the right position to obtain a cross formed by the two projected line lasers. The center of the cross corresponded the center of the camera.

The alignment of cameras was a two-step process. First the alignment of the center of the two cameras was performed (Fig. 3-b) followed by the alignment of camera axes (Fig. 3-c).

### 2.2 Experimental process

A one single healthy male volunteer participated in this study. The volunteer was laid supine on a bed that was placed under the structure with the

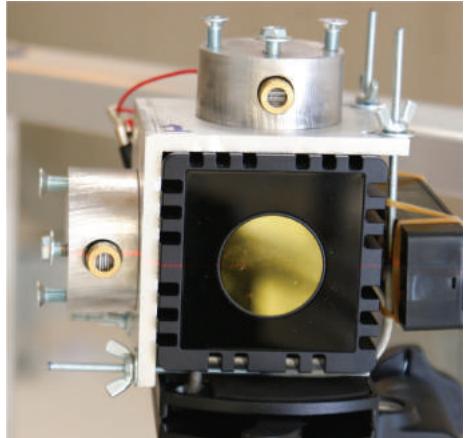


Figure 2. Laser in their box with two degrees of freedom in rotation, which permitted to align the  $\bar{x}$  and  $\bar{y}$  of cameras and obtain the center of camera by the cross performed by them.

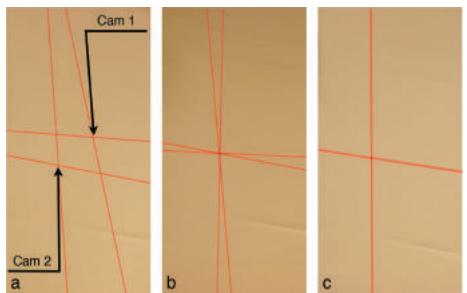


Figure 3. Method to align the two cameras together. a) Cameras not aligned. b) Center of cameras aligned. c) Axes and center of cameras aligned.

knee located directly under the plan of arc. Using this set-up, nine tests were performed using a combination of  $\alpha$  and  $\theta$  in degrees:

$$\alpha = \{60; 90; 120\} \quad \text{and} \quad \theta = \{0; 30; 60\}.$$

In each test, two consecutive flexion movements were performed. To obtain the same coordinate system for the two cameras views, rigid transformations were applied. One rotation of  $\alpha$  angle to obtain in the same coordinate system clouds points and a translation corresponding to the translation between the two cameras.

### 2.3 Vicon system measurement

In addition to the ToF camera measurements, the volunteer was instrumented with marker visibility to follow the flexion movement with the Vicon optoelectronic system. The Vicon system was as our gold standard. Markers were placed on thigh (3 femur markers), calf (3 tibia markers), ankle (2 markers), knee (2 condyle markers) and hip (2 markers). The frequency of acquisition was 120 Hz.

### 2.4 Synchronization of the two systems

Because of difficulties to synchronize the *Vicon* system with any other systems which is not designed by the same company, an in-house system for synchronization was created. Inspired from the synchronization method in movie studio in order to synchronize sound and image, an inflatable balloon was used for this purpose. A reflective marker was stuck in the center of the hand of a manipulator and the balloon was put on it. The manipulator placed his hand with the marker and the balloon under the structure. The balloon was visible by ToF cameras and covered the marker visibility from *Vicon* cameras. As soon as the volunteer started moving his legs, the manipulator poked the balloon with a pin. The balloon disappeared from the ToF camera views and the reflective marker became visible by the *Vicon* system. With this method, an accurate start signal was obtained for each acquisition system and their synchronization was possible.

## 3 RESULTS

### 3.1 Data from the Vicon system

Using skin markers of tibia and femur, the position and orientation of both tibia and femur segments were known (Fig. 4). Using this information the angle of flexion  $\alpha_{vicon}$  was plotted against time.

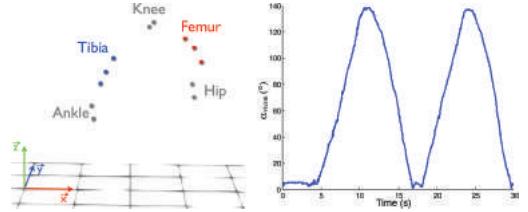


Figure 4. Left: visualization and identification of markers position captured by *Vicon* system. Right: flexion angles calculated with *Vicon*: ( $\alpha_{vicon}$ ).

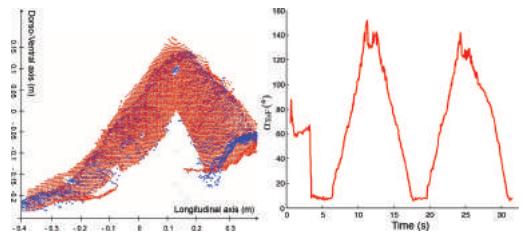


Figure 5. Shows, in the sagittal plan, these two points clouds in the same coordinate system, after filtering to delete uninteresting points. Two colors appear, corresponding to points acquired by each camera. Left: view of the two points clouds in the sagittal plan. Camera 1 in red and Camera 2 in blue. Right: flexion angles along time calculated with two-ToF cameras: ( $\alpha_{ToF}$ ).

### 3.2 Data from the two-ToF cameras

ToF cameras gave two different views of the leg during flexion movement. Each view was represented by a point cloud. After rigid rotation and translation, depending on  $\alpha$  and  $\theta$  values, two point clouds in the same coordinate system were obtained. A first filtering was applied to these sets of points to only keep points corresponding to the moving leg. By this mean we removed all other points, corresponding to the bed, the ground, the ladders, and the other leg which did not move (Fig. 5).

A second filtering was applied on the point cloud corresponding to the leg. This point cloud was split in two different point clouds set corresponding to the femur and the tibia. From each of them the principal inertial axis was determined followed by the angle between these two axes (corresponding to femur and tibia) to obtain flexion angles measured by ToF cameras,  $\alpha_{ToF}$ .

### 3.3 Measurement errors

Using the synchronization method between the *Vicon* system and ToF camera system, it was possible to compare the flexion curves as calculated by two systems (Fig. 6). The curve of differences

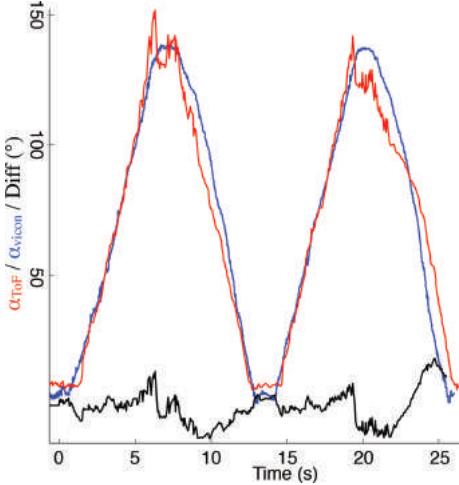


Figure 6. Comparison of flexion angles measured,  $\alpha=60^\circ$  and  $\theta=0^\circ$ .

Table 1. Error and RMS error calculated between ToF cameras system and Vicon system for different positions of ToF cameras.

$(\alpha; \theta)$	[Min; Max] (deg)	RMS Error (deg)
(60; 0)	[-11,82; 14,36]	6.19
(60; 30)	[-11,67; 15,21]	9.64
(60; 60)	[-13,05; 14,21]	9.10
(90; 0)	[-10,75; 12,53]	5.73
(90; 30)	[-11,26; 14,49]	6.22
(90; 60)	[-10,89; 14,19]	7.15
(120; 0)	[-8,66; 10,32]	5.82
(120; 30)	[-11,10; 13,74]	7.32
(120; 60)	[-9,21; 9,83]	5.61

between  $\alpha_{ToF}$  and  $\alpha_{Vicon}$  was also plotted (Fig. 6) for each test.

Furthermore, the difference error at each time interval and the RMS error for the entire flexion curves was also determined, which was more representative of accuracy of our system.

This analysis was performed for each test with different values of  $\alpha$  and  $\theta$  angles between ToF cameras and between cameras and patient. The results are summarized in Table 1.

#### 4 DISCUSSION

This study successfully implemented an integrated ToF system to determine knee flexion angle and in doing so illustrated the feasibility of using multiple ToF camera system as a viable, quick, and

economical option for future clinical or surgical procedures. RMS errors obtained were highly dependent on camera positions and provided valuable insights in determining the optimal camera positions during in vivo assessments. When  $\{\alpha; \theta\} = \{120; 60\}$ , the camera number 1 was on the left side of the leg, which resulted in lowest RMS error. A low angle of incidence between infrared lights of the ToF and the acquired surface decreased the accuracy of measurements [6]. This phenomenon was confirmed by Chibrando and colleagues [8] who reported unacceptable angle measurements for incident angles smaller than  $45^\circ$ . When a camera was on one side of the leg, incidences of rays were constant during all the movement and close of  $90^\circ$ . When a camera was over the leg with a flexion close to  $140^\circ$  of flexion, the incidence of light rays became smaller than  $20^\circ$  due to the orientation of femur and tibia and the angle measurements were less accurate: for example, at  $\{\alpha; \theta\}$  equal to  $\{60; 30\}$ . Thus, keeping the ToF camera on one side of the leg permitted us to obtain accurate results. The second camera was needed to increase the acquired surface and to increase the measurement accuracy.

Minimum and maximum errors varied from  $-13.05^\circ$  to  $15.21^\circ$  but they were extrema values. No correlation was found relative to the cameras position. As it appeared on Figure 6, the high values of error appeared when the maximum of flexion was reached ( $140^\circ$ ) and when the leg returned to the extension. To understand the errors better, knee movement was split into flexion and extension cycles and data in each cycle was analyzed separately. For all the nine tests, the minimum and maximum errors calculated during the flexion phase ( $0^\circ$  to  $140^\circ$ ) were  $-7.34^\circ$  and  $6.42^\circ$ , respectively for test  $\{\alpha; \theta\} = \{90; 30\}$  and  $\{60, 60\}$ . These values were smaller than those obtained during the extension (Min and Max in Table 1 were obtained during extension cycles). We justify this difference between flexion and extension phases, by the muscle contraction effect. During the movement, quadricep was active during extension phase and the hamstring was passive. During the flexion, hamstring became active and quadricep passive. Thus, the surface shape was greatly modified between both the movement phases. The difference in the point clouds shape could probably have impacted the calculus of femur and tibia axes. For the Vicon system, the skin markers were probably impacted by the soft tissues artifacts.

#### 5 CONCLUSIONS

This markerless ToF camera system provides valuable insights and feasibility in its use as a rehabilitative diagnostic tool. The best RMS

errors for this system were low when best camera positions were considered. These errors (5.61) were in fact in line with the known accuracy of the optoelectronic system (5.3) with skin motion artifact [7]. Thus, results found in this study encouraged us to perform test on cadaveric leg with intracortical pins in femur and tibia followed by a *Vicon* system. This way, the gold standard system will be the best in order to acquire the actual positions of bones and to compare with flexion angles measured by our two-ToF synchronized cameras system.

## ACKNOWLEDGEMENT

We would like to thank Bhushan Borotikar, PhD for the assistance in the language corrections and to Mathieu Lempereur, PhD for the use of the *Vicon* system.

## REFERENCES

- [1] M. Windolf, N. Gtzen, M. Morlock (2008), “Systematic accuracy and precision analysis of video motion capturing systems-exemplified on the Vicon-460 system”, *Journal of Biomechanics*, 41(12):2776–2780.
- [2] F. Leboeuf, C. Schwartz, S. Brochard, M. Lempereur, V. Burdin, O. Remy-Neris (2012), “Relevance of skin deformation to track the scapula during forward humeral elevations”, *Computer Methods in Biomechanics and Biomedical Engineering*, 15(1):365–367.
- [3] A. Leardini, L. Chiari, U. Della Croce, A. Cappozzo (2005), “Human movement analysis using stereophotogrammetry, Part 3: Soft tissue artifact assessment and compensation”, *Gait & posture*, 21(2):212–225.
- [4] German K.M. Cheung, Simon Baker, and Takeo Kanade (2003), “Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture”, In *Proceedings of the IEEE computer society conference on Computer Vision and Pattern Recognition (CVPR'03)*, Washington DC, 77–84.
- [5] C. Rocchini, P. Cignoni, C. Montani, P. Pingu and R. Scopigno (2001), “A low cost 3D scanner based on structured light”, *Computer Graphics Forum*, 20(3):299–308. DOI: 10.1111/1467-8659.00522.
- [6] E. Veron-Tocquet, J. Leboucher, V. Burdin, J. Savean, O. Remy-Neris (2014), “A study of accuracy for a single Time of Flight camera capturing knee flexion movement”, In *Proceedings of the IEEE Healthcare Innovation Conference (HIC)*, Seattle WA, 339–342. DOI: 10.1109/HIC.2014.7038944.
- [7] C. Reinschmidt, A.J. Van Den Bogert, B.M. Nigg, A. Lundberg, N. Murphy (1997), “Effect of skin movement on the analysis of skeletal knee joint motion during running”, *Journal of Biomechanics*, 30(7):729–732.
- [8] F. Chiabrando, D. Piatti, F. Rinaudo (2010), “SR-4000 ToF camera: further experimental tests and first applications to metric surveys”, *Intern. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 38(5):149–154.

This page intentionally left blank

# Mechanics-based analysis of the left atrium via echocardiographic imaging

S. Gabriele, L. Teresi & V. Varano

*Università Roma Tre, Roma, Italy*

P. Nardinocchi, P. Piras, G. Esposito, P.E. Puddu & C. Torromeo

*Sapienza Università di Roma, Roma, Italy*

A. Evangelista

*Ospedale San Giovanni Calibita Fatebenefratelli—Isola Tiberina, Roma, Italy*

**ABSTRACT:** A mechanics-based analysis of data from three-dimensional speckle tracking echocardiography can help to infer useful clinical data about the state of myocardium. Precisely, we recently conjectured that principal strain lines in the left ventricle change in presence of heart disease, and used data from healthy subjects and patients to support our conjecture. Here, we set the first steps of our analysis relative to the left atrium, describing our method of acquisition, and discussing the key mechanical characteristics delivered by the echocardiographic device for healthy subjects and patients affected by hypertrophic cardiomyopathy.

## 1 INTRODUCTION

The recent techniques of visualization realized by Three-Dimensional Speckle Tracking-based motion-detecting Echocardiography (3DSTE) make possible, non invasively, to generate a class of data which can be studied through a post-processing mechanics-based analysis. In the last years, our group proposed two different methods of analysis, which may also be integrated one with the other: the first generates a Matlab-based tool denominated S-Lab which can automatically analyze a large amount of data and identify both principal strain lines and values in the epicardial and endocardial layers of the Left Ventricle (LV); the second borrows tools from Modern Shape Analysis and investigates human left ventricular heart morphological changes in time among a group of subjects [1, 2, 3, 4]. Both the analyses aim to identify indicators of early LV dysfunctions, which are only based on non-invasive techniques.

As it is well known, also Left Atrium (LA) function is an important clinical variable, and recent studies conferred credit to echocardiographically derived LA mechanics, to non invasively assess LA compliance, and asked for the development of specific software for LA analysis [5].

We wish to make some progress towards addressing this question, extending the analysis proposed

and discussed for the human LV to the corresponding LA. Precisely, as for any LA of our database, we also have the corresponding LV, and both echocardiographic data are acquired in the same cardiac cycle, we can plan a crossed analysis which can take advantage of the knowledge of the mechanics of both LA and LV chambers of the heart. We focus the analysis on LA data corresponding to healthy subjects and Hypertrophic Cardiomyopathy (HCM) patients, and preliminarily discuss the mechanics of LA on the basis of the mechanical variables delivered by the echocardiographic device.

Next steps involve the application of our S-Lab tool on the acquired LA, to set a crossed analysis based on the endocardial and epicardial principal strain values in healthy subjects and HCM patients.

## 2 METHOD

The technique is based on data obtained by 3DSTE which is an application of pattern-matching technology to ultrasound cine data, based on the tracking of the ‘speckles’ in a 3D volume; speckles are disturbances in ultrasounds caused by reflections in the ultrasound beam: each structure in the body has a unique speckle pattern that moves with tissue (see [4, 6] for further details). As when LV is the object of the study, the operator manually

approves LA shape and identifies the endocardial border starting from the mitral annulus. From here, the acquisition process may start automatically through the identification of the epicardial-border of the LA wall, and of the 16 segments which portend the LA. The segments numbered from 1 to 6 correspond to the basal segments: Basal Antero-Septum (BAS), Basal Anterior (BA), Basal Lateral (BL), Basal Posterior (BP), Basal Inferior (BI), Basal infero-Septum (BS); the segments from 7 to 12 to the middle segments: Middle Antero-Septum (MAS), Middle Anterior (MA), Middle Lateral (ML), Middle Posterior (MP), Middle Inferior (MI), Middle infero-Septum (MS); the segments from 13 to 16 to the apical segments: Apical Anterior (AA), Apical Lateral (AL), Apical Inferior (AI), Apical Septal (AS).

## 2.1 Subjects

The study was performed under the approval of the review board of “Dipartimento di Scienze Cardiovascolari, Respiratorie, Nefrologiche, Anestesiologiche e Geriatriche, Sapienza Università di Roma”, and in accordance with the ethical guidelines of the Declaration of Helsinki. Written informed consent was obtained from each subject. From April 2012 to May 2014, a total of 67 individuals were enrolled and divided into two main groups (see Table 1 for descriptive statistics). The first group consisted of 45 healthy subjects without a history of hypertension or cardiac disease; they all had normal ECG and blood pressure below 140/90 mmHg. The second group included 22 HCM patients (see Table 1 for 3DSTE details). In all 67 individuals enrolled, echocardiographic examinations were performed by an Aprio-Artida ultrasound system (Toshiba Medical Systems Co, Tochigi, Japan). Full-volume ECG-gated 3D data sets were acquired from apical positions using a 1–4 MHz 3D matrix array transducer to visualize the entire LA in a volumetric image (see [7] and Fig. 1).

Table 1. Descriptive statistics  $\pm$  standard deviation for the groups under study.

Descriptive parameter	Control = 45	HCM = 22
Age	$39.2 \pm 8.4$	$48.2 \pm 12.6$
Sex (males/females)	29/15	14/8
Beat rate	$77.3 \pm 15$	$73 \pm 12$
EDV	$39.67 \pm 14.2$	$90.1 \pm 51.6$
ESV	$20.1 \pm 7.9$	$59.2 \pm 48.6$
EF	$49.3 \pm 0.07$	$39.6 \pm 0.12$
Interventricular septum	$8.55 \pm 1.4$	$18.13 \pm 4$

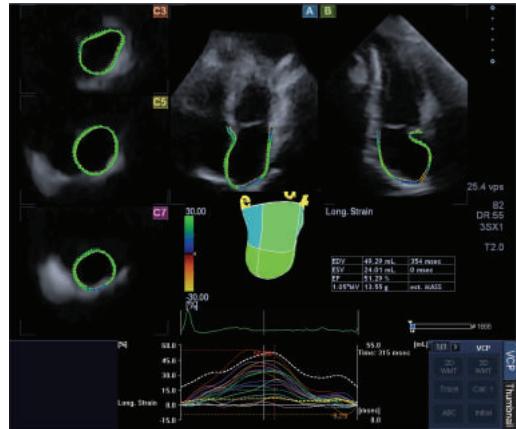


Figure 1. Wall-motion tracking. The position of many speckle patterns is tracked along a cardiac cycle; those data are then processed to evaluate specific characteristic of the overall LA motion. Here, the field of longitudinal strain is displayed in a 5 plane view; color-code scale represents longitudinal strain, positive.

## 2.2 The protocol of measurement

The protocol of measurement is based on a discretisation of the LA analogous to the one adopted for the LV: the LA is divided through 36 planes perpendicular to the longitudinal axis of the LA; each plane is then divided into 36 sections along the circumference; this subdivision gives rise to  $36 \times 36$  points on both the endocardial and the epicardial surfaces of the LA and determines, for any subject, the distance between estimation points.

It has to be noted that endocardial and epicardial surfaces of LA are extremely close one to another; due to it, we expect to find almost the same kind of mechanical information from the analysis of the corresponding data. At the end, the LA geometry is identified by a cloud of  $1296 \times 2$  points i.e. for endocardial and epicardial geometry, whose motion is followed along the cardiac cycle (see Fig. 1).

## 3 ANALYSIS

Here we set the keystones for a mechanics-based analysis of the left chambers, based on the ecocardiographic data collected for healthy subjects and HCM. LA deformations (longitudinal, circumferential and radial) during cardiac revolution have been studied in the Literature [5, 8, 9] via 3DSTE. However, there were relatively few 3DSTE investigations whereby LA function was assessed in both normal and HCM patients.

Interestingly, for any LA of our database, we also have the corresponding LV, and both ecocardiographic data are acquired in the same cardiac cycle. This situation allows us to plan an analysis which can take advantage of the knowledge of the mechanics of both LA and LV chambers of the heart.

### 3.1 Collective behaviour at homologous times

To estimate the collective behaviour of the group, we need firstly to refer to echocardiography frames corresponding to the same mechanically and electrically meaningful times, that is, to homologous times. In Figure 2, we show a typical LA endocardial volume (red line) vs. time pattern so as it appears in an echocardiographic image, whereas black line shows a typical ECG pattern; both of them are associated to any 3DSTE registration. Blue line shows the corresponding LV endocardial volume of the same subject. Black dots on the red and blue curves identify the homologous times, which are used to make physiologically appropriate comparisons among subjects.

#### 3.1.1 Crossed analysis for healthy subjects and HCM patients

The values of LA circumferential and longitudinal strains in our control and HCM groups have been measured through 3DSTE as averages on all the 16 LA segments, at the six homologous times. Then, using these data, we evaluated an average value vs. homologous times for each subject, only considering the middle segments, and represented it in Figures 3 and 4, with the left panels

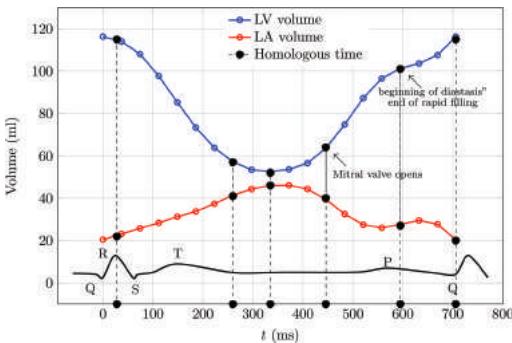


Figure 2. Six electrophysiologically homologous times are identified along the cardiac cycle: R peak, end T wave, ESV, mitral valve opening, onset of diastasis, Q wave. Blue and red lines show a typical time course of endocardial volume for LV and LA, respectively, and black line shows a typical ECG pattern so as they appear in a 3DSTE registration.

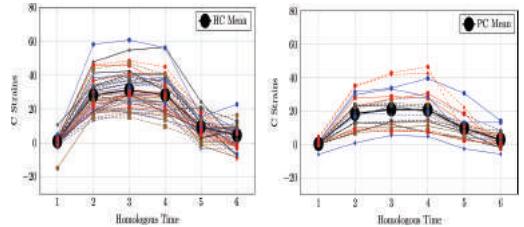


Figure 3. Average values of LA circumferential strains vs. homologous times for each healthy subjects (left) and HCM patients.

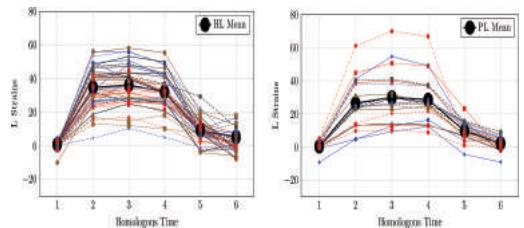


Figure 4. Average values of LA longitudinal strains vs. homologous times for each healthy subjects (left) and HCM patients.

corresponding to the control group and the right panel to the HCM group.

We remark as the intra-group comparisons of the strains vs. time patterns are significative only due to the established correspondence of the same mechanically and electrically meaningful times, which identify the homologous times.

The direct comparison of the mean value of circumferential and longitudinal strains for the control and HCM group is shown in Figure 5.

## 4 CONCLUSIONS AND FUTURE DIRECTIONS

We presented a mechanics-based analysis of LA 3D data echocardiographically obtained from healthy subjects and HCM patients. Interestingly, for any LA of our database, we also had the corresponding LV, and both ecocardiographic data were acquired in the same cardiac cycle. This situation will allow us in the next future to plan a crossed analysis which can take advantage of the knowledge of the mechanics of both LA and LV chambers of the heart as well as to detect the collective behaviour of the two groups of subjects using the S-Lab tool proposed and discussed in (4).

Therein we conjectured that the inflation-induced dilation due to blood pressure is more

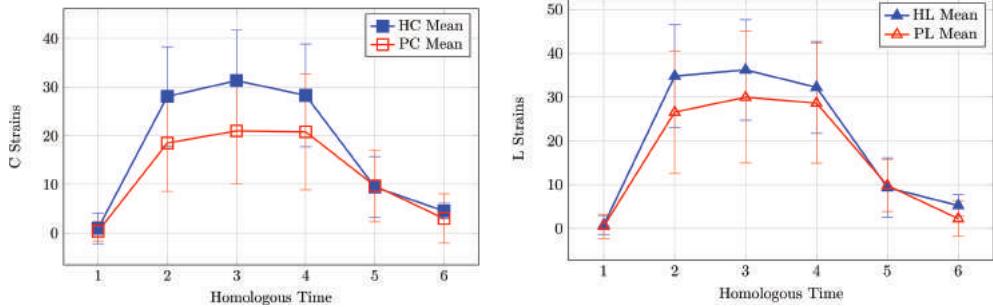


Figure 5. Six electrophysiologically homologous times are identified along the cardiac cycle: R peak, end T wave, ESV, mitral valve opening, onset of diastasis, Q wave.

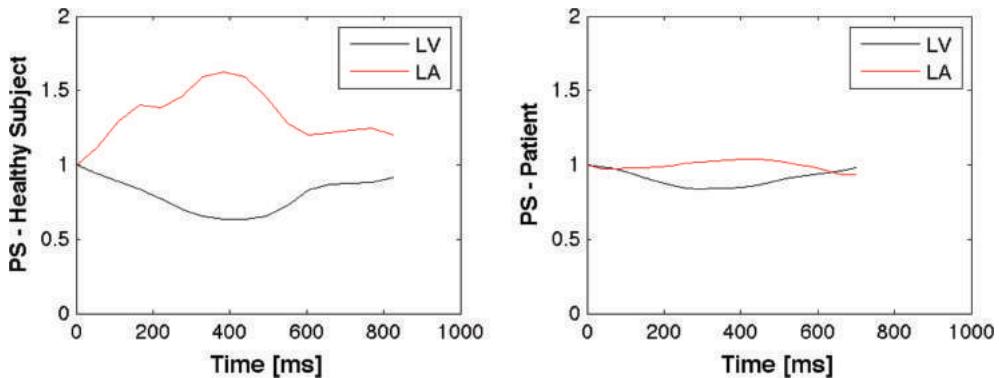


Figure 6. Primary Strains (PS) measured by 3DSTE along the cardiac cycle in one representative healthy subject (left panel) from our series of 45 subjects and in one patient with primary pressure overload (right panel) from our series of 22 HCM patients at endocardial level: it is evident that at end-systolic homologous time (around 400 ms) the patient presents greatly reduced PS in both LV and LA levels, indicating a reduced capacity to contrast intra-chamber pressures, a possible consequence of reduced elasticity due to intrinsic heart muscle pathology typical of HCM. Interestingly enough, in comparison to the representative healthy subject of the left panel, the reduced elasticity indexed by greatly lower PS of the endocardium is dimensionally similar at LV and LA levels in the representative HCM patient shown in the right panel.

effective for subendocardial layers, which dilating reduce the circumferential shortening induced by muscle contraction. Better is the capacity of the elastic response of the endocardial surface, smaller is the dilation induced by blood pressure. This concept is well represented by endocardial Primary Strain (PS) lines and values, which represent at a point of the endocardial surface the directions where the smallest are to be expected. In this sense, we interpreted the primary strain values measured by 3DSTE along the cardiac cycle as shown in Figure 6. our representative healthy subject shows a PS value which denotes a larger shortening than the corresponding value shown by the representative HCM patient. It means that the capacity to contrast blood pressure is reduced in patients with primary pressure overload due to intrinsic heart muscle pathology, hence the PS

values which correspond to the circumferential PSL are smaller.

## REFERENCES

- [1] Piras, P., Evangelista, A., Gabriele, S., Nardinocchi, P., Teresi, L., Torromeo, C., Schiariti, M., Varano, V., Puddu, P.E., 2014. 4D-analysis of Left Ventricular Heart Cycle Using Procrustes Motion Analysis. *PLoS ONE* 9(1): e86896. (doi:10.1371/journal.pone.0086896).
- [2] Evangelista, A., Gabriele, S., Nardinocchi, P., Piras, P., Puddu, P.E., Teresi, L., Torromeo, C., Varano, V., 2014. A comparative analysis of the strain-line pattern in the human left ventricle: experiments vs modeling. *Computer Methods in Biomechanics and Biomedical Engineering/Imaging & Visualization*. Published online: Jun 23, 2014. (doi:10.1080/21681163.2014.927741).

- [3] Gabriele, S., Nardinocchi, P., Varano, V., 2013. Evaluation of the strain-line patterns in a human left ventricle: a simulation study. *Computer Methods in Biomechanics and Biomedical Engineering* 18(7), 790–798.
- [4] Evangelista, A., Gabriele, S., Nardinocchi, P., Piras, P., Puddu, P.E., Teresi, L., Torromeo, C., Varano, V., 2015. Non-invasive assessment of functional strain lines in the real human left ventricle via speckle tracking echocardiography. *Journal of Biomechanics* 48, 465–471.
- [5] Vieira, M.J., Teixeira, R., Goncalves, L., Gersh, B.J., 2014. Left Atrial Mechanics: Echocardiographic Assessment and Clinical Implications. *Journal of the American Society of Echocardiography* 27(5), 463–478.
- [6] Yeung, F., Levinson, S.F., Parker, K.J., 1998. Multilevel and motion model-based ultrasonic speckle tracking algorithms. *Ultrasound Med. Biol.* 24, 427–441.
- [7] Evangelista, A., Nardinocchi, P., Puddu, P.E., Teresi, L., Torromeo, C., Varano, V., 2011. Torsion of the human left ventricle: Experimental analysis and computational modeling. *Progress in Biophysics and Molecular Biology*, 107(1), 112–121.
- [8] Onishi, N., Kawasaki, M., Tanaka, R. et al., 2014. Comparison between left atrial features in well-controlled Hypertensive patients and normal subjects assessed by three dimensional speckle tracking echocardiography. *Journal of Cardiology* 63, 291–295.
- [9] Ancona, R., Comenale Pinto, S., Caso, P. et al., 2014. Left atrium by echocardiography in clinical practice: from conventional methods to new echocardiographic. *Scientif World Journal* 2014, 451042.
- [10] Pedrizzetti, G., Kraigher-Krainer, E., De Luca, A., Caracciolo, G., Mangual, J.O., Shah, A., Toncelli, L., Domenichini, F., Tonti, G., Galanti, G., Sengupta, P.P., Narula, J., Solomon, S., 2012. Functional Strain-line Pattern in the Human Left Ventricle. *Phys. Rev. Lett.* 109, 048103.
- [11] Goffinet, C., Chenot, F., Robert, A., Pouleur, A.C., le Polain de Waroux, J.B., Vancrayenest, D., Gerard, O., Pasquet, A., Gerber, B.L. and Vanoverschelde, J.L., 2009. Assessment of subendocardial vs. subepicardial left ventricular rotation and twist using two dimensional speckle tracking echocardiography comparison with tagged cardiac magnetic resonance. *Eur. Heart J.* 30, 608–617.
- [12] Helle-Valle, T., Crosby, J., Edvardsen, T., Lyseggen, E., Amundsen, B.H., Smith, H.J., Rosen, B.D., Lima, J.A.C., Torp, H., Ihlen, H. and Smiseth, O.A., 2005. New noninvasive method for assessment of left ventricular rotation: speckle tracking echocardiography. *Circulation* 112, 3149–3156.
- [13] Maffessanti, F., Nesser, H.J., Weinert, L., Steringer-Mascherbauer, R., Niel, J., Gorissend, W., Sugeng, L., Lang, R.M. and Mor-Avi, V., 2009. Quantitative evaluation of regional left ventricular function using three-dimensional speckle tracking echocardiography in patients with and without heart disease. *Am. J. Cardiol.* 104, 1755–1762.
- [14] Cerqueira, M.D., Weissman, N.J., Dilsizian, V., Jacobs, A.K., Kaul, S., Laskey, W.K., Pennell, D.J., Rumberger, J.A., Ryan, T., Verani, M.S., 2002. Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart: a statement for healthcare professionals from the Cardiac Imaging Committee of the Council on Clinical Cardiology of the American Heart Association. *Circulation* 105, 539–542.
- [15] Madeo, A., Piras, P., Re, F., Gabriele, S., Nardinocchi, P., Teresi, L., Torromeo, C., Chialastri, C., Schiariti, M., Giura, G., Evangelista, A., Dominici, T., Varano, V., Zachara, E., Puddu, P.E., 2015. A new 4D trajectory-based approach unveils abnormal LV revolution dynamics in hypertrophic cardiomyopathy. *PLoS One* 10(4), e0122376.
- [16] Torromeo, C., Evangelista, A., Pandian, N.G., Nardinocchi, P., Varano, V., Gabriele, S., Schiariti, M., Teresi, L., Piras, P., Puddu, P.E., 2014. Torsional correlates for End Systolic Volume Index in adult healthy subjects. *International Journal of Applied Science and Technology* 4(4), 11–23.

This page intentionally left blank

# A new thick-walled conical model of the Left Ventricle

B. Leśniak-Plewińska, S. Cygan, J. Żmigrodzki & K. Kałużyński

*Institute of Metrology and Biomedical Engineering, Faculty of Mechatronics, Warsaw University of Technology, Warsaw, Poland*

**ABSTRACT:** In this paper, we proposed a numerical left ventricle model with simplified conical geometry. The numerical simulation of the model deformation is carried out using the Yeoh constitutive model, hexahedral elements, and nonlinear geometry (Abaqus CAE, Simulia). To validate the model utility, the displacements and strains of some representative points are compared with clinical data. The proposed computational LV model can simulate the LV wall deformations to some extent. Although, it cannot fully mimic the strain distribution and absolute peak strain values obtained in clinical practice, a physical model approximating the LV shape with a cone and deformed in a dedicated measurement stand will be a profitable source of reliable data for the validation of myocardial strain imaging modalities.

## 1 INTRODUCTION

The heart has been described as a “complex three-dimensional fibre-wound structure with mechanical properties that are nonlinear, anisotropic, time varying and spatially inhomogeneous” (Greenbaum 1981) attributing to the complex multidimensional deformation field in the Left Ventricle (LV), with the apex rotating counterclockwise (when viewed from the apex) relatively to the atrioventricular plane (the base) and the base moving along the long axis (Sengupta et al. 2006).

Characterization of global and regional left ventricular deformation is the target of numerous studies (D’hooge et al. 2002, Jasaitite et al. 2013, Langeland et al. 2004). Most of them consider the entire myocardial wall thickness without distinction between any particular layers of the myocardium. However, based on experimental and clinical observations, transmural heterogeneity of a myocardial mechanical function is commonly accepted. The mechanisms for this are not fully understood. Changes in myocardial electro-mechanical properties, fiber orientations as well as ventricular geometry and shape may be significant. Numerous studies have measured the characteristics of the transmural distribution of myocardial deformation. The focus of the studies has been mainly on the transmural gradient of the longitudinal, radial and circumferential deformation, demonstrating the lowest strain in subepicardial and the highest in subendocardial layer (Adamu et al. 2009, D’hooge et al. 2006, Leitman et al. 2010, McCulloch et al. 1991). Fewer studies concerned the transmural characteristics of the human left ventricular twist (defined here as angular rotation relative to the base of

the heart) and its variation from one LV site to the other. Measurements with implanted markers have demonstrated the existence of both circumferential transmural and longitudinal transmural shear strains (Waldman et al. 1985). Furthermore, the Magnetic Resonance Imaging (MRI)-based measurements have shown that the magnitude of the endocardial twist is significantly higher than the epicardial twist (Buchalter et al. 1990). These findings can result from the structure of the myocardium and the geometry of the LV.

Despite the development of noninvasive myocardial strain imaging techniques, the “ground truth” about the cardiac deformation pattern is still not well known, as all measurement techniques have limitations. Therefore, validation of existing and new strain imaging methods in the *in vivo* settings is very difficult. Models capable of reproducing deformation patterns similar to those encountered in clinical practice may provide reliable data for the development and evaluation of the myocardial strain imaging modalities. With the advances in numerical modeling and simulations as well as non-invasive imaging modalities, patient-specific models of ventricular biomechanics can be created to validate the strain imaging modalities. However, many challenges still have to be overcome to estimate model parameters from available patient data (the heart geometry in unloaded state, the myofibre anatomy, and the material properties on patients (Krishnamurthy et al. 2013).

Although physical LV models have its limitations, an experimental setup and phantom capable of producing deformation patterns similar to those encountered in clinical practice is still a profitable source of reliable data for the validation

of myocardial strain imaging modalities as well as numerical simulations. The physical model and laboratory stand may deliver actual data for FE modeling, that, on the other hand, may be cross-referenced with deformation pattern estimated based on physical experiments.

Various heart phantoms have been used to study and validate cardiac measurements (Claessens et al. 2004, Langeland et al. 2004, Jia et al. (2006), Lesniak-Plewinska et al. 2010, Smith & Rinaldi 1989; Visser et al. 2003, Zhu et al. 2014). These phantoms have many limitations. They are all passively deformed. Moreover, they are either made of inappropriate material, thin-walled or using the experimental conditions (pressure waveform, stroke volume, and heart rate) which are far from physiological. Only one of them can be twisted (Zhu et al. 2014). Unfortunately, the turnable model has a mechanical connection to the apex that makes the apical views unfeasible. Moreover, it has limited applicability to model physiological heart motion—the stress pattern reported for the model bears no analogy to that of the physiological ventricle (Zhu et al. 2014).

The authors have already developed a measurement stand with a dual chamber heart model shaped as thick-walled truncated ellipsoid (Lesniak-Plewinska et al. 2010). Unfortunately, the twisting motion has not been implemented in the system. LV models with simplified geometry usually have cylindrical, spherical or ellipsoidal shape. Azhari et al. (1992) and Chaudhry et al. (1996) reported that a thick conical shell can better approximate the non-uniformity of heart motion, especially the transmural and longitudinal nonuniformity of the twisting motion. Therefore, the authors decided to construct new measurement stand with the LV model simplified to the conical shape.

This work is a part of the ongoing project on the effect of selected echocardiographic data acquisition parameters on accuracy of the myocardial deformation estimation. The paper presents preliminary results of FE simulations aimed to investigate the transmural deformation patterns for model with conical geometry.

## 2 MATERIAL AND METHODS

### 2.1 *LV model*

The LV model was designed as a thick circular-conical shell with two additional fixing elements: a cylindrical collar at the base and circular flange at the apex end (Fig. 1).

The physical model will be made of Poly(vinyl alcohol) hydrogel (PVA) prepared by freezing-thawing cyclic processing. The strain-stress relation for the PVA cryogel is nonlinear. To obtain a reliable material model of the PVA cryogel, mechanical

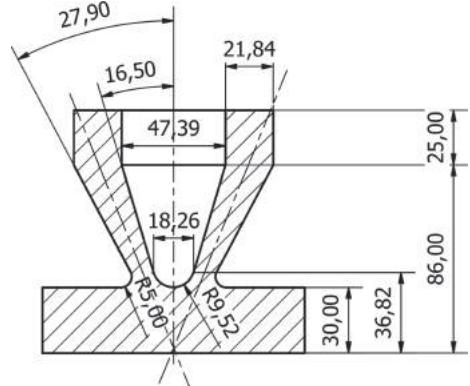


Figure 1. Model geometry and dimensions (in mm).

examinations were carried out. A cylindrical sample with 34.26 mm in diameter and 28.85 mm in height was made from a liquid solution of 10 mass% (m%) PVA (Sigma Aldrich, St. Louis, USA) and 10 m% glycerin (Chempur, Piekary Śląskie, Poland). This solution was physically solidified in a laboratory freezer (ZLN 74, Pol-Eko, Poland) by two cycles of freezing at  $-25^{\circ}\text{C}$  for 24 hours and thawing for the next 24 hours. The mechanical examinations were carried out on an MTS QTest/10 material testing machine (MTS, USA) with deformation ranging up to 30% of the original sample size.

### 2.2 *Simulation*

Numerical model of the phantom geometry (Fig. 1), created using the Autodesk Inventor 2012 software (Autodesk Inc., USA), was exported in CAD format to the FE simulation software Abaqus 6.13 (Simulia, USA). The model was meshed with the seed (maximum element edge length) of 2 mm, which resulted in 69890 linear hexahedral elements of type C3D8H, defined by 76347 nodes.

A fixed cylindrical coordinate system was created by specifying the origin in the center of the circular surface of the flange (point O), Z-axis running along the main symmetry axis of the model and the R-axis perpendicular to the Z-axis (Figs. 2 and 4).

The strain-stress relationship of PVA cryogel was modeled with 3-rd order Yeoh strain energy potential definition as an isotropic hyperelastic material based on strain-stress relation derived from the experiments and with Poisson's ratio of 0.45. Evaluation of the chosen material constitutive model has been carried out using built-in functions of Abaqus software, providing the model and verifying the stability of the material model.

Deformation of the conical part was forced by applying: (i) pressure load to the internal surface

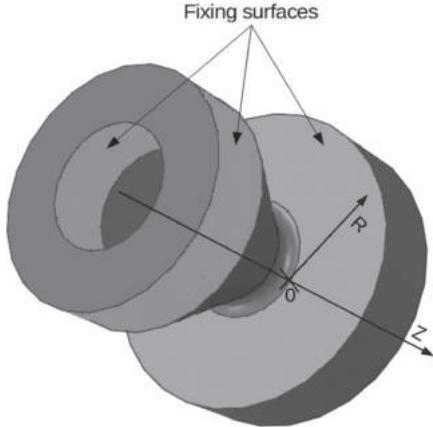


Figure 2. Coordinate system and fixing surfaces.

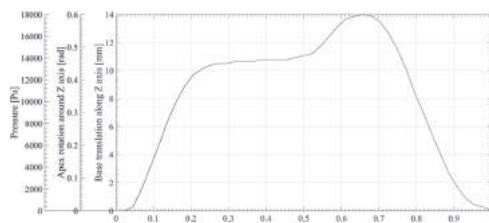


Figure 3. Loads time curve: base translation along Z axis (in mm), apex rotation around Z axis (in radians), and pressure (in Pa).

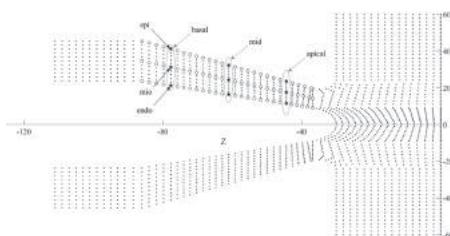


Figure 4. Nodes of the mesh for the plane of symmetry in cylindrical coordinates. White circle—nodes used to calculate twist along the model wall. Thick black dots—nodes selected as corresponding to particular layers in segments of myocardium e.g. epi-, mio-, and endocardial layers in basal, mid and apical segment.

of the this part, (ii) base translation along main axis of the model symmetry (Z axis) by linear shift of rigid tubes glued to the inner and outer surface of the cylindrical collar (Fig. 2), and (iii) apex rotation around main axis of the model symmetry (Z axis) by rotation of one rigid ring glued to the circular surfaces of the flange (Fig. 2). The loads

varied in time, and their value changes were represented by the same relative amplitude curve that shape has been derived from the LV volume curve (Konturek 2013) (Fig. 2). The maximal amplitudes of the loads were: 18000 Pa (pressure), 14 mm (base translation along Z axis), 0.6 rad (apex rotation around Z axis).

The simulation was carried out in steps using non-linear geometry option turned on (NLgeom) since the process includes large deformations.

As the result of FE simulation, values of nominal strains at all nodes in cylindrical coordinate systems were returned.

According to the Abaqus documentation the nominal strain is defined as:

$$\varepsilon^N = \mathbf{V} - \mathbf{I} = \sum_{i=1}^3 (\lambda_i - 1) \cdot \mathbf{n}_i \cdot \mathbf{n}_i^T \quad (1)$$

where  $\mathbf{V} = \sqrt{(\mathbf{F} \cdot \mathbf{F}^T)}$  is the left stretch tensor,  $\lambda_i$  are the principal stretches, and  $\mathbf{n}_i$  are the principal stretch directions in the current configuration. The principal values of nominal strain are the ratios of change in length to length in the reference configuration in the principal directions (Abaqus 6.13 documentation, Abaqus Analysis User's Manual, 1.2.2 Conventions).

A set of nodes lying in the plane of symmetry of model in the unloaded state was selected (Fig. 4) as corresponding to particular layers in segments of myocardium e.g. epi-, mio-, and endocardial layers in basal, mid and apical segment.

Three nominal strain components in cylindrical coordinate system: axial, circumferential, and radial as well as twist in the nine selected were exported from Abaqus for further processing in Matlab 2014b (Mathworks, USA). Angular displacements of nodes lying on the outer (epi), middle (mio) and inner (endo) conical surface were used to calculate the twist as difference in rotation magnitude between the given node and the corresponding basal node.

### 3 RESULTS

The time curves of the radial, circumferential, and axial strain components resulting from FE simulations for nodes corresponding to the epi-, mio-, and endocardial layers in basal, mid, and apical segments are presented in Figure 5.

A transmural gradient may be seen for circumferential and radial strain components with the lowest values at the epicardial and highest at the endocardial nodes for each segment (basal, mid, and apical) during the entire cycle.

Axial strain component increases from the epicardial to the endocardial layer only in the case

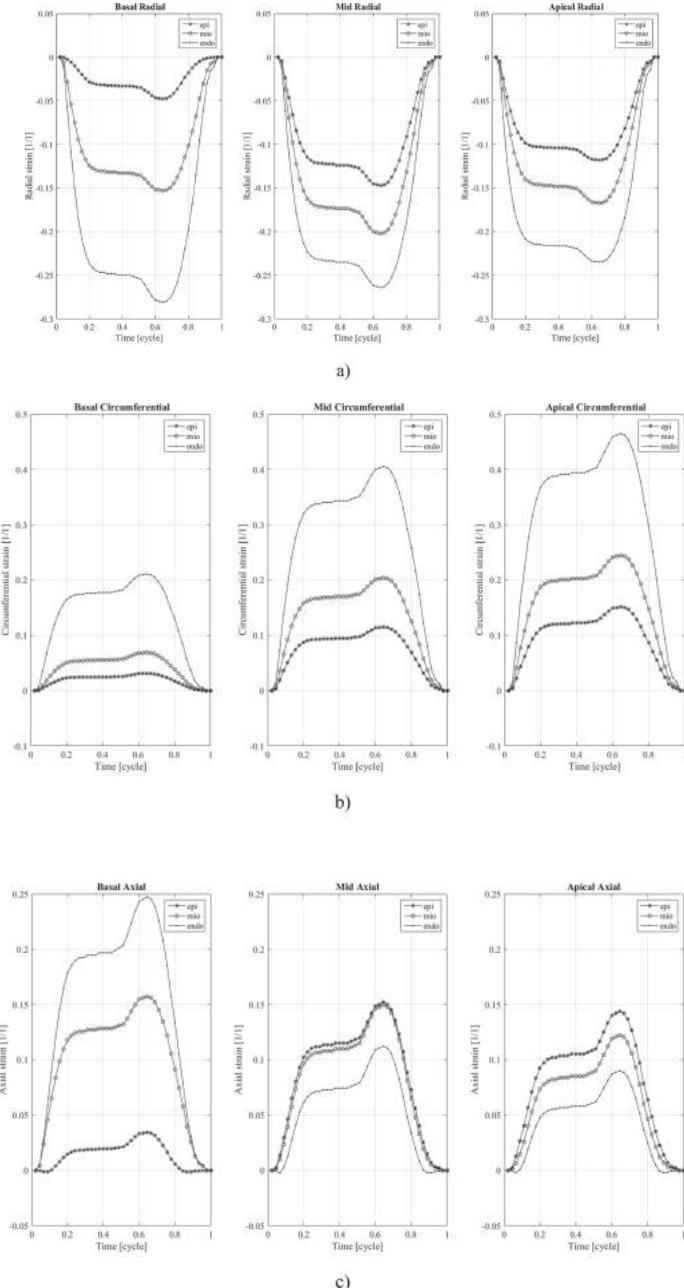


Figure 5. Time curves of radial (a), circumferential (b), and axial (c) strain components resulting from FE simulations for nodes corresponding to epi- (\*), mio- (○), and endocardial (●) layers in basal (left), mid (central), and apical (right) segments. Abscissa normalized to one cycle, ordinate unitless.

of apical segment. The opposite trend may be observed for the basal segment. In the mid segment, axial strain component for the epi- and myocardial layers are almost identical and higher than for the endocardial layer.

The twist along the model wall for the nodes located at the outer (epi) and inner (endo) surface of the conical part is presented in Figure 6.

A twist gradient may be observed across the model wall with twist values for the endocardial layer higher

that for the epicardial one over the entire length of the LV model wall. The twist gradient increases from the base towards the apex of the model (Fig. 6).

The maximal values of the nominal strain components obtained in the selected nodes and the twist for basal nodes are summarized in Table 1.

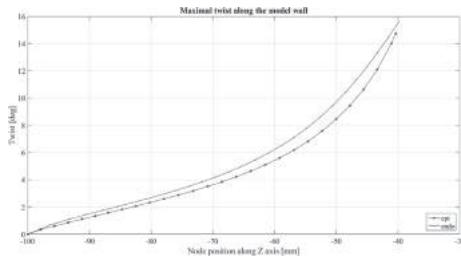


Figure 6. The maximal twist along the *in silico* model wall for the nodes located at the epi- (-\*) and endocardial (-•) layer.

Table 1. Maximal absolute values of the nominal strain components and twist values obtained in the selected nodes.

Strain component [%]	Wall segment	FE simulation		
		Epi	Mio	Endo
Axial	Basal	3.4	15.7	24.7
	Mid	15.2	14.9	11.2
	Apical	14.4	12.2	9.0
Circumferential	Basal	3.1	6.9	21.1
	Mid	11.5	20.4	40.5
	Apical	15.2	24.5	46.5
Radial	Basal	4.8	15.3	28.1
	Mid	15.7	20.2	26.4
	Apical	11.8	16.8	23.5
Twist [°]		10.6	11.0	12.3

## 4 CONCLUSIONS

In this study, we have shown that the proposed computational LV model can simulate the LV wall deformations to some extent.

Comparing the obtained results to that obtained in clinical practice with the absolute peak strain values summarized in Table 2, the main differences were observed between the *in situ* axial and the *in vivo* longitudinal component. This may result from the fact that in clinical applications the longitudinal component is calculated as tangential to the wall curvature while the axial component is calculated as parallel to model main symmetry axis (axis Z).

The *in silico* absolute peak values of circumferential and radial strain components for basal segment are much lower than the results reported by physicians, what may be explained by the fact that the outer and inner surfaces of the fixing flange where immobilized. At the same time, in the case of mid and apical segments good quantitative agreement was found between the numerical model and the clinical reports.

In addition, the proposed conical model exhibits the transmural twist gradient increasing from base towards apex, with the endocardial twist exceeding the epicardial one, what is in agreement with the other studies (Buchalter et al. 1990).

Although, the proposed model cannot fully mimic the strain distribution and absolute peak strain values obtained in clinical practice, a physical model approximating the LV shape with a cone and deformed in a dedicated measurement stand will be a profitable source of reliable data for the validation of myocardial strain imaging modalities. However, further investigations are necessary to optimize the way the physical model will be fastened in the measurement stand, as it seems to be one of the major reasons for the non-compliance with the clinics.

Table 2. Maximal absolute strain and twist values obtained in clinical practice.

Strain component [%]	Wall segment	Clinical data*		
		Epi	Mio	Endo
Longitudinal	Basal	$19.2 \pm 3.1$	$19.8 \pm 3.2$	$20.3 \pm 4.8$
	Mid	$19.4 \pm 3.0$	$21.3 \pm 3.1$	$23.5 \pm 3.6$
	Apical	$18.9 \pm 3.9$	$25.0 \pm 4.5$	$33.5 \pm 5.3$
Circumferential	Basal	$15.6 \pm 6.7$	$22.0 \pm 8.0$	$32.4 \pm 8.4$
	Mid	$19.4 \pm 6.4$	$21.3 \pm 6.1$	$23.5 \pm 9.1$
	Apical	$18.9 \pm 7.6$	$25.0 \pm 4.9$	$33.5 \pm 11.2$
Radial	Basal	$39.0 \pm 12.9$	—	$47.8 \pm 16.9$
	Mid	—	—	—
	Apical	$32.6 \pm 12.5$	—	$39.4 \pm 16.4$
Twist [°]		$8.4 \pm 3.7$	$9.7 \pm 4.1$	$12.5 \pm 5.7$

\*Longitudinal, Circumferential (Leitman et al. 2010), Radial (Chan-Dewar et al. 2010), Twist (Goffinet et al. 2009).

## ACKNOWLEDGEMENTS

This study was supported by the Norwegian Research Programme operated by the National Centre for Research and Development in Poland under the Norwegian Financial Mechanism 2009-14 (Project Contract No: Pol-Nor/209584/19/2013).

## REFERENCES

- Adamu, U., Schmitz, F., Becker, M., Kelm, M., Hoffmann, R. 2009. Advanced speckle tracking echocardiography allowing a three-myocardial layer-specific analysis of deformation parameters. *Eur J Echocardiogr* 10: 303–308.
- Azhari, H., Buchalter, M., Sideman, S., Shapiro, E., Beyar, R. 1992. A conical model to describe the nonuniformity of the left ventricular twisting motion. *Ann Biomed Eng* 20(2): 149–65.
- Buchalter, M.B., Weiss, J.L., Rogers, W.J., Zerhouni, E.A., Weisfeldt, M.L., Beyar, R., Shapiro, E.P. 1990. Noninvasive quantification of left ventricular rotational deformation in normal humans using magnetic resonance imaging myocardial tagging. *Circulation*. 81(4): 1236–44.
- Chan-Dewar, F., Oxborough, D., Shave, R., Gregson, W., Whyte, G., George, K. 2010. Left ventricular myocardial strain and strain rates in sub-endocardial and sub-epicardial layers before and after a marathon *Eur J Appl Physiol*. 109(6): 1191–6.
- Chaudhry, H.R., Bukiet, B. and Davis, A.M. 1996. Stresses and strains in the left ventricular wall approximated as a thick conical shell using large deformation theory. *J Biol Syst.* 4: 353–72.
- Claessens, T., Segers, P., Verdonck, P. 2004. A new thick-walled hydraulic model of the left heart for the assessment of blood-wall interaction using ultrasound. *IEEE Proc Computers in Cardiology* 31: 333 6.
- D'hooge, J., Konofagou, E., Jamal, F., Heimdal, A., Barrios, L., Bijnens, B., Thoen, J., Van de Werf, F., Sutherland, G., Suetens, P. 2002 Two-dimensional ultrasonic strain rate measurement of the human heart *in vivo* *IEEE Trans UFFC*. 49: 281–6.
- D'hooge, J., Missant, C., Claus, P., Rex, S., Eroglu, S., Rademakers, F.E., Wouters, P. 2006. Gradients in Radial Strain Across the Cardiac Wall Studied with Epicardial Ultrasound *IEEE Ultrasonics Symposium 2–6 Oct. 2006*. Vancouver, BC.
- Goffinet, C., Chenot, F., Robert, A., Pouleur, A-C., le Polain de Waroux, J-B., Vancrayenest, D., Gerard, O., Pasquet, A., Gerber, B.L., Vanoverschelde, J-L. 2009. Assessment of subendo- cardial vs. Subepicardial left ventricular rotation and twist using twodimensional speckle tracking echocardiography: comparison with tagged cardiac magnetic resonance *Eur Heart J* 30: 608–17.
- Greenbaum, R.A., Ho, S.Y., Gibson, D.G., Becker, A.E., Anderson, R.H. 1981. Left ventricular fiber architecture in man. *Brit Heart J.* 45(3): 248–63.
- Jasaityte, R., Heyde, B., D'Hooge, J. 2013. Current state of three-dimensional myocardial strain estimation using echocardiography *J Am Soc Echocardiog*. 26(1): 15–28.
- Jia, C., Kim, K., Kolias, T.J., Rubin, J.M., O'Donnell, M. 2006. Left ventricular phantom with pulsatile circulation for ultrasound strain rate Imaging. *IEEE Ultrasonics Symposium 2–6 Oct. 2006*. Vancouver, BC.
- Konturek, S. 2013 Konturek Fizjologia człowieka. Podręcznik dla studentów medycyny Elsevier Urban & Partner. Wrocław.
- Krishnamurthy, A., Villongco, C.T., Chuang, J., Frank, L.R., Nigam, V., Belezzuoli, E., Stark, P., Krummen, D.E., Narayan S., Omens, J.H., McCulloch, A.D., Kerckhoffs, R.C.P. 2013. Patient-specific models of cardiac biomechanics. *J Comput Phys*. 244: 4–21.
- Langeland, S., D'hooge, J., Claessens, T., Claus, P., Verdonck, P., Suetens, P., Sutherland, G.R., Bijnens, B. 2004 RF-based two-dimensional cardiac strain estimation: a validation study in a tissue-mimicking phantom *IEEE Trans. UFFC*. 51: 1537–46.
- Leitman, M., Lysiansky, M., Lysiansky, P., Friedman, Z., Tyomkin, V., Fuchs, T., Adam, D., Krakover, R., Vered, Z. 2010. Circumferential and Longitudinal Strain in 3 Myocardial Layers in Normal Subjects and in Patients with Regional Left Ventricular Dysfunction *J Am Soc Echocardiog* 23(1): 64:70.
- Lesniak-Plewińska, B., Cygan, S., Kaluzynski, K., D'hooge, J., Zmigrodzki, J., Kowalik, E., Kordybach, M., Kowalski, M. 2010. A dual-chamber, thick-walled cardiac phantom for use in cardiac motion and deformation imaging by ultrasound. *Ultrasound Med Biol.* 36(7): 1145–56.
- McCulloch, A.D., Omens, J.H. 1991. Non-homogeneous analysis of 3-dimensional transmural finite deformation in canine ventricular myocardium. *J Biomech* 24(7): 539–48.
- Sengupta, P.P., Korinek, J., Belohlavek, M., Narula, J., Vannan, M.A., Jahangir, A., Khandheria, B.L. 2006. Left ventricular structure and function: basic science for cardiac imaging. *J Am Coll Cardiol.* 48: 1988–2001.
- Smith, S.W., Rinaldi, J.E. 1989. Anthropomorphic cardiac ultrasound phantom. *IEEE Trans BME* 10: 1055–8.
- Visser, J.J., Sokole, E.B., Verberne, H.J., Habraken, J.B., van de Stadt, H.J., Jaspers, J.E., Shehata, M., Heeman, P.M., van Eck-Smit, B.L. 2003. A realistic 3-D gated cardiac phantom for quality control of gated myocardial perfusion SPECT: the Amsterdam gated (AGATE) cardiac phantom. *Eur J Nuclear Med Molecular Imag* 31: 222–8.
- Waldman, L.K., Fung, Y.C., Covell, J.W. 1985. Transmural myocardial deformation in the canine left ventricle. Normal *in vivo* three-dimensional finite strains. *Circ Res.* 57(1): 152–63.
- Zhu, Y., Luo, X.Y., Gao, H., McComb, C., Berry, C. (2014) A numerical study of a heart phantom model, *International J Comput Math.* 91: 7, 1535–1551.

## Augmented Reality in radiofrequency ablation of the liver tumors

Lucio Tommaso De Paolis & Francesco Ricciardi

*Department of Engineering for Innovation, University of Salento, Lecce, Italy*

**ABSTRACT:** Minimally Invasive Surgery is a surgery technique that provides evident advantages for the patients, but also some difficulties for the surgeons. In medicine, the Augmented Reality (AR) technology allows surgeons to have a sort of “X-ray” vision of the patient’s body and can help them during the surgical procedures. In this paper we present an application of Augmented Reality that could be used as support for a more accurate preoperative surgical planning and also for an image-guided surgery. The AR can support the surgeon during the needle insertion for the Radiofrequency Ablation of the liver tumours in order to guide the needle and to have an accurate placement of the surgical instrument within the lesion. The augmented visualization can avoid as much as possible to destroy healthy cells of the liver.

### 1 INTRODUCTION

The actual trend in surgery is the transition from open procedures to minimally invasive interventions. Using this surgery technique visual feedback to the surgeon is only possible through the laparoscopic camera. Furthermore direct palpation of organs is not possible. Minimally Invasive Surgery (MIS), such as laparoscopy or endoscopy, has changed the way to practice the surgery. It is a promising technique and the use of this surgical approach is nowadays widely accepted and adopted as a valid alternative to classical procedures.

The use of MIS offers to surgeons the possibility of reaching the patient’s internal anatomy in a less invasive way. This reduces the surgical trauma for the patient. The diseased area is reached by means of small incisions made on the patient body called ports. Specific instruments and a camera are inserted through these ports. The surgeon uses a monitor to see what happens in the surgical field inside the patient body.

With the use of MIS the patient has a shorter hospitalizations, faster bowel function return, fewer wound-related complications and a more rapid return to normal activities. These benefits have contributed to accept these surgical procedures. If the advantages of this surgical method are evident on the patients, this technique has some limitations for the surgeons. First of all, the imagery is in 2D and with this limitation the surgeon needs to develop new skills and dexterity in order to estimate the distance from the anatomical structures. In addition, we have to consider that he had to work in a very limited workspace. Some modern systems use two cameras to offer a 3D view to the surgeon but this technology is useful when the working volume is not too small.

Medical images (CT or MRI) associated to the latest medical image processing techniques could provide an accurate knowledge of the patient’s anatomy and pathologies. This information can be used to guide surgeons during the surgical procedure in order to improve the patient care.

The Augmented Reality (AR) technology has the potential to improve the surgeon’s visualization during a MIS surgical procedure. AR technology refers to a perception of a physical real environment whose elements are merged with virtual computer-generated objects in order to create a mixed reality. The merging of virtual and real objects in an AR application has to run in real time. Virtual objects have to be aligned (registered) with real world structures. Both of these requirements guarantee that the dynamics of real world environments remain unchanged after virtual data has been added.

In medicine, the use of AR technology makes possible to overlay virtual medical images of the organs on the real patient. This allows the surgeon to have a sort of “X-ray vision” of the patient’s internal anatomy. This technology can “augment” surgeon’s perception of the real world with the use of information gathered from patient’s medical images. The use of AR in surgery produces a better spatial perception and a reduction in the duration of the surgical procedure.

In order to obtain a correct alignment of virtual and real objects is needed a registration procedure that can be accomplished using an optical tracker.

Many research groups are now focusing on the development of systems that assists surgeons during the treatment of the liver tumours.

Nicolau et al. present a real-time superimposition of virtual models over the patient’s abdomen in order to have a three dimensional view of the internal anatomy. The authors have used the

developed system in an operating room and to reduce the influence of the liver breathing motion they have tried to simulate the patient's breathing cycle (Nicolau et al. 2009).

LiverPlanner is a virtual liver surgery planning system developed at Graz University of Technology that combines image analysis and computer graphics in order to simplify the clinical process of planning liver tumour resections. The treatment planning stage enables the surgeon to elaborate a detailed strategy for the surgical intervention and the outcome of pre-operative planning can then be used directly for the surgical intervention (Reitinger et al. 2004).

Maier-Hein et al. present a system developed for computer-assisted needle placement that uses a set of fiducial needles to compensate for organ motion in real time; the purpose of this study was to assess the accuracy of the system *in vivo* (Maier-Hein et al. 2008).

Stüdeli et al. present a system that provides surgeon, during placement and insertion of RFA needle, with information from pre-operative CT images and real-time tracking data (Stüdeli et al. 2008).

The aim of this paper is to present an AR system that could be used as support for a more accurate surgical preoperative planning and also for image-guided surgery. The application can support the surgeon during the needle insertion in the radiofrequency ablation of the liver tumours.

## 2 3D MODEL OF THE ORGANS

A reconstruction of the 3D model of the anatomical and pathological structures of the patient is required in order to improve the standard slice view. An efficient 3D reconstruction of the patient's organs is generated by applying some segmentation and classification algorithms to medical images (CT or MRI). The grey levels in the medical images are replaced by colors associated to the different organs (Yoo 2004).

Nowadays there are different software used in medicine for the visualization and the analysis of scientific images and the 3D modeling of human organs; Mimics ([www.materialise.com](http://www.materialise.com)), 3D Slicer ([www.slicer.org](http://www.slicer.org)), ParaView ([www.paraview.org](http://www.paraview.org)), and OsiriX ([www.osirix-viewer.com](http://www.osirix-viewer.com)) play an important role among these tools.

In our application we have used 3D Slicer for the building of the 3D model of the patient's organ; 3D Slicer is a multi-platform open-source software package for visualization and image analysis.

In the developed AR platform it is also necessary to use an optical tracker that is able to detect some intentionally introduced retro-reflective spheres (fiducials) placed on the surgical tools. In this way

it is possible to know the right position and the orientation of the surgical instruments used during the surgical procedure. The tracking system is also used in order to permit the overlapping of the virtual organs on the real ones providing an augmented visualization of the scene.

## 3 AR FOR THE RF ABLATION OF THE LIVER TUMORS

Hepatic cancer is one of the most common solid cancers in the world. Hepatocellular Carcinoma (HCC) is the most common primary hepatic cancer.

The Liver Radiofrequency Ablation (RFA) is a minimally invasive treatment for liver cancer used since 1980's. It consists in the placement of a needle inside the liver parenchyma in order to reach the centre of the tumour lesion. When the lesion center is reached, an array of electrodes is extracted from the tip of the needle and it is expanded in the tumour tissue. From these electrodes is injected in the tumour tissue a radiofrequency current that causes tumour cell necrosis for hyperthermia (the local temperature is higher than 60°C and cancer cells are more sensitive to heat than normal cells).

One problem in using radiofrequency tumour ablation technique is the correct placement of the needle. To ensure a maximum efficacy of the treatment the needle has to reach the centre of the tumour.

Today surgeons use ultrasound, CT or MNR images acquired during the needle placement in order to correctly direct the needle towards the centre of the tumour. The use of these two-dimensional images makes the insertion procedure very difficult and requires sometimes more than one insertion.

In addition, the surgeon, in order to destroy all tumoral cells, applies the RFA on an extended area of the liver. In this way a large number of healthy cells are also destroyed. This practice can cause to the patient a number of other different consequences. To reduce this problem is of primary importance to reach the centre of the tumour.

A guidance system of the needle in tumour ablation procedures can be obtained using Augmented Reality technology. The purpose of this AR application is to provide a guidance system that can help the surgeon during the needle insertion in liver RFA. The position and orientation of the ablation tool are measured using some reflective spheres that are detected by an optical tracker in order to measure in real time position and orientation of the real ablation tool.

With the superimposition of the virtual models of the patient's anatomy (liver, vessels, biliary ducts, cancer, etc.) exactly where are the real ones, it is possible to make the needle placement task

less difficult. In this way the surgeon has a sort of x-ray 3D vision of the patient internal anatomy.

To achieve a correct augmentation it is necessary to have a perfect correspondence between the virtual organs and the real ones. This is very important in an image-guided surgery application because a very small error in the registration phase can cause serious consequences on the patient due to a bad alignment of the virtual organs on the real ones.

The registration phase is one of the most delicate steps in an AR system for surgery and is obtained using some fiducial points. These points have been defined on the patient body and identified on the patient tomographic images. In this way the application can establish the transformation between the “real” world and “virtual” one.

If the patient position can change during surgery it is necessary to fix on the patient body a tracker tool that permits to estimate its new positions and orientations.

#### 4 DEVELOPED APPLICATION

The application is provided of an user interface designed to be simple and functional at the same time. In the left side of that interface we placed the application control panel. On the right-top window we show 3D models and the augmented reality scene. On the right-bottom there are the three smaller windows where axial, coronal and sagittal views of CT dataset are placed.

The application offers to surgeon the possibility to study the case study before going in operating room. There is the possibility to apply a clipping modality that permits the surgeon to dissect the model and study its internal structure changing the opacity of the organs. The dissection could be made along the three principal axes of CT images.

The application features described till now are part of what we consider the pre-operative planning task. During this task the surgeon can use the application to study the pathology in a more simple and natural way than that provided by simple CT slice visualization.

For the navigation and augmentation task are devoted to the surgery room. Here the surgeon needs to use the optical tracker and to carry out the registration task. When the registration process is complete, a virtual ablation tool is shown in 3D view. It is coupled with the real ones and follows its movements.

In the developed AR platform we use an optical tracker. This tool is able to detect some retro-reflective spheres intentionally introduced in the surgical scene. These spheres are placed on the surgical tools. They provide within a defined coordinate system the real-time spatial measurements of

the location and orientation of the surgical instruments used during the surgical procedure.

In the first prototype of the system we have used the Polaris Vicra optical tracker, a commercial device composed from two IR cameras that provides precise and real-time spatial measurements of the location and orientation of an object or tool within a defined coordinate system (NDI [www.ndigital.com](http://www.ndigital.com)). After the test in the operating room we changed the tracking system to ensure surgeons more freedom of movement. For this reason we decided to use the Bonita Vicon tracking system (VICON [www.vicon.com](http://www.vicon.com)). In particular, we used four IR cameras positioned at the four edges of the operating bed. A specific library has been developed, tested and integrated inside IGSTK framework for interfacing with the Bonita Vicon tracker.

At the base of the functioning of IGSTK there is the use of a state machine that allows increasing the safety and robustness of the toolkit (Cleary 2009).

The use of a state machine, in fact, allows limiting and controlling the possible behaviours of application in order to ensure that it is always in a formal state planned in the design phase. This guarantees a reproducible and deterministic behaviour that eliminates the risk of application design formal errors.

When a user sends a request for an action through a call of a public method, this request is translated into an input for the state machine. The state machine taking into account the current state and the sent input changes its state as expected in the design phase. In any case, the object will always be in a known state because each type of behaviour has been programmed.

The IGSTK tracker component communicates with the tracker to obtain the position and orientation relative to each tool that is present in the acquisition volume.

#### 5 APPLICATION TESTING

The first test of application was made on a dummy and we tried to overlap the 3D models of a real patient’s organs on the dummy. This test was not successful because there was a misalignment between the dummy and the virtual models due to a different thoracic girth between patient and dummy. In Figure 1 is shown an augmented visualization over a dummy.

To overcome this problem we decided to make a preliminary qualitative test of application in the operating theatre. This test is designed to evaluate the application uncertainty and also to highlight all the possible issues related to a live use of the developed platform in the operating room.

The test was carried out on a patient operated in open surgery with the tumour located on the

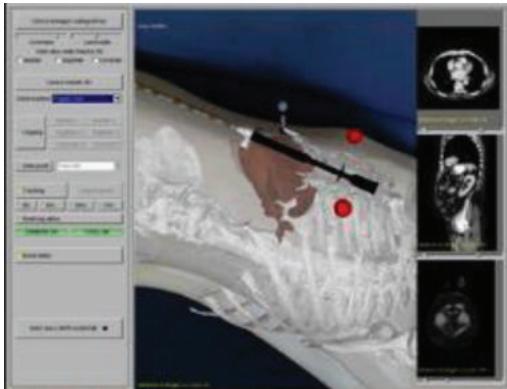


Figure 1. Augmented visualization over a dummy.

surface of the liver; in that surgical case the surgeon was able to see the correct position of the tumour and touch it.

Before a traditional needle insertion, we carried out the registration phase in order to overlap the virtual organs on the real ones. The surgeon verified that a correct overlap with tumour and real organs was obtained.

The test doesn't add risks for the patient, if compared with traditional RFA surgery, because the application was not used to guide the needle insertion. Anyway the patient was informed on the nature of the experiment and signed an informed consent form.

## 6 CONCLUSIONS AND FUTURE WORK

In this paper we present a guidance system for needle placement in radiofrequency ablation of the liver tumour.

The application has been tested previously on a dummy and afterwards a first test has been carried out in the operating room during an open surgery procedure for the liver tumour resection.

The results of the laboratory test were not as expected because there was a difference in thoracic girth between patient and dummy.

We decided to move to the operating room to test the application in a qualitative manner on a real patient. The test in the operating room has also the aim to highlight all the possible issues of the augmented visualization in the operating scenario and to evaluate the correct overlap of virtual models of the organs on real ones.

The test in the operating room produced interesting results for all the aspects related to the constraints of the particular environment. Operating rooms are overloaded of systems and devices and an efficient use of space is mandatory. In this first

test the correct position of the devices close to the operating table has been defined and a precise definition of the more appropriate fiducial points used for the registration phase has been decided.

Surgeons judged that the volume of interaction of the NDI Vicra optical tracker was not sufficient. For this reason we decided to use the VICON Bonita tracker in order to obtain a bigger volume of interaction. A specific library was developed and tested for interfacing with this new tracker. This library was integrated within the IGSTK framework. This new tracker should be tested in the laboratory to evaluate its measure uncertainty and the in the operating room to evaluate if the new volume of interaction can be sufficient for the surgeons.

Anyway surgeons were excited for the obtained results and for the potentiality of the augmented visualization in terms of precision of the needle placement.

We are planning to design a quantitative test in operating room to evaluate and measure the uncertainty of the needle placement. We are also taking into account the possibility to include in the developed system the simulation of the virtual model deformations due to the breathing of the patient.

## REFERENCES

- Cleary, K., P. Cheng, A. Enquobahrie, Z. Yaniv, "IGSTK: The Book", May 29, 2009.
- Maier-Hein, L., A. Tekbas, A. Seitel, et al., "In vivo accuracy assessment of a needlebased navigation system for CT-guided radiofrequency ablation of the liver", Medical Physics, 2008, Vol. 35, No. 12, 5385-5396, 0094-2405.
- Mimics Medical Imaging Software, Materialise Group, <http://www.materialise.com>.
- Nicolau, S., A. Garcia, X. Pennec, L. Soler, X. Buy, A. Gangi, N. Ayache, and J. Marescaux, "An augmented reality system for liver thermal ablation: Design and evaluation on clinical cases", Elsevier, 2009.
- NDI Polaris Vicra, <http://www.ndigital.com>.
- OsiriX Imaging Software, <http://www.osirix-viewer.com>.
- ParaView, <http://www.paraview.org>.
- Reitinger, B., A. Bornik, R. Beichel, G. Werkgartner, and E. Sorantin, "Tools for augmented reality based liver resection planning", Proceedings of the SPIE Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display, pages 88-99, San Diego, February 2004.
- Stüdeli, T., D. Kalkofen, P. Risholm, et al., "Visualization tool for improved accuracy in needle placement during percutaneous radio-frequency ablation of liver tumors", Medical Imaging 2008: Visualization, Image-Guided Procedures, and Modeling, Pts 1 and 2, Vol. 6918, B9180-B9180, 0277-786X.
- Slicer, <http://www.slicer.org>.
- VICON, Bonita, <http://www.vicon.com/products/bonita.html>.
- Yoo, T.S., "Insight into Images: Principles and Practice for Segmentation, Registration, and Image Analysis", A.K. Peters, Ltd, 2004.

## Anthropomorphic robot forefinger virtually simulated

H.J. Rabiela, B.V. González & G.D.A. Miranda

*Metropolitan Autonomous University, Azcapotzalco, México City, México*

**ABSTRACT:** In this paper the virtual simulation of an anthropomorphic robot forefinger, analyzed and synthesized as a four-link mechanism open chain is proposed. The metacarpal bone is considered fixed; the proximal, medial and distal phalanges are links in movement of said mechanism. The five degrees of freedom of the system are reproduced without interference between the links at the corresponding kinematic pairs and using a unique angular parameter to generate the different geometries. The simulation is parameterized using a spreadsheet and graphic soft ware. After describing the advantages of virtual simulation their results are presented, which show the possibility of studying the movement of the system in different circumstances without higher costs. Additionally, application possibilities are shown in the care of diseases of the fingers.

### 1 INTRODUCTION

In recent years, technological development has made possible a new era in the progress of science. The consolidation of robotic design is an example. In the last century, the idea of building a mechanical system capable of emulating the movement of parts of the human body has changed from dream to reality; progress in creating new mechanisms that mimic human locomotion, like: Shadow Robot Company, Robonaut Arm, Robonaut Hand, Dexhand, LWR Hand and others it show. Virtual reality, important technology in the life cycle management of a product, offers solutions in reduced time including “testing and tuning”. It is in the virtual environment in which the kinematics simulation of the forefinger is presented, considering five degrees of freedom. The viability of the system is verified through the kinematic and static analysis for different positions of the system components ensuring no interference between them.

### 2 NOTATION

The kinematic model of the forefinger of the right hand it is shown, with the longitudinal axis of the bones in the direction of the x axis, in the Figure 1. Figure 2 shows the hand, corresponding to the forefinger shown in Figure 1.

To differentiate the forefinger bones, each line shown with different thickness. The kinematic lengths of the bones (links) are parameterized. The parameters used for the kinematic lengths of each link (bone) are:

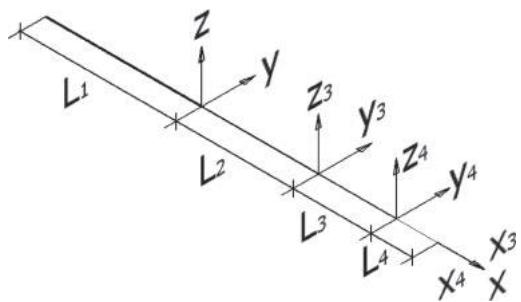


Figure 1. Forefinger.



Figure 2. Hand for the forefinger shown in Figure 1.

$L_1$  = length kinematic link 1 (metacarpal bone).

A value of 60 mm was used.

$L_2$  = length kinematic link 2 (proximal phalanx bone). A value of 45 mm was used.

$L_3$  = length kinematic link 3 (middle phalanx bone). A value of 30 mm was used.

$L_4$  = length kinematic link 4 (distal phalanx bone).

A value of 16 mm was used.

The angles of rotation between the links converging on a kinematic pair are a function of  $\alpha$ .

$\alpha$  = Minimum angle of rotation in a kinematic pair. A value of 0.01 degrees for each and every one of the five degrees of freedom was used. The rotation angles are multiples of  $\alpha$ ; for example, to indicate an angle of 45 degrees is used:  $4500 * \alpha$ .

Other variables used are:

$r$  = Radio container. A value of 50 mm was used.

$\Theta_3$  = Angle of the longitudinal axis of the link 3 with respect to the x axis.

### 3 COORDINATE SYSTEMS

Whether the inertial coordinate system ( $x$ ,  $y$ ,  $z$ ) located at the beginning of the link 2, see Figures 1 and 2. A non-inertial reference system ( $x_3$ ,  $y_3$ ,  $z_3$ ) is fixed at the beginning of the link 3. Other non-inertial reference system ( $x_4$ ,  $y_4$ ,  $z_4$ ) is fixed at the start of link 4.

Coordinate system ( $x_3$ ,  $y_3$ ,  $z_3$ ) and ( $x_4$ ,  $y_4$ ,  $z_4$ ) move respectively with links 3 and 4.

### 4 DEGREES OF FREEDOM

Defined coordinate systems, the five degrees of freedom of the system are [1, 2, 3]:

- Rotation of the phalanges around the  $x$  axis.
- Flexion-extension of the proximal phalanx with respect to metacarpal bone around the  $y$  axis, with an average range of 80 degrees.
- Flexion-extension of the medial phalanx with respect to the proximal phalanx about the  $y_3$  axis, with an average range of 100 degrees.
- Flexion-extension of the distal phalanx with respect to the medial phalanx about the  $y_4$  axis, with an average range of 65 degrees.
- Abduction-adduction of the phalanges around the  $z$  axis, with an average range of 20 degrees.

### 5 SYSTEM COMPONENTS

For the design of system components, it was tried to reproduce the shape of the actual bones of a forefinger, with minor variations to achieve greater

ranges of angular motion. In Figures 3, 4, 5 and 6 links system (bones) are shown.

The links have cylindrical or spherical ends to allow the five degrees of freedom.

Cover the bones, not included, it was considered of 12 mm thickness along the  $y$  axis.



Figure 3. Metacarpal bone link 1.



Figure 4. Proximal phalanx link 2.



Figure 5. Medial phalanx link 3.



Figure 6. Distal phalanx link 4.

The bones were designed with a maximum diameter of 10 mm.

## 6 SIMULATION OF THE DEGREES OF FREEDOM

The system in its initial position is shown in Figure 7. The system, after a rotation about x axis is shown in Figure 8. The system, after three flexion-extension movements (rotation about  $y_4$ ,  $y_3$  and  $y$  axes), is shown in Figure 9. The system, after an abduction-adduction motion (rotation around the z axis) between the phalanges and the bone metacarpal is shown in Figure 10.

## 7 PRACTICAL APPLICATION

When a person suffers an accident, the effects are unpredictable. In case of loss of the phalanges of all the fingers of his right hand, with prosthesis of the phalanges of his forefinger should grab a container.

The prosthesis must be programmed to achieve this objective. In Figure 11 the prosthesis with flexion between phalanges and between the proximal phalanx and the metacarpal bone is shown.

If we consider a container 38 mm in diameter, See Figures 11 and 12, the lines representing each link should grab a container of 25 mm radius, due to the thickness of the bones and its cover.

If the slope of the link 1 is zero and the slope of the link 2 is equal to infinity; the slope of the links 3 and 4 can be calculated. The results are shown in the following equations:

$$m_3 = \frac{F^2 - r^2}{2rF} \quad (1)$$



Figure 7. System in its initial position.



Figure 8. Rotation of the phalanges around the x axis.

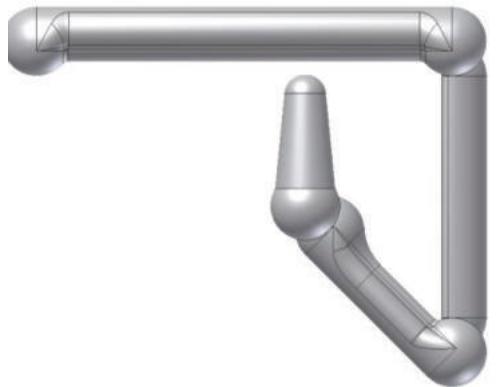


Figure 9. Movements of flexion-extension.

$$F = r - L_2 \quad (2)$$

$$m_4 = \frac{G(x_D + r) + r\sqrt{G^2 + (x_D)^2 + 2rx_D}}{(x_D)^2 + 2rx_D} \quad (3)$$

$$G = z_D + r \quad (4)$$

$$z_D = -L_2 - L_3 \sin \theta_3 \quad (5)$$

$$x_D = -L_3 \cos \theta_3 \quad (6)$$



Figure 10. Movement abduction-adduction.

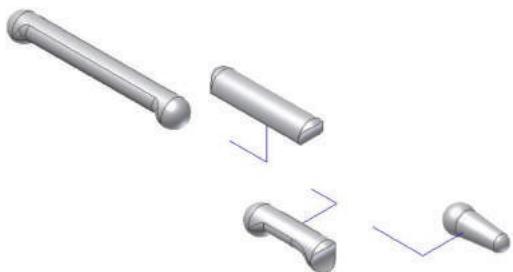


Figure 13. Explode system.



Figure 11. Finger around a container.

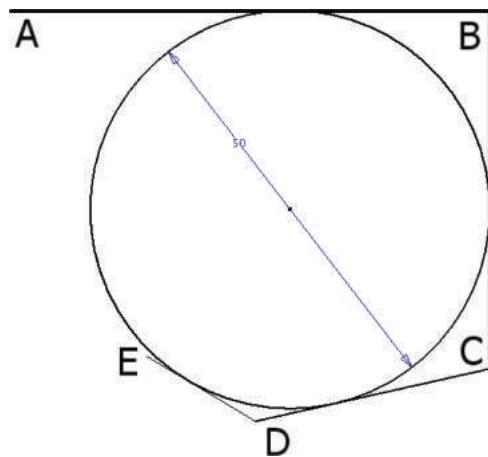


Figure 12. Forefinger around of a container.

$$\theta_3 = \operatorname{arctg}(m_3) \quad (7)$$

Since the spreadsheet is checking files (graphics) components and assembly; independent parameters can be varied to analyze their impact on the results.

## 8 ADVANTAGES OF VIRTUAL SIMULATION

The advantages of virtual simulation systems are:

- Makes it possible to reproduce the dynamic response.
- Makes it possible to reproduce the corresponding processes of assembly and disassembly. See Figure 13.
- Enables the prior analysis of their operation.
- Facilitates analysis and synthesis from the point of view of kinematics.
- Analogous systems with different parameters can be simulated.
- You can check the compliance of the design constraints in its stages of manufacture, operation and maintenance.

## 9 ANALYSIS OF RESULTS

The simulation of different degrees of freedom was possible without interference between the links of the system in its different images of positions, limiting these to the following values:

- Rotation around the x axis of  $\pm 45$  degrees. Figure 8 illustrates one of these movements.
- Abduction-adduction of  $\pm 30$  degrees. Figure 10 illustrates one of these movements.
- Flexion-extension between metacarpal bone and proximal phalanx of  $+90$  degrees and  $-30$  degrees. In Figure 9 the flexion of  $+90$  degrees is illustrated.
- Flexion-extension between the proximal phalanx and the middle phalanx of  $+135$  degrees and zero degrees. In Figure 9 the flexion the  $+135$  degrees is illustrated. The negative flexion between these two phalanges was considered humanly impossible.
- Flexion-extension between the medial phalanx and the distal phalanx of  $+45$  degrees and  $-30$  degrees. In Figure 9 the flexion the  $+45$  degrees is illustrated.

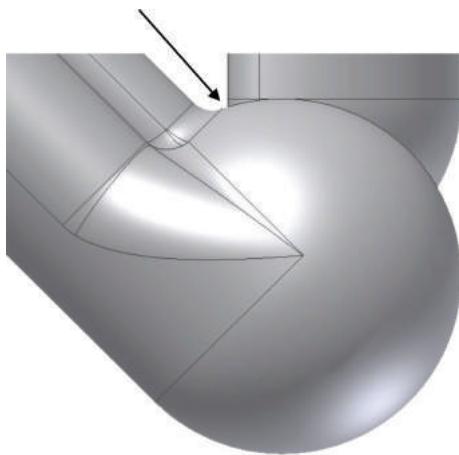


Figure 14. Detail of Figure 9.



Figure 15. Boutonniere deformity.



Figure 16. Swan Neck deformity.

An iteratively refined design was necessary to make possible the above limits, without interference between the links converging on different kinematic pairs. The limits obtained are higher than average limits of a person. In Figure 14, detail of Figure 9 in the kinematic pair between the proximal and medial phalanges, it can be seen that a larger positive flexion angle of 135 degrees may cause interference between these bones.

Deeper knowledge of the behaviour from the point of view of the kinematic joints of the

forefinger bones enables better diagnosis and therefore a better prognosis.

Boutonniere deformity and swan neck deformity can be resolved with treatments of modern medicine and, in critical cases, complemented by solutions of general mechanics and kinematics in particular. See Figures 15 and 16.

## 10 CONCLUSIONS

- The movements in the kinematic pairs can be programmed, for any position of the system components in real time.
- The three-dimensional virtual and dynamic simulation of a system enables deeper analysis, to synthesize and optimize it.
- Forefinger design will enable the capacity of this subsystem prosthesis body exceed those of a human subsystem average real.
- Additionally, these capabilities may be maintained for longer time, depending on the proper selection of materials also made with computer support.

## 11 FUTURE WORK

Some of future works are given below:

- Generating physical prototypes, of the component parts of the system.
- Assembly, the generated component parts, to get the actual prototype system.
- Instrumentation and control the real system.

## ACKNOWLEDGEMENTS

The authors acknowledge the institutional support of the Autonomous Metropolitan University and the Azcapotzalco Unit in general and of the Division of Basic Sciences and Engineering and the Department of Energy in particular.

## REFERENCES

- [1] Neil A. Davidoff, Andris Freivalds (August 4, 1993). "A graphic model of the human hand using CATIA". Department of Industrial and Management Systems Engineering, Pennsylvania State University, University Park, PA 16802, USA.
- [2] Jung Won Kang, Hyun Seok Hong, Bong Sung Kim, Myung Jin Chung, "Work Assistive Mobile Robots Assisting the Disabled in a Real Manufacturing Environment", International Journal of Assistive Robotics and Mechatronics, Volume 8 Number 3, September 2007.
- [3] J.K. Salisbury, M.T. Mason, "Robot Hands and the Mechanics of Manipulation", MIT Press, Cambridge, MA, 1985.

This page intentionally left blank

# CT based identification problem for the multicompartiment model of blood perfusion

E. Rohan, V. Lukeš & J. Brašnová

*Department of Mechanics and NTIS, Faculty of Applied Sciences, The European Centre of Excellence, University of West Bohemia in Pilsen, Plzeň, Czech Republic*

**ABSTRACT:** This paper deals with modelling the tissue blood perfusion and focuses on the identification of the model parameters using the patient-specific data obtained using the standard CT, or MRI investigations. The research is aimed to develop a complex model which would assist in planning the liver surgery. Recently we proposed a multicompartiment model of the blood perfusion in liver which serves as the feed back for simulations of dynamic CT investigation (Rohan, Jonášová, & Lukeš 2014). The flow can be characterized at several scales for which different models are used. Flow in larger branching vessels is described using a simple 1D model based of the Bernoulli equation with correction terms respecting the pressure losses due to the dissipation. This model is coupled through point sources/sinks with a 3D model describing multicompartiment flows at the lower hierarchies of the perfusion trees penetrating to the parenchyma. The compartments in the liver tissue are associated with segments which confine the flow to subdomains within the organ, and hierarchies which reflect the flow complexity on branching vascular trees of the portal and hepatic veins. For simulations of the CT perfusion test, a model of the dynamic transport of the contrast fluid in the compartments was developed; the time-space distribution of the so-called tissue density can be computed and compared with the measured data obtained from the CT. To find suitable values of the perfusion model parameters, we formulate an optimization problem where the objective function expresses the difference between the standard CT images and the corresponding perfusion maps of the contrast agent concentration. This is computed by solving the perfusion problem to obtain a steady distribution of the blood pressure at all the compartments, and the transport problem describing the contrast fluid saturation. The optimization problem can be solved using a suitable gradient-based method. For this the sensitivity analysis formulae were derived using the adjoint system method. A simplified identification problem was implemented and solved numerically to show viability of the proposed approach.

## 1 INTRODUCTION

The blood flow is featured by a complex hierarchical system of blood vessels comprising arteries, smaller arteries, arterioles, pre-capillary vessels and capillaries. There the blood is filtrated and participates in the metabolism process, giving the oxygen to the tissue and collecting the waste. The ascending hierarchical system of veins is arranged similarly to the arterial system, so that the blood circulation at the tissue level can be represented by two penetrating but separated trees which are connected only at the capillary level. Detailed modelling of this part can be based on the homogenization approach, as proposed by Rohan & Cimrman (2010). This is an idealization which neglects shunts existing between the two trees, however, it is the essential scheme of the blood flow which should be reflected by any model of the blood perfusion. In the case of liver tissue, vascular geometry at the microlevel was studied by Debbaut et al. (2014).

## 2 FORMULATION OF THE STATE PROBLEM

### 2.1 Structure of compartments in a perfused tissue

We consider a part of tissue occupying the domain  $\Omega \subset \mathbb{R}^3$ . The compartments are established as connected parts of the vasculature with a given range of characteristic vessel cross-sections; such a range is indicated by a hierarchy index. Each compartment is associated with a segment  $\Omega_e \subset \Omega$  at the level of the whole organ (such as liver). Moreover, different groups of the vascular system are distinguished, namely the arterial and venous trees form separate groups. To summarize, any compartment belongs to either the arterial, or the venous group, and is associated with a hierarchy and with a segment.

We approximate the perfusion at the level of tissue parenchyma using a macroscopic model

describing parallel flows in multiple compartments, (Cimrman & Rohan 2007).

The model involves pressures  $\{p^i\}_i$ , associated to each compartment  $i = 1, \dots, \bar{i}$ . Any  $i$ -th compartment occupying domain  $\Omega_i$  can be saturated from an external source (or drained by a sink); in general, the local source/sink flux  $F^i$  can be defined which describes the fluid exchange with the upper (macroscopic) hierarchies of the perfusion trees. Alternatively, the pressure can be prescribed in a given subdomain  $\Sigma_i \subset \Omega_i$  which may represent junctions with upper hierarchies of the perfusion system described by the 1D flow model on branching networks; in practice,  $\Sigma_i$  can be formed by “small” balls,  $|\Sigma_i| \ll |\Omega_i|$ .

## 2.2 Flow problem

For the weak formulation of the multicompartment Darcy flow problem the following admissibility sets are needed:  $\mathcal{V}^i = \{q \in H^1(\Omega_i) | q = \bar{p}_i \text{ on } \partial\Sigma_i\}$  the space of test pressures  $\mathcal{V}_0^i = \{q \in H^1(\Omega_i) | q = 0 \text{ on } \partial\Sigma_i\}$ , and  $\mathcal{V} = \mathcal{V}^1 \oplus \dots \oplus \mathcal{V}^{\bar{i}}$ . The problem reads, as follows: Find  $\mathbf{p} = (p^1, p^2, \dots, p^{\bar{i}})$  whereby  $p^j \in \mathcal{V}^j$ ,  $j = 1, \dots, \bar{i}$  satisfy

$$\begin{aligned} & \int_{\Omega_i \setminus \Sigma_i} \mathbf{K}^i \nabla p^i \cdot \nabla q^i + \int_{\Omega_i \setminus \Sigma_i} \sum_j G_j^i (p^i - p^j) q^i \\ &= \int_{\Omega_i \setminus \Sigma_i} F^i q^i, \quad \forall q^i \in \mathcal{V}_0^i, \end{aligned} \quad (1)$$

for all compartments  $i = 1, \dots, \bar{i}$ . The summation in the second term takes only over nonvanishing perfusion coefficients,  $G_j^i$ . The perfusion velocity in one compartment is given by the Darcy law and involving the permeability tensors  $\mathbf{K}^i = (K_{kl}^i)$ , thus,  $\mathbf{w}^i = -\mathbf{K}^i \nabla p^i$ . The flux between different compartments is computed as  $\mathcal{J}_j^i = G_j^i (p^i - p^j)$ . All parameters  $\mathbf{K}^i(x)$  and  $G_j^i(x)$  depend on the local volume fractions  $\phi_i(x)$ .

## 2.3 Contrast fluid transport

The blood saturation by the tracer, denoted by  $S^i$ , is associated with the  $i$ -the compartment. Clearly, values of  $S^i$  are restricted by  $S^i \in [0, 1]$ . The tracer partial concentration is introduced using the volume fraction  $\phi^i$  of the vessels for each compartment  $i$  by  $c^i = \phi^i S^i$  (no summation). This is the equivalent quantity of the so-called tissue density measured by the CT investigation. The total apparent concentration (the gray level) is then given as

$$C = \sum_i c^i = \sum_i \phi^i S^i, \quad (2)$$

where the summation is taken over all locally overlapping compartments. We shall use the abbreviation  $\mathbf{S} = (S_1, \dots, S_{\bar{i}})$ .

The transport of the contrast fluid is described as the solution of the following problem: Given  $\{\mathbf{w}^i\}_i$  and  $\{p^i\}_i$ , for a given initial conditions  $\{S^i(t=0, x)\}_i = \{S_0^i(x)\}_i$  given for  $x \in \Omega_i$ , find  $\{S^i(t, x)\}_i$  such that, for all compartments  $i = 1, \dots, \bar{i}$ ,

$$\begin{aligned} & \phi^i \frac{\partial S^i}{\partial t} + \nabla \cdot (S^i \mathbf{w}^i) + \sum_j Z_j^i(S) \mathcal{J}_j^i \\ &= S_{\text{in}} f_+^i + S^i f_-^i \quad x \in \Omega_i, \quad t > 0 \end{aligned} \quad (3)$$

where

$$Z_j^i(S) = \begin{cases} S^i & \text{if } \mathcal{J}_j^i > 0, \\ S^j & \text{if } \mathcal{J}_j^i \leq 0. \end{cases} \quad (4)$$

To simplify expressions we may introduce corresponding index sets  $\mathcal{I}_+^i = \{j \neq i | \mathcal{J}_j^i > 0\}$  and  $\mathcal{I}_-^i = \{j \neq i | \mathcal{J}_j^i \leq 0\}$  which, however, are defined point-wise in  $\Omega$  according to the variation of fluxes  $\mathcal{J}_j^i(x), x \in \Omega_i \cap \Omega_j$ .

## 3 IDENTIFICATION PROBLEM BASED ON KNOWN CONCENTRATION

We assume that the total concentration defined in (2) can be measured, i.e.  $\bar{c}(x, t)$  is known for  $x \in \Omega$  and  $t \in [0, T]$ . We define the objective function

$$\Phi(\alpha, \mathbf{p}, \mathbf{S}) = \int_0^T \int_{\Omega} \left| \sum_i \phi^i S^i - \bar{c} \right|^2 dx dt. \quad (5)$$

In fact,  $\partial_{\alpha} \Phi(\alpha, \mathbf{p}, \mathbf{S}) = 0$ , since  $\Phi$  in (5) depends explicitly just on  $\mathbf{S}$ , thus  $\Phi = \Phi(\mathbf{S})$ .

The identification problem is formulated as an optimization problem, see (6), where the unknown parameters to determine are optimization variables  $\alpha = (\alpha_k)$ ,  $k = 1, \dots, m+n$ . We shall consider  $\alpha$  being associated with the unknown parameters  $(\{G\}_k(x), \{\mathbf{K}\}_l(x))$ ,  $x \in \Omega$ . Obviously we assume the problem is discretized using the Finite Element (FE) method, so that  $x$  is confined to a nodal positions of the FE mesh. Above  $k = k(i, j, x)$  is the multiindex numbering of coupled compartments with enabled flux between  $i$  and  $j$  at a position  $x \in \Omega_i$ ;  $l = l(i, x)$  is associated with the compartment index  $i$  and  $x \in \Omega_i$ . We assume either  $G_i^j \equiv 0$ , or  $G_i^j = G_i^j(\alpha_k)$ . We consider only such couples  $(i, j)$ , for which  $G_i^j \neq 0$ . A natural choice is to assume  $G_i^j = \bar{G} \alpha_k$ ,  $k = 1, \dots, m$ , with a constant  $\bar{G} > 0$  and  $\alpha_k(x) \in \mathbb{R}$  is a Lipschitz continuous function defined for  $x \in (\Omega_i \setminus \Sigma_i) \cap (\Omega_j \setminus \Sigma_j)$ . In analogy,  $\mathbf{K}^i = \bar{\mathbf{K}}^i \alpha_l$ ,  $l = m+1, \dots, m+n$ . The “reference parameters” are estimated using the analysis of the vascular trees according to a method

pursuing the works by Huyghe & Campen (1995b) and Hyde et al. (2013).

### 3.1 The optimization problem

Find  $\alpha \in \mathcal{A} \subset \mathbb{R}^{m+n}$  such that:

$$\begin{aligned}\Phi(\alpha, p, S) \rightarrow \min \\ p = p(\alpha) \quad \text{solves (1)} \\ S = S(p) \quad \text{solves (3)}\end{aligned}\tag{6}$$

where  $\mathcal{A}$  is a given admissibility set.

The following bilinear forms and the linear functional are employed:

$$\begin{aligned}a^i(p, q) &= \int_{\Omega_i \setminus \Sigma_i} \mathbf{K}^i \nabla p \cdot \nabla q, \\ b^{ij}(p, q) &= \int_{\Omega_i \setminus \Sigma_i} G_j^i p q, \quad g^i(q) = \int_{\Omega_i \setminus \Sigma_i} F^i q.\end{aligned}$$

Further, by  $b_\alpha^{ij}(p, q)$  the dependence of  $G_j^i(x)$  on  $\alpha_k$  is indicated; recall that index  $k = k(i, j, x)$  is associated with the two compartments. Similarly,  $a_\alpha^i(p, q)$  depends on  $\alpha_l$  which is associated with  $\mathbf{K}_l^i(x)$ .

The Lagrange function associated with (6) is defined by

$$\begin{aligned}L(\alpha, (p, S), (\lambda, \psi)) &= \Phi(\alpha, p, S) \\ &+ \sum_i \left( a_\alpha^i(p^i, \lambda^i) + \sum_{j \neq i} b_\alpha^{ij}(p^i - p^j, \lambda^i) - g^i(\lambda^i) \right) \\ &+ \sum_i \int_{\Omega_i} \int_0^T \left( \frac{\partial S^i}{\partial t} + \nabla \cdot (\mathbf{w}^i S^i) \right) \psi^i dt dx \\ &+ \sum_i \sum_{j \neq i} \int_0^T b_\alpha^{ij}(p^i - p^j, Z_j^i(S) \psi^i) dt \\ &- \sum_i \int_0^T \int_{\Omega_i} (S_{in}^i f_+^i + S^i f_-^i) \psi^i dx dt,\end{aligned}$$

where  $\lambda^i$  and  $\psi^i$  are the Lagrange multipliers;  $\lambda^i \in \mathcal{V}_0^i$  and  $\psi^i = 0$  at  $t = T$ .

### 3.2 Optimality conditions

The optimality conditions associated with feasibility of the state problem read:

$$\delta_S L \circ \delta S = 0, \quad \delta_p L \circ \delta p = 0, \tag{7}$$

for admissible variations  $\delta S$  and  $\delta p$ . From (7) we obtain the adjoint problem which can be solved in two steps:

1. Find  $\{\psi^j\}$ ,  $j = 1, \dots, \bar{i}$  such that  $\psi^j(T, \cdot) = 0$  and  $\{\psi^j\}$  satisfy (for  $x \in \Omega, t \in ]0, T[$ ),

$$\begin{aligned}\frac{\partial \psi^j}{\partial t} + \mathbf{w}^j \cdot \nabla \psi^j + f_-^j \psi^j - \sum_{i \in \mathcal{I}_+^j} \mathcal{J}_i^j(\psi^j - \psi^i) \\ = \phi^j \left( \sum_k \phi^k S^k - \bar{c} \right),\end{aligned}$$

recalling  $\mathcal{J}_i^j = G_i^j(p^j - p^i)$  and  $\mathbf{w}^i = -\mathbf{K}^i \nabla p^i$ .

2. Find  $\lambda^j \in \mathcal{V}_0^j$ ,  $j = 1, \dots, \bar{i}$  such that

$$\begin{aligned}a_\alpha^j(q, \lambda^j) + \sum_{i \neq j} b_\alpha^{ij}(q, \lambda^j - \lambda^i) &= - \int_0^T a_\alpha^j(q, S^j \psi^j) dt \\ &- \sum_{i \neq j} \int_0^T b_\alpha^{ij}(q, \psi^j Z_i^j(S) - \psi^i Z_j^i(S)) dt,\end{aligned}$$

for all  $q \in \mathcal{V}_0^j$ .

The total sensitivity of  $\Phi$  w.r.t. to the optimization parameters  $\alpha$  for any admissible state  $(p(\alpha), S(p)) \in \mathcal{V} \times \mathcal{S}$  can now be evaluated using the following formula:

$$\begin{aligned}\partial_\alpha^{\text{tot}} L(\alpha, (p, S), (\lambda, \psi)) &= \partial_\alpha^{\text{tot}} \Phi(\alpha, p, S) \\ &= \sum_i \left( \partial_\alpha a_\alpha^i(p^i, \lambda^i) + \sum_{j \neq i} \partial_\alpha b_\alpha^{ij}(p^i - p^j, \lambda^i) \right) \\ &\quad + \sum_i \int_0^T \sum_{j \neq i} \partial_\alpha b_\alpha^{ij}(p^i - p^j, Z_j^i(S) \psi^i) dt \\ &\quad - \int_0^T \partial_\alpha a_\alpha^i(p^i, S^i \psi^i) dt.\end{aligned}\tag{8}$$

## 4 SIMPLIFIED PROBLEM

We consider a simplified problem which was implemented and solved to examine some potential difficulties arising in this kind of inverse problems. The liver perfusion is described by a reduced multicompartment model which incorporates the vascular tree of the portal vein only, Figure 1(a). Upon decomposing the tree in 2 compartments, the hierarchical flow is represented by three compartments which are defined in terms of two permeabilities,  $\mathbf{K}^1, \mathbf{K}^2$  and two inter-compartment perfusion coefficients,  $G_1^2$  and  $G_2^3$ . The 3rd compartment is degenerate, i.e. the pressure  $p^3 = 0$  is prescribed. The flow is described by only two pressures,  $p = (p^1, p^2)$ , such that  $p^j \in \mathcal{V}_0^j$ ,  $j = 1, 2$ . Thus, the state problem is given by the following 2 equations:

$$\begin{aligned}\int_{\Omega} \mathbf{K}^1 \nabla p^1 \cdot \nabla q^1 + \int_{\Omega} G_2^1(p^1 - p^2) q^1 &= \int_{\Omega} F^1 q^1, \\ \int_{\Omega} \mathbf{K}^2 \nabla p^2 \cdot \nabla q^2 + \int_{\Omega} (G_1^2(p^2 - p^1) + G_3^2 p^2) q^2 &= 0,\end{aligned}\tag{9}$$

for all  $q^1, q^2 \in \mathcal{V}_0 = L^2(\Omega)$ . No Dirichlet conditions are needed, so that also  $\mathcal{V}^j = L^2(\Omega)$ ,  $j = 1, 2$ .

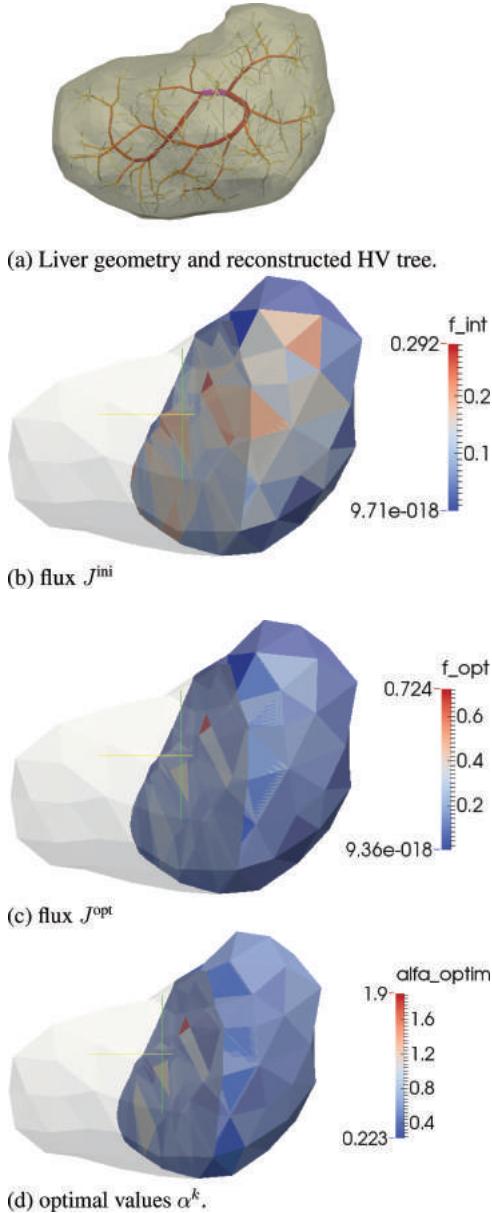


Figure 1. Identification of spatial distribution of perfusion coefficients  $G_2^1(x)$  using parameters  $\alpha$ . Perfusion tree of the hepatic vein (a), FE model (b, c, d)—a part displayed (planar cut). Initial (b) and optimized (c) fluxes  $J(x)$  at finite elements. Optimal parameters  $\alpha_k^{\text{opt}}$  at elements  $\Omega_k$  (d).

This model assumes a given constant pressure associated with the parenchyma sinusoids, which enables to solve a “half-problem”; only the flow through the saturating vasculature of the portal vein is concerned. Let us recall that  $F^1$  represents

the distribution of fluid sources of the 1st compartment; in fact,  $F^1$  is a given function with localized supports to small subdomains  $\omega^k$  which are situated at connection points of terminal branches of a truncated perfusion tree describing the upper hierarchies of the portal vein.

#### 4.1 Identification problem for $G_2^1$

The permeability tensors  $K^i$ , given for any compartment  $i$  can be computed “pointwise” by the averaging procedure proposed by Vankampen et al. 1997 using the theory derived in (Huyghe & Campen 1995a), (Hyde et al. 2013). Although  $G_i^j$  can be computed as well, in this case the procedure is more cumbersome, requiring additional fluid problems to be solved. Therefore, we propose to obtain relevant values of  $G_i^j$  by solving an inverse problem. In our simplified problem we shall assume that the correct inter-compartmental fluxes  $\bar{J}$  can be measured, so that  $G_2^1$  is computed by minimizing the following objective function:

$$\phi(\alpha, p) = \int_{\Omega} |G_2^1(\alpha)(p^i - p^j) - \bar{J}|^2,$$

where  $\bar{J}$  is the desired data, i.e. a “measured” flux between the compartments 1 and 2. The identification problem reads, as follows: Find  $\alpha \in \mathcal{A}$  such that

$$\begin{aligned} \phi(\alpha, p) &\rightarrow \min \\ p &= p(\alpha) \text{ solves (9),} \end{aligned} \quad (10)$$

where  $\mathcal{A}$  is a given admissibility set. The state problem (9) is approximated using the FE method. Therefore  $\alpha = (\alpha_k)$  is a piece-wise constant function, given by values  $\alpha_k$  on each element  $\Omega_k \subset \Omega$ ,  $k = 1, \dots, m$ . Then,  $G_2^1(x) = \bar{G}_k \alpha_k$  for  $x \in \Omega_k$  is established using  $\bar{G}_k$  which for a given tree, see Figure 1, decomposed in two compartments is computed using the averaging method (Hyde et al. 2013). Using the same approach, the permeability tensors  $K^1, K^2$  are defined by an averaging procedure at each element  $\Omega_k$ , however, these parameters are fixed, not considered for the identification. For a given FE approximation and  $\bar{J}^k$  given by constants on elements  $\Omega_k$ , the objective function can be rewritten, as follows:

$$\phi(\alpha, p) = \sum_{k=1}^m \int_{\Omega_k} |J^k(x) - \bar{J}^k|^2, \quad (11)$$

$$\text{where } J^k(x) = \alpha_k \bar{G}^k(p^1(x) - p^2(x)).$$

#### 4.2 Sensitivity analysis

Since the contrast fluid transport is not involved in (10), also the sensitivity analysis simplifies. The adjoint state  $(\lambda^1, \lambda^2) \in (L^2(\Omega))^2$  is the solution of

$$\begin{aligned}
& \int_{\Omega} \mathbf{K}^1 \nabla \lambda^1 \cdot \nabla q^1 + \int_{\Omega} G_2^1 (\lambda^1 - \lambda^2) q^1 \\
&= -2 \int_{\Omega} (G_2^1 (p^1 - p^2) - \bar{J}) G_2^1 q^1, \\
& \int_{\Omega} \mathbf{K}^2 \nabla \lambda^2 \cdot \nabla q^2 + \int_{\Omega} (G_1^2 (\lambda^2 - \lambda^1) + G_3^2 \lambda^2) q^2 \\
&= 2 \int_{\Omega} (G_2^1 (p^1 - p^2) - \bar{J}) G_2^1 q^2,
\end{aligned} \tag{12}$$

for all  $q^1, q^2 \in \mathcal{V}_0 = L^2(\Omega)$ . We recall that  $G_2^1 = G_1^2$  depends on  $\alpha$ , whereas  $\mathbf{K}^1, \mathbf{K}^2$  and  $G_3^2$  are fixed. The total gradient of  $\phi$  defined in (11) is computed by

$$\begin{aligned}
d_{\alpha_k}^{\text{tot}} \phi(\alpha, p) &= 2 \int_{\Omega_k} (\alpha_k \bar{G}_k (p^i - p^j) - \bar{J}_k) \bar{G}_k (p^i - p^j) \\
&\quad + \int_{\Omega_k} \bar{G}_k (p^1 - p^2) (\lambda^1 - \lambda^2),
\end{aligned}$$

so that the  $k$ -th component is evaluated using integrals on  $\Omega_k$ .

### 4.3 Numerical results

The problems (9) and (12) have been implemented in the SfePy code (Cimrman and et al. 2011) using the FE method. The liver geometry was reconstructed using anatomically correct data, however the portal venous tree was generated using the Global Constructive Optimization method, as reported in (Lukeš, Jiřík, Jonášová, Rohan, Bublík, & Cimrman 2014) using the algorithm by Georg, Preusser, & Hahn. The FE model included 322 elements. Using the approach employed by Hyde et al. (2013) the initial guess of the perfusion coefficients  $G_2^1 = \bar{G}$  was obtained for each element. For this auxiliary flow problems were solved on segments of the perfusion tree. The desired fluxes  $\bar{J}$  were computed by solving a reduced flow problem on the perfusion tree, see e.g. (Jonášová, Bublík, & Vimmer 2014). The initial values  $\alpha_k^{\text{ini}} := 0.5$ , thus the

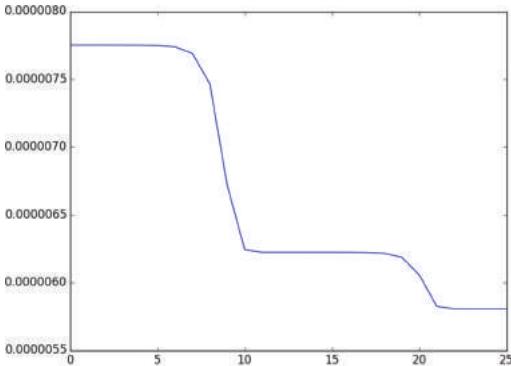


Figure 2. Optimization problem (10): objective function decreasing in iterations.

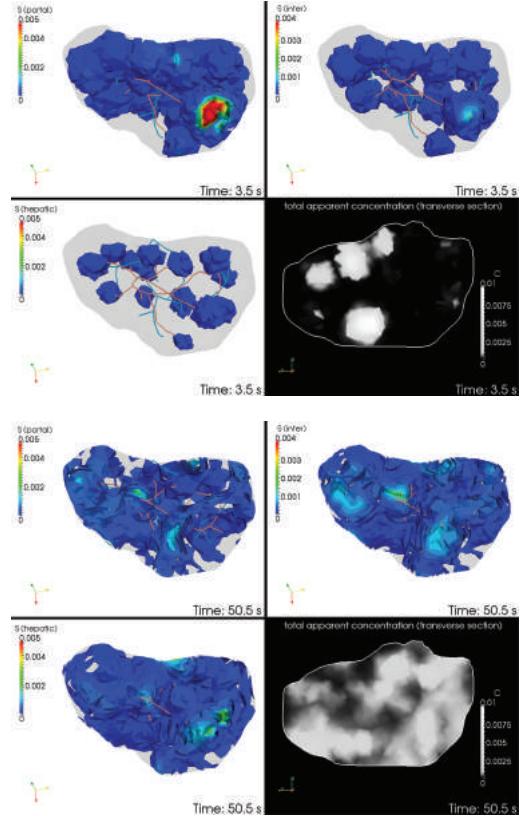


Figure 3. Simulation of the contrast fluid saturation  $S^i(\Omega, t)$  in the 3 compartments,  $i = 1, 2, 3$ , for  $t = 3.5$  sec. The black-and-white pictures are the reconstructed concentrations which correspond to the “tissue density” measured by the CT test.

optimal values were expected  $\alpha_k^{\text{opt}} \approx 1$ , although, obviously, the fluxes  $J$  defined in (11)<sub>2</sub> in the sense of optimization problem (10) may not be coherent with the definition of  $\bar{J}$  based on a different flow problem. Thus, the example reported in this paper was intended rather to study some numerical aspects of the identification problem.

In Figure 1(b,c), the initial and optimized fluxes defined in (11)<sub>2</sub> are displayed, whereas the optimal values of  $\alpha_k$  on elements  $\Omega_k$  are depicted in Figure 1(c). The objective function was reduced by  $\approx 25\%$  of its initial value, only 25 iterations by a gradient SQP method were needed, see Figure 2.

## 5 CONCLUSIONS

The complete model describing the CT examination of the liver perfusion, thus, including the contrast agent transport, has been implemented

in the SfePy code (see (Cimrman & et al. 2011)). Both the portal and hepatic vein vasculatures were reconstructed using artificial trees (Georg et al.). So defined portal and hepatic vein segments enabled simulation of the upper-level hierarchical flow using the “1D” model (Jonášová, Bublík, & Vimmr 2014). In Figure 3, we illustrate the flow and the tracer distributions governed by the equations (1) and (3). The flow in the lower-level hierarchies was approximated using the Darcy-type flow model with 3 compartments: one associated with the portal vein structure, the second with the hepatic vein structure, and the third one representing the lower-most level of the parenchyma corresponding to the liver lobes. The mesh of the FEM discretization comprised 11019 tetrahedral elements, approximately 7300 degrees of freedom associated with the pressure in all compartments were involved in the computations. The model was coupled with the upper-level “1D” flow through 28 junctions of which 14 were the sinks and 14 were the sources defining the domain  $\Sigma_l$  and functions  $F^l$ .

We have formulated the optimization problem aimed at identification of the unknown model parameters defined pointwise in the whole domain, (Cimrman & Rohan 2012), namely the permeabilities  $K^i$  for each compartment and the perfusion parameters  $G_i^j$  controlling the fluid flow between different compartments. The state problem consists of two parts: the perfusion problem, involving the above mentioned unknown parameters, and the contrast fluid transport problem which involves the perfusion velocities. The sensitivity formula (8) provides the gradient of the objective function w.r.t. all parameters; since the adjoint state technique is used, the numerical complexity of the sensitivity does not increase proportionally to the number of the optimization variables, (Rohan & Cimrman 2002).

The optimization problem (6) is nonconvex, therefore, one may hope to compute local minimizers only. This deteriorates efficiency of the identification, although the simplified identification problem (10) reported in section 4 proved viability of the proposed approach. The implementation of the full problem (6) is now being tested, new numerical issues appear and will have to be sorted out.

## ACKNOWLEDGMENT

This research is supported by project NT 13326 of the Ministry of Health of the Czech Republic (CR), in part by project GAČR 13-00863S of the Sci. Foundation of the CR. Publication enabled due to project LO1506 (MŠMT) of CR.

## REFERENCES

- Cimrman, R. & et al. (2011). SfePy home page. <http://sfePy.org>. Software, finite element code and applications.
- Cimrman, R. & E. Rohan (2007). On modelling the parallel diffusion flow in deforming porous media. *Math. and Computers in Simulation* 76, 34–43.
- Cimrman, R. & E. Rohan (2012). On identification of the arterial model parameters from experiments applicable “in vivo”. *Mathematics and Computers in Simulation* 80, 1232–1245.
- Debbaut, C., P. Segers, P. Cornillie, C. Casteleyn, M. Dierick, W. Laleman, & D. Monbaliu (2014). Analyzing the human liver vascular architecture by combining vascular corrosion casting and micro-ct scanning: a feasibility study. *J. of Anatomy* 224(4), 509–517.
- Georg, M., T. Preusser, & H.K. Hahn. Global constructive optimization of vascular systems. *Technical Report: Washington University in St. Louis*.
- Huyghe, J.M. & D.H. Campen (1995a). Finite deformation theory of hierarchically arranged porous solids—i. balance of mass and momentum. *Int. J. Engrg. Sci.* 33(13), 1861–1871.
- Huyghe, J.M. & D.H. Campen (1995b). Finite deformation theory of hierarchically arranged porous solids, part I., II. *Int. J. Engrg. Sci.* 33(13), 1861–1886.
- Hyde, E.R., C. Michler, J. Lee, A.N. Cookson, R. Chabinio, D.A. Nordsletten, & N.P. Smith (2013). Parameterisation of multi-scale continuum perfusion models from discrete vascular networks. *Med Biol Eng Comput* 51, 557–570.
- Jonášová, A., O. Bublík, & J. Vimmr (2014). A comparative study of 1d and 3d hemodynamics in patient-specific hepatic portal vein networks. *Appl. and Comp. Mech.* 8, 177–186.
- Lukeš, V., M. Jiřík, A. Jonášová, E. Rohan, O. Bublík, & R. Cimrman (2014). Numerical simulation of liver perfusion: from CT scans to FE model. *Proc. of the 6th European Conference on Python in Science (EuroSciPy 2014)*.
- Rohan, E. & R. Cimrman (2002). Sensitivity analysis and material identification for activated smooth muscle. *Computer Assisted Mechanics and Engineering Science* 9, 519–541.
- Rohan, E. & R. Cimrman (2010). Two-scale modelling of tissue perfusion problem using homogenization of dual porous media. *Int. J. for Multiscale Comput. Engrg.* 8, 81–102.
- Rohan, E., A. Jonášová, & V. Lukeš (2014). Complex hierarchical modeling of the dynamic perfusion test: application to liver. In *11th World Congress on Computational Mechanics (WCCM XI)*.
- Vankan, W., J. Huyghe, J. Janssen, A. Huson, W. Hacking, & W. Schreiner (1997). Finite element analysis of blood flow through biological tissue. *I. Jour of Engrg. Sci.* 35(4), 375–385.

# A study on discrimination of SIFT feature applied to binary images

Insaf Setitra

*Research Center on Scientific and Technical Information Cerist, Algeria  
University of Science and Technology Houari Boumediene, Algeria*

Slimane Larabi

*University of Science and Technology Houari Boumediene, Algeria*

**ABSTRACT:** Scale Invariant Feature Transform (SIFT) since its first apparition in 2004 has been (and still is) extensively used in computer vision to classify and match objects in RGB and grey level images and videos. However, since the descriptor used in SIFT approach is based on gradient magnitude and orientation, it has always been considered as texture feature and received less interest when treating binary images. In this work we investigate the power of discrimination of SIFT applied to binary images. A theoretical and experimental studies show that SIFT can still describe shapes and can be used to distinguish objects of several classes.

## 1 INTRODUCTION

Description of a shape is very important for many applications. Usually, accuracy of further processing such as matching and classification relies on the description of the shape. Morphological features are usually used to describe binary shapes. Shape Contexts [10], Graph Transduction [14], Inner Distance [11], and full and partial matching [12] are some examples of descriptors. In this work we analyze a descriptor that has been extensively used for gray level shapes but not for binary shapes. Scale Invariant Feature Transform proposed by David Lowe (2004) [7] is commonly known to be a texture feature and is used to describe gray level shapes in terms of gradient orientation and magnitude. However, previous to computing gradient orientation and magnitude, the shape is convoluted several times so that to keep strong edges. This cue is very important to describe binary shapes and makes the feature suitable for them.

The most relevant work dealing with SIFT on binary images is devoted to hand gestures classification [13] by the use of SIFT feature applied to binary masks. Authors claim to improve accuracy of classification and to decrease time processing since the descriptor is not sensitive to illumination changes and less information is needed to quantify the keypoints.

In this work, we enlarge the concept of discrimination of SIFT feature on several shapes. We first give a review of SIFT computation and study its effect on binary image. Experiments are conducted in order to verify robustness of the descriptor on both retrieval and classification.

The remainder of the paper is organized as follows: We describe in section 2 the descriptor, in section 3 we describe our shape retrieval approach followed in section 4 by the shape classification. We present and discuss results in section 3.

## 2 SHAPE DESCRIPTOR

In [7], scale invariant feature transform has been proved to be robust to affine transformations (scale and orientation) and slight changes in illumination. In [8] its robustness has been proved mathematically. Although the descriptor was successfully applied to tasks of classification, matching, and reconstruction for color and gray level images, it lacks application to binary images. Theoretical background and experiments shows that SIFT can also be applied to binary images. We describe in what follows steps of computation of the descriptor and the effect of each step on a binary image description.

### 2.1 Keypoint localization

The most distinctive feature in a binary image is its contour, and more precisely curvatures usually contained in the edges of the image. In order to describe binary shapes we first localize these features or keypoints.

To do so, let us consider an original binary image  $I(n \times m)$ . The image is convolved  $c$  times using a Laplacian of Gaussian filter (we describe how  $c$  is computed below). The expression of Laplacian of Gaussian Filter is given by:

$$LoG(x, y; \sigma) = -\frac{1}{\pi\sigma^4} \left[ 1 - \frac{x^2 + y^2}{2\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \right] \quad (1)$$

where  $x$  and  $y$  are respectively row and column coordinates in the shape to be convolved and  $\sigma$  is the standard deviation of the Gaussian distribution.

Using the Laplacian of Gaussian besides being an accurate edge detector due to its precision in detecting zeros-crossings, has many other advantages. First, the Laplacian of Gaussian is the second derivative of an image convolved with a Gaussian which prevents from noise in the image. Secondly, as the filter is pre-calculated in advance, processing time is reduced. Convolving the image several times allows to define the zeros-crossing over different scales.

Number of convolutions  $c$  is computed as follows:

$$c = \lceil \frac{\log \sigma_c - \log \sigma}{\log k} \rceil \quad (2)$$

where  $\lceil$  is the ceiling function which maps a real number to its smallest following integer,  $\sigma$  and  $\sigma_c$  are respectively initial and final standard deviation of the Laplacian of Gaussian Filter and are given as input parameters.  $\log$  is the natural logarithm of base 2 and  $k$  is the scale step.

At each of the  $c$  convolutions applied to  $I$ ,  $\sigma$  is updated as follows:

$$\sigma_i = \sigma \times k^{i-1} \quad (3)$$

where  $1 \leq i \leq c$  is a counter ranging from 1 to  $c$ . The initial image  $I$  is convoluted with the new  $\sigma$  using the formula 1.

Once the initial image  $I$  is convoluted  $c$  times, local maxima are defined and their coordinates (row, column indices and radius) are returned. The maxima is defined as follows:

Let  $L_i$ ,  $L_{i+1}$  and  $L_{i+2}$  be three adjacent images computed as follow:

$$L_i = LoG(x, y; \sigma_i) * I, \quad L_{i+1} = LoG(x, y; \sigma_{i+1}) * I$$

$$L_{i+2} = LoG(x, y; \sigma_{i+2}) * I$$

The local maxima is first computed on the neighborhood of  $L_{i+1}$  which is the center of the current  $3 \times 3$  tuples of convoluted images ( $L_i, L_{i+1}, L_{i+2}$ ). More precisely, each pixel  $p(x, y)$  in  $L_{i+1}$  is compared to its 8 neighbors. If  $p$  is either minimum or maximum, it is considered as a potential maxima of the inner scale and its coordinates are stored.

Then,  $p$  is compared to its 18 neighbors in a  $3 \times 3$  windows in  $L_i$  and  $L_{i+2}$  centered at  $(x, y)$ .

Comparisons are done the same way as in the inner scale i.e. if the central pixel in  $L_{i+1}$  is either

maximum or minimum over its 18 neighbors defined above, it is considered potential maxima and its coordinates are saved.

Once potential maxima over all scales defined, a further refinement is performed in order to eliminate weak edges. This is done by suppressing from each convoluted image  $L_{i+1}$  potential maxima which are less than an input threshold (measure of stability).

Maxima detected in this step are returned as output of this phase and are called keypoints. Each maxima is defined by its row coordinate  $r$ , column coordinate  $c$  and its radius  $rad$  which is computed as follows:

$$rad = \sigma \times k^{i-1} \times \sqrt{2} \quad (4)$$

Changes in scale allows to see the image across different resolutions. The next step is to describe keypoints detected in this step.

## 2.2 Keypoint description

A binary image can be defined as a matrix which can have only two possible values typically relative to black and white. Depending on the image representation, those two values can be either 0 and 255 or 0 and 1. In a sake of simplicity we use the second representation in this paper.

Description of keypoints localized in the first step is done by first, computing the gradient magnitude and direction of the original image  $I$  around its keypoints and quantifying orientations weighted by the magnitude to form the final descriptor. We describe the whole process in what follows.

First gradient of the original image  $I$ :  $(I_x = (\partial I / \partial x), I_y = (\partial I / \partial y))$  is computed. Hence, Gradient of an image can have either -1, 0 or 1 value in both directions.

Gradient magnitude and Gradient orientation are computed. As gradient orientation is based on the gradient magnitude, 8 Gradient orientations can then take up to eight values:  $-\frac{3\pi}{4}, -\frac{\pi}{2}, -\frac{\pi}{4}, \pi, 0, \frac{3\pi}{4}, \frac{\pi}{2}, \frac{\pi}{4}$ .

Once orientations of the gradient for the whole image computed, the second step is to get for each keypoint its window computed as follows:

$$x_o = \lceil \max(1, r - rad) \rceil, \quad y_o = \lceil \max(1, c - rad) \rceil$$

$$x_e = \lceil \min(r + rad, n) \rceil, \quad y_e = \lceil \min(c + rad, m) \rceil$$

where  $x_o, y_o, x_e, y_e$  are respectively the (row, column) of the beginning respectively end of the keypoint window.

$r, c$  and  $rad$  are row, column coordinates and radius of the keypoint.  $n$  and  $m$  are the two dimension sizes of the image  $I$ .

The next step is to divide that window into 16 blocks arranged in a 4×4 manner (4 blocks horizontally and 4 blocks vertically). Number of pixels in each block depends on the size of the window and is rounded to its nearest integer. Histogram orientation of each block is then computed.

### 3 SHAPE RETRIEVAL

Previous steps allow us to describe binary shapes in a form of a set of keypoints. Each keypoint is a 128×1 vector. Matching two shapes is then performed by matching each keypoint of the first image to one keypoint of the second image by computing minimal distance between them. In the following are the notation used for matching followed by the matching process.

#### 3.1 Notation used

Let  $B$  be a  $n \times 128$  matrix representing all  $n$  keypoints of an image query  $I$  and  $B_a$  be a  $n_a \times 128$  matrix representing all  $n_a$  keypoints of an image  $I_a$  of a dataset. In order to get the best match of  $I$  over all  $I_a$ 's images where  $a$  ranges from 1 to  $s$  ( $s$  is number of images in the dataset), best match of  $B$  to one  $B_a$  is computed and the process is repeated to all remaining images.

#### 3.2 Retrieval by matching keypoints independently

**Step 1:** The first match of  $B$  to  $B_a$  is done by first computing the angles between each keypoint of  $B$  to each keypoint of  $B_a$  and get the minimum value for a best match. Angles between keypoints is computed as follow:

$$\text{Angles}_a = \text{acos}(B * B_a^T) \quad (5)$$

where  $\text{acos}$  is the inverse cosine,  $B_a^T$  is the transpose of  $B_a$  and  $*$  is the dot product of the two matrices.

**Step 2:** Angles will be a  $n \times n_a$  matrix where each line is relative to a keypoint in  $B$  and each column is relative to a keypoint in  $B_a$ . The matrix  $\text{Angles}$  is then sorted column wise in increasing order. The minimum distance is then the first column of  $\text{Angles}$ .

**Step 3:** Step 1 and 2 are repeated to all  $B_a$  where  $a$  ranges from 1 to  $s$  (number of images in the dataset). For each  $B_a$ ,  $\text{Angles}_a$  are concatenated vertically so that to form the matrix  $D$  of all minimum angles between keypoints of  $B$  and  $B_a$  ( $1 \leq a \leq s$ ).  $D$  is then

$$D = \text{Angles}_1 \| \text{Angles}_2 \| \dots \| \text{Angles}_s \quad (6)$$

where  $\|$  is a concatenation operator. Each line in  $D$  represents a keypoint and each column the minimum angle with a keypoint in the  $a$ th image.

**Step 4:** Construct the vote matrix  $F$ .  $F$  takes indices of column for each line in  $D$  where  $D$  is minimum. The vector  $F$  reflects then to which image in the dataset, a keypoint in  $B$  looks more similar. Number of lines in  $F$  is equal to  $n$  ( $n$  is number of keypoints of  $B$ ) and number of columns to  $1$ .

**Step 5:** The last step is to compute number of occurrences of each image index in  $F$  (values in  $F$  which range from 1 to  $s$ ).  $V$  is then a  $s \times 1$  matrix which contains for each image index, number of its occurrences in  $F$ .  $V$  is then sorted in a increasing order and the first index represent the best matching image among all  $s$  images.

**Step 6:** The ancient dataset is updated where the best matching image found in step 5 is removed. If the dataset is empty or number of images( $y$ ) which should be matched is reached, then the retrieval is stopped. Otherwise steps from 1 to 6 are repeated for the query shape and the new dataset.

### 4 SHAPE CLASSIFICATION

Classification aims to classify patterns from a dataset using a set of patterns for training. In our context, the dataset is comprised of images and is divided into two subsets: a training set and a test set. The training set is used to train the classifier and the test set is used for tests. Classification of images using SIFT is very difficult because of the high dimensionality of SIFT feature and the non-fixed feature representation. Classification using a majority vote such as done previously is very time and memory consuming since each key point of each image is stored and compared over all key points of all images of the training set. A good compromise between accuracy and computations will be to have a new representation of image features with a fixed length. The bag of features method seems to be a good solution. In what follows, we describe briefly how we use Bag of Features in classifying binary images using sift. The first step is to quantize SIFT features and build visual vocabulary using K-means and is described in subsection 4.1

Then, for a new instance, the second step is to classify it using one of supervised learning classifiers. We choose the Nearest Neighbor classification and describe it in subsection 4.2.

#### 4.1 Building visual vocabulary using K-means

SIFT features of images of training are concatenated into one  $128 \times d$  matrix  $M$  where  $d = \sum_{i=1}^t n_i$ ,  $n_i$  is number of keypoints of image  $I_i$  and  $t$  is the number of images of the training set.

Rows of the matrix  $M$  are then clustered using  $K - means$  clustering algorithm described in algorithm 1.

Result of clustering is then  $l$  stabilized clusters, each cluster is represented by its center which is a one of the 128 dimensional vector of keypoints and represents the visual vocabulary. Each image of the training and test set is then represented by a histogram where number of bins is  $l$  and each bin stands for number of keypoints nearest to one of the  $l$  centers. Figure 1 represents an example of Bag of Words representation.

#### 4.2 Training and classification using bag of words

After having Bag of Features descriptor of each of training and test images, the distance between Bag of Features of test images and training images is computed, for each test image, the minimal distance is chosen, and label of the nearest training

---

#### Algorithm 1 $K - means$ Algorithm

---

```

Begin -Input  $l$  number of clusters
-Input number of iterations
-Choose initial center of the  $l$  clusters
Repeat
-Compute Euclidean Distance between the feature
and each center
-Assign the feature to the cluster which minimizes the
Euclidean Distance
-Recompute the center of each cluster
UNTIL stabilization of the centers is observed or
number of iterations is reached
End.
```

---

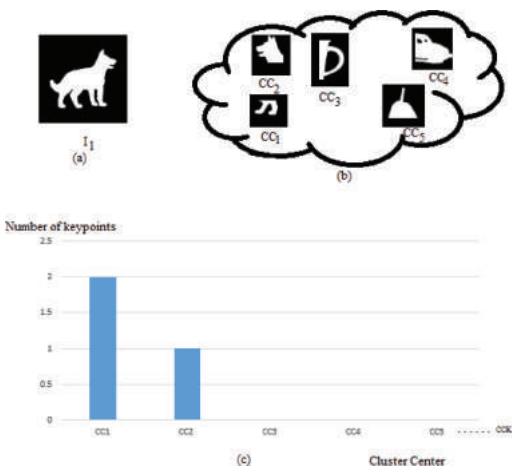


Figure 1. Example of Bag of Features representation of an image  $I$  (a) the image  $I$  to be represented by the visual vocabulary, (b) Visual Vocabulary represented by its centers, (c) Bag of Features of the image  $I$ .

image is attributed to the test image. We use two distances, Euclidean and  $\chi^2$ .  $\chi^2$  distance is described as follows:

Let  $HI_i$  be the feature vector representing the image  $I_i$ . Note that This feature if the normalized histogram of visual words of the image  $I_i$ . Let  $HT_i$  be one of the feature vector representing on of the training examples.  $\chi^2$  distance is described as:

$$\chi^2 = \sum_{i=1}^l \frac{|HI_i - HT_i|^2}{HT_i} \quad (7)$$

where  $l$  is the dimension of the feature vector equal to number of clusters of the  $K - means$  algorithm.

## 5 EXPERIMENTAL RESULTS

### 5.1 Shape retrieval accuracy of the dataset ETH-80

Retrieval experiments are done on the dataset ETH-80 [9]. The dataset comprises 8 categories each of which has 400 images. We choose randomly 10 images of each category which results in 80 images in the dataset prepared for retrieval. Then, we chose randomly an image from each category for retrieval. We use the bullseye score for retrieval accuracy measured as the percent of correct retrievals in the top 20 ranks. Table 1 shows retrieval accuracy of retrieval. The first column represents the query shape, the second column represents shapes retrieved in the order of similarity i.e. similar shapes start in the top left of the column and are ordered line wise. Names of shapes are the same as in the original database<sup>1</sup>. The third column stands for the processing time. The table shows that best rates are for shapes with few curvatures whereas weak rates are observed for shapes with high curvatures. This is because keypoints in curvatures are repetitive in many shapes. However, processing time for shapes with few curvatures is higher. Besides, shapes with few curvatures are most often matched to shapes with few curvatures, examples include: apple, pear and tomato. However, some shapes of high curvatures are confused with shapes with few curvatures. The reason is that, many keypoints in a shape can be matched to only one keypoint in another shape and distance is minimal. The opposite is also true. One solution could be to match a keypoint in a query shape to only one keypoint of a shape in the dataset. We let this improvement to a future work.

<sup>1</sup><http://people.csail.mit.edu/jjl/libpmk/samples/eth.html>

Table 1. Retrieval accuracy on ETH-80 dataset.

Query image	Bullseye score (%)	Processing time (ms)
Apple20	100	135.58
Car4	50	53.69
Cow2	50	55.67
Cup15	100	133.21
Dog31	50	88.36
Horse20	80	82.72
Pear20	100	87.36
Tomato36	100	107.81

Table 2. Overall classification accuracy on ETH-80 dataset by varying size of training set and test set.

Training %	Test %	Accuracy %	
		Binary	RGB
20	80	83.01	90.22
40	60	93.91	96.79
60	40	99.03	98.71

Table 3. Overall classification accuracy on ETH-80 dataset by varying size of training set and fixing size of test set.

Training %	Test %	Accuracy %	
		Binary	RGB
10	30	77.77	84.61
20	30	84.82	88.35
30	30	90.49	93.05
40	30	92.20	94.87
50	30	94.23	95.51
60	30	96.68	97.75
70	30	99.14	99.25

## 5.2 Classification accuracy of the dataset ETH-80

Experiments on the dataset ETH-80 are divided into two parts. The first part concern classification accuracy of RGB images, the second part concerns classification accuracy of the same binary images.

Extensive experiments of ETH-80 are done by dividing the whole dataset into training and classification. Division is done following several percentages from 10 to 70% where this percentage represents ratio of training set with the whole dataset. In a first experiment, we vary size of both training and test set where sum of their two ratios is equal to 100%. By doing this, we insure not to include images of the training set into the test set. Table 2 represents accuracy of this variation.

In a second experiment, we choose a small ratio of the test set (30%) and fix it for all percentages of training. Table 3 represents this experiment. Size of the training set in both cases do not exceed 70% which represent the classical percentage used in most classification experiments (2/3 training and 1/3 test). We analyze classification accuracy of RGB images and their binary masks (binary images) using the same parameters of the code book (same number of iterations of K-means clustering, same number of features in the visual vocabulary, same distance measure).

Table 2 and 3 show that classification accuracy of RGB images in most cases outperform the one of the binary images of the original RGB images. However, the difference is slight and in one case, classification of binary images outperformed classification of RGB images. In order to analyze more deeply the classification results, we moreover generate for each ratio confusion matrices. We generate more than 60 confusion matrices by varying the size of both training and test set, however, we kept only the most significant ones which correspond to ratios of Table 2. Confusion matrices show that classification accuracy is enhanced in RGB images compared to binary images, however, the difference is still slight. Also, in some cases, classification of RGB images. Some interesting points can be seen in the confusion matrices: Confusion of RGB images can occur even when the shape is different, example of such a confusion can be seen in Figure 2 where horses have been confused with cows. This confusion is less apparent in the binary images and accuracy of horse classification can reach 100% for binary images while it does not for RGB images. Such errors occurred for horses with same colors as cows and dogs and when some animals have same skin (texture). The same observation can be made for classes “Dog” and “Cow”. For all other classes, classification accuracy of RGB images is slightly higher than classification accuracy of the same binary images.

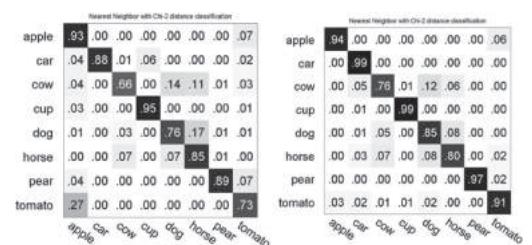


Figure 2. Confusion matrices for ETH-80 with training size 20%, test size 80% (right matrix binary images, left matrix RGB images).

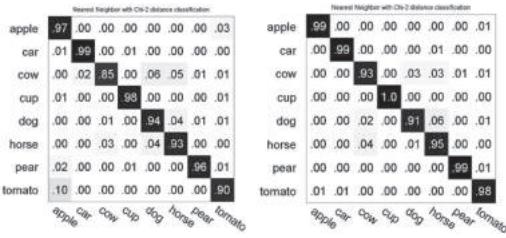


Figure 3. Confusion matrices for ETH-80 with training size 40%, test size 60% (right matrix binary images, left matrix RGB images).

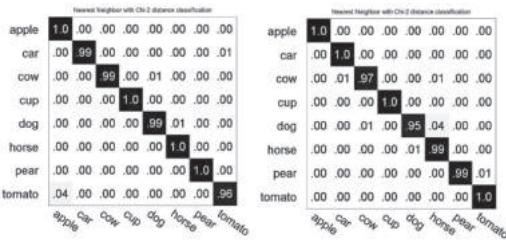


Figure 4. Confusion matrices for ETH-80 with training size 60%, test size 40% (right matrix binary images, left matrix RGB images).

## 6 CONCLUSION

We have presented in this work the Scale Invariant Feature Transform SIFT feature and its application to binary images. We showed theoretically how the feature can be applied to binary image and enhance the study with experiments both on retrieval and classification. The study concludes that SIFT can be applied to binary shapes. We aim in further works to improve the matching and apply the approach on binary masks derived from background subtraction (15).

## REFERENCES

- [1] Veltkamp, R.C., Shape matching: similarity measures and algorithms, Shape Modeling and Applications, SMI 2001 International Conference on. 2001.

- [2] Learning Context-Sensitive Shape Similarity by Graph Transduction, Xiang Bai and Yang, Xingwei and Latecki, L.J. and Wenyu Liu and Zhuowen Tu, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2010.
- [3] Xiang Bai and Cong Rao and Xinggang Wang, Shape Vocabulary: A Robust and Efficient Shape Representation for Shape Matching, Image Processing, IEEE Transactions on, 2014.
- [4] Haibin Ling and Jacobs, D.W., Using the inner-distance for classification of articulated shapes, Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on,
- [5] J.F. Nunes, P.M. Moreira, J.M.R.S. Tavares., Simple and fast shape based image retrieval, Computational Vision and Medical Image Processing: ViPIMAGE 2011.
- [6] Bouagar, Saliha and Larabi, Slimane, Efficient descriptor for full and partial shape matching Multimedia Tools and Applications, 2014.
- [7] Lowe, David G., Distinctive Image Features from Scale-Invariant Keypoints, Int. J. Comput. Vision, 2004.
- [8] Jean-Michel Morel and Guoshen Yu, Is SIFT scale invariant?, Inverse Problems and Imaging, 2011.
- [9] Leibe, B. and Schiele, B., Analyzing appearance and contour based methods for object categorization Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on.
- [10] Belongie, S. and Malik, J. and Puzicha, J., Shape matching and object recognition using shape contexts, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2002.
- [11] Haibin Ling and Jacobs, D.W., Shape Classification Using the Inner-Distance, Pattern Analysis and Machine Intelligence, IEEE Transactions on, 2007,
- [12] Bouagar, Saliha and Larabi, Slimane, Efficient descriptor for full and partial shape matching, Multimedia Tools and Applications, 2014.
- [13] Lin, Wei-Syun and Wu, Yi-Leh and Hung, Wei-Chih and Tang, Cheng-Yuan, A Study of Real-Time Hand Gesture Recognition Using SIFT on Binary Images, Advances in Intelligent Systems and Applications—Volume 2, 2013.
- [14] Xiang Bai and Xingwei Yang and Longin Jan Latecki and Wenyu Liu and et al., Learning Context-Sensitive Shape Similarity by Graph Transduction, 2010.
- [15] Setitria, I. and Larabi, S., Pattern Recognition (ICPR), 2014 22nd International Conference on, Background Subtraction Algorithms with Post-processing: A Review, 2014.

# Time-constrained detection of colored objects on raw Bayer data

António J.R. Neves, Alina Trifan & José Luís Azevedo

*Department of Electronics and Telecommunications Engineering, Intelligent Systems and Intelligent Robotics Laboratory, Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro, Portugal*

**ABSTRACT:** In most image processing applications, the raw data acquired by an image sensor passes through a demosaicing algorithm in order to obtain a full color image. The demosaicing step can take time and resources that are of great importance when developing real-time applications. Most of the times, this is done with single purpose of rendering the images into a viewable format for humans. However, most of nowadays sensors allow the retrieval of images in raw format and processing the data as is. In this paper we present a study on the direct usage of raw Bayer data for real-time color object detection, in the scenario of autonomous robot soccer. The experimental results that we provide prove that the efficiency of the chosen object detection algorithm is not lowered and that several improvements of the autonomous robotic vision system used in this study can be noted: the bandwidth needed to transmit the data between the digital camera and the PC is reduced, allowing, in some models, to increase the maximum frame rate. Moreover, a decrease in the delay between the perception and action of the autonomous robot was measured, when using raw image data for processing the environment.

## 1 INTRODUCTION

In digital cameras, a Color Filter Array (CFA) is a mosaic of tiny color filters placed over the pixels of an image sensor to capture color information. They are needed because the typical photosensors detect light intensity with little or no wavelength specificity, and therefore cannot separate color information (Nakamura 2005).

A Bayer filter mosaic is a type of CFA for arranging RGB color filters on a square grid of photosensors. Its particular arrangement is used in most single-chip digital image sensors used in digital cameras, camcorders, and scanners to create a color image. The filter pattern is 50% green, 25% red and 25% blue, usually called BGGR, RGBG, GRGB, RGGB, etc. depending on the position of the filters.

In most of the image processing applications, the raw data acquired by the image sensor passes through a demosaicing algorithm in order to obtain a full color image (Kimmel 1999). This is a digital image process used to reconstruct a full color image from the incomplete color samples output from an image sensor overlaid with a CFA. There are several well known algorithms, namely Simple Interpolation (Malvar et al. 2004), Variable Number of Gradients (Chang et al. 1999), Pixel Grouping (Chang and Tan 2006) or Adaptive homogeneity-directed (Hirakawa et al. 2005).

Most modern digital cameras acquire images using a single image sensor overlaid with a CFA, so demosaicing is part of the processing pipeline required to render these images into a

viewable format. However, in most of them it is possible to retrieve images in a raw format allowing the user to demosaic them using software, rather than using the camera's built-in firmware.

An example of a raw image acquired by the camera and the corresponding interpolated RGB version is presented in Figure 1. As we can see in the image on the left, the raw image data is a single channel image where the value of each pixel corresponds to the information about a specific color, depending on the position of the filters in the sensor, as described before.

In this paper we study the effect of using directly the raw data from an image sensor for real-time

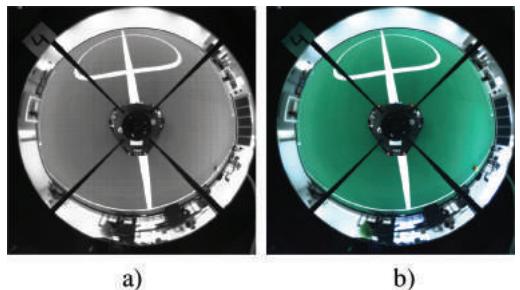


Figure 1. In a) an example of a RAW image acquired by the camera (grayscale image containing the RGB information directly from the sensor—Bayer pattern). In b) the image a) after the interpolation of the missing information for each pixel in order to obtain a complete RGB image.

color object detection. The experiments are conducted on an autonomous soccer robot available at our research unit (Neves et al. 2010, Silva et al. 2011)<sup>1</sup>. We provide algorithms for image processing and object detection using the raw data directly included in a computer vision library developed within our research group and denoted UAVision (Trifan et al. 2015)<sup>2</sup>. Based on the experimental results obtained and presented in this paper, we can conclude that the performance of the object detection is not affected by the use of the raw image data. Moreover, by using directly the raw data, we reduce the bandwidth needed to transmit the data between the digital camera and the processing unit, allowing, in some camera models, to increase the maximum frame rate available due to processing and bandwidth limitations when the demosaicing is performed on the camera. We provide also experimental results showing a decrease in the delay between the perception and action of the autonomous robot, when using raw image data for object detection.

This paper is structured in 4 sections, first of them being this Introduction. Section 2 presents the algorithm implemented for colored object detection used in this study. We present experimental results in Section 3. Section 4 concludes the paper.

## 2 COLORED BALL DETECTION ALGORITHM

In order to implement a time efficient vision system it is necessary to develop all the algorithms and data structures to be used. Taking this into consideration, we developed a computer vision library denoted UAVision (Trifan et al. 2015) that incorporates software for image acquisition from digital cameras supporting different technologies, as well as several modules for camera calibration, blob formation, which stands at the basis of the object detection, and image transfer using TCP communications.

For fast color classification, color classes are defined through the use of a Look-Up Table (LUT). A LUT represents a data structure, in this case an array, used for replacing a runtime computation by a basic array indexing operation.

This approach has been chosen in order to save significant processing time. The LUT can be applied to images in full RGB color or raw image data (Bayer format) and allows the conversion to an index image (image of labels).

The table consists of 16,777,216 entries ( $2^{24}$ , 8 bits for R, 8 bits for G and 8 bits for B) with

one byte each. For example, considering the RGB color space, the index of the table to be chosen for a specific triplet is obtained as

$$idx = R \ll 16 | G \ll 8 | B,$$

being the  $\ll$  the left shift operator and  $|$  the bitwise *OR*. The raw data (Bayer pattern) is also RGB information, however, the positions in the image where the R, G and B information are located change. These positions are also calculated only once using a LUT in order to obtain a faster access during color classification.

Each bit in the table entries expresses if one of the colors of interest (in this case: white, green, blue, yellow, orange, red, blue sky, black, gray—no color) is within the corresponding class or not. A given color can be assigned to multiple classes at the same time. For classifying a pixel, first the value of the color of the pixel is read and then used as an index into the table. The 8-bit value then read from the table is called the “color mask” of the pixel.

Taking into consideration a CFA BGGR, the LUT for the position of the R, G and B components for each pixel is built based on the following algorithm:

```
// p is the linear position on the image
for(p = 0 ; p < cols * rows ; p++)
{
    row = p / cols
    col = p % cols;
    if(row % 2 == 0) // even lines
    {
        if(col % 2 == 0) // odd columns
        {
            R = p;
            G = p + 1;
            B = p + cols + 1;
        }
        else // even columns
        {
            G = p;
            r = p - 1;
            b = p + cols;
        }
    }
    else // odd lines
    {
        if(col % 2 == 0) // odd columns
        {
            g = p ;
            r = p - cols;
            b = p + 1;
        }
        else // even columns
        {
            g = p - 1;
            r = p - cols - 1;
            b = p;
        }
    }
}
```

<sup>1</sup><http://robotica.ua.pt/CAMBADA>

<sup>2</sup><http://sweet.ua.pt/an/uavision/>

The pipeline of the ball detection procedure can be resumed as follows: after having acquired an image, this is transformed into an image of labels using the LUTs previously built. This image of color labels, also designated as index image, will be the basis of all the processing that follows.

The index image is scanned using one of three types of scan lines or search lines (circular, radial or linear). The information about colors of interest or transitions between colors of interest is Run Length Encoded (RLE). Blobs are formed by merging adjacent RLEs of a color of interest.

The blobs that are of the color of the ball have to meet the following validation criteria before being labeled as ball. First, a mapping function, that has been experimentally designed, is used for verifying a size-distance ratio of the blob (from the robot). This is complemented by a solidity measure and a width-height ratio validation, taking into consideration that the ball has to be a round blob. The validation is done taking into consideration the detection of the ball even when it is partially occluded. More details about all the steps of this algorithm can be found in (Trifan et al. 2015).

### 3 EXPERIMENTAL RESULTS

In order to experimentally study the effect of using directly the raw data from an image sensor to detect colored objects, several experiments were conducted using an autonomous soccer robot. These robots use a catadioptric vision system formed by an Ethernet camera point towards a mirror. This vision system should detect, in useful time, the ball, the limits of the field and the field lines, the goal posts and the other robots that are on the field.

In Figure 2 we present the ball position detected by the developed vision system after the robot performs a kick, considering two different configurations. In one case, the vision system was processing directly the raw image data acquired from the camera and in the other one the full RGB data was processed. As we can see in both graphics, the ball is correctly detected for all its trajectory, both when we process the raw image data and when the processing is performed in the full RGB image. Moreover, we can see that the robot can detect the ball until more than 10 meters, which is an interesting and important result in this application (see an example of the images acquired by the robots with some marks over the detected objects in Fig. 3).

Two game scenarios have been tested using the CAMBADA autonomous mobile robots. In each one, two experiments have been made. In one experiment, the vision system processed the full RGB image after demosicing for object detection (balls, field lines used in robot localization

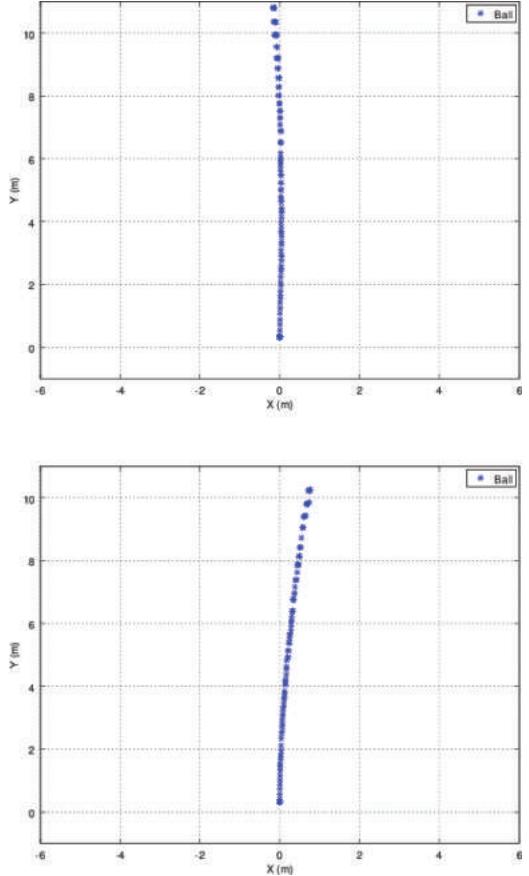


Figure 2. Ball detection after being kicked by the robot. On top, a graphic showing the ball detection on full RGB images after demosicing and on bottom the ball detection on raw image data.

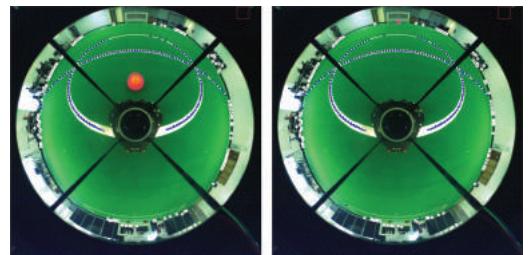


Figure 3. Images acquired by the omnidirectional vision system with the result of the color-coded object detection. On the left, the ball was at one meter of the robot. On the right, the ball was at 10 m from the robot. Blue circles mark the white lines, the white circles mark the black obstacles and the magenta circles mark the orange blobs that passed the validation thresholds and are considered ball.

and obstacles). In the other experiment, the raw image data was processed.

In Figure 4 we present a graphic with the result of the ball detection when the ball is stopped at a given position (the central point of the field, in this case) while the robot is moving. The graphic shows a consistent ball detection with both configurations while the robot is performing a tour around the field. The field lines are also properly detected, as it is proven by the correct localization of the robot in all the experiments.

The second scenario that has been tested is illustrated in Figure 5, where both the robot and the ball are moving. The robot is making a tour around the soccer field, while the ball is being sent across the field. In all these experiments, no false positives were observed and the ball has been detected in more than 90% of the images acquired

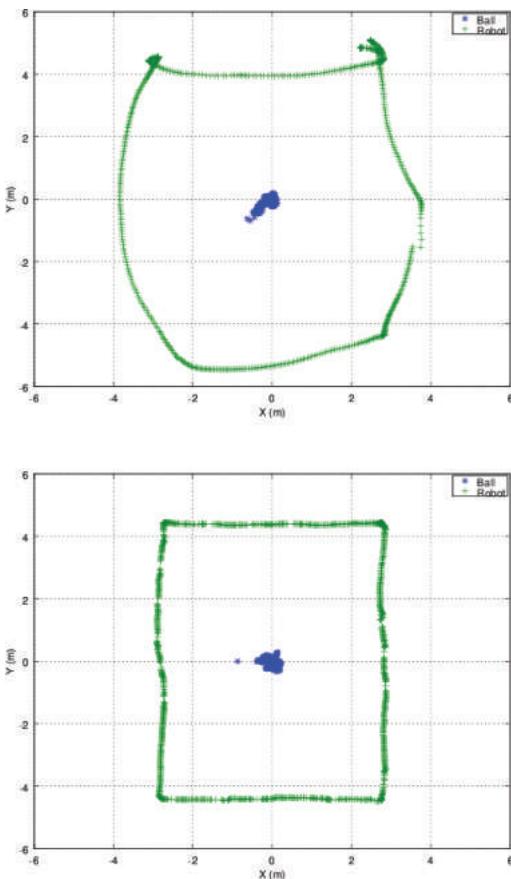


Figure 4. Ball detection with the robot moving around the field and the ball stopped in the middle of the field. On top, a graphic showing the ball detection on full RGB images after demosicing and on bottom the ball detection on raw image data.

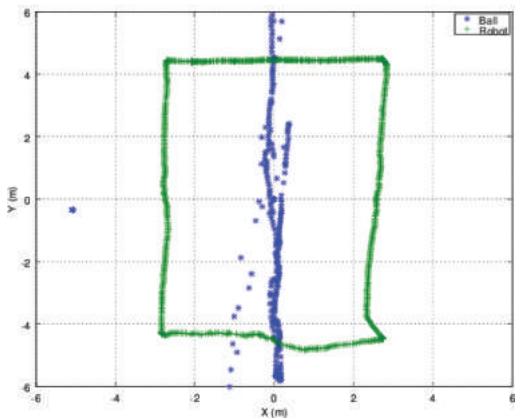
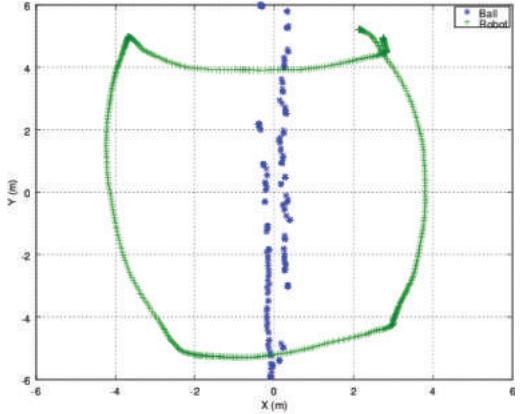


Figure 5. Ball detection when both the robot and the ball are moving. On top, a graphic showing the ball detection on full RGB images after demosicing and on bottom the ball detection on raw image data.

by the robot. Most of the times when the ball was not detected was due to the fact that it was hidden by the bars that hold the mirror of the omnidirectional vision system.

We can observe a different robot trajectory in the tours performed by the robots when the vision system processed the full RGB images, when comparing with the scenario when the vision system processed the raw image data. Since the control parameters were adjusted to the vision system processing the raw image data, the robot has a different behavior in this situation. This is also an important observation that we will explain right after.

In addition to the good performance in the detection of objects, both in terms of number of times that an object is visible and detected and in terms of error in its position, the vision system must also

perform well in minimizing the delay between the perception of the environment and the reaction of the robot. It is obvious that this delay depends on several factors, namely the type of the sensor used, the processing unit, the communication channels and the actuators, among others.

Some digital cameras do not provide frame rates above 30 fps in RGB, but allow the access to the raw data of the sensor at higher frame rates, mostly due to processing and bandwidth restrictions when the demosaicing algorithm runs on the camera. Moreover, it is expected that the conversion between the raw sensor data and a specific color space in the camera or in the device driver on the computer side will take some time. Therefore, we performed the following experiments for measuring the delay between perception and reaction of the robot accessing directly the raw data.

The delay measurement system used in this work is based on a microcontroller board connected to the vision processing PC through a USB port (Fig. 6). The microcontroller periodically turns on a LED, during a fixed period of time, and measures the time that the whole vision system takes to detect it in the acquired image. The measured time includes camera acquisition time, image data transfer time and vision processing time. In the setup used for these experiments, the camera is configured with an exposure time of 5 ms and Gigabit Ethernet connection that can provide for our setup (cable, network card, computer and operating system) 80 MB/s, far from the theoretical 125 MB/s, has been used. The vision software is a simplification of the one described in (Trifan et al. 2015) since we do not have to detect the obstacles and white lines and takes, on average, 6 ms to detect the led. The vision system detects the led on and when it happens, sends a message to the microcontroller.

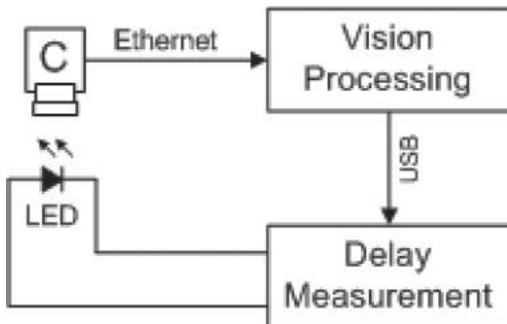


Figure 6. The main blocks used in our measurement setup. The microcontroller periodically turns on a LED, during a fixed period of time, and measures the time that the whole vision system takes to detect it in the acquired image.

This is the normal working mode of every robot that uses vision as main sensor.

An important result is the delay between the perception and reaction of the robot. We compared the delay when acquiring frames in RGB and when accessing the raw data. Based on the graphics presented in Figure 7, we measured an improvement of 23% of the reaction time at 50 fps when using raw data. If the camera used by the robot could deliver raw data, these results prove that it is more advantageous.

Using the LUTs described below and the scanlines approach for blob detection as described in (Trifan et al. 2015) the delay will not be influenced by the scene complexity, as presented in the results of that work. Using a different algorithm

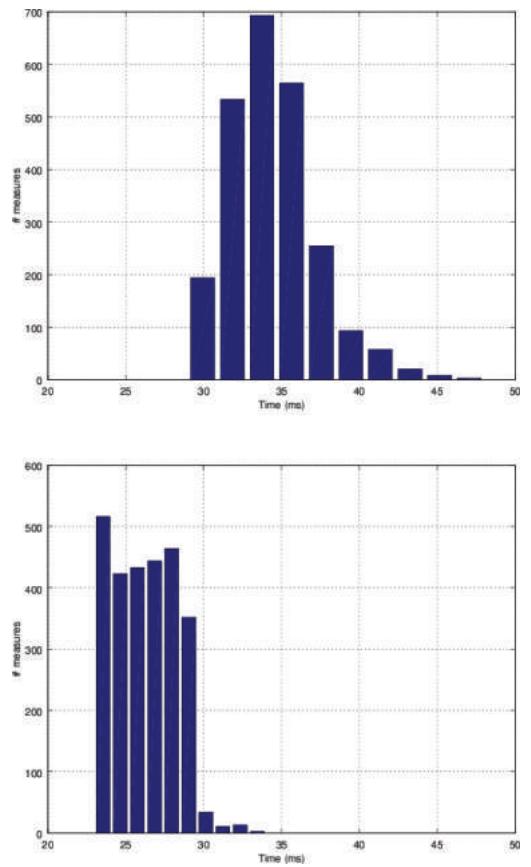


Figure 7. Histograms showing the delay between perception and action on the vision system of the CAMBADA robots. On top, the camera is acquiring RGB frames and is working at 50 fps (average = 34.2 ms, max = 48 ms, min = 29 ms). On the bottom, the camera acquiring single channel frames, returning the raw data of the sensor (Bayer pattern) (average = 26.4 ms, max = 34 ms, min = 23 ms).

in another application, the delay time will be proportional to the processing time of the algorithm used and the scene complexity. Regarding external conditions, namely illumination or other factors, we use in our robots algorithms that guarantee the same properties of the acquired images (Neves et al. 2014).

The processing time of each captured image, both in raw data and full RGB color, using the developed vision system built using the UAVision library is considerably low. The full execution of the pipeline, from image acquisition to object detection after validation (balls, lines and obstacles) only takes at most 10 ms, allowing thus a framerate of 100 fps if the digital camera supports it and if other control or artificial intelligence processes can run in parallel. The time results have been obtained on a computer with a Intel Core i5-3340M CPU @ 2.70 GHz 4 processor, processing images with resolutions of  $1024 \times 1024$  pixels. In the implementation of these vision systems we did not use multi-threading. However, image classification as well as the next steps can be parallelized if needed.

## 4 CONCLUSIONS

Most of digital cameras used in robotic applications allow the retrieval of images in raw format and in some other camera models the demosaicing algorithm is performed by the device driver in the processing unit. However, as far as we know, no previous study have been done before exploring the direct usage of raw data in colored object detection.

In this paper we have presented a study on the direct usage of raw Bayer data for real-time color object detection and the experimental results that we provide prove that the efficiency of colored object detection algorithm is not lowered and several improvements can be noted: the bandwidth needed to transmit the data between the digital camera and the PC is reduced; a the delay between the perception and action of the autonomous robot, when using raw image data for processing the environment was improved by more than 20%.

In what concerns future work, the next step will be to explore the usage of raw Bayer data in feature detection and classification of objects that do not have a dominant color.

## ACKNOWLEDGEMENTS

This work was developed in the Institute of Electronic and Telematic Engineering of University of

Aveiro and was partially supported by FEDER through the Operational Program Competitiveness Factors—COMPETE and by National Funds through FCT—Foundation for Science and Technology in a context of a PhD Grant (FCT reference SFRH/BD/85855/2012) and by National Funds through FCT—Foundation for Science and Technology, in the context of the projects UID/CEC/00127/2013 and Incentivo/EEI/UI0127/2014.



## REFERENCES

- Chang, E., S. Cheung, & D.Y. Pan (1999). Color filter array recovery using a threshold-based variable number of gradients. In *Proceedings of SPIE*, Volume 3650, pp. 36.
- Chang, L. & Y. Tan (2006). Hybrid color filter array demosaicing for effective artifact suppression. *J. Electronic Imaging* 15(1), 013003.
- Hirakawa, K., S. Member, & T.W. Parks (2005). Adaptive homogeneity-directed demosaicing algorithm. *IEEE Trans. Image Processing* 14, 360–369.
- Kimmel, R. (1999). Demosaicing: image reconstruction from color ccd samples. *Image Processing, IEEE Transactions on*.
- Malvar, H.S., L. wei He, & R. Cutler (2004). High-quality linear interpolation for demosaicing of bayer-patterned color images. In *Proceedings of the IEEE International Conference on Speech, Acoustics, and Signal Processing*.
- Nakamura, J. (2005). *Image Sensors and Signal Processing for Digital Still Cameras*. Boca Raton, FL, USA: CRC Press, Inc.
- Neves, A., J. Azevedo, N.L.B. Cunha, J. Silva, F. Santos, G. Corrente, D.A. Martins, N. Figueiredo, A. Pereira, L. Almeida, L.S. Lopes, & P. Pedreiras (2010, January). *CAMBADA soccer team: from robot architecture to multiagent coordination*, Chapter 2, pp. 19–45. Vienna, Austria: I-Tech Education and Publishing.
- Neves, A.J.R., A. Trifan, & B. Cunha (2014). Self-calibration of colormetric parameters in vision systems for autonomous soccer robots. In S. Behnke, M.M. Veloso, A. Visser, and R. Xiong (Eds.), *RoboCup 2013: Robot Soccer World Cup XVII*, Volume 8371 of *Lecture Notes in Computer Science*, pp. 183–194. Springer.
- Silva, J., N. Lau, A.J.R. Neves, J. Rodrigues, & J.L. Azevedo (2011, March). World modeling on an MSL robotic soccer team. *Mechatronics* 21(2), 411–422.
- Trifan, A., A.J.R. Neves, B. Cunha, & J.L. Azevedo (2015). Uavision: a modular time-constrained library for soccer robots. In R.A.C. Bianchi, H.L. Akin, S. Ramamoorthy, and K. Sugiura (Eds.), *RoboCup 2014: Robot Soccer World Cup XVIII*, Volume 8992 of *Lecture Notes in Computer Science*, pp. 490–501. Springer.

# Image scanning techniques for speeded-up color object detection

Alina Trifan, António J.R. Neves & Bernardo Cunha

*Department of Electronics and Telecommunications Engineering, Intelligent Systems and Intelligent Robotics Laboratory, Institute of Electronics and Informatics Engineering of Aveiro, University of Aveiro, Portugal*

**ABSTRACT:** Nowadays there are many computer vision algorithms dedicated to solve the problem of object detection, from many different perspectives. Many of these algorithms take a considerable processing time even for low resolution images. Research areas that are more concerned about real-time capabilities of a system, such as the case of robotics for example, are focusing on adapting the existing algorithms or developing new ones that cope with time constraints. Color object detection algorithms are often used in time-constrained application since the color of an object is an important clue for its detection. In addition, color segmentation algorithms have a smaller computational cost than generic object detection algorithms. Most of the color object detection algorithms include a step of image scanning in search of pixels of the color of interest. In this paper we present a study on the influence of an image scanning pattern while searching for pixels of a certain color versus the image resolution in the results of three color object detection algorithms. We present results comparing different scanning approaches versus different resolution images for color object detection algorithms based on blob formation, contours detection and region growing clustering. Our experiments show that using search lines for color segmentation, followed by blob formation provides faster and more accurate results.

## 1 INTRODUCTION

Object detection with small computational cost and processing time is a necessity of some of the most diverse domains, such as: robotics, video surveillance and traffic analysis, among others. With current advances in technology and the decrease of prices in image sensors and video cameras, nowadays digital vision systems can acquire images at higher resolutions than 1 Megapixel and can work at high frame rates. This implies a considerable data size that has to be processed in a very small period of time when real-time operations and data processing are needed. For example, for a camera working at 50 fps and acquiring images of  $1024 \times 1024$  pixels, the total processing time of any given real-time application should not exceed 20 ms.

A common solution for reducing the processing time of an object detection algorithm is to sample the data, to process only parts of an image or to acquire and process smaller resolution images. However, these procedures come with a cost in the accuracy of the object detection results. By not processing all the available information, one might leave out important data and introduce assumptions that can tackle the final result of the detection algorithms. In this paper we present a study based on the real-time scenario of robotic soccer. Using sequences acquired by a robotic soccer player

(Neves et al. 2010)<sup>1</sup> and the digital vision system implemented using the UAVision computer vision library<sup>2</sup> (Trifan et al. 2014), we present results of color object detection algorithms using optimal configurations of scan lines versus different image resolutions. Moreover, we complement these results by comparing their detection efficiency and time processing with two other color detection algorithms based on contours detection and region growing clustering, for which some of the methods of the OpenCV Library<sup>3</sup> have been used.

This paper is structured in 5 sections, first of them being this Introduction. Section 2 presents the approach taken for image scanning. In Section 3 we succinctly present the algorithms and computer vision libraries used in this study. We present experimental results in Section 4 and Section 5 concludes the paper.

## 2 SCANLINES

Most algorithms for color object detection, just as the name implies, perform a color segmentation step in which pixels of the same color are grouped together.

---

<sup>1</sup><http://robotica.ua.pt/CAMBADA>.

<sup>2</sup><http://sweet.ua.pt/an/uavision/>.

<sup>3</sup><http://docs.opencv.org>.

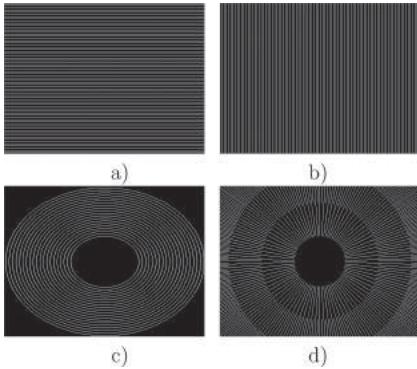


Figure 1. Examples of different types of scanlines: a) horizontal scanlines; b) vertical scanlines; c) circular scanlines; d) radial scanlines.

For color segmenting an image, most algorithms use the approach of scanlines. A scanline, or a search line, is a list of pixel positions in the image, whose shape depends on its type. That is, an image can be scanned in search for a color of interest under different directions, for example either horizontally, vertically, radially or circularly. In the robotic soccer competitions promoted by the RoboCup Federation<sup>4</sup> the environment is color coded and many participating research teams use the approach of scanlines while searching for one of the colors of interest (Trifan et al. 2012), (Neves et al. 2008), (Rofer et al. 2011), (Garcia et al. 2012), (Akin et al. 2014).

For the purpose of this study, we have used sequences acquired by the vision system of the robotic soccer agents of CAMBADA team and the UAVision library (Trifan et al. 2014). UAVision is a modular computer vision library developed within our research group for color object detection. We have used the **Scanlines Module** of the library in order to test the impact of the scanning pattern and sparseness in the result of the color object detection algorithms.

In Figure 1 the three different types of scanlines supported by this module are illustrated: linear, circular and radial.

### 3 ALGORITHMS

Three different algorithms for color object detection have been used for the purpose of this study. First, we used the **Blob Module** of the UAVision library, that uses scanlines to scan the image and then the color information is run-length encoded by the **RLE Module**. The **Scanlines Module** allows the definition of an object that is a vector of scanlines.

<sup>4</sup><http://www.robocup.org/>.

The amount and characteristics of these scanlines depend on the parameters (start and end positions, distance—vertical, horizontal or angular between scanlines, etc).

Each *scanline* object is used by the **RLE Module** that produces a list of occurrences of a specific color. Depending on the number of *scanline* objects created in a specific vision system, we will have the same number of RLE lists produced. Then, the **Blob Module** is able to use all these lists to combine RLEs and form blobs based on several features that are presented with more detail in (Trifan et al. 2014).

The OpenCV library (Bradski and Kaehler 2008) provides two alternative algorithms that have been used in this study for comparison of the results. The first algorithm that has been used is a region growing algorithm (Gonzalez and Woods 2001), in which, using methods from the OpenCV library, the image is scanned horizontally and every time a pixel of the color of interest is found, it is considered to be the seed for the clustering algorithm. The second approach was using the same library for contour detection (Canny 1986) and scanning the pixels inside each contour. If the pixels belonging to a given contour are of the color of interest, that contour is validated as being a blob of the color of interest.

For all of the three algorithms, for fast color classification, color classes are defined through the use of a Look-Up Table (LUT). A LUT represents a data structure, in this case an array, used for replacing a runtime computation by a basic array indexing operation. Each bit in the table entries expresses if one of the colors of interest (in the case of robotic soccer: white, green, blue, yellow, orange, red, blue sky, black, gray—no color) is within the corresponding class or not. A given color can be assigned to multiple classes at the same time. For classifying a pixel, first the value of the color of the pixel is read and then used as an index into the table. The 8-bit value then read from the table is called the “color mask” of the pixel. This image of color labels (Fig. 2 (b)), will be the basis of all the processing that has been previously described. Using an image of labels not only saves computational time but it also introduces less noise than using for example a grayscale version of the original image.

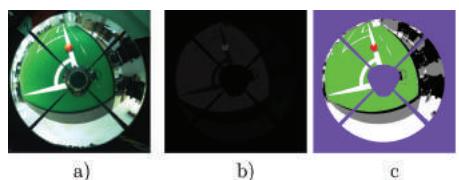


Figure 2. a) Original image. b) Corresponding 8-bit image of color labels. c) “Painted” image of labels for visual understanding purpose only.

## 4 EXPERIMENTAL RESULTS

In this study we present results of the three algorithms used for color object detection and we evaluate the effect of the scanning pattern and sparseness, as well as image resolution in the accuracy and processing time of the detection.

We focus the results on the ball detection since the obstacle and field line detections are only supported by the UAVision library. The region growing and contour detection algorithms are not suitable for more generic object detection since for example field lines do not have a contiguous contour, nor can be merged into a colored cluster.

In Table 1 we present the time processing of each of the algorithms previously described. We have tested several sequences at three different resolutions:  $1024 \times 1024$ ,  $512 \times 512$  and  $256 \times 256$  pixels. After having performed several iterative runs of the object detection algorithm based on

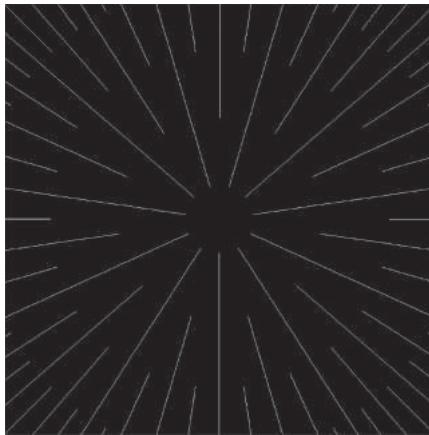


Figure 3. Illustration of the four levels of radial scanlines. In this image, there are only 90 scanlines distributed on four levels for a better understanding of each of the levels. 2880 scanlines are used for the detection of the ball in this study, but with this number of scanlines, the reader cannot clearly distinguish the four levels.

Table 1. Average and maximum time processing, in milliseconds, for each of the three algorithms studied in this paper.

	Blobs		Contours		Region growing	
	Mean	Max	Mean	Max	Mean	Max
$1024 \times 1024$	8.72	10	17.16	35	8.86	21
$512 \times 512$	2.9	7	5.16	16	2.03	7
$256 \times 256$	1	5	1.02	3	1	1

blob detection, we have reached the conclusion that the use of radial scan lines provide a better result in terms of ball accuracy. For this experiment, we have used a configuration of the radial scan lines of 2880 scanlines, disposed on four levels 3. This configuration has been considered optimal by the authors since it is the best compromise between number of visited pixels and number of repeated pixels. For the other two algorithms, we perform a linear (horizontal) scan in search for pixels of interest.

The table that we are presenting shows that at any of the chosen resolutions, the blob detection algorithm performs faster and the maximum processing time of a frame that was measured was 10 ms, while for example, the contour detection algorithms can take a maximum time of 35 ms per frame.

Table 2 shows the maximum distances at which the ball is being detected by each of these algorithms. The blob detection algorithms achieves correct results for balls found at a distance of up to 12 m, even for the smallest chosen resolution, while the results of the contours detection algorithm and region growing clustering algorithm slightly degrade with distance.

Table 3 presents the percentages of correct ball detection for each of the three algorithms. In the video sequences that have been used for tests the robot rotates around itself and the ball is placed in front of it at distances from 1 m to 12 m. These sequences have a total of 5040 frames, out of which, the ball is present in 4580 frames. The situation in which the ball is not present in a frame is due to the hardware architecture of the vision system (Fig. 4).

When the ball is aligned with any of the four supportive bars of the vision system, it cannot be seen in the image as it is fully hidden by the bar. The blob detection algorithm has a detection rate of 100% for full resolution images and the decrease of the resolution is followed by a slight decrease in the performance

Table 2. Maximum distance from the robot at which the ball is correctly detected by each of the algorithms.

	Blobs	Contours	Region growing
$1024 \times 1024$	12 m	12 m	11 m
$512 \times 512$	12 m	12 m	11 m
$256 \times 256$	12 m	11 m	11 m

Table 3. Percentages of correct ball detection for the three algorithms.

	Blobs	Contours	Region growing
$1024 \times 1024$	100%	68.5%	73.1%
$512 \times 512$	95.7%	63.7%	69%
$256 \times 256$	85.7%	24.6%	53.9%



Figure 4. Close-up of the vision system hardware of the robots.

Table 4. Percentages of correct ball detection for the three algorithms for balls found at 6 m or more from the robot.

	Blobs	Contours	Region growing
1024 × 1024	100%	63.8%	70%
512 × 512	91%	62.8%	64.7%
256 × 256	70%	16.3%	22.9%

of the algorithm. Out of the three algorithms, the one based on contour detection is not only the heaviest from a computational point of view, but it also performs worse than the other two. Moreover, for the smallest resolution, the contours algorithm can only detect the ball in a quarter of the frames.

All three algorithms can detect the ball up to distances of 11 m (12 m in some cases). However, for some of the algorithms, the detection rate at longer distances is very small. Tables 4 and 5 present the detection accuracy for balls that are found further than 6 m and 9 m, respectively, from the robot.

Figures 5, 6, 7 show the ball detection accuracy for the three algorithms, for each of the three resolutions. The graphics illustrate the position of the detected balls, relative to the robot, in world coordinate at the ground level. The algorithm used for converting ball position in pixels to world coordinate has been developed within our research group and can be found in (Neves et al. 2011).

The four “valleys” that can be noted in all of the graphics represent four bars that support the

Table 5. Percentages of correct ball detection for the three algorithms for balls found at 9 m or more from the robot.

	Blobs	Contours	Region growing
1024 × 1024	100%	67.6%	68.8%
512 × 512	78.4%	58.8%	66.2%
256 × 256	44.5%	11.5%	11.2%

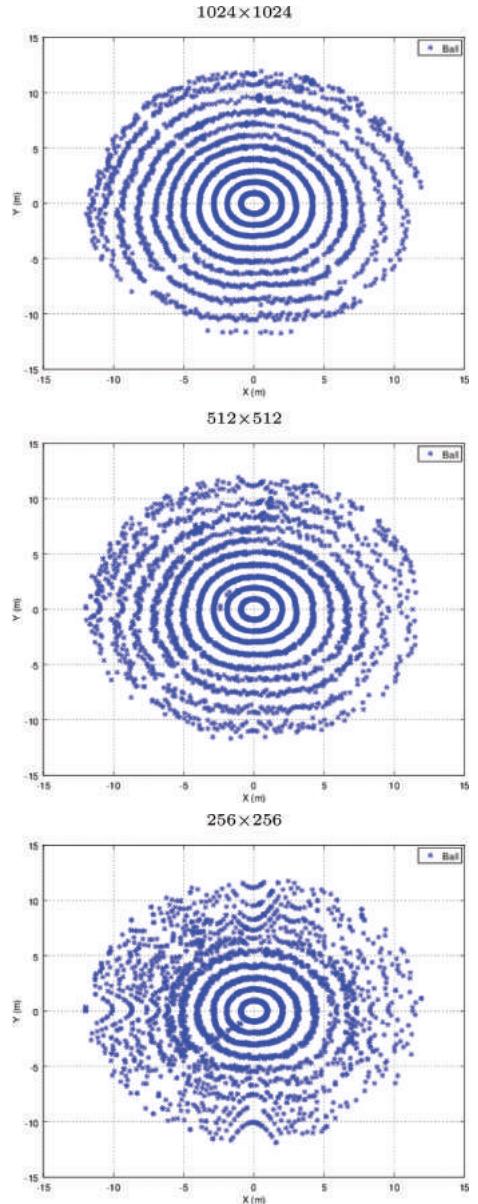


Figure 5. Graphics representing the ball detection for the three resolutions using the blob algorithm.

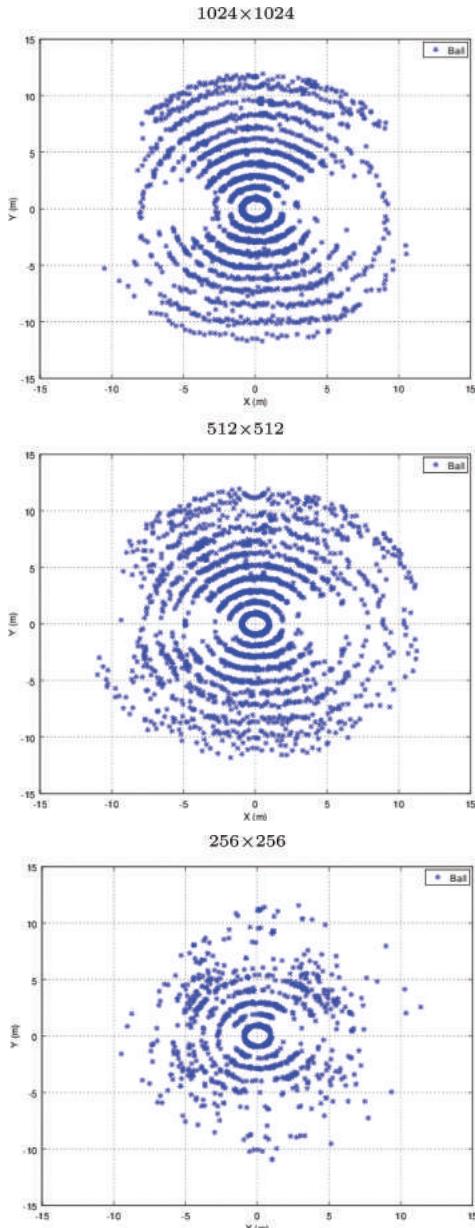


Figure 6. Graphics representing the ball detection for the three resolutions using the contours detection algorithm.

vision system in the hardware architecture of the robot. When the ball is aligned with any of these bars, the ball is not visible.

For images of  $256 \times 256$  pixels, the contours detection algorithms provides quite poor results. When the ball is found at 6 m or more from the robot, it is only detected in 16% of the cases. The

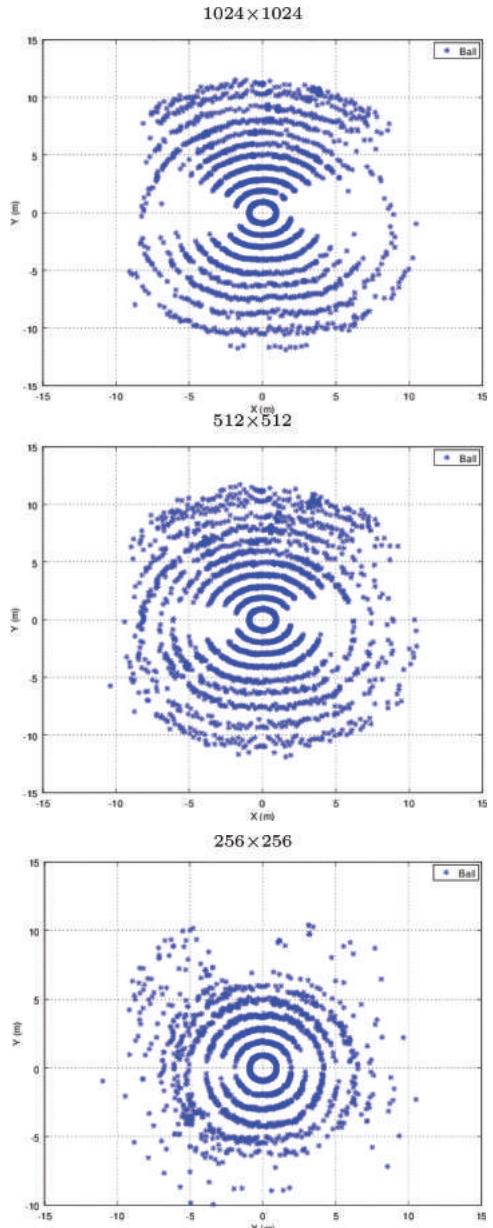


Figure 7. Graphics representing the ball detection for the three resolutions using the region growing clustering algorithm.

results are even less reliable for balls found at 9 m or more from the robot. In this case, only 10% of the balls are correctly detected.

The region growing clustering algorithm delivers slightly better results, compared to the contours detection algorithm but it is still far from achieving the detection rate of the blob detection algorithm.

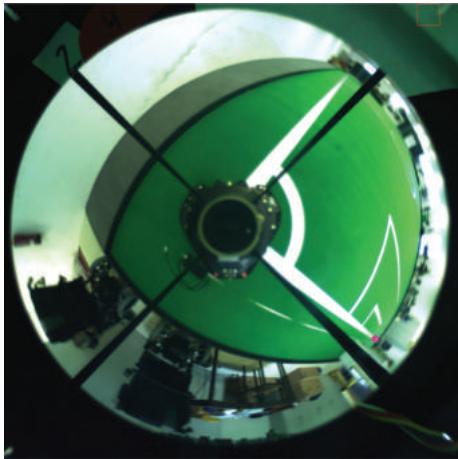


Figure 8. Ball detection using the blob algorithm at 12 m. The detected ball is marked with a magenta circle.

These results prove that working with images at full resolution can provide optimal results in a small amount of time when using an appropriate scanning pattern of the image. For applications where detection accuracy and total lack of false positives is required, the algorithm based on radial scanlines and blob formation is a suitable solution since its detection rate is 100% and it can be used with sensors working at 100 fps, if the hardware allows it. A smaller resolution can lead to an even smaller processing time, for applications where fast performance is more crucial than having 100% correct detection.

In Figure 8 we present an image with the ball correctly detected by the blob algorithm at a distance of 12 m from the robot.

The results that have been presented in this paper have been obtained on a computer with an Intel Core i5-3340M CPU @ 2.70 GHz 4 processor, which is the main processing unit of the soccer robots used in the experiments.

## 5 CONCLUSIONS

In this paper we have presented a study about the influence of image scanning in the pipeline of a color object detection digital vision system. Using the real-time scenario of robotic soccer and the digital vision system of an autonomous robot, we presented results for three different algorithms for object detection, using different scanning patterns and different image resolutions. The results presented show that when reducing the image resolution, object detection algorithms tend to lower their performance, due to loss of information.

## ACKNOWLEDGEMENTS

This work was developed in the Institute of Electronic and Telematic Engineering of University of Aveiro and was partially supported by FEDER through the Operational Program Competitiveness Factors—COMPETE and by National Funds through FCT—Foundation for Science and Technology in a context of a PhD Grant (FCT reference SFRH/BD/85855/2012) and in the context of the projects UID/CEC/00127/2013 and Incentivo/EEI/UI0127/2014.



## REFERENCES

- Akin, H.L., T. Mericli, E. Ozukur, C. Kavaklıoglu, & B.Gokce(2014).Cerberus Team Description. RoboCup 2014, Joao Pessoa, Brazil.
- Bradski, G. & A. Kaehler (2008, September). *Learning OpenCV* (First ed.). O'Reilly.
- Canny, J.F. (1986, November). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8 (6).
- Garcia, D., E. Carbajal, C. Quintana, E. Torres, I. Gonzalez, C. Bustamante, & L. Garrido (2012). Borregos Team Description. RoboCup 2012, Mexico City, Mexico.
- Gonzalez, R.C. & R.E. Woods (2001). *Digital Image Processing* (2 ed.). Addison-Wesley Longman Publishing Co.
- Neves, A., J. Azevedo, N.L.B. Cunha, J. Silva, F. Santos, G. Corrente, D.A. Martins, N. Figueiredo, A. Pereira, L. Almeida, L.S. Lopes, & P. Pedreiras (In Vladan Papic (Ed.), Robot Soccer, 2010). *CAMBADA soccer team: from robot architecture to multiagent coordination*, Chapter 2. I-Tech Education and Publishing, Vienna, Austria.
- Neves, A.J.R., D.A. Martins, & A.J. Pinho (2008, April). A hybrid vision system for soccer robots using radial search lines. In *Proc. of the 8th Conference on Autonomous Robot Systems and Competitions, Portuguese Robotics Open—ROBOT-ICA2008*, Aveiro, Portugal, pp. 51–55.
- Neves, A.J.R., A.J. Pinho, D.A. Martins, & B. Cunha (2011, mar). An efficient omni directional vision system for soccer robots: from calibration to object detection. *Mechatronics* 21 (2), 399–410.
- Röfer, T., T. Laue, C. Graf, T. Kastner, A. Fabisch, & C. Thedieck (2011). B-Human Team Description. RoboCup 2011, Istanbul, Turkey.
- Trifan, A., A. Neves, N. Lau, & B. Cunha (2012). A modular real-time vision system for humanoid robots. In *Proceedings of IS&T/SPIE Electronic Imaging 2012*.
- Trifan, A., A.J.R. Neves, B. Cunha, & J.L. Azevedo (2014). Uavision: a modular time-constrained library for soccer robots. In *Proc. of RoboCup 2014*.

# Improving deep neural networks classification by preprocessing images

H. Erdmann

*Department of Research and Technology, Boolabs, São Paulo, Brazil*

F.T. Ito

*Department of Computer Science, Universidade Federal de São Carlos, Brazil*

D. Takabayashi

*Department of Research and Technology, Boolabs, São Paulo, Brazil*

D.N. dos Santos

*Department of Research and Technology, Boolabs, São Paulo, Brazil*

*Department of Electrical Engineering, Centro Universitário da FEI, São Paulo, Brazil*

**ABSTRACT:** The amount of images used to train a classifier has a great impact in the algorithm's performance. In the domain of e-commerce, most of the dataset is formed by studio images and we demonstrate in this work that by multiplying the amount of images with different transformations can provide a significant boost on the overall performance of the classifier. Moreover, we list the challenges encountered in such task and present the improvements obtained with a classifier trained with additional images that were synthetically generated by applying several transformations on the original dataset.

## 1 INTRODUCTION

To classify an image correctly is a task with applications in several areas related to digital image processing. It can be done by extracting features that describe the image and compare them with the mean features of the known classes. The simplest classification techniques are binary classifiers, which are used to define if an image belongs or not to a certain class. Moreover, image classification algorithms are usually developed for a specific set of images because the objective of these algorithms is strongly related to their applications. In this work we focus on a multi-classification algorithm using state-of-the-art deep neural networks. It is widely known that the amount of images used to train a classifier has a direct effect on its final results (Fuxman et al. 2009). In the domain of e-commerce, most of the dataset is formed by studio images and we demonstrate how we improved our classifier by multiplying the amount of images with different transformations.

The task of object recognition is a great challenge in computer vision and has been the focus of thorough research by many authors for many years. There are many applications and algorithms towards this goal. Mainly, these applications make use of natural images. There are many image datasets that researchers can use to compare their work, such

as MNIST (Lecun and Cortes, LeCun et al. 2001), Image Net (Deng et al. 2009, Russakovsky et al. 2014) and LabelMe (Russell et al. 2008), composed of, mostly, natural images. However, the common images found in the market and e-commerce sites are doctored ones, with white backgrounds and perfect illumination. Approaches such as Global Contrast Normalization (GCN) (Graham and Sutter 2000) and Zero Components Analysis (ZCA) Whitening (Bell and Sejnowski 1997b, Bell and Sejnowski 1997a) fail in these situations, and cannot be used directly.

Deep Learning (DL) is an area of machine learning that relies on high-level, hierarchized abstractions to create models for specific functions (Krizhevsky et al. 2012a, Schmidhuber 2014). The theory behind Deep Learning gained traction circa 2000/2003, with the advent of powerful graphic cards and new works on the field. One of the prominent algorithms for Image Classification fits the DL profile: the Convolutional Neural Network (Sermanet et al. 2012, Smirnov et al. 2014). In this paper, we discuss the aforementioned approaches to preprocessing and their results in studio images, along with other techniques to expand image datasets and to find meaningful features to use in classifiers. A Convolutional Neural Network is used to get practical results from these preprocessed techniques on an image dataset composed by e-commerce products.

## 2 RELATED WORK

The preprocessing of image data for machine learning application has been extensively covered by several different authors (Kotsiantis and Kanellopoulos 2006, Förstner 2000), as using the image itself pixel by pixel as an instance can be unwieldy at times. A  $256 \times 256$  colored image, for example, yields 196608 intensity values: training a classifier with this data and no Graphic Processing Unit (GPU) takes weeks.

Finding a shorter representation without losing too much information is the goal of some preprocessing algorithms (Bell and Sejnowski 1997b, Krizhevsky and Hinton 2011, Farabet et al. 2013). However, with the help of a GPU, we can fully utilize the abstraction levels provided by CNNs and Deep Learning with complete images. In Farabet et al. (Farabet et al. 2013), horizontal flipping, rotation and scaling were used to decrease the bias that low-level features can bring to deep classification.

This is further explored in Simard et al. (Simard et al. 2003), where distortions and translations were used to expand the MNIST database and optimize classification in many algorithms. Finding appropriate features for image matching and classification is one of the challenges that each dataset poses to the task (Dollar et al. 2007), as distribution, characteristics and size influence the speed and quality of it. Studio and product images have few works in the literature, and preprocessing and feature mining on these is a relatively unexplored area.

The Deep Learning classifier we use is well known in image classification (Sigaud and Wilson 2007, Arel et al. 2010, Zeng et al. 2013). Optimal accuracy was reached in many competitions by Deep Learning related approaches, and the Large Scale Visual Recognition Challenge (ILSVRC) thrives with similar classification techniques. The specific architecture we use, the Convolutional Neural Network, has numerous applications and detailed papers and results (Ranzato et al. 2011, Wu et al. 2014, Le et al. 2011, Krizhevsky et al. 2012a, Le 2013).

## 3 METHODOLOGY

The methodology used on this work consists of some preprocessing techniques followed by the classification at the end of the pipeline. The preprocessing stage consists of four transformations that are detailed here: a. Affine transformations; b. Perspective transformations; c. Image whitening; d. Image normalization.

Affine and perspective transformations can be used to increase generality in image classifiers where a common approach is to alter the referential camera position to distort the image and

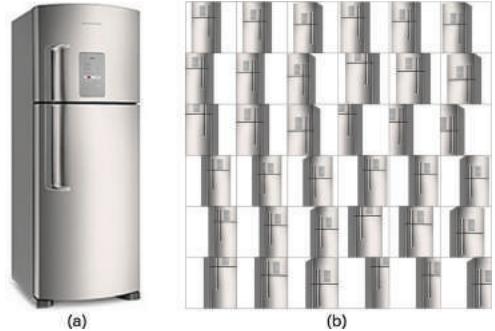


Figure 1. In (a): original image; In (b): 36 different images obtained with perspective transformations applied to the original image.

simulate different situations (Yaeger et al. 1996, Simard et al. 2003). Moreover, affine transformations keep distance ratios (midpoints of lines and their relative proportions are unaltered) and collinearity (points residing in lines keep their former positions on them). Composed by the translation, rotation, scaling and shearing operations, this group of transforms can focus on different parts of the image to increase the classification accuracy by a few percents. An example of perspective transformations can be viewed in Figure 1.

On the other hand, the perspective transformation projects points onto the image plane along lines that emanate from a single point, called the center of projection, changing the focus of the image. This type of transform does not preserve line parallelism, and is used to create distortions in images. These techniques are used in a variety of image classification tasks that involve low-level features (i.e. raw pixels) to increase the feature space and reduce bias.

In an image dataset, instances can have high contrast and illumination variation: this damages the classification process and can incur unnecessary bias. To soften the difference between images, matrix normalization and whitening algorithms can be applied before training. In Deep Learning, they are essential for optimization, but in non-DL algorithms these have little to no effect on classification.

Normalization is a process that decreases the variation in ranges between matrixes. In image processing, it is used to create consistency in datasets with natural scenes in a wide spectrum of illumination sources (Cirean et al. 2011, Makwana et al. 2011). Effectively, this process spreads out the pixel ranges on images to enhance contrast and edges, amplifying corners and form features. Many normalization techniques exist in computer vision. In this paper we use the Global Contrast Normalization (GCN), a common preprocessing approach to image classification (Goodfellow et al. 2013,

Springenberg and Riedmiller 2013). It consists in subtracting from each pixel intensity value the mean of all pixels in the image, then dividing by their standard deviation. Whitening algorithms are processes that decorrelate neighboring pixels by transforming the data until the covariance matrix is equal to the identity matrix. The pixels from an image are usually highly correlated, and thus many points are redundant in training and classification. Whitening is vital in the preprocessing step of algorithms that use raw images as data to enhance the identification of higher-order features.

In Bell et al. (Bell and Sejnowski 1997b, Bell and Sejnowski 1997a), several whitening techniques are discussed and explained. Mainly, they find a decorrelation matrix and multiply the desired data by it to transform the covariation matrix into the diagonal, then factoring it to produce the identity. Among them, the Principal Component Analysis (PCA) and the Zero-phase Component Analysis (ZCA) have been used in classification tasks by various authors.

The PCA is applied to a matrix  $X$  by using the following values from the covariance matrix: multiplying it by the transpose of the eigenvectors, scaled by the inverse square root of the eigenvalues. However, this transform is not unique as multiplying the whitened data by an orthogonal matrix will yield a result with covariance matrix equal to the identity. The ZCA consists in finding the canonical whitening by multiplying the PCA transform by an orthonormal matrix. This transform is unique and shows the closest resemblance to the original data. On the other hand, the ZCA consists in finding the canonical whitening by multiplying the PCA transform by an orthonormal matrix. This transform is unique and shows the closest resemblance to the original data. Both PCA and ZCA approaches can be viewed in Figure 2.

After all the preprocessing, the images are subjected to a Convolutional Neural Network (CNN) classifier. Neural networks have been used in several classification tasks (Hunt et al. 1992, Zhang

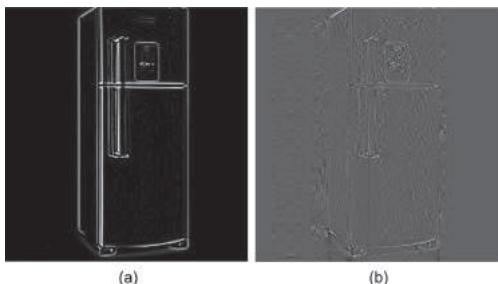


Figure 2. In (a): image after application of PCA transform; In (b): image after application of the ZCA transform.

2000, Arel et al. 2010, Schmidhuber 2015), and the convolutional neural network in particular shows great results in visual classification. These hierarchical networks reduce the image by filtering and reducing it in each iteration, differing from the usual neural network approach by sharing weights between nodes in the same layer and spatial pooling for simplification.

A conventional Neural Network (NN) consists of many simple processors connected, called neurons, each one producing a sequence of values of activations. Input neurons are activated by sensors perceiving the environment, other neurons are activated by means of weighted connections from previous activated neurons. Some neurons can influence the network by firing actions. Learning or credit assignment is about finding weights that make the NN exhibit desired behavior, such as driving a car. Depending on the problem and how the neurons are connected, such behavior may require long causal chains of computational stages, where each stage transforms (often in a non-linear way) the aggregate activation of the network. Deep Learning is about accurately assigning credit across many such stages (Schmidhuber 2015).

A Convolutional Neural Network (CNN) has several different layers for each operation, but the basic model has a convolutional layer that separates the image in many sub-matrixes called feature maps, a pooling layer that groups pixels and an activation layer. These feature maps are each the result of a filter passed throughout the input matrix, and the activation layer can be chosen from a variety of functions, such as sigmoids or a Rectified Linear Unit. This network is fed with squared pixel patches surrounding each pixel in each image. The label for each patch is the label of the center pixel.

## 4 EXPERIMENTS

In our experiments, we used an image object database with roughly two million instances obtained from e-commerce sites and web crawlers. Numerous products, from kitchen appliances to tires and cell-phone accessories, were gathered to form it. To reduce the scope of classification, we restrained the size of this database to include only appliances and related items, totalling 10 classes. All images were standardized to  $256 \times 256$  pixels, with 3 color channels. To create a Convolutional Neural Network, we used the Deep Learning framework Caffe (Jia et al. 2014). The network layers are based on the reference CaffeNet (Krizhevsky et al. 2012b).

On each essay, we used a different preprocessing approach. First, we use the database without any augmentation. Then, a boosted database with affine transformations (rotations, translations

with zoom and horizontal flipping) applied to all images. Finally, we use the ZCA whitening algorithm in all images in the boosted database and train with that. A fourth database containing other classes of objects was used to further measure the image augmented classification accuracy.

The classifiers are first tested for assertivity using a random sampling from the database. Then, with a second database composed of images mined with web crawlers and classified by hand, test the model for the top-1 and top-5 error rates. We separated resulting classifiers by their number of iterations, from 1000 to 5000 in 1000 steps. The first experiment was done in order to set a baseline for comparing our techniques results and has a test sampling set composed of 7,5% of the total images. On the following experiments this quantity was raised to 25%.

Tables 1, 3 and 5 refer to the training data and the accuracy results on each number of training iterations. The classifier takes snapshots of

Table 1. Classifier's performance for the training data.

Original database—training data	
Iterations	Assertivity
1000	82.00%
2000	87.02%
3000	89.70%
4000	90.75%
5000	90.67%

Table 2. Classifier's performance for the testing data.

Original database—testing data		
Iterations	Top-1	Top-5
1000	30.32%	76.03%
2000	30.33%	76.21%
3000	32.01%	73.33%
4000	35.24%	76.19%
5000	34.02%	80.44%

Table 3. Classifier's performance for the training data with boosted dataset.

Boosted database—training data	
Iterations	Assertivity
1000	77.09%
2000	83.00%
3000	84.53%
4000	86.00%
5000	87.00%

itself and measures the error rate of the preliminary results in 1000, 2000, 3000, 4000 and 5000 iterations, giving the accuracy value (100%—error rate).

Tables 2, 4 and 6 refer to the test data taken from web crawlers and made of natural scenes. This test uses the snapshot classifiers that were obtained in the training phase and measures the error rate when choosing the prediction of the most likely classes. The top-1 error rate takes in account only one prediction and the top-5, a set with the five most probable classes.

Table 7 is an elaborate description of the results from the last classifier by class. Each class has a super-class (parent), the number that represents the class and the name of the class. We measured the error rates of the last classifier with 5000 iterations and separated the accuracy and recall per class.

Table 4. Classifier's performance for the testing data with boosted dataset. The gain in both top-1 and top-5 metrics is around 8%.

Boosted database—testing data		
Iterations	Top-1	Top-5
1000	25.12%	76.04%
2000	44.09%	84.37%
3000	40.37%	86.11%
4000	47.31%	85.19%
5000	47.04%	88.44%

Table 5. Classifier's performance for the training data with boosted and whitened dataset.

Boosted + ZCA database—training data	
Iterations	Assertivity
1000	75.09%
2000	83.26%
3000	85.48%
4000	86.33%
5000	87.29%

Table 6. Classifier's performance for the testing data with boosted and whitened dataset. A slight improvement was obtained.

Boosted + ZCA database—testing data		
Iterations	Top-1	Top-5
1000	39.25%	89.19%
2000	42.92%	88.31%
3000	47.70%	94.17%
4000	52.18%	90.95%
5000	44.43%	92.42%

Table 7. Detailed information describing accuracy and recall for Top-1 and Top-5 for each class.

Boosted database with extra classes—detailed test data				
Class	Top-1	Top-3	Recall	
			Top-1	Top-3
Tires	0.78%	0.78%	1.00%	1.00%
Car mats	0.40%	0.60%	0.40%	0.60%
Wine cellars	0.73%	0.82%	0.82%	0.91%
Freezers	0.82%	1.00%	0.82%	0.91%
Microwaves	0.64%	0.73%	0.73%	1.00%
Dishwashers	0.55%	0.73%	1.00%	1.00%
Ovens	0.45%	0.82%	0.73%	1.00%
Cooktops	0.50%	0.50%	0.38%	0.62%
Stoves	0.30%	0.40%	0.90%	1.00%
Washmachines	0.00%	0.18%	0.64%	1.00%
Drywashers	0.45%	0.73%	1.00%	1.00%
Dryers	0.55%	1.00%	0.64%	1.00%
Centrifuges	0.30%	0.60%	0.90%	0.90%
Refrigerators	0.20%	0.50%	0.70%	1.00%
Watches	0.12%	0.36%	0.27%	0.55%
Suitcases	0.30%	0.30%	1.00%	1.00%
Keys	0.64%	0.64%	1.00%	1.00%

The first preprocessing stage led to significant increase in accuracy, but in the training of the last classifier for the database with the preprocessing plus the ZCA transformations the classifier did not converge in the maximum 5000 iterations. Testing the same approach with a database composed of natural images, however, yielded better results than the second classifier.

## 5 CONCLUSIONS

In our experiments, we first evaluated the classifiers performance against a image dataset without any preprocessing. In this scenario, we obtained an accuracy of 80% within the top-5 predictions, which means that one of the 5 best predictions yielded by the classifier for a image is correct. On the other hand, for the top-1, the classifier performed with 34% accuracy. By training the classifier again with the preprocessing techniques described earlier, we were able to improve the classification to a 92% accuracy on the top-5 and 44% accuracy on the top-1. This means an improvement of 15% in the top-5 and 30% in the top-1 predictions.

The first and second classifiers yielded results as expected, as primary preprocessing algorithms such as affine transformations do not heavily depend on the color composition of the image and can give the classifier a good amount of generalization. But the third classifier did not even converge in the maximum number of iterations previously set, without any tangible results. The dataset without studio images reached great accuracy rating. We state here that studio images cannot be doctored with the current famous preprocessing techniques and increase a Convolutional Neural Network classifier accuracy directly: extra steps must be made to ensure that the white background does not interfere.

in the maximum number of iterations previously set, without any tangible results.

The dataset without studio images reached great accuracy rating. We state here that studio images cannot be doctored with the current famous preprocessing techniques and increase a Convolutional Neural Network classifier accuracy directly: extra steps must be made to ensure that the white background does not interfere.

Some possible ways of minimizing the background's importance in the picture are the use of image segmentation algorithms such as selective search (Sande 2011, Uijlings et al. 2013) or DPM (Felzenszwalb et al. 2010, Trulls et al. 2014) to extract only the object from the image and create features from it; Or inserting dummy backgrounds from natural images behind objects. This can lead to some bias if the background list is small; Also, edge detection filters may eliminate some of the white interference, as regions with that background will have no edges and no useful information. Illumination will also be discarded. Another improvement is the creation of a tree of classifiers where the top most classifiers are more generic as the bottom most classifiers are more specific. Also, using binary classifiers for each class in conjunction with the main classifier as a pool can add to the generality of the solution.

We first evaluated the classifiers performance against a image dataset without any preprocessing. In this scenario, we obtained an accuracy of 80% within the top-5 predictions, which means that one of the 5 best predictions yielded by the classifier for a image is correct. On the other hand, for the top-1, the classifier performed with 34% accuracy. By training the classifier again with the preprocessing techniques described earlier, we were able to improve the classification to a 92% accuracy on the top-5 and 44% accuracy on the top-1. This means an improvement of 15% in the top-5 and 30% in the top-1 predictions.

The first and second classifiers yielded results as expected, as primary preprocessing algorithms such as affine transformations do not heavily depend on the color composition of the image and can give the classifier a good amount of generalization. But the third classifier did not even converge in the maximum number of iterations previously set, without any tangible results. The dataset without studio images reached great accuracy rating. We state here that studio images cannot be doctored with the current famous preprocessing techniques and increase a Convolutional Neural Network classifier accuracy directly: extra steps must be made to ensure that the white background does not interfere.

Some possible ways of minimizing the background's importance in the picture are the use of image segmentation algorithms such as selective

search (Sande 2011, Uijlings et al. 2013) or DPM (Felzenszwalb et al. 2010, Trulls et al. 2014) to extract only the object from the image and create features from it; Or inserting dummy backgrounds from natural images behind objects. This can lead to some bias if the background list is small; Also, edge detection filters may eliminate some of the white interference, as regions with that background will have no edges and no useful information. Illumination will also be discarded. Another possible improvement that we are currently working on is the creation of a tree of classifiers where the top most classifiers are more generic and the following classifiers on the tree are more specific. Also, using binary classifiers for each class in conjunction with the main classifier as a pool can add to the generality of the solution.

## REFERENCES

- Arel, I., D.C. Rose, & T.P. Karnowski (2010). Deep Machine Learning -A New Frontier in Artificial Intelligence Research [Research Frontier].
- Bell, A. & T. Sejnowski (1997a). Edges are the “independent components” of natural scenes. *Advances in neural information processing systems*, 831–837.
- Bell, A.J. & T.J. Sejnowski (1997b). The ‘independent components’ of natural scenes are edge filters. *Vision Research* 37(23), 3327–3338.
- Cirean, D., U. Meier, & J. Schmidhuber (2011). A committee of neural networks for traffic sign classification. In *Proceedings of the International Joint Conference on Neural Networks*, pp. 1918–1921.
- Deng, J., W. Dong, R. Socher, L.-J. Li, K. Li, & L. Fei-Fei (2009). ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*.
- Dollar, P., Z. Tu, H. Tao, & S. Belongie (2007). Feature mining for image classification. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- Farabet, C., C. Couprie, L. Najman, & Y. Lecun (2013). Learning hierarchical features for scene labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(8), 1915–1929.
- Felzenszwalb, P.F., R.B. Girshick, & D. McAllester (2010). Cascade object detection with deformable part models. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2241–2248.
- Förstner, W. (2000). Image Preprocessing for Feature Extraction in Digital Intensity, Color and Range Images. In *Geomatic Method for the Analysis of Data in the Earth Sciences*, Volume 95, pp. 165–189.
- Fuxman, A., A. Kannan, A.B. Goldberg, R. Agrawal, P. Tsaparas, & J. Shafer (2009). Improving classification accuracy using automatically extracted training data. *Proceedings of the 15th ACM International Conference on Knowledge Discovery and Data Mining* (2009) 33(2), 1145.
- Goodfellow, I.J., D. Warde-Farley, M. Mirza, A. Courville, & Y. Bengio (2013). Maxout Networks. *arXiv preprint*, 1319–1327.
- Graham, N. & A. Sutter (2000). Normalization: Contrast-gain control in simple (Fourier) and complex (non-Fourier) pathways of pattern vision.
- Hunt, K.J., D. Sbarbaro, R. Zbikowski, & P.J. Gawthrop (1992). Neural networks for control systems—A survey.
- Jia, Y., E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, & T. Darrell (2014). Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*.
- Kotsiantis, S.B. & D. Kanellopoulos (2006). Data preprocessing for supervised learning. *International Journal of ...*, 1–7.
- Krizhevsky, A. & G. Hinton (2011). Using Very Deep Autoencoders for Content-Based Image Retrieval. *ESANN*.
- Krizhevsky, A., I. Sutskever, & G.E. Hinton (2012a). ImageNet Classification with Deep Convolutional Neural Networks. *Advances In Neural Information Processing Systems*, 1–9.
- Krizhevsky, A., I. Sutskever, & G.E. Hinton (2012b). Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger (Eds.), *Advances in Neural Information Processing Systems* 25, pp. 1097–1105. Curran Associates, Inc.
- Le, Q.V. (2013). Building high-level features using large scale unsupervised learning. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 8595–8598.
- Le, Q.V., R. Monga, M. Devin, G. Corrado, K. Chen, M.A. Ranzato, J. Dean, & A.Y. Ng (2011). Building high-level features using large scale unsupervised learning Dataset constructions. *ICML*, 1–10.
- LeCun, Y., L. Bottou, Y. Bengio, & P. Haffner (2001). Gradient-based learning applied to document recognition. In *Intelligent Signal Processing*, pp. 306–351. IEEE Press.
- Lecun, Y. & C. Cortes. The MNIST database of handwritten digits.
- Makwana, R., V. Thakar, & N. Chauhan (2011, Nov). Evaluation and analysis of illumination normalization methods for face recognition. In *Image Information Processing (ICIIP), 2011 International Conference on*, pp. 1–6.
- Ranzato, M., Q.V. Le, R. Monga, M. Devin, K. Chen, G.S. Corrado, J. Dean, & A.Y. Ng (2011). Building high-level features using large scale unsupervised learning. In *International Conference in Machine Learning*, pp. 38115.
- Russakovsky, O., J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, & L. Fei-Fei (2014). Imagenet large scale visual recognition challenge.
- Russell, B., A. Torralba, K. Murphy, & W. Freeman (2008). Labelme: A database and web-based tool for image annotation. *International Journal of Computer Vision* 77(1–3), 157–173.
- Sande, K.V.D. (2011). Segmentation as selective search for object recognition. ... *Vision(ICCV), 2011...(2)*, 1879–1886.
- Schmidhuber, J. (2014). Deep Learning in Neural Networks: An Overview. *arXiv preprint arXiv:1404.7828*, 1–66.

- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks* 61, 85–117. Published online 2014; based on TR arXiv:1404.7828 [cs.NE].
- Sermanet, P., S. Chintala, & Y. LeCun (2012). Convolutional neural networks applied to house numbers digit classification. *Proceedings of International Conference on Pattern Recognition ICPR12*, 10–13.
- Sigaud, O. & S.W. Wilson (2007). Learning classifier systems: A survey. *Soft Computing* 11(11), 1065–1078.
- Simard, P., D. Steinkraus, & J. Platt (2003). Best practices for convolutional neural networks applied to visual document analysis. *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.*
- Smirnov, E.A., D.M. Timoshenko, & S.N. Andrianov (2014). Comparison of Regularization Methods for ImageNet Classification with Deep Convolutional Neural Networks. *AASRI Procedia* 6, 89–94.
- Springenberg, J. & M. Riedmiller (2013). Improving Deep Neural Networks with Probabilistic Maxout Units. *arXiv:1312.6116*, 1–10.
- Trulls, E., S. Tsogkas, I. Kokkinos, A. Sanfeliu, & F. Moreno-Noguer (2014). Segmentation-aware deformable part models. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conferenceon*, pp. 168–175. IEEE.
- Uijlings, J.R.R., K.E.A. Van De Sande, T. Gevers, & A.W.M. Smeulders (2013). Selective search for object recognition. *International Journal of Computer Vision* 104(2), 154–171.
- Wu, C., W. Fan, Y. He, J. Sun, & S. Naoi (2014). Handwritten Character Recognition by Alternately Trained Relaxation Convolutional Neural Network. *2014 14th International Conference on Frontiers in Handwriting Recognition*, 291–296.
- Yaeger, L., R. Lyon, & B. Webb (1996). Effective Training of a Neural Network Character Classifier for Word Recognition. *Nips* 9, 807–813.
- Zeng, X., W. Ouyang, & X. Wang (2013). Multi-stage Contextual Deep Learning for Pedestrian Detection. *2013 IEEE International Conference on Computer Vision*, 121–128.
- Zhang, G.P. (2000). Neural networks for classification: a survey. In *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, Volume 30, pp. 451–462.

This page intentionally left blank

# Surface reconstruction of bone fragments: A comparative study

F. Paulano, J.J. Jiménez & J.R. Jiménez

*Graphics and Geomatics Group of Jaén, Universidad de Jaén, Jaén, Spain*

**ABSTRACT:** The generation of surface models of bone fragments is useful for many medical applications. On the one hand, bone fragment models can be visualized in a more effective way. On the other hand, simulation and analysis techniques usually need surface models as input. Simulation and analysis processes require surface models to have certain properties and features that depend on the methods used for these processes. Therefore, the suitability of the method utilized to generate the surface will depend on the simulation or analysis performed. In this paper, the suitability of two algorithms that generate bone fragment surface models from point clouds is tested. A proven 2D region growing based method enables the segmentation of bone fragments from CT stacks. Then a new approach to extract contour points and estimated normals is used in a pre-process step. The experiments performed allow extracting the benefits and drawbacks of each method applied to different bones and fracture types. Moreover, the quality of the obtained results is analyzed focusing on the special features of the data obtained from CT stacks representing bone fractures. To the best of our knowledge, nowadays studied methods are not widespread in the generation of surfaces to represent bone fragments.

**Keywords:** bone fragments; surface reconstruction; 3D medical images; bone fracture reduction; segmentation; contour extraction; poisson; ball-pivoting

## 1 INTRODUCTION

Computational techniques are frequently used to help the specialist in the treatment of bone injuries and diseases (Jaramaz, Hafez, & DiGioia 2006). These techniques usually require 3D models of bone fragments as input. In most cases the Marching Cubes (MC) algorithm (Lorensen & Cline 1987) is used in order to obtain isosurfaces from CT scans. Due to the complexity showing the bone tissue in a CT stack (Paulano, Jiménez, & Pulido 2014), the models obtained are usually huge and just have topological information. Therefore, these models often need to be post-processed before being used. Alternatively, other methods can be used to generate mesh surfaces. The selection of a method depends on the problem to be solved and the features that the surface is expected to have.

The generation of surface models from the information available in medical images is a complex task. In addition, the morphology of bone fragments increases the complexity of the problem. The inner part of the bone is composed by trabecular tissue. This type of tissue is very heterogeneous, thus segmentation methods usually obtain holes. To make the process more complicated the shape of the bone fragments is very irregular and trabecular tissue appears in the outer part of the bone due to the fracture (Paulano, Jiménez, & Pulido 2013)

(Paulano, Jiménez, & Pulido 2014). Thus the contours obtained for each slice are very irregular in fracture zones. All these features influence the choice of a method for generating surfaces.

In this paper, a study is performed to test the suitability of surface generation methods to model bone fragments. In a first step, bone fragments are segmented from CT scans. Then contours are extracted from the segmented regions. The point clouds used as input for surface generation are composed of the points of the contours. Finally surface generation methods are tested and their main advantages and disadvantages are shown. Specifically, Ball-pivoting (Bernardini, Mittleman, Rushmeier, Silva, & Taubin 1999) and Poisson (Kazhdan, Bolitho, & Hoppe 2006) surface generation methods are considered. The results obtained show the performance of each method depending on the bone and the fracture type. In addition, the adequacy and quality of the results are analyzed focusing on the special features of the bone fracture data obtained from CT stacks. Specifically, the performed study focuses on point distribution and fracture zones. The paper is structured as follows. Section 2 describes the studied methods and the procedure followed in order to generate the surface models. Section 3 shows the results of the study using different criteria. Finally section 4 discusses the obtained results and concludes the paper.

## 2 METHODOLOGY

Nowadays, most of the published studies that work with bone fragment models use MC to generate them. These models are usually huge and contain noise due to the morphology of trabecular tissue. As an alternative, two different methods have been considered for this study: Ball-pivoting (Bernardini, Mittleman, Rushmeier, Silva, & Taubin 1999) and Poisson surface generation (Kazhdan, Bolitho, & Hoppe 2006).

Ball-pivoting is a geometric algorithm that triangulates a point cloud by traversing a ball of a specific size over the surface. The algorithm starts with an initial triangle and test if the ball is able to reach a point from each triangle edge. If so, the point reached is added to the model. This procedure is repeated until the ball is not able to reach any points from the current edges. This method performs well if the distance between points is homogeneous. Otherwise some points could not be triangulated and the obtained model would be incomplete.

Poisson surface reconstruction requires as input a set of points and their associated normals. This method formalizes the surface approximation as a Poisson problem. Both the points and normals are interpreted as samples of an indicator function that is 0 outside a model and 1 inside it. In order to approximate the indicator function trivariate B-splines are used. This algorithm is robust against noise in the input data. Nevertheless the approximation causes that the original points of the contours are not preserved.

The generation of surfaces requires point clouds that represent the outer part of each bone fragment. Therefore these point clouds need to be extracted from medical images. Most of the proposed works use MC to generate isosurfaces from medical images. This method could be used to generate point clouds and estimate the normal associated to each point. However, MC generates many

internal triangles due to the anatomy of trabecular tissue. As an alternative, we propose to generate contours after the segmentation process and then remove internal contours in 2D. Finally normals are estimated using the topological information of the contours. This process is explained in detail in next paragraphs.

To begin with, all the bone fragments are segmented from CT scans. For this purpose, a semi-automatic method is used (Paulano, Jiménez, & Pulido 2014). The method is based on 2D region growing and only requires that the user places a seed in the first slice in which each bone fragment appears. The method is also able to deal with erroneously joined fragments. In order to split them the method uses an isolated region growing algorithm, hence additional seeds can be required for this special case.

The segmentation algorithm generates a region for each bone fragment in each slice. Then the contour of each region is extracted. For this purpose, an approach similar to the proposed by Pulido et al. is applied (Pulido, Paulano, & Jiménez 214). First, the marching squares algorithm is utilized in order to generate contours from the segmented regions (Ho, Wu, Chen, Chuang, & Ouhyoung 2005). This algorithm can be considered a 2D adaptation of MC. Using a divide-and-conquer approach, the algorithm studies each cell composed of four points in a slice to detect boundaries. The algorithm does not ensure that the generated contours are closed, but this is not relevant in this case since the goal is to generate point clouds.

Due to the noise present in the regions because of the trabecular tissue, inner contours can be obtained. These inner contours are removed using the point in polygon algorithm by (Feito & Torres 1997), hence only the outer contour of each fragment is extracted in each slice. This algorithm allows determining if a contour is inside other contours without the need to perform complex calculations, such as solving a system of equations.

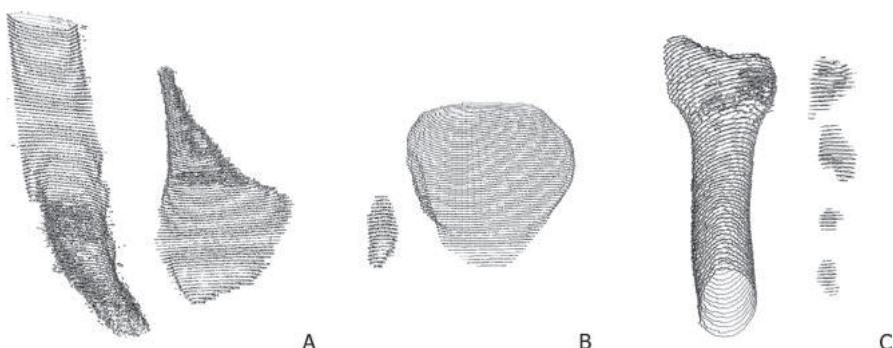


Figure 1. Point clouds extracted from CT scans. These points are used as input for the surface reconstruction algorithms. A—Spiral fracture of distal fibula. B—Simple fracture of patella. C—Comminuted fracture of radius.

After removing internal contours, the points of each external contour are joined in order to form the point cloud used as input by the surface generation algorithms (see Fig. 1).

In addition to the outer contours, the Poisson surface reconstruction algorithm also requires the normal of each point. These normals are estimated in 2D for each point using topological information from contours. For each three consecutive points  $P_{i-1}$ ,  $P_i$  and  $P_{i+1}$  in a contour, the normal associated to  $P_i$  is estimated as the sum of the normals of the two segments formed by  $P_{i-1} - P_i$  and  $P_i - P_{i+1}$ . The third component of each normal is defined by the plane of the slice. Despite of this estimation, normals are accurate enough to estimate the surface curvature. Once point clouds have been generated, surface generation methods can be applied.

### 3 RESULTS

In this study the suitability of two different methods to generate surfaces of bone fragments is tested: Ball-pivoting and Poisson. These methods have been widely used to generate surfaces for different applications (Campos, Garcia, & Nicosevici 2011) (Wiemann, Annuth, Lingemann, & Hertzberg 2015). Nevertheless, their use is not widespread in the generation of surfaces to represent bone fragments.

In order to test the performance of the two studied methods, they were used to generate surfaces from 6 CT stacks previously processed to extract bone fragments contours in each slice

(see Fig. 1). These CT stacks contain fragments of different bones and fracture types. The resolution of all images is  $512 \times 512$ . The results show the performance of each method depending on the features of both the bone and the fracture.

The first CT stack shows a spiral fracture of distal fibula (see Fig. 1, A). The fracture generated two fragments of similar size. The stack has 173 slices with dimensions  $159 \times 159$  mm and 0.625 mm spacing. The next three datasets contain simple fractures in which a small and a big fragment were generated. The second stack represents a simple fracture of patella (see Fig. 1, B) and contains 157 slices. The dimensions of each slice are  $156 \times 156$  mm and the separation between slices is 0.625 mm. The third stack represents a simple fracture of calcaneus and contains 205 slices. The separation between slices is 0.625 mm and the dimensions of each slice are  $173 \times 173$  mm. The fourth stack contains a simple fracture of astragalus and it is formed by 156 slices with dimensions  $150 \times 150$  mm and with 0.625 mm spacing. The fifth CT stack contains a comminuted fracture of distal tibia and it is formed by 173 slices with dimensions  $159 \times 159$  mm and with 0.625 mm spacing. The fracture generated 3 fragments of irregular shape. The last stack shows a comminuted fracture of radius that contains 5 fragments (see Fig. 1, C). The stack has 134 slices with dimensions  $96 \times 96$  mm and 1.25 mm spacing.

Table 1 shows the size of the models obtained using the two studied methods. Ball-pivoting does not modify the original points of the contours. Although there is no excessive difference in size between the models generated by the two methods,

Table 1. Number of vertices and faces generated for each of the datasets.

Dataset	Fragment	Original point cloud		Poisson reconstruction		Ball-pivoting	
		Points		Points	Triangles	Points	Triangles
Fibula	1	18919		13500	26988	18919	20837
Fibula	2	16501		14863	29722	16501	19924
Patella	1	11734		8177	16350	11734	21558
Patella	2	994		999	1994	994	1099
Calcaneus	1	24315		23546	47088	24315	42952
Calcaneus	2	1866		1599	3194	1866	3283
Astragalus	1	22053		22764	45508	22053	37677
Astragalus	2	4740		4087	8162	4740	8603
Tibia	1	89889		74475	149094	89889	101292
Tibia	2	19696		17879	35742	19696	23890
Tibia	3	30320		26000	52012	30320	35464
Radius	1	21965		16829	33650	21965	26860
Radius	2	2352		1608	3212	2352	4117
Radius	3	1772		1337	2670	1772	2068
Radius	4	529		610	1212	529	646
Radius	5	717		1011	2018	717	949

the number of points of the models generated by the Poisson reconstruction method is slightly lower but the number of triangles is higher in most cases in comparison with Ball-pivoting. Both methods obtain good results in the reconstruction of simple fractures in which bone fragments have similar size. Poisson reconstruction generates better models from a visual point of view (see Fig. 3, A). However, Ball-pivoting preserves the original features of the model since does not modify the original points of the contours (see Fig. 2, A). This is an advantage when models are used in processes that utilize bone features, such as virtual bone fracture reduction. Other simple fractures lead to the generation of a small fragment and a big fragment. Avulsed fracture is one of these cases. Poisson reconstruction method does not respect the shape of the small bone fragment. In addition, the method has difficulties in respecting the shape of the bone at the top and the bottom of the big fragment (see Fig. 3, B). On the other hand, Ball-pivoting retains the shape of small fragments but sometimes does not generate correct triangles in small fragments due to the

proximity of the points in the same slice (see Fig. 2, B). Comminuted fractures generate more than two fragments. These fractures are commonly caused by high-energy traumas, hence bone can be deformed and small fragments are usually generated. Therefore, both methods have the same problems before mentioned for small fragments. Furthermore, one of the stacks tested has 1.25 mm spacing. This causes that the distance between points is quite heterogeneous. In this case, Ball-pivoting is not able to connect many of the triangles, hence the generated models contain holes and sometimes are incomplete (see Fig. 2, C). Poisson also has difficulty in modeling the most detailed areas when the distance between slices is high (see Fig. 3, C).

Additionally, the adequacy and quality of the obtained results are analyzed depending on the special features of the data obtained from CT scans, which differs from the data not-related with bone fractures.

The distribution of the points generated from a CT stack depends on the resolution and the size of the 2D images, and the distance between slices. Since

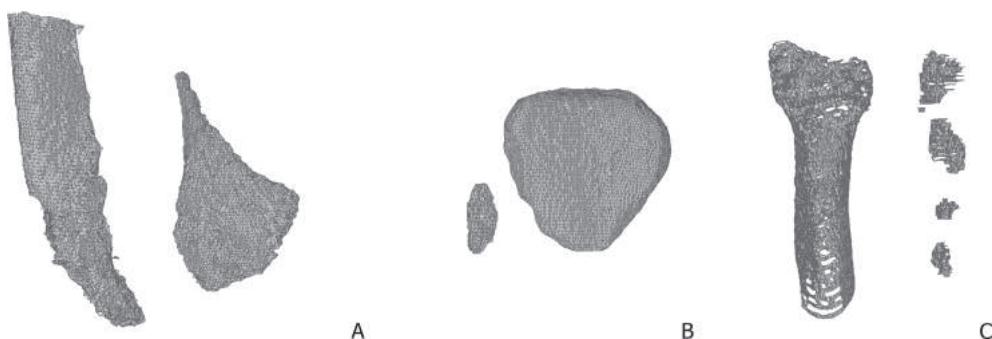


Figure 2. Surfaces generated using the Ball-pivoting algorithm. A—Spiral fracture of distal fibula. B—Simple fracture of patella. C—Comminuted fracture of radius.

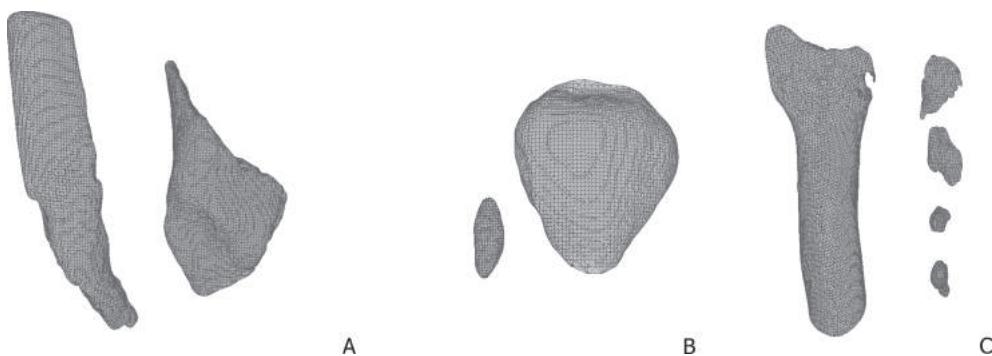


Figure 3. Surfaces generated using the Poisson algorithm. A—Spiral fracture of distal fibula. B—Simple fracture of patella. C—Comminuted fracture of radius.

Table 2. Point distribution for each CT stack tested. Resolution of all slices is  $512 \times 512$ .

	Size of slice (mm)	Distance between points (mm)	Spacing between slices (mm)	Ratio
Fibula	$159 \times 159$	0.3105	0.625	2.0126
Patella	$156 \times 156$	0.3047	0.625	2.0513
Calcaneus	$173 \times 173$	0.3379	0.625	1.8497
Astragalus	$150 \times 150$	0.293	0.625	2.1333
Tibia	$159 \times 159$	0.3105	0.625	2.0126
Radius	$96 \times 96$	0.1875	1.25	6.6667

the resolution of all the tested images is  $512 \times 512$ , the only variables are the size of the images and the distance between slices. Table 2 shows the value of both variables for all tested CT stacks. For the first 5 CT stacks the distance between slices approximately doubles the distance between points in the same slice. Despite the fact that both algorithms perform better if the points are uniformly distributed, they obtain acceptable results in these 5 cases. The distribution of the last CT stack is more heterogeneous because the distance between slices is more than 6 times higher than the distance between points in the same slice. In this situation, Poisson reconstruction performs better than Ball-pivoting. Poisson only has difficulties in the more detailed areas, such as the fracture zone. Otherwise, Ball-pivoting is not able to properly triangulate in most cases, hence the obtained models contain many holes and incorrect triangles.

One of the most important applications of the generation of bone fragment models from CT images is the simulation of a bone fracture reduction. In recent years, some methods have been proposed in order to help specialists in the preoperative planning of this surgery (Fürnstahl, Székely, Gerber, Hodler, Snedeker, & Harders 2012). Using these methods, specialists can know the number of fragments and their correct position before the real intervention; hence surgery time can be reduced (Sugano 2003). Some of the proposed methods try to virtually reduce bone fractures by calculating, matching and registering fracture zones (Zhou, Willis, Sui, Anderson, Brown, & Thomas 2009). Therefore, it is important that the reconstruction method does not modify the shape of the fracture zone. Since Poisson reconstruction generate new points and does not respect the original points of the contour (see Fig. 4), these models are not suitable for being used in a virtual fracture reduction process. On the other hand, Ball-pivoting not only keeps the original points of the contours, but also adds topological information to the models that can be helpful during the bone fracture reduction process.

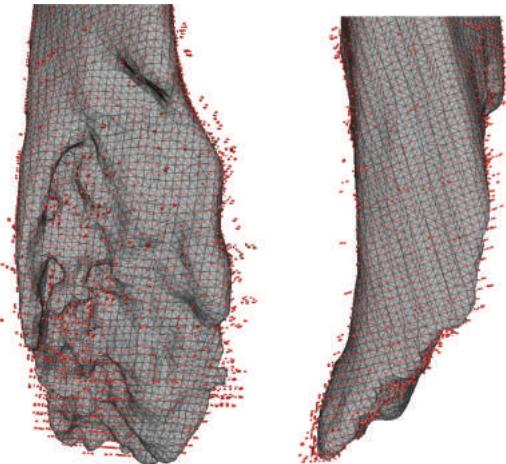


Figure 4. Two views of the same fracture zone. Model was generated using Poisson reconstruction. Original points are in red. Differences in position can be appreciated between original and generated points.

#### 4 DISCUSSION AND CONCLUSIONS

In this paper a comparison between two well-known methods to generate surfaces of bone fragments has been performed. This experiment has previously required extracting point clouds representing the outer part of each bone fragment from CT scans. In order to segment bone fragments from CT stacks, the method proposed in (Paulano, Jiménez, & Pulido 2014) has been used. The method has been able to segment bone regions and to separate bone fragments. Then contours have been extracted and point normals have been estimated using a new approach. Finally, some tests have been performed with the aim of knowing the benefits and shortcomings of each method when generating models of different bones and fracture types.

The performed tests have shown the advantages and disadvantages of each method when generating

surfaces of fractured bones. Poisson surface reconstruction obtained visually good results, but does not respect the original shape of the bone fragments. This can be a drawback in simulation processes that require accuracy, such as virtual fracture reduction. In addition small details in bone fragments can be removed. Ball pivoting obtained visually acceptable results in most cases and did not modify the original position of the points. However, the method could not assure that all the points are used to build the mesh because points are not always uniformly distributed. Therefore, incomplete results could be obtained.

According to the obtained results, the performance of the two methods is related to the quality of the CT stacks. Both methods obtain better results if the points are homogeneously distributed. For that purpose, the distance between slices should be smaller, increasing the radiation to the patient. Therefore, the development of new data acquisition techniques that reduce radiation would make improve the results obtained by both reconstruction algorithms. On the other hand, results of the segmentation also influence the obtained models. Due to the fracture, trabecular tissue is in the outer part of the bone. The irregularity of this type of tissue and the low resolution of the CT stack make difficult to correctly identify the boundary of bone fragments, especially when two fragments appear joined by their fracture zones in the CT image. This may lead to the generation of wrong surfaces in the fracture area.

## ACKNOWLEDGEMENTS

This work has been partially supported by the Ministerio de Economía y Competitividad and the European Union (via ERDF funds) through the research project TIN2011-25259.

## REFERENCES

- Bernardini, F., J. Mittleman, H. Rushmeier, C. Silva, & G. Taubin (1999, Oct). The ball-pivoting algorithm for surface reconstruction. *Visualization and Computer Graphics, IEEE Transactions on* 5(4), 349–359.
- Campos, R., R. Garcia, & T. Nicosevici (2011, June). Surface reconstruction methods for the recovery of 3d models from underwater interest areas. In *OCEANS, 2011 IEEE—Spain*, pp. 1–10.
- Feito, F.R. & J.C. Torres (1997). Inclusion test for general polyhedra. *Computers & Graphics* 21(1), 23–30.
- Fürnstahl, P., G. Székely, C. Gerber, J. Hodler, J.G. Snedeker, & M. Harders (2012, April). Computer assisted reconstruction of complex proximal humerus fractures for preoperative planning. *Medical image analysis* 16(3), 704–20.
- Ho, C.-c., F.-c. Wu, B.-Y. Chen, Y.-Y. Chuang, & M. Ouh young (2005). Cubical marching squares: adaptive feature preserving surface extraction from volume data. *Computer Graphics Forum* 24(3), 537–545.
- Jaramaz, B., M. Hafez, & A. DiGiioia (2006, Sept). Computer-assisted orthopaedic surgery. *Proceedings of the IEEE* 94(9), 1689–1695.
- Kazhdan, M., M. Bolitho, & H. Hoppe (2006). Poisson surface reconstruction. In *Proceedings of the fourth Eurographics symposium on Geometry processing*, Cagliari, Sardinia, Italy, pp. 61–70. Eurographics Association.
- Lorensen, W.E. & H.E. Cline (1987). Marching cubes: A high resolution 3D surface construction algorithm. In *ACM Siggraph Computer Graphics*, Volume 21, pp. 163–169. ACM.
- Paulano, F., J. Jiménez, & R. Pulido (2013). *Trends on identification of fractured bone tissue from CT images*, pp. 263–269. CRC Press.
- Paulano, F., J.J. Jiménez, & R. Pulido (2014, May). Dsegmentation and labeling of fractured bone from CT images. *The Visual Computer* 30(6–8), 939–948.
- Paulano, F., J.J. Jiménez, & R. Pulido (2014). Identification of fractured bone tissue from ct images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*.
- Pulido, R., F. Paulano, & J.J. Jiménez (214). Reconstruction & interaction with 3d simplified bone models. In *WSCG 2014 Communication Papers Proceedings*, pp. 321–327.
- Sugano, N. (2003). Computer-assisted orthopedic surgery. *Journal of Orthopaedic Science* 8(3), 442–448.
- Wiemann, T., H. Annuth, K. Lingemann, & J. Hertzberg (2015). An extended evaluation of open source surface reconstruction software for robotic applications. *Journal of Intelligent & Robotic Systems* 77(1), 149–170.
- Zhou, B., A. Willis, Y. Sui, D. Anderson, T. Brown, & T. Thomas (2009, September). Virtual 3D bone fracture reconstruction via inter-fragmentary surface alignment. In *IEEE Workshop on 3D Digital Imaging and Modeling (3DIM)*, pp. 1809–1816. IEEE.

# Towards a robust patch based multi-view stereo technique for textureless and occluded 3D reconstruction

B. Haines & L. Bai

*Department of Computer Science, University of Nottingham, Nottingham, UK*

**ABSTRACT:** Patch based reconstruction methods have been and still are one of the top performing approaches to 3D reconstruction to date. Their local approach to refining the position and orientation of a patch, free of global minimisation and independent of surface smoothness, make patch based methods extremely powerful in recovering fine grained detail of an objects surface. However, patch based approaches still fail to faithfully reconstruct textureless or highly occluded surface regions; thus, though performing well under lab conditions, they deteriorate in industrial or real world situations. They are also computationally expensive. Current patch based methods generate point clouds with holes in textureless or occluded regions that require expensive energy minimisation techniques to fill and interpolate a high fidelity reconstruction. Such shortcomings hinder the adaptation of the methods for industrial applications where object surfaces are often highly textureless and the speed of reconstruction is an important factor. This paper presents on-going work towards a multi-resolution approach to address the problems, utilising particle swarm optimisation to reconstruct high fidelity geometry, and increasing robustness to textureless features through an adapted approach to the normalised cross correlation. The work also aims to speed up the reconstruction using advances in GPU technologies and remove the need for costly feature based initialisation and expansion. Through the combination of these enhancements, it is the intention of this work to create denser patch clouds even in textureless regions within a reasonable time. Initial results show the potential of such an approach to construct denser point clouds with a comparable accuracy to that of the current top-performing algorithms.

## 1 INTRODUCTION

High performance 3D reconstruction via Multi-View Stereo (MVS) has received much interest from both academia and industry. Seitz et al. (2006a) and Jensen et al. (2014) both showed that MVS algorithms can achieve reconstruction accuracies paralleling that of prohibitively expensive laser scanners. However, to achieve robustness for a wide range of input imagery, such works demand high processing times limiting their adoption into the public domain.

Most MVS algorithms define a surface and check its fidelity via its photo-consistency value throughout the image set. This value is normally computed using similarity metrics such as the Normalised Cross Correlation (NCC) or the Sum of Squared Differences (SSD). The issue plaguing reconstruction algorithms of this sort is that the search space for the lowest cost photo-consistent surface is not only extremely large but also extremely costly to evaluate. Large high resolution image sets, with textureless regions, occlusion and lighting assumptions make the matching process highly complex.

3D reconstruction via MVS can roughly be divided into two categories, global and local.

Global based methods perform reconstruction by defining a global objective on the surface; hence, minimising an energy functional based on the photo-consistency score over the entire surface. Vogiatzis et al. (2007), model the problem as a graph cut, extracting the globally optimal solution to a weighted minimal surface. The energy functional is biased towards the minimal surface, smoothing out high frequency details in favour of lower surface area, as such a user defined ballooning term must be added. Such graph cut methods struggle to capture high frequency surface variations whilst also reconstructing thin structures due to the over-smoothing induced by the ballooning term, volume resolution and minimal surface requirement.

Level set methods presented by Cremers and Kolev (2011), or mesh based deformation frameworks of Esteban and Schmitt (2004) and Hiép et al. (2009), extract the locally minimal solution to the globally defined energy functional by iteratively deforming the surface based on its local derivatives. If the energy function is convex then these can also converge to the globally optimal solution. These methods evolve based on two terms in the energy functional, a data-term measuring

the photo-consistency and a regularisation term, enforcing smoothness constraints on the surface. Along with high computational cost, the regularisation term often leads to an overly-smoothed solution lacking sharp features.

Local methods focus on refining only a small patch of the surface. A planar patch,  $p$ , is created with position,  $\mathbf{c}(p)$ , and normal,  $\mathbf{n}(p)$ . The view with the shallowest angle to the patch normal observes this patch best, becoming the patches reference view,  $R(p)$ . This patch is then confined to lie on a viewing ray through the reference views camera centre.

To refine the patch to better fit the surface its back projection in the image set is sampled over a  $\mu * \mu$  pixel window, and its photo-consistency score is evaluated. The goal of refinement is to find the optimal position and normal in which the photo-consistency score is lowest. To optimise the patch parameters Habbecke and Kobbelt (2007) suggests using the Gauss-Newton algorithm to solve the non-linear least squares problem, whereas Furukawa and Ponce (2010a) suggests using a conjugate gradient method. Both methods rely on derivative based optimisation and hence a precise initialisation of the patch is very important to avoid convergence to an incorrect local minima.

Furukawa and Ponce (2010a) proposed the well-known PMVS algorithm which begins by initialising a sparse patch cloud through feature matching within the image set. These matches can be robustly refined and then expanded to initialise new patches close to those just created. Following expansion a filtering step is performed to remove erroneous patches that occur at depth discontinuities, occlusion or textureless areas. The reliance on feature based initialisation, expansion and filtering limits the extent at which patches can expand into textureless regions; therefore, highly textured regions are densely covered and textureless regions, or regions seen only by a few views, are sparsely covered, if at all. Chen et al. (2013) replaced the derivative based refinement stage with a Particle Swarm Optimisation (PSO) following a similar matching, expansion and filtering process. This technique showed promise in filling textureless areas but still does not allow for fully dense patch cloud creation for objects with large textureless areas, where no matching can be obtained, or datasets with a low number of input views.

This paper proposes an alternate approach to patch based MVS with the following contributions:

- An improved NCC photo-detail-consistency metric allowing better handling of textureless image channels.
- A PSO approach to patch refinement avoiding the need for precise feature based initialization and expansion.

- A multi-resolution approach creating denser patch clouds in textureless image regions.

## 2 NORMALISED CROSS CORRELATION

The NCC metric returns the inner product between two image patches, reflecting a measure of similarity between them. The metric is very popular for comparing image patches due to its invariance to both additive and multiplicative global illumination changes of the form  $I \mapsto \alpha I + \beta$ , where  $\alpha$  is an additive and  $\beta$  a multiplicative global illumination change. This allows for two images that were taken at different times under different illuminations to be compared as if they were taken instantaneously under the same lighting conditions.

$$\gamma(p) = 1 - \frac{1}{3|V(p)/R(p)|} \sum_{v \in V(p)/R(p)} g(I_{R(p)}, I_{v,c}) \quad (1)$$

The photo-consistency value  $\gamma(p)$  of a patch  $p$  visible in the set of images  $V(p)$ , with  $R(p)$  as the reference view, is shown in Equation 1. The NCC for the patch window projected into images  $I_1$  and  $I_2$ , for the colour channel,  $c$ , is denoted by  $g(I_1, I_2, c)$ . Each colour channel computes a NCC score, and each view computes a score with the reference view. The score is then averaged out and converted to a minimisation problem to compute the final photo-consistency score.

### 2.1 Limitations of NCC

The NCC performs very well in highly textured circumstances due to its invariance to illumination changes, however when using it to compare continuous coloured patches, its value is undefined. Given the presence of image noise it is unlikely that textureless surface patches will be completely constantly coloured in either the reference or comparison views. However, in such cases the NCC still fails to return useful information relating to similarity of the patches. Rather it returns a measure of the noise correlation that does not relate to the underlying patch similarity. Using an NCC based similarity metric to refine textureless patches fails, as the minimum score does not relate to the optimum patch parameters.

### 2.2 Photo-detail-consistency

The photo-consistency measure defined in Equation 1 can suffer similar anomalous results when comparing highly textured patches as it does

in the textureless case. If the NCC is computed per image channel, and one or more of the image channels have low texture, these channel will return incorrect measures. This effectively skews the similarity score when summed and averaged with the results from the highly textured channels. Further to this, when a patch is projected into an image where the angle between the patch and image plane is high, bilinear interpolation is used to sample the patch. In cases where the patch axes fail to cover sufficiently many pixels in the image based matching window, bilinear interpolation returns similar values, creating a textureless effect.

$$\gamma(p) = 1 - \frac{1}{3|V(p)/R(p)|} \sum_{v \in V(p)/R(p)} w_v \sum_{c \in \{r, g, b\}} w_c g(I_{R(p)}, I_v, c) \quad (2)$$

Equation 2 proposes adaptively weighting the NCC score based on weights  $w_c$  and  $w_v$ . We define this new similarity metric as the *photo-detail-consistency*.  $w_c$  computes the amount of texture per image channel, and  $w_v$  weights the angle of the image plane to the patch normal.  $w_c$  prevents textureless colour channels contributing to the similarity metric. Thus it is close to zero when the texture on the image channel is low and high when there is sufficient texture.

$$w_c = 1 - e^{-t^2/2\sigma_c^2} \quad (3)$$

where  $t = \|\nabla I_{R(p),c}(x, y)\|$

In Equation 3,  $w_c$  is computed per patch,  $p$ , where  $\nabla I_{R(p),c}(x, y)$  is the gradient of the colour channel,  $c$ , for its reference image,  $I_{R(p)}$ . This is taken at the image point  $(x, y)$ , that is the projection of the patch center,  $\mathbf{c}(p)$ , into its reference image. In practice the detail coefficients of the Haar Wavelet Transform from the multi-resolution approach presented in Section 4 are used to approximate the image gradient.

$$w_v = \begin{cases} 0 & \text{if } \mathbf{n}(p) \cdot \mathbf{vd}(v) > 0 \\ 1 - e^{-\mathbf{n}(p) \cdot \mathbf{vd}(v)/2\sigma_v^2} & \text{if } \mathbf{n}(p) \cdot \mathbf{vd}(v) \leq 0 \end{cases} \quad (4)$$

To help reduce the influence of views with sharper viewing angles,  $w_v$  is calculated per view using the patches normal,  $\mathbf{n}(p)$ , and the current views principle view direction,  $\mathbf{vd}(v)$ . Factors  $\sigma_c$  and  $\sigma_v$  from Equations 3 and 4 respectively are tunable parameters controlling the strength of the filtering; smaller values represent stricter filtering. In our experiment  $\sigma_c$  and  $\sigma_v$  are fixed at 0.2.

### 3 PARTICLE SWARM OPTIMISATION

Derivative based optimisation methods such as gradient or conjugate gradient decent, Newton and Quasi-Newton methods are popular techniques for optimising a functions parameters. They are often highly effective when correctly initialised, in most cases requiring only a small number of iterations to converge. However, in cases of poor initialisation or a noisy cost function such methods suffer from a tendency to get stuck in local minima, where the derivatives of the cost function become zero. As such a careful and costly initialisation or the need to run the method over multiple seed points is required, reducing their accuracy and efficiency.

PSO is a derivative free population based stochastic optimisation method for continuous non-linear functions, introduced by Kennedy (2010). Inspired by the social behaviours of animals such as birds flocking or fish schooling, it can be used to iteratively minimise or maximise the solution to a given candidate function, using a population based search strategy with deterministic and probabilistic rules. All particles in a swarm move through the search space looking for the best solution and updating their position using a velocity defined in Equation 5.

$$\begin{aligned} \mathbf{v}^i(t+1) = & \omega * \mathbf{v}^i(t) + \vartheta_p * \text{rand}() * (\mathbf{pbest}^i - \mathbf{x}^i(t)) \\ & + \vartheta_g * \text{rand}() * (\mathbf{gbest} - \mathbf{x}^i(t)) \\ & + \vartheta_l * \text{rand}() * (\mathbf{lbest}^i - \mathbf{x}^i(t)) \\ & + \vartheta_n * \text{rand}() * (\mathbf{nbest}^i - \mathbf{x}^i(t)) \end{aligned} \quad (5)$$

The  $i$ -th particle of the swarm, at the  $t$ -th iteration, is updated based on its current position,  $\mathbf{x}^i(t)$ , and velocity,  $\mathbf{v}^i(t)$ . An inertia factor,  $\omega$ , is used to help stop the effects of stagnation. Stagnation occurs when a particles velocity becomes zero as its position coincides with the current best solution, all other particles will be drawn towards it. If no other particle improves its solution on the way towards the current particle this may lead to premature convergence, where no further improvement can be made.

The evolution of each particle in the swarm is also controlled by different social factors: its personal best visited location,  $\mathbf{pbest}^i$ , the global best location found over the entire swarm,  $\mathbf{gbest}$ , the best position found by its local neighbourhood in terms of its index into the swarm,  $\mathbf{lbest}^i$ , and finally the best position found in terms of its nearest spatial neighbours,  $\mathbf{nbest}^i$ . The influence of the personal, global, local and nearest social factors are weighted by  $\vartheta_p$ ,  $\vartheta_g$ ,  $\vartheta_l$  and  $\vartheta_n$  respectively. Finally  $\text{rand}()$  simulates the stochastic nature of flocking, generating a random number in the range  $[-1, 1]$ .

Setting  $\vartheta_p$ ,  $\vartheta_g = 2$  and  $\vartheta_l$ ,  $\vartheta_n = 0$ , results in the traditional particle swarm optimisation introduced by Kennedy (2010). Setting  $\vartheta_p$ ,  $\vartheta_l = 2.05$  and  $\vartheta_g$ ,  $\vartheta_n = 0$ , defines the improved Standard Particle Swarm Optimisation (SPSO) presented by Bratton and Kennedy (2007). Setting  $\vartheta_p$ ,  $\vartheta_g$ ,  $\vartheta_l$ ,  $\vartheta_n = 1$ , results in Pongcharerks and Kachitvichyanukul (2005) Global Local Nearest PSO (GLN-PSO) structure. Mutation operations from Wang et al. (2007) can be applied, regardless of the social structures used, to further avoid the effects of stagnation.

### 3.1 PSOs application in MVS

In the context of MVS the cost function in PSO will become the per-patch photo-detail-consistency measure. As this is computationally expensive to evaluate we implement our PSO algorithm on the GPU using OpenCL to achieve real time performance, similarly to the synchronous version presented by Zhou and Tan (2009).

Chen et al. (2013) originally proposed the use of GLN-PSO to optimise a patches parameters inside the framework of PMVS. By using the sparsely matched features as initialisation they were able to restrict the optimisation search space to nearby depth and normal values, thus performing optimisation similarly to gradient decent. Their results proved effective but due to their reliance on feature detection, expansion and filtering their approach is limited in terms of expanding into larger textureless regions. We look to build on their work by removing the reliance on feature matching and widening the search space of PSO, hoping to achieve a similar result without the requirement of feature matching.

Although GLN-PSO can have superior performance in terms of finding the optimal solution in less iterations, its reliance on the global best particle and neighbouring best particles positions are not easily adapted to a GPU based algorithm. Knowledge of the global best particle would require running a reduction operation on all the particles to find the current best solution per iteration of the swarm. Then each particle would require a slow global memory read to access this global best position over each kernel. Further to this calculating a particles spacial neighbours also requires high levels of computation that is not easily parallelized to the GPU. Therefore we opt to excluded these high computational force operations only using personal and local forces in the form of SPSO. Mutation operations are also randomly applied to add further robustness.

The search space for patch refinement can be relatively large, but it is of low dimensionality, one for the patch depth,  $\lambda$ , along its viewing ray and two for the patch normal expressed in spherical

coordinates,  $\theta$  and  $\phi$ . In order to constrain the search space without the requirements of feature matching, the depth range is set to the patches intersection of its viewing ray with the scenes bounding volume. This is easily found using a simplex algorithm to compute the convex hull of the viewing volume. If silhouettes are available the visual hull can be obtained and the depth range can be further restricted to lie within the front and back intersection of the viewing ray to the visual hull. This information can be extracted for every pixel, regardless of whether it contains a feature match. This also guarantees that the search space must contain the correct depth position, regardless of how deep any surface concavity may be.

The patch normal search space, expressed in spherical coordinates relative to its reference view, is restricted such that  $\theta, \phi \in [0, \pi]$ . This ensures that the patch does not become back facing in the reference view. Having such wide search space on the patch normal allows for sharp grazing angles of the patch against the reference view. Scaling up the patch to a  $\mu * \mu$  matching window in the reference view should go some way to account for this effect, whilst still allowing correct patch refinement for these conditions.

Given that PSO refinement of a patch only requires knowledge of an approximate depth and normal range, there is no need to perform the matching and expansion procedures needed for precise initialisation of derivative based methods. Instead for each pixel in the silhouette we can automatically generate a search space and refine a patch, following a patch per pixel approach. From only a coarse search space initialisation, in highly textured areas, PSO performs comparatively to derivative based methods, converging on the same optimia.

Using our implementation of GPU based PSO we are able to refine a single  $5 \times 5$  patch in roughly 2–4 ms, depending on the number of iterations of the swarm. No doubt this speed could be decreased by a more heavily optimised kernel, however this has served well as a proof of concept. Increasing the patch size deteriorates the run time and also begins to over-smooth the patch cloud. As larger patches approximate a larger surface area, reducing the patches ability to reconstruct high frequency surface details.

Figure 1 compares the results of running PMVS, the current top performing patch based method, relying on gradient decent and expansion from precisely matched features, against our PSO based approach for the sparse ‘Dino’ benchmark available from Seitz et al. (2006b). Source code for the PMVS algorithm is available from Furukawa and Ponce (2010b).

Figure 1(a) shows that PMVS failed to create as dense a set of patches, hence is unable to express

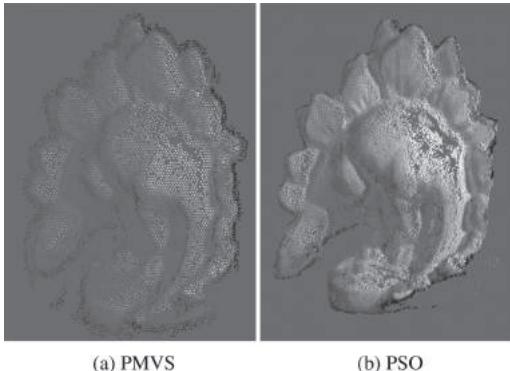


Figure 1. Reconstruction results for the sparse Dino dataset.

high frequency surface details such as the variation on the fins of the ‘Dino’ model. In low textured areas the patch cloud contains large holes, where patches are filtered out due to incorrect refinement, particularly note the dinosaur’s back and legs. Figure 1(b) shows a much denser patch cloud created by our PSO refinement, using the patch per pixel approach. In highly textured areas it is able to capture the same degree of accuracy but at a significantly higher resolution. However in low textured areas where the variation over the patch only expresses noise, our method also produces incorrectly refined patches. Rather than filter these out and create holes, to help resolve incorrect refinement a multi-resolution approach is proposed in the following section.

#### 4 MULTI-RESOLUTION

The lifted Haar wavelet transform, Sweldens and Schröder (2000), is a popular tool for signal processing and multi-resolution analysis. It takes an image and removes high frequency components storing them in detail coefficients. The high frequency components are then removed from the input image, forming a lower resolution approximation image, which in turn is stored in approximation coefficients. The original image can be reconstructed by applying the inverse wavelet transform, combining detail coefficients with the approximation coefficients. The approximation image can also be further decomposed until a desired level is reached forming a full multi-resolution decomposition. Classically this transform is applied using high and low pass filters in the frequency domain, however in a second generation lifted wavelet framework, these filters can be expressed in the spatial domain. This allows the transform to be performed faster

as well as in-place, requiring no extra memory than is already required to store the images, which is important when dealing with large, high resolution image sets.

##### 4.1 Wavelets for PSO patch refinement

In theory patch based methods are only limited in accuracy by the resolution of the input imagery. However, in practice the size of the patch and its corresponding matching window in image space can impact its ability to reconstruct high frequency surface details. Klowsky et al. (2012) presented a study on how patch size affects the corresponding frequencies that can be reconstructed. They show that adapting the patch size to the surface frequency is incredibly important in defining surface details. Smaller patches require less samples and computation, and can reconstruct higher frequency details, however, they’re less stable and susceptible to errors in refinement due to lack of texture variance over the samples of the patch. This limitation occurred in the PSO method presented above, where refining small patches over textureless regions meant that the global minimum of the photo-detail-consistency function no longer corresponds to the correct patch parameters.

Increasing the patch size creates a more stable and less noisy patch cloud filling larger textureless regions, however, this comes at a cost of deteriorating the run time and over smoothing the patch cloud. Therefore, using the correct patch size for accuracy and computational gains is a core problem in a patch based method, choosing an optimal patch size is a challenging problem. Habbecke and Kobbelt (2007) suggest using the texture variance to adapt the patch size, however, this can lead to high computational cost as the number of surface samples grows with the patch size. Here we adopt a multi-resolution approach to achieve a similar effect, without the cost of increasing the surface samples.

The Haar wavelet transform compresses large textureless areas into smaller ones, reducing the number of textureless pixels that describe the underlying surface of a patch. Hence, lower resolution approximation images induce textured regions and a stabilising effect in a similar way to increasing the extent of the patch, without increasing the cost of sampling and computing the photo-detail-consistency metric. Therefore, the PSO algorithm can be run at a lower resolution, where patches are more robust to textureless areas. This can be seen in Figure 2, showing the results of PSO refinement at multiple resolutions. In textureless regions, patch clouds generated from lower resolution images do not normally suffer the same problems as their higher resolution counterparts. In particular, note how the highly textureless region over the dinosaurs back is stably recovered at

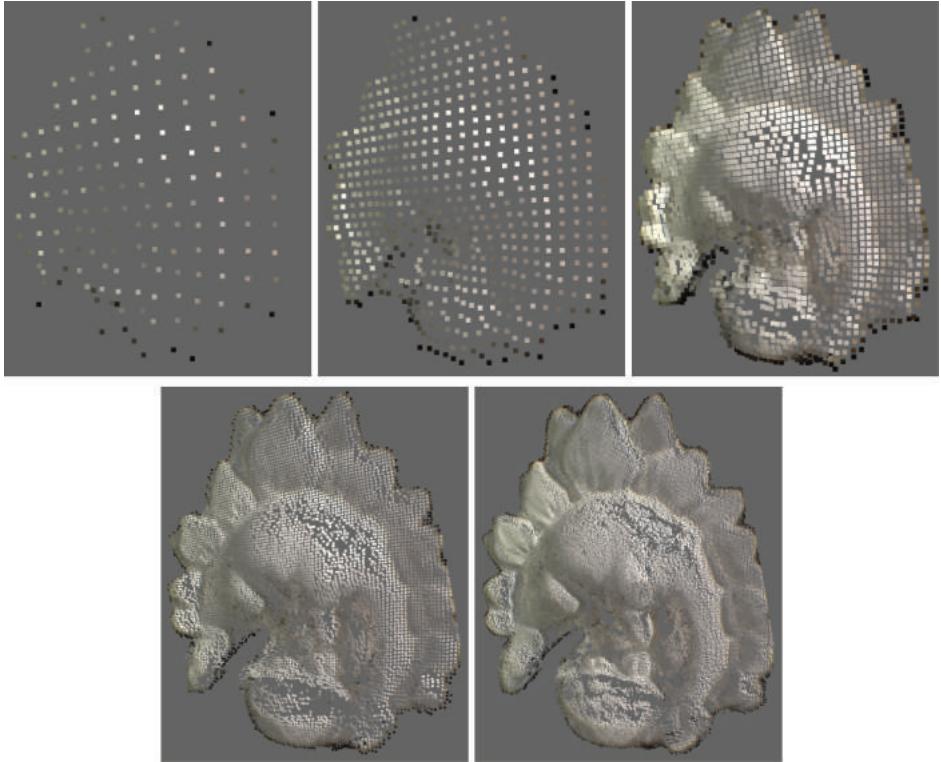


Figure 2. PSO results over the multi-resolution scales from coarse to fine resolutions respectively. Initial low resolution scales show robustness to large textureless regions whereas fine scales show the ability to reconstruct high frequency information but also suffers from erroneously refined patches.

lower but not higher resolutions, where this region becomes steadily more uniform as the resolution increases.

Once the patch cloud is stably initialised at the lowest level the image resolution can be raised and the results propagated upwards. PSO is run at the higher resolution using the previous resolutions optimised result as a velocity factor added to Equation 5. This allows for reconstruction of surface details not expressed in the lower resolution images. Further to this we can decide not to run PSO refinement for a higher resolution pixel if the detail coefficients of the wavelet transform reflect a non-textured patch. In these cases, rather than run PSO, we smooth the patch parameters to obtain a consistent patch cloud. Figure 2 shows promise that with correct propagation the refinement of patches in textureless regions can be improved significantly.

## 5 CONCLUSIONS

This paper has presented on-going work towards a multi-resolution patch based approach to MVS.

To improve the performance of the NCC, a photo-detail consistency measure is defined, weighting the NCC based on the patches texture and viewing angle. To increase the point clouds density a PSO patch refinement scheme is run per pixel allowing for only approximate initialisation and removing the need for feature based matching and expansion. Finally a multi-resolution scheme using the wavelet transform is proposed, propagating refined patch results for textureless surfaces up through the resolutions, where consistency measures start to fail. The experimental results show the improved patch clouds and the potential for more accurate MVS reconstruction in textureless regions through propagation of data from lower resolutions.

## REFERENCES

- Bratton, D. & J. Kennedy (2007). Defining a standard for particle swarm optimization. In *Swarm Intelligence Symposium, 2007. SIS 2007. IEEE*, pp. 120–127. IEEE.

- Chen, Z., W.-C. Chen, & P.-Y. Sung (2013). A novel 3d dense reconstruction with high accuracy and completeness. In *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, pp. 1–4. IEEE.
- Cremers, D. & K. Kolev (2011). Multiview stereo and silhouette consistency via convex functionals over convex domains. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 33(6), 1161–1174.
- Esteban, C.H. & F. Schmitt (2004). Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Under-standing* 96(3), 367–392.
- Furukawa, Y. & J. Ponce (2010a). Accurate, dense, and robust multiview stereopsis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 32(8), 1362–1376.
- Furukawa, Y. & J. Ponce (2010b). Patch based multi-view stereo.
- Habbecke, M. & L. Kobbelt (2007). A surface-growing approach to multi-view stereo reconstruction. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pp. 1–8. IEEE.
- Hiep, V.H., R. Keriven, P. Labatut, & J.-P. Pons (2009). Towards high-resolution large-scale multi-view stereo. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 1430–1437. IEEE.
- Jensen, R., A. Dahl, G. Vogiatzis, E. Tola, & H. Aanæs (2014). Large scale multi-view stereopsis evaluation. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pp. 406–413. IEEE.
- Kennedy, J. (2010). Particle swarm optimization. In *Encyclopedia of Machine Learning*, pp. 760–766. Springer.
- Klowsky, R., A. Kuijper, & M. Goesele (2012). Modulation transfer function of patch-based stereo systems. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 1386–1393. IEEE.
- Pongcharerks, P. & V. Kachitvichyanukul (2005). Non-homogenous particle swarm optimization with multiple social structures.
- Seitz, S.M., B. Curless, J. Diebel, D. Scharstein, & R. Szeliski (2006a). A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*, Volume 1, pp. 519–528. IEEE.
- Seitz, S.M., B. Curless, J. Diebel, D. Scharstein, & R. Szeliski (2006b). Multi-view stereo evaluation web page, <http://vision.middlebury.edu/mview/data>.
- Sweldens, W. & P. Schröder (2000). Building your own wavelets at home. In *Wavelets in the Geosciences*, pp. 72–107. Springer.
- Vogiatzis, G., C. Hernández, P.H. Torr, & R. Cipolla (2007). Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29(12), 2241–2246.
- Wang, H., Y. Liu, & S. Zeng (2007). A hybrid particle swarm algorithm with cauchy mutation. In *Swarm Intelligence Symposium, 2007. SIS 2007. IEEE*, pp. 356–360. IEEE.
- Zhou, Y. & Y. Tan (2009). Gpu-based parallel particle swarm optimization. In *Evolutionary Computation, 2009. CEC'09. IEEE Congress on*, pp. 1493–1500. IEEE.

This page intentionally left blank

# Forming tool inspection using fiber-optic sensor heads

S. Matthias, M. Kästner & E. Reithmeier

*Institute of Measurement and Control, Leibniz Universität Hannover, Hannover, Germany*

P. Sieczkarek & A.E. Tekkaya

*Institut für Umformtechnik und Leichtbau, Technische Universität Dortmund, Dortmund, Germany*

**ABSTRACT:** In-situ measurements of tools allow inspecting a running metal forming process. This enables a fast adaption of the process and helps reducing costs. However, the requirements on the measurement systems are high, as the measurement duration needs to be low and the available space inside the machines is limited. Commercially available measurement systems are unable to operate in these conditions. An application in metal forming is presented followed by the introduction of a novel design for an optical measurement system based on the fringe projection profilometry technique. The new design employs flexible fiber bundles to transport the images of the fringe projection system to a compact sensor head which is based around gradient-index lenses. The pattern generation unit and camera are housed in a base unit, which can be placed outside of the machine. While this new approach allows flexible positioning of the sensor head, it creates challenges for the pattern evaluation algorithms and the optical design of the system. The laser-based optical setup will be presented followed by measurements of the tool of the aforementioned forming process. To evaluate the applicability of the new measurement system, the determined tool parameters will be compared to reference measurements obtained by a tactile Hommel T8000 system. In combination with a positioning device, which is currently being developed, the new measurement system will be used to obtain wear information of the running forming process in order to control the process quality loop in the future.

## 1 INTRODUCTION

A new metal forming process is being researched within the CRC/TR 73 collaborative research center funded by the German Research Foundation (DFG) (Merklein, Allwood, Behrens, Brosius, Hagenah, Kuzman, Mori, Tekkaya, & Weckenmann 2012). The aim of the sheet-bulk metal forming process is to reduce the required number of forming steps to manufacture complex geometries. Additionally, the production of more sophisticated parts with reduced weight or improved stability is made possible. The scope of research covers the evaluation of different materials, surface treatments and tool geometries. For the production of gearings or similar elements, high forces are necessary in order to achieve proper molding. To be able to monitor wear on the forming elements in the limited space of the forming presses, measurement devices are needed which can be positioned in the limited space of the forming machines. Many commercially available tactile and optical measurement devices are unable to operate under these conditions. In order to overcome these limitations, a combination of endoscopy techniques with contact-free optical measurement principles can be used in order to

acquire geometry data. Different approaches are currently being researched, which use passive stereometry, structured light, confocal techniques or the time-of-flight principle (Geng & Xie 2014). This paper describes an exemplary sheet-bulk metal forming process based on a multi-axis press, followed by the introduction of a new endoscopic measurement system. Based on reconstruction by triangulation and structured light, the new fringe projection system is able to obtain areal measurements of a specimen's geometry in the form of a 3D point cloud. Based on this data, further analysis, such as comparisons against CAD reference data, can be carried out. The new system features a laser and an adjustable pattern generator for fast measurements, as well as exchangeable sensor heads to be able to adapt to different measurement tasks. Based on the description of the forming tool for the presented process, measurements of tool features are shown. Results from the new system are compared to reference data obtained with a tactile profile sensor. Potential benefits of obtaining real-time information on tool wear with the new system for the evaluation of important process parameters, such as lubrication or forming trajectories, are discussed.

## 2 FORMING PROCESS

The development of flexible forming technologies aims to manufacture load-adapted and weight-optimized components by a high integration of functional elements. A suitable approach to reach this goal is the incremental sheet-bulk metal forming process. In addition to reducing the forming forces by decreasing the surface contact between tool and workpiece, the process offers high flexibility. The incremental procedure allows the manufacturing of near-net-shape parts with small functional elements, such as gearings. However, high positioning accuracies are required under load. With a specially developed multi-axis press, the application of different processes, such as embossing or rolling, is made possible with almost arbitrary trajectories (Sieczkarek, Kwiatkowski, Ben Khalifa, & Tekkaya 2013).

Shaping the gearings and tooth structures by a bulk metal forming process offers significant advantages over conventional machining. A continuous fiber orientation in the formed teeth leads to improved mechanical properties. In addition to the positively usable effect of strain hardening, no pores are exposed in the material, which are often the origin of cracks.

The incremental forming of tooth structures on the sheet edge (Fig. 1) requires thin forming tools, which need to be of the same or only slightly higher thickness compared to the sheet-metal parts. This is related to the necessary chambering of the workpiece to prevent a lateral material flow and to achieve the best possible shape of the tooth structure. The high local contact pressure leads to a reduced tool life as well as to a tool deformation (Sieczkarek, Kwiatkowski, Tekkaya, Krebs, Kersting, Tillmann, & Herper 2013). In order to achieve higher accuracies as well as stable forming tools,

the current research focusses on a reduction of the high tool wear.

With the objective of a near-net-shape production, a forming process is dependent on reliable, dimensionally accurate tools. The desired processing capabilities are targeting a value of  $10 \mu\text{m}$  (IT7). The knowledge of the process-related high tool wear suggests that in-situ measurements of the tool geometry are desirable for a possible industrial application in the future. Particularly in the context of incremental manufacturing, which features a high flexibility for the prototype segment or small batch production, much higher cycle times occur. These provide enough time for a tool measurement during the running process. Besides the control of the tool geometry, also a complete inspection of the process would be conceivable. The aspired geometrical accuracy of the produced components is affected by disturbances, such as strain, lubrication, number of forming stages, elastic spring-back of the setup and different material properties. With in-situ measurements, a target-performance comparison of wear effects could be conducted. In the future, direct adjustment of the targeted command variables of the quality control loop can be achieved with this additional information.

## 3 MEASUREMENT SYSTEM

In order to be able to perform measurements inside the limited space of a forming press, very compact dimensions of the sensor head are required. The newly developed endoscopic fringe projection achieves a very compact design by separating the sensor head from the pattern generation and camera unit with fiber bundles. These imaging fibers are supplied by Fujikura and feature a length of  $1000 \text{ mm}$  at a diameter of  $1.7 \text{ mm}$  and a resolution of  $100,000 \text{ pixels}$ . A minimum bending radius of  $200 \text{ mm}$  allows independent positioning of the sensor head and the base station.

The base station is designed with overlapping sides to safely enclose the laser illumination for the pattern generator. A single-mode laser with a wavelength of  $532 \text{ nm}$  is used for its good properties for fiber incoupling, as well as high light intensities with an optical output power of  $40 \text{ mW}$ . Challenges arise from interference patterns, the so called speckle, when measuring technical surfaces with a laser based projector. The contrast of these interference patterns can easily exceed the contrast of the fringe patterns, leading to high noise in the measurements. In order to reduce the effect of speckle, a former design of the projector beam path employed a rotating diffuse or plate, leading to temporally changing speckle patterns. By integrating over the camera's exposure duration, speckle can

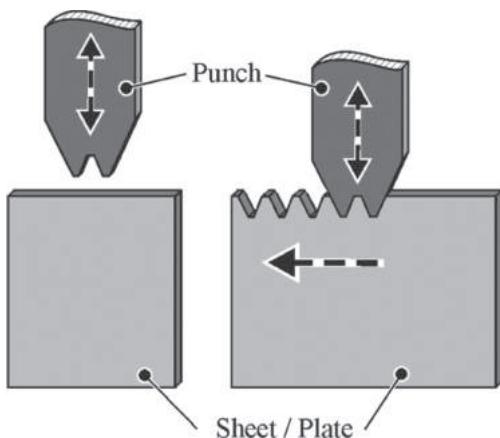


Figure 1. Schematic of the incremental forming process.

be mostly compensated. However, the motor for the rotating motion initiates vibrations. Therefore, the rotating diffusor has been displaced by a laser speckle reducer from Optotune, which performs the movement of the diffusor plate with electro-active polymers, resulting in a significant reduction of undesired vibrations (Blum, Büeler, Grätzel, Giger, & Aschwanden 2012). Another factor to be considered for the projector is the uniformity of brightness of the fringe patterns. The single-mode laser exhibits a typical Gauss-beam profile. For the purpose of generating a homogenous profile, two micro-lens arrays supplied by SUSS MicroOptics are used (Voelkel & Weible 2008). A mirror reflects the laser beam onto the active pixel matrix of a Digital Micro-mirror Device (DMD) (Hornbeck 1996). Due to the switching angle of  $\pm 12^\circ$  of the individual active mirrors, accurate adjustment of the beam path is necessary in order to focus the generated fringe patterns into the optical fiber bundles.

The detector part of the base unit consists of a monochrome Point Grey CCD-camera with a sensitivity maximum at about  $530\text{ nm}$  to complement the laser light source. Precise synchronisation of camera and projector is achieved by triggering image capture with the hardware output of the DMD-controller.

At the far end of the projector and camera fiber bundle, gradient-index lenses are directly attached to the planar fiber surface. The imaging rod lenses are manufactured by GRINTECH and are  $2\text{ mm}$  in diameter and less than  $6\text{ mm}$  in length. Depending on the desired measurement volume and working distance, lenses with a focal length of either  $10\text{ mm}$  or  $20\text{ mm}$  are used. For the  $10\text{ mm}$  optics a measurement volume of about  $3 \times 3 \times 3\text{ mm}^3$  can be achieved. A view of the tool through the endoscope with a  $20\text{ mm}$  lens can be seen in Figure 2.

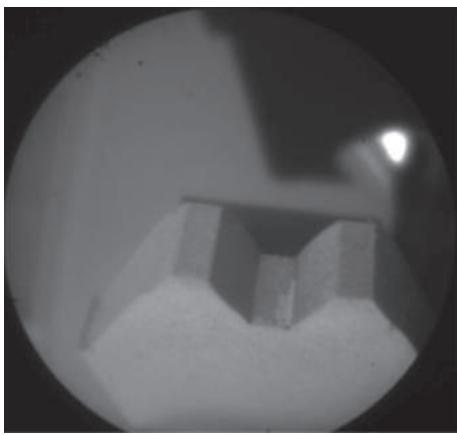


Figure 2. Photo of the tool captured through the endoscope.

In order to perform metric measurements, the system is calibrated with a planar calibration pattern according to a mixed pinhole and black-box model (Matthias, Kästner, & Reithmeier 2014). Measurements with 12 phase-shift patterns for the reconstruction of an absolute phasemap currently require a projection time of less than 3 seconds. Using the inverse fringe projection principle, geometry deviations from a reference object can be detected with a single image (Pösch, Vynnyk, & Reithmeier 2013). This reduces the needed time for tool inspection to less than one second.

#### 4 TOOL MEASUREMENTS

Figure 3 shows a drawing of the forming tool with the features to be captured. To obtain reference surface data, the forming tools are first measured using a Hommel T8000 tactile measurement device. By moving a ruby ball over the specimen's surface, a very accurate 2D-profile is acquired. The feature parameters are derived from the surface profiles by fitting a line to the data points of the tooth sides, base and top. In comparison, the endoscopic fringe projection system captures 3D surface data using absolute phase measurements. As 3D point data is acquired, planes are fitted to the respective tool surfaces to derive the angles as well as the tooth height. To improve the optical characteristics of the tool's surface, the tool is covered with a thin layer of diffuse spray for the fringe projection measurements.

Table 1 lists the results measured for two different forming tools and the difference of the results obtained with the fringe projection system compared with the tactile reference results. Only

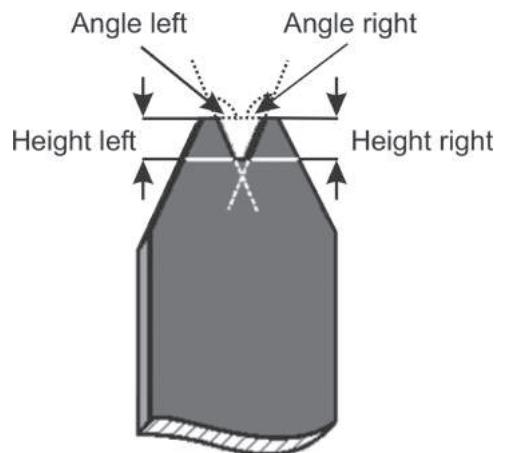


Figure 3. Drawing of the forming tool features.

Table 1. Comparison of feature measurements of the forming tool.

Feature	Reference		Fringe projection system		Tactile Hommel T8000		Difference optical/tactile	
	Tool 1	Tool 2	Tool 1	Tool 2	Tool 1	Tool 2	Tool 1	Tool 2
Height left	3.38 mm	3.38 mm	3.42 mm	3.36 mm	3.42 mm	3.33 mm	0.00 mm	0.03 mm
Height right	3.38 mm	3.38 mm	3.40 mm	3.38 mm	3.45 mm	3.38 mm	-0.05 mm	0.00 mm
Angle left	110.0°	110.0°	110.5°	109.6°	110.1°	109.9°	0.4°	-0.3°
Angle right	110.0°	110.0°	110.6°	110.8°	109.9°	109.8°	0.7°	1.0°

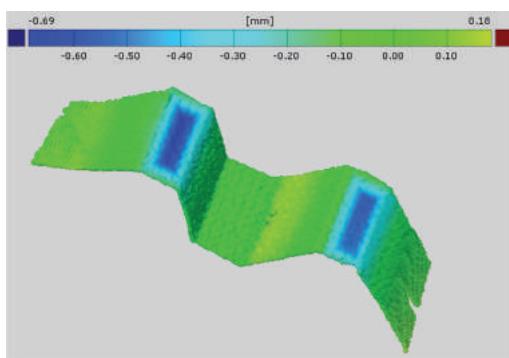


Figure 4. Deviation analysis of a damaged forming tool.

small disparities between the two devices are visible, with a maximum difference of  $50 \mu\text{m}$  for the tooth height and a maximum deviation of  $1.0^\circ$  for the tooth angles. Apart from measurement uncertainties, the differences may arise from the varying measurement principles, with the tactile device capturing 2D-profile data compared to 3D-surface data obtained by the fringe projection system. For example deformations on the tooth tip plane or a tilt to the sides can not be registered with a 2D profile. Additionally, some small differences might arise from the diffuse spray used to cover the tool surface to improve the optical measurements. Therefore, the performance of the system will be assessed by measurements of a geometry standard with precise tolerances in the future.

Figure 4 shows a deviation analysis of a measurement of a damaged tool compared to the CAD reference data with the evaluation software GOM Inspect. The color encodes the distance to the next reference surface. The number of data points in this measurement is 18657. It can be seen that the tip of the gearing elements shows large deformations. The height of the teeth is reduced by  $0.6 \text{ mm}$ , resulting in a push of material to the sides. By obtaining 3D information about the deformation of the tool, analysis of material flow is made possible. Also it

is possible to detect and assess the extend of wear on the tool to decide a tool change in the running process based on these measurement, once characteristic deformations are categorized.

## 5 CONCLUSION AND OUTLOOK

An application for in-situ measurements in metal forming has been introduced. To overcome the high requirements in terms of measurement time and space restrictions, the design of a new measurement system based on fringe projection profilometry and endoscopy techniques has been introduced. The capabilities of the device have been demonstrated by tool measurements of the presented forming processes. It could be shown that the system meets the accuracy requirements by comparisons to reference data obtained by tactile measurements.

To enable operation of the fiber-based fringe projection system in the forming press, a positioning system is being devised. This will enable capturing real-time measurement data from the running process. In order to further assess the applicability of the endoscopic measurement system, comparisons with a geometry standard will be performed. Analysis of deformations of characteristic tool features will help to decide tool changes if certain thresholds are surpassed.

## ACKNOWLEDGEMENT

The authors would like to thank the German Research Foundation (DFG) for funding the projects B6 “Endoscopic geometry inspection” and A4 “Incremental Forming” within the Collaborative Research Center (CRC)/TR 73.

## REFERENCES

- Blum, M., M. Büeler, C. Grätzel, J. Giger, & M. Aschwanden (2012). Optotune focus tunable lenses and laser speckle reduction based on electroactive polymers. In *SPIE MOEMS-MEMS*, pp. 825207–825207. International Society for Optics and Photonics.

- Geng, J. & J. Xie (2014). Review of 3-d endoscopic surface imaging techniques. *Sensors Journal, IEEE* 14(4), 945–960.
- Hornbeck, L.J. (1996). Digital light processing and mems: An overview. In *Advanced Applications of Lasers in Materials Processing/Broadband Optical Networks/ Smart Pixels/Optical MEMs and Their Applications. IEEE/LEOS 1996 Summer Topical Meetings*:., pp. 7–8. IEEE.
- Matthias, S., M. Kästner, & E. Reithmeier (2014). Camera and projector calibration of an endoscopic fringe projection system. In *Proceedings of 11th IMEKO Symposium LMPMI2014*.
- Merklein, M., J. Allwood, B.-A. Behrens, A. Brosius, H. Hagenah, K. Kuzman, K. Mori, A. Tekkaya, & A. Weckemann (2012). Bulk forming of sheet metal. *CIRP Annals-Manufacturing Technology* 61(2), 725–745.
- Pösch, A., T. Vynnyk, & E. Reithmeier (2013). Fast detection of geometry defects on free form surfaces using inverse fringe projection. *Journal of the CMSC* 8(1), 10–13.
- Sieczkarek, P., L. Kwiatkowski, N. Ben Khalifa, & A.E. Tekkaya (2013). Novel five-axis forming press for the incremental sheet-bulk metal forming. *Key Engineering Materials* 554, 1478–1483.
- Sieczkarek, P., L. Kwiatkowski, A.E. Tekkaya, E. Krebs, P. Kersting, W. Tillmann, & J. Herper (2013). Innovative tools to improve incremental bulk forming processes. *Key Engineering Materials* 554, 1490–1497.
- Voelkel, R. & K.J. Weible (2008). Laser beam homogenizing: limitations and constraints. In *Optical Systems Design*, pp. 71020J–71020J. International Society for Optics and Photonics.

This page intentionally left blank

# Navigation of robotics platform using advanced image processing navigation methods

L. Beran, P. Chmelař & L. Rejsek

*Faculty of Electrical Engineering and Informatics, University of Pardubice, Pardubice, Czech Republic*

**ABSTRACT:** This paper deals with methods suitable for movable robotics platform navigation. We can divide these methods into suitability for indoor navigation, outdoor navigation or its combination. At present, the navigation methods of robotics platforms, suitable for outdoor spaces, are based on GNSS (Global Navigation Satellite System). Accurate position estimation is a big advantage for this type of navigation. Unfortunately this type of navigation is not suitable for indoor spaces and industrial areas, mainly because of the weak signal and low position accuracy. An alternative navigation system is necessary for these solutions—system independent on satellite navigation. Our research is focused on movable robotics platform navigation using advanced image processing methods. Descriptions of these sophisticated methods are presented in this paper.

**Keywords:** monocular camera vision; image processing; visual odometry; robot navigation; surface detection

## 1 INTRODUCTION

At present, the systems for automatic space mapping are widely distributed in many fields of robotic navigation. This automatic mapping system needs a precision navigation system. In the most cases, places are unknown or dangerous; like ruins or caves for autonomous space mapping. Navigation types for these platforms are different for indoor or outdoor spaces. Navigation based on Global Navigation Satellite System (GNSS) with an example such as GPS, GLONASS and in future Galileo, as precision navigation system are mainly used. The benefit of using GNSS navigation systems is the achieved accuracy, typically better than 15 meters. When the augmentation system is used, the accuracy is typically better than 3 meters in position determination. Such an augmentation system is called European Geostationary Navigation Overlay Service (EGNOS) and in the US Wide Area Augmentation Service (WAAS). When the GPS signal phase measurement is used, the accuracy is at the centimetre level (Mcnamara, 2004), (Jonas & Mandlik, 2013). Satellite navigation is unsuitable for our purpose, because the signal has a lot of drop outs in buildings, tunnels, or mines and we are focused on indoor navigation.

It is possible to use a navigation system based on Inertial Navigation System (INS) navigation in an indoor space. This type of navigation system is based on the fusion of two different types

of sensors—accelerometers and gyroscopes. Non-stationary offset and a high noise level is typical for cheap Inertial Measurement Units (IMU). These properties lead to inaccurate position estimation in an integration process. For estimated path correction, it is necessary to use some correction system. In the paper (Beran et al, 2014) a correction system using a geometrical transformation based on Direct Cosine Matrix (DCM) is shown. Even a small offset error leads to a big mistake in position determination which grows quickly with time. This problem was described in detail in paper (Beran et al, 2014). This type of navigation can be used for navigation of a robotics platform.

A modern method of navigation called Visual Odometry using advanced image processing can be used for navigation of a mobile platform (Yang et al, 2005), (Campbell et al, 2005). The visual odometry method is known from the eighties, but massive penetration of this method is connected with development of fast Digital Signal Processors (DSP). In our research, we are focused on the application of these modern advanced image processing methods. In our case, the Scale Invariant Feature Transform (SIFT) method (Lowe, D.G., 1999) and Speeded-Up Robust Features (Rodriguez & Aouf, 2010), (Bay et al, 2008), (Nemra et al, 2013), can be used. Both modern methods are based on extraction of pairs of interesting points from image and template. It is possible to calculate the pixel distances from differences between these pairs of

interesting points. The SURF method is optimized and faster than the SIFT method. We used a result from the paper (Suaib et al, 2014) and we focus on the SURF method in our research.

## 2 ADVANCED IMAGE PROCESSING METHOD

### 2.1 Speed Up Robust Features (SURF)

The SURF method is based on extraction of interesting points from an image. This method is also used for description of these points. The big advantage in comparison with the SIFT method is speed (Suaib et al. 2014). Higher speed processing is achieved by using the integral convolution method and approximation of the Gaussian second order derivative. This approximation allows speed-up. The equation of the Gaussian function is shown in (1).

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (1)$$

where  $x$  and  $y$  are pixel coordinates in the image,  $\sigma$  is a parameter of Gaussian function.

In Figure 1 a discrete and approximate second-order derivation of the Gaussian function is shown. The second-order partial derivation of Gaussian function derive in the  $y$ -direction is shown on left side; the Gaussian function derive in the  $xy$ -direction is shown on right side.

Every interesting point in integral image  $I_{\Sigma(x)}$  at position  $x = (x, y)$  is defined as a sum the all

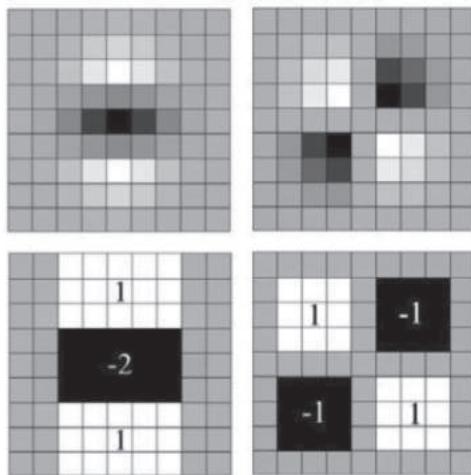


Figure 1. Discrete (up) and approximation (down) of second order derivation of Gaussian Function (Bay et al, 2008).

pixels of input image  $I(i, j)$ . This sum is defined as a square between point  $x$  and its origin. This integral image can be calculated using (2).

$$I_{\Sigma}(x) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j) \quad (2)$$

Interesting points in the image are extracted by using a Hessian Matrix. This matrix is created like a second-order partial derivation of function. This matrix is calculated using equation (3).

$$H(x, y, \sigma) = \begin{bmatrix} L_{xx}(x, y, \sigma) & L_{xy}(x, y, \sigma) \\ L_{xy}(x, y, \sigma) & L_{yy}(x, y, \sigma) \end{bmatrix} \quad (3)$$

where  $x$  is coordinates of pixel;  $\sigma$  is pixel scale;  $L_{xx}(x, y, \sigma)$  is convolution of the second-order derivation of the Gaussian function with original image  $I$  in pixel  $x$

$$\text{Det}[H(x, y, \sigma)] = L_{xx}(x, y, \sigma) \cdot L_{yy}(x, y, \sigma) - L_{xy}(x, y, \sigma)^2 \quad (4)$$

where  $\text{Det}[H(x, y, \sigma)]$  is a determinant of the Hessian matrix.

The descriptors must be independent in rotation. This rule is provided by setting the orientation in all important points. The decision about orientation is provided by using a sum of Haar Wavelet responses. The response is written into a histogram with six bin length (1 bin = 60°). The orientation is chosen by the maximal value in this histogram. It is possible to write the descriptor of each interesting point using (5).

$$z = (\sum dx, \sum dy, \sum |dx|, \sum |dy|) \quad (5)$$

where  $z$  is the important point;  $dx$  and  $dy$  are sub-regions.

## 3 EXPERIMENTAL RESULT

In our research we construct two workplaces. The first workplace has a static camera perpendicular to the ground. This camera has a static scale and we are moving the background. The second workplace was constructed as a user-controlled movable platform. On both platforms a SURF algorithm was implemented.

### 3.1 Static platform

The static platform consists of a stationary construction with fixed camera and a movable background. During our testing we mounted a camera on the print head of 3D printer. The 3D

printer has high precision of position estimation. The maximal error of position estimation is 0.25 mm. This value is sufficient for our purpose.

The SURF algorithm is used for template tracking. The template is created from the first captured image by cutting out the centre of image. This template is tracked in the field of view. The trace of tracking and results are shown in Figure 2. The black colour is used for a tracked trace and the white square with white dot in the centre, is the cropped template from the first image.

The more information about static platform and results of measurement is presented in (Beran et al, 2015).

### 3.2 Movable platform

In our research we construct a mobile robotics platform. At present, the robotics platform is radio-controlled by a user, but in future, we are planning autonomous control of this platform. The movable platform is based on a Raspberry PI. The parameters of this platform are written in Table 1. At present, the Raspberry is on a vehicle designed only for capturing a video. The camera is recording a floor with defined angle. Captured images are stored into an SD-Card for post-processing. Post-processing is provided by Matlab Software. The template is continuously cropped from the center of the previous image.

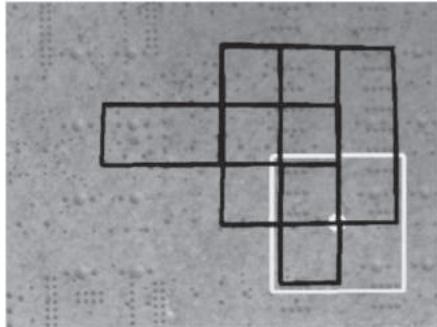


Figure 2. Tracked template using SURF algorithm.

Table 1. Parameters of movable platform.

Type	Parameters
Raspberry PI 2	Broadcom BCM2836 ARM Cortex A7, 4 × 900 MHz, 1 GB LPDDR2 SDRAM
Camera	OV05647-G04A, Pixel size 1.4 × 1.4 $\mu\text{m}$ , Frame rate: 1080p (30 fps), 960p (45 fps), 720p (60 fps), 640 × 480 (90 fps), 320 × 240 (120 fps), lens chiev array (24°).

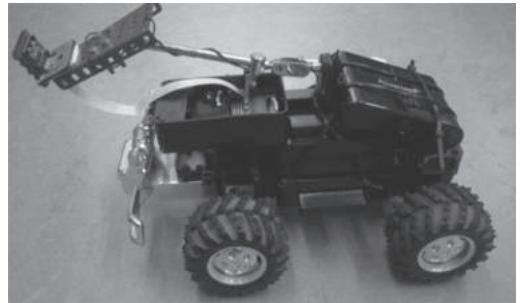


Figure 3. Radio-controlled platform for testing visual odometry.

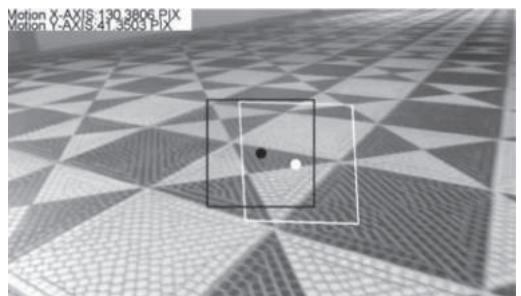


Figure 4. Demonstration of detection left curve using visual odometry.

In Figure 3, our radio-controlled robotic platform is shown. The algorithm sequentially detects a position. The template is cropped from the centre of the previous image. The algorithm calculates interesting points in the template and also in the image, where many pairs of these interesting points are found. The pixel position is calculated, from distance between centre of previous image and searched centre, by algorithm.

In Figure 4, the detected trace using the SURF algorithm is shown. The white square with white point in the centre is detected position of the template in the next image. The black square with black point is the original position of the template. The pixel distance is calculated as a difference between these points.

## 4 MAPPING CAMERA COORDINATES

The geometrical transformation is a transformation from the camera coordinates to real coordinates. This transformation is based on simply mathematics equations. First, it is necessary to measure a distance between ground and camera. Also it is necessary to know a camera's vertical

field of view. This vertical field is calculated from trigonometric equation.

In Figure 5, our camera physical settings are depicted. At the beginning it is possible measure these variables:  $H$ ,  $\alpha$  and  $L$ .

When a height  $H$  of camera was measured, we can measure camera view angle  $\alpha$

$$\alpha = \tan^{-1}\left(\frac{H}{L}\right) \quad (6)$$

where  $\alpha$  is an angle between the principal ray and ground;  $H$  is a distance between the camera and ground;  $L$  is a distance between camera and principal ray of camera mapped to the ground.

Also it is possible calculate an angle  $\beta$  using angle  $VFOV$  (Vertical Field of View). The relationship between focal length  $f$  and field of view is shown in Figure 6.

$$\tan(\beta) = \frac{Pv - (V/2)}{f} \quad (7)$$

where  $\beta$  is an angle between captured point  $P$  at CCD and the principal ray;  $Pv$  is vertical position of point  $P$  at CCD;  $V$  is vertical resolution of chip (in our case 240 rows);  $f$  is focal length.

When the Horizontal Field of View ( $HFOV$ ) is not known, this angle can be calculated using (8).

$$HFOV = 2 \cdot \tan^{-1}\left(\tan\left(\frac{VFOV}{2}\right) \cdot \frac{\text{Width}}{\text{Height}}\right) \quad (8)$$

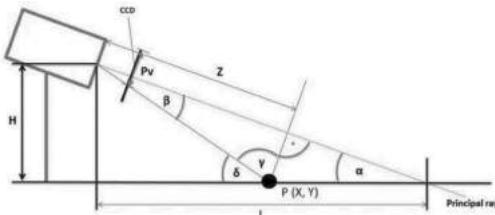


Figure 5. Mapping a camera coordinates to ground.

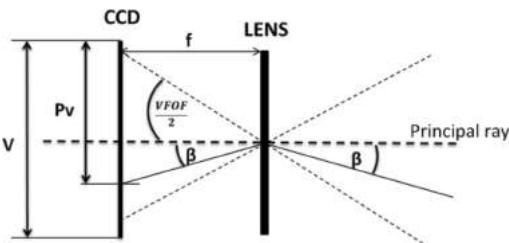


Figure 6. Relationship between angle  $\beta$  and focal length  $f$ .

where width and height is resolution of camera

$$f = \frac{V}{2} \cdot \frac{1}{\tan(VFOV/2)} \quad (9)$$

$$\tan(\beta) = \frac{(Pv - (V/2)) \cdot 2 \cdot \tan(VFOV/2)}{V} \quad (10)$$

When focal length is not known, angle  $\beta$  can be established using  $VFOV$ . These relationships are written in equations (7), (9) and (10).

$$\delta = \alpha + \beta \quad (11)$$

$$y = \frac{H}{\tan \delta} \quad (12)$$

$$z = \frac{H \cdot \cos(\beta)}{\sin(\delta)} \quad (13)$$

where  $y$  is coordinates of point  $P$ ;  $z$  is depth of observed point.

## 5 RESULTS

In our research we are testing two platforms. The first platform is statically constructed and the results from this platform are shown in (Beran et al, 2015). The second platform is movable. In this platform the scene is captured by the camera with the defined angle. All parameters of configuration of our system are written in Table 2.

In Table 3 and Table 4, the sample of measurements are written. All measurements have fifteen repetitions of the same path. For better readability, the first five results of these measurements are written. In this research a five different path lengths were tested and results are written in Table 5.

In Figure 7, the corrected estimated path using geometrical transformation is shown. In the image, the black square is a template and the white square

Table 2. Configuration of movable platform.

Parameter	Value
$VFOV$	60°
$HFOV$	35, 98°
$H$	17.5 cm
$L$	44 cm
$\alpha$	21.68°
Width × height	320 × 240 pix

Table 3. Results of estimated path—length 3 m.

Direct drive—path length 3 m						
Measure nr.	1	2	3	4	5	Std. dev.
X [m]	0,05	-0,07	0,05	0,02	0,08	0,0328 0,052
Y [m]	3,05	2,97	3,05	3,09	3,11	3,05 0,114

Table 4. Results of estimated path—length 6 m.

Direct drive—path length 6 m						
Measure nr.	1	2	3	4	5	Std. dev.
X [m]	0,02	-0,05	-0,03	0,1	0,04	0,026 0,099
Y [m]	6,05	5,96	6,15	6,13	6,09	6,07 0,157

Table 5. Statistical results of estimated path.

Path length	1.5 m		3 m		5 m	
	X	Y	X	Y	X	Y
Mean	0,052	1,52	0,0328	3,05	0,13	5,1
Std.dev.	0,0056	0,09	0,052	0,114	0,088	0,14

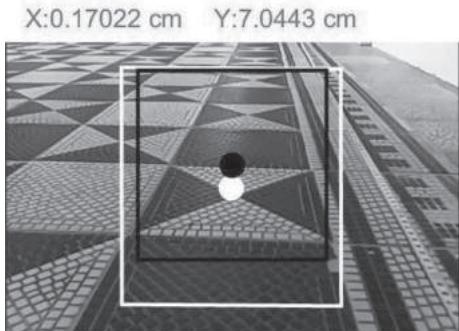


Figure 7. Demonstration of estimated corrected path.

is a detected previous position of the template. At the top of this image, estimated differences between black and white centres are shown in centimetres. At first, this distance is calculated in pixel distance and later, it is transformed to real coordinates.

## 6 CONCLUSION

In this paper a modern navigation method based on an advanced image processing method called

SURF was presented. Also an implementation of this method on static and movable robotics platforms is described. The first platform works with a static camera and movable background. The second platform is user-controlled platform. We are planning to construct an autonomous robotics platform in the future. Our platform is based on a camera and Raspberry PI. The scene is captured into an internal memory, and all calculations are evaluated as post-processing in Matlab software. The algorithm searches interesting points of the template in an image. For this purpose it is necessary to have a sharp image using a camera with fast shutter. The pixel distance is calculated from pairs of these interesting points. Every fifth to fifteenth image is selected in Matlab to calculate interesting points. The frame step is selected by the real speed of a movable platform, at present, manually. In this case we are using a high camera framerate (120 fps) and image decimation which will decrease time consumption. The algorithm also provides geometrical transformations which, after recalculation, transform this pixel distance to real distance in centimeters. From the results, it is possible to recognize a higher estimated value than real distance. In following research this higher value must be corrected. At this time, the algorithm calculates with direct drive only. But in future the some correction system will be implemented in all types of paths. This correction will be similar to the Direct Cosine Matrix transformation as described in paper (Beran et al, 2014). In this paper geometrical transformation for direct paths was described. In the near future, a correction system to get a corrected estimated path will be implemented. For this purpose is necessary to use a fusion of visual odometry with high precision gyroscopes. This system, based on visual odometry, will be implemented as a fusion of Inertial Measurement Unit (IMU) to Autonomous Research Exploration System (ARES). Also, it will cooperate with a system called Optical Measurement Device (Chmelar & Dobrovolny, 2013), (Chmelar et al, 2014).

## ACKNOWLEDGMENT

The research was supported by the Internal Grant Agency of University of Pardubice, the project No. SGSFEI 2015003.

## REFERENCES

- Bay, H. et al, 2008, Speeded-Up Robust Features (SURF), In: LEONARDIS, Computer Vision and Image Understanding: 9th European Conference on Computer Vision, 2008, s. 346–359.

- Beran, L. & Chmelar, P. & Dobrovolny, M., 2014, Navigation of robotic platform with using inertial measurement unit and Direct Cosine Matrix, Proceedings Elmar—International Symposium Electronics in Marine 2014, pp. 87–90.
- Beran, L. & Chmelar, P. & Rejsek L., 2015, Navigation of Robotics Platform using Monocular Visual Odometry, Radioelektronika (RADIOELEKTRONIKA), 2015 25th International Conference.
- Campbell, J. et al, 2005, A Robust Visual Odometry and Precipice Detection System Using Consumer-grade Monocular Vision, Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005, vol., no., pp. 3421, 3427.
- Chmelar, P. & Dobrovolny, M., 2013 The Fusion of Ultrasonic and Optical Measurement Devices for Autonomous Mapping, Radioelektronika (RADIOELEKTRONIKA), 2013 23rd International Conference, pp. 292,296.
- Chmelar, P. & Beran, L., & Kudriavtseva, N., 2014, Projection of point cloud for basic object detection, ELMAR (ELMAR), 2014 56th International Symposium
- Jonas, M. & Mandlik, M., 2013 “Influence of the GNSS satellite elevation to the position error,” Radioelektronika 2013 23rd International Conference, pp. 119, 124.
- Lowe, D.G., 1999, Object recognition from local scale-invariant features, Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference, vol.2, no., pp. 1150, 1157 vol. 2.
- Mcnamara, J., 2004, GPS for dummies. Hoboken, NJ: Wiley, p. 57.
- Nemra, A. et al, 2013, Adaptive Iterative Closest SURF for visual scan matching, application to Visual odometry,” Systems and Control (ICSC), 2013 3rd International Conference, pp. 927.
- Rodriguez, D. & Aouf, N., 2010, Robust Harris-SURF features for robotic vision based navigation,” Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference, pp. 1160.
- Suaib, N.M. et al, 2014, Performance evaluation of feature detection and feature matching for stereo visual odometry using SIFT and SURF,” Region 10 Symposium, 2014 IEEE, vol., no., pp. 200, 203.
- Yang, C.H. & Maimone, M. & Matthies, L., 2005 “Visual odometry on the Mars Exploration Rovers,” Systems, Man and Cybernetics, vol. 1, no., pp. 903.

## SNIP: Smile—neutral facial display intensity predictor

K. Nurzynska

*Institute of Informatics, Silesian University of Technology, Poland*

B. Smolka

*Institute of Automatic Control, Silesian University of Technology, Poland*

**ABSTRACT:** Many affective disorders can influence the capacity of emotion expression and understanding. For instance, in the case of depression, the frequency and intensity of smiling is diminishing when the disease proceeds. Analysing these changes by a psychologist could bring information about the effectiveness of the prescribed medical treatment. Therefore, this work describes a system which automatically analyses video recordings from sessions with patients and determines the length and intensity of smile.

The proposed system exploits a standard database to train the classifier, which is later used to distinguish between the smiling and neutral facial expression depicted in the video frames. The obtained information is presented graphically, indicating the class membership and temporal smile intensity. The performed experiments revealed that the classification accuracy exceeds 85%. Moreover, the case studies show, that the variations in smile magnitude enable better insight into the smile analysis.

### 1 INTRODUCTION

Patients suffering from some mood disorders have problems to correctly recognize or express emotions. When schizophrenia is considered, the accurate understanding of others emotions is troublesome (Kohler et al. 2008, Souto et al. 2013). The ability of autistic persons to recognize the genuine and posed smile is impaired, because they usually tend to omit the eye contact (Boraston et al. 2008). Next, the research presented in (Scherer et al. 2013, Girard et al. 2014) show that in the case of depression, the changes in smile intensity and its frequency correlate with the patient's mental state and can be used as a measure of effectiveness of the therapy. For comparison, the measurements of cardiovascular changes, e.g. cardiac interbeat interval, finger pulse amplitude, etc., bring no effect (Tsai et al. 2003).

Recording sessions with patients to analyse the behavioural changes were commonly used in research described by (Reed et al. 2007, Girard et al. 2014, Hurley and Frank 2011). The gathered material proved to be very useful in diagnosis process and mental state evaluation of the patient. The biggest problem is the necessity to annotate each frame in the video, which is a tiresome effort. Moreover, this work should be performed by the trained experts with expertise in the Facial Action Coding System (FACS), which was developed by (Ekman and Friesen 1978) and became very

popular in the medicine and psychology. Hence, no matter how useful the findings are, it is impossible to follow the proposed methodology in everyday use. Therefore, a system, which automatically annotates the frames whether they refer to smiling or neutral face with additional information about smile intensity is of high practical importance.

The paper addresses the problem of automatic classification between smiling and neutral facial displays with additional estimation of the smile intensity. The SNIP system uses pre-trained linear Support Vector Machine (SVM) to decide to which class each frame from the video can be assigned,

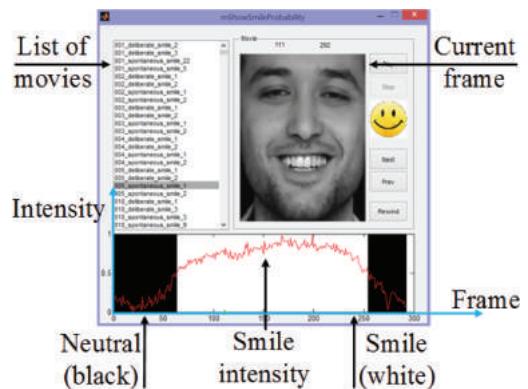


Figure 1. Main window of the SNIP system.

then basing on the response from classifier the smile intensity is determined. The main software window graphically presents all the relevant information in Figure 1.

The paper is structured as follows. Section 2 discussed previous work concerning the recognition between smiling and neutral facial displays. Next in Section 3, the Local Binary Patterns (LBP) is presented as a tool to describe the facial expression. The system work flow is given in Section 4 and results of the performed experiments are presented in Section 5. Finally, the conclusions are drawn in Section 6.

## 2 STATE OF THE ART

The emotional conditions depicted in human face are changing dynamically, therefore the first attempt to the smiling and neutral facial display recognition were based on the facial muscles movements recorded in the image sequence and described by the optical flow (Mase 1990, Bartlett et al. 1999, Essa and Pentland 1997). Many approaches use texture operators for frame description in order to discriminate between basic emotions (for instance sadness, anger, fear, happiness, etc.) (Cohn et al. 1999, Cohn et al. 1998, Pantic and Rothkrantz 2000, Shan et al. 2009). But there are also works which exploit extreme learning machine as a classifier (An et al. 2014) or Gabor filters and GentleSVM (Wu et al. 2011).

Different approaches do not concentrate on the whole facial display, but use only some part of it. The active shape model was applied to follow the movement of the lips in the image sequence (Lanitis et al. 1997). Similarly, the research of (Freire et al. 2009, Freire et al. 2009) concentrates on the lips describing the changes by texture operators, while (Dibeklioğlu et al. 2010) examines the eyes.

The attempt to estimate the smile intensity was made previously by (Shimada et al. 2010) and was also mentioned in the review of smile detection (Whitehill et al. 2009). Although, the presented solution is more complex when compared to our approach, the results enable to distinguish the moment of transition between emotions, yet smile intensities measured during smile period are almost constant. In contrast, the SNIP system estimates the smile intensity more accurately, which enables to characterize the character and variation of smile.

## 3 LOCAL BINARY PATTERNS

Local Binary Patterns (LBP) (Ojala et al. 2001, Ojala et al. 1996) is a type of textural feature

which summarises an image content in a form of a histogram. A histogram bin contains the number of similar texture patches in the image, whereas the texture characteristics itself are used to define which bin it belongs to.

In order to calculate the descriptive value, a monochromatic image, denoted as  $I$ , is considered. For each pixel with intensity  $g_c = I(x, y)$  a circular neighbourhood is sampled in  $P$  points on a radius  $R$  following the equation:

$$x_p = x + R \cos(2\pi p/P), y_p = y - R \sin(2\pi p/P), \quad (1)$$

where  $p = 0, \dots, P-1$ , resulting in an ordered sequence of grey-scale values  $g_p$ . More information may be derived from the differences between the sampled points  $g_p$  and the central value  $g_c$ , however that would result in a big range of necessary codes to describe the textural information. Therefore, only the sign is exploited in the definition of the generic LBP:

$$LBP_{P,R}(x, y) = \sum_{p=0}^{P-1} s(g_p - g_c) \cdot 2^p, \quad (2)$$

where

$$s(z) = \begin{cases} 1, & z \geq 0, \\ 0, & z < 0. \end{cases} \quad (3)$$

This operator proved to be very descriptive and found many applications in image understanding domain (Brahnam et al. 2014), yet in some cases the large length of the generated histogram is a severe drawback of this method (for  $P = 8$  the LBP histogram length is 256) as it influences the computation costs remarkably. Moreover, it was noticed (Pietikäinen et al. 2011) that some patterns comprise more information than others. The distinction is based on the number of transition between 0s and 1s when treating the LBP code as a sequence of bits. The discerning codes, called ‘uniform’ patterns (ULBP), have up to two such transitions, e.g. 00110000, 01111100, and are stored in separate histogram bins. All other codes share a common bin. This upgrade shortens the histogram length to  $P(P-1) + 3$ .

## 4 SNIP SYSTEM SET-UP

The SNIP determines between the smiling and neutral facial displays in a sequence of images. The classifier is trained once, using the standard facial display database and then applied for each face detected in the video. The training and testing processes both require to perform following steps:

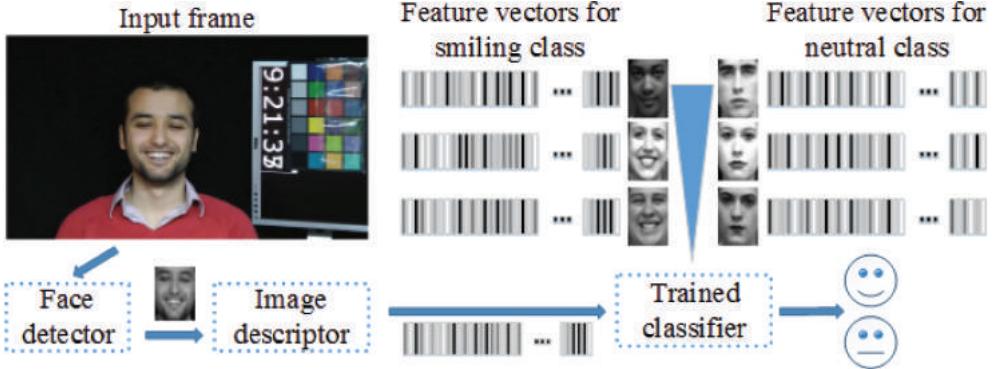


Figure 2. Data flow in the SNIP system. On the right, the feature vectors from the standard *Cohn-Kanade* database used to train the classifier are presented. On the left, is the input frame from the *UVA-NEMO* database. The arrows show successive processing steps.

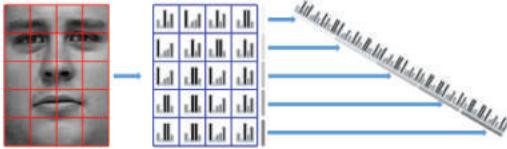


Figure 3. Descriptor calculation. From the left: original image, with marked sub-image grid; ULBP histograms retrieved from each sub-image; ULBP concatenated histogram obtained from the sub-images.

- Convert image/frame to a grey-scale.
- Determine a region of the face in the image/frame using the face detector described in (Kawulok and Szymanek 2012).
- Resize the area to the standard resolution of  $112 \times 150$  pixels.
- Calculate the descriptor for facial display.

The data flow in the SNIP system is depicted in Figure 2.

#### 4.1 Descriptor calculation

A lot of work concerning the smile description was conducted by (Ekman and Friesen 1978), who concluded that the movement of two muscles is sufficient to generate a Duchenne smile (the only one of smile types assumed to reflect pleasure) on the human face. These are the orbicularis oculi, which raises the cheek and gathers the skin inwards from around the eye socket and the zygomatic major which pulls the lip corners upwards towards the cheekbones. Since the movement is present only in several parts of the face, constructing a descriptor based on the whole image would not be satisfactory, as the crucial information gets blended by the regions where no change is noticeable. Therefore, the

presented approach assumes a division of the image into several sub-images as it is depicted in Figure 3. According to previous research (Nurzynska and Smolka 2014) the best results are achieved when the original image is divided into 7 columns and 10 rows and the final descriptor is a concatenation of ULBP histograms obtained for each sub-image.

#### 4.2 Training database

Since it is possible to determine single muscle activation during expression of each emotion, (Ekman and Friesen 1978) defined Action Units (AU), which correspond to the movement of the smallest possible set of muscles activated simultaneously. FACS defines which set of AUs is necessary to present a particular facial gesture. As mentioned before, this approach is commonly used to annotate the experimental data, therefore a natural consequence is its application for training purposes. The database prepared according to these demands is the *Cohn-Kanade AU-Coded Facial Expression Database* (Cohn et al. 1999). It contains image sequences from neutral to target display of basic emotions like joy, surprise, anger, fear, disgust, and happiness. The image resolution is  $640 \times 480$  pixels with 8-bit precision for grey-scale values. For training, frames from 82 sequences were chosen which depict the neutral and smiling expression. In the paper we refer to this data as *Cohn-Kanade* database.

#### 4.3 Support vector machine

Support Vector Machine (SVM) method was applied for the recognition between the smiling and neutral facial displays. Using the training data, a model for two-class problem solution is created. The exemplary objects are treated as points in the space in

which a dividing gap is looked for. If necessary, the points are transformed by non-linear kernel in order to better discriminate between objects from different categories. However, for considered problem, the linear kernel proved to achieve the best results (Nurzynska and Smolka 2014). The goal is to obtain a gap as broad as possible, while the division plane is placed in the middle. Then in the testing phase, the examples are mapped to the same space. Depending on which side of the cutting plane they are situated, its membership to the category is defined. Additionally, the distance from the division plane defines the strength of belonging to one category. These values are normalized in the range 0–1 to plot the smile intensity variation over the time.

## 5 RESULTS

Verification of the classification accuracy and correlation of intensity estimation by the system with visual data was performed on three datasets. First of them (UVA-NEMO dataset) concentrates on providing movies for smile truthfulness description, therefore only basic test was possible. The second (FEEDB dataset) consists of posed expression and is treated as a reference for results gathered for UVA-NEMO. Whereas the third dataset supplied movies with varying intensity of presented emotion, what enabled better insight into correct smile intensity prediction.

### 5.1 Case I: UVA-NEMO database

The experiments were performed on the *UVA-NEMO* database (Dibeklioğlu et al. 2012) which consists of 1240 videos presenting the spontaneous and deliberate smiles. The data were collected for 400 subjects (56% male and 44% female) with age ranging from 8 to 76. The frame resolution is 1920 × 1080 pixels for color images, recorded at the speed of 50 frames per second. The number of frames in videos vary depending on the facial gesture length. Moreover, it is not assured that each image sequence starts and finishes with the neutral facial display.

In order to verify whether the elaborated linear SVM classifier is accurate for smiling and neutral facial display recognition, its classification results were compared to frames manually annotated by a person without prior experience in emotion description. For the experiment 100 videos were selected, their length varied from 91 to 540 frames. For total 21,022 frames the face was correctly detected in 20,642 images, where 16,249 correspond to the smiling facial display and 4,394 describe the neutral emotion. In general, the system accuracy achieved 91.41% of correct emotion classification. Analysis of the results showed that

the highest error ratio was noticed for the change between the emotional states (96.56% of the wrong classification) where frames labelled with neutral facial display were recognized as smiling expression. The annotating person reported that in these parts of the videos it was difficult to clearly decide which emotion is presented, therefore it might be assumed that some wrong annotations took place. To conclude, the system enables satisfactory classification between the emotions, the only problem is the estimation of point in time where the change between the neutral and smiling facial display occurs, as it might be shifted in some videos when compared to others.

Figure 4 shows the results for a video presenting the change between neutral, smiling, and once again neutral expressions for spontaneously smiling person. In the middle part of the image, the graph presents the smile intensity indicated by a red line and category annotation resulting from the classifier exhibited by black (neutral) and white (smile) colors. The top row shows every second frame recorded when the transition between neutral and smiling expression took place, whereas the bottom row collects images describing the change from the smiling into the neutral facial display. Additionally, frames for the most neutral expression (frame No. 19) and the highest intensity of smile (frame No. 199) are presented. This case study confirms the accuracy in classification between the emotional states.

### 5.2 Case II: FEEDB—multimodal database of facial expressions and emotions

The subset of 62 movies presenting happy and laugh emotion from Multimodal database of Facial Expressions and Emotion (*FEEDB*) (Szwoch 2013,

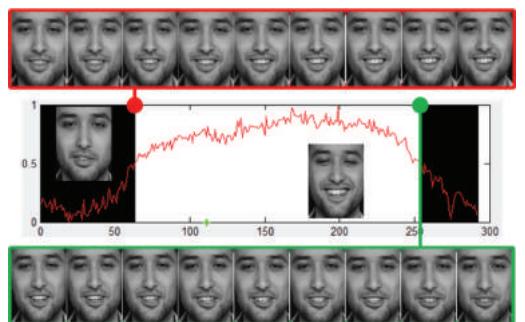


Figure 4. Video presenting transition between neutral—smiling—neutral expressions. The middle part visualizes the change of the smile intensity over the time (red line), with additional annotation for neutral (black background) and smiling expressions (white). Top and bottom rows present every second frame for the moment of shift between emotions.

Szwoch 2014) was exploited for verification of the SNIP system performance. The database consists of movies representing 30 different emotions per subject presented by the undergraduates and graduates students of both genders. The recordings contain not only the color visual information stored with  $640 \times 480$  pixel resolution frames gathered with 25–30 frame per second speed but also the depth map. Yet, in presented research only the visual information is exploited. The length of the considered movies varied from 11 to 220 frames with average movie length of 120 frames (6560 frame count in total).

The aim of this experiment was to repeat the procedure described in case study I using different dataset. Hence, similarly the results of automatic classification were compared with the manual annotations to give the overview of SNIP system accuracy. For these data the correct classification ratio was 89.76% and as in previous example most errors occurred for frames depicting transition from neutral to the smiling state and vice versa. This high performance confirms the ability of emotional state recognition by the presented system.

### 5.3 Case III: DaFEx database

In this experiment the *DaFEx* database (Battocchi et al. 2005, Battocchi et al. 2005) was used. It consists of recordings prepared by 8 actors (4 women, and 4 men) performing 6 Ekman's emotions (anger, disgust, fear, happiness, sadness, and surprise) along with a neutral facial display recording. Each emotion was acquired in 6 separate sessions, where two of them presented the emotion only, while in four additional sentence was uttered by the subject. Moreover, each session consisted of three recordings with varying emotion strength: low, medium, and high. To perform the experiment, sequences depicting only happiness were exploited, that gave 144 movies in total (35870 frames). The frame resolution is  $360 \times 288$  pixels for color images. The number of frames in videos vary from 147 to 532 depending on the facial gesture length.

Similarly as in previous case studies, the result of classification of the frames into smiling and neutral groups is compared with manual annotations performed by the person without prior experience. Table 1 presents achieved correct classification accuracy ratio considering some characteristics of the dataset. First of all, the general smiling versus neutral facial display classification efficiency is given in the top row. The score 85% is satisfactory, however it is lower than the result in the previous cases. There are several reasons which could deteriorate the classification accuracy: Firstly, the low image resolution. Secondly, introduced utterance in the sequences. Finally, unclear transition

between emotional states during speaking mostly occurring in case of presentation of emotion with low intensity. Rows 2nd and 3rd in Table 1 show the influence of incorporating speech into the emotion recognition. One can see that the data with uttered sentences caused more errors, yet the difference is not significant. Finally, last three rows summarize the influence of smile intensity on the emotion recognition. As it was supposed the most difficult case is when the emotion display has low intensity (83%), whereas for medium and high emotion expression 86% and 87% were recorded respectively. Yet, it should be pointed out, than in many cases most frames of the video presenting the low intensity of happiness was assigned to neutral class by the system and annotating person. Another interesting phenomenon, which was noticed due to the possibility of smile intensity visualization was

Table 1. Correct emotion classification into smiling and neutral one for movies depicting happiness in *DaFEx* database.

Description	Efficiency [%]	No. of movies
All data	85.48	144
No utterance	86.44	48
Utterance	84.94	96
High	87.17	48
Medium	86.00	48
Low	83.08	48

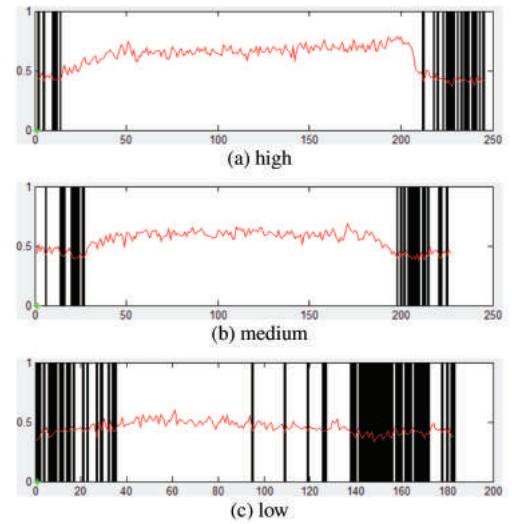


Figure 5. Examples of smile intensity estimation (red line) for time series of various smile intensity presented without utterances by subject No. 5 in series No. 3. Black background corresponds to neutral expression, while white to smiling one.

the subjectivity of the emotion intensity presentation. For one actor the highest smile intensity is on the level of medium or even low level of the other subject. On the other hand, when comparing the same person in similar circumstances, the assumed change in smile intensity is easily visible on the plots generated by SNIP system as is presented in Figure 5.

## 6 CONCLUSIONS

This work presents a system for automatic classification between smiling and neutral facial display with estimation of the smile intensity. The system needs a standard database for classifier training before its application to automatically annotate each frame of the input video. As a result, the user receives information about the smile intensity changes within the whole video with additional classification of each frame for those presenting smile or not.

The performed research revealed that the correct classification ratio between neutral and smiling facial display is above 85% for *DaFEx*, 89% for *FEEDB*, and 91% for *UVA-NEMO* databases. However, in the last case when excluding from the calculation situation of transition between the emotion display it reaches almost 99%. The case studies show, that it is possible to derive the smile intensity information from the SVM classifier. The performed experiments confirmed that not only binary distinction between smiling and neutral facial display is possible but also the graduation in smile intensity can be expressed in this manner.

The authors expect, that the presented functionality of the system will find an application in medical treatment of patients with mood disorders and facilitate the work of psychologists and psychiatrists.

## ACKNOWLEDGEMENTS

This work has been supported by the Polish National Science Centre (NCN) under the Grant: DEC-2012/07/B/ST6/01227 and was performed using the infrastructure supported by POIG.02.03.01-24-099/13 grant: GeCONiI—Upper-Silesian Center for Scientific Computation.

## REFERENCES

- An, L., S. Yang, & B. Bhanu (2014). Efficient smile detection by extreme learning machine. *Neurocomputing* 149, 354–363.
- Bartlett, M., J. Hager, P. Ekman, & T. Sejnowski (1999). Measuring facial expressions by computer image analysis. *Psychophysiology* 36(2), 253–263.
- Battocchi, A., F. Pianesi, & D. Goren-Bar (2005). A first evaluation study of a database of kinetic facial expressions (DaFEx). In *Proc. 7th Intern. Conf. Multi modal Interfaces*, ICMI '05, New York, NY, USA, pp. 214–221. ACM.
- Battocchi, A., F. Pianesi, & D. Goren-Bar (2005, Oct.). The properties of DaFEx, a database of kinetic facial expressions. In *Proc. 1st Intern. Conf. Affective Computing and Intelligent Interaction*, pp. 558–565.
- Boraston, Z., B. Corden, L. Miles, D. Skuse, & S. Blakemore (2008). Brief report: Perception of genuine and posed smiles by individuals with autism. *Journal of Autism and Developmental Disorders* 38(3), 574–580.
- Brahnam, S., L. Jain, L. Nanni, & A. Lumini (2014). *Local binary patterns: new variants and applications*. Springer.
- Cohn, J. A., Zlochower, J. Lien, & T. Kanade (1999). Automated face analysis by feature point tracking has high concurrent validity with manual FACS coding. *Psychophysiology* 36(2), 35–43.
- Cohn, J.F., A.J. Zlochower, J.J. Lien, & T. Kanade (1998). Feature-point tracking by optical flow discriminates subtle differences in facial expression. In *Proc. of the 3rd. Intern. Conf. on Face & Gesture Recog.*, Washington, DC, USA, pp. 396–401.
- Dibeklioğlu, H., A. Salah, & T. Gevers (2012). Are you really smiling at me? Spontaneous versus posed enjoyment smiles. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid (Eds.), *Computer Vision ECCV 2012*, Volume 7574 of *Lecture Notes in Computer Science*, pp. 525–538. Springer.
- Dibeklioğlu, H., R. Valenti, A.A. Salah, & T. Gevers (2010). Eyes do not lie: Spontaneous versus posed smiles. In *Proc. of the Intern. Conf. on Multimedia*, MM '10, New York, NY, USA, pp. 703–706. ACM.
- Ekman, P. & W. Friesen (1978). Facial action coding system: a technique for the measurement of facial movement. *Consulting Psychologists Press*.
- Essa, I. & A. Pentland (1997, Jul). Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Trans Pattern Anal Mach Intell* 19(7), 757–763.
- Freire, D., L. Anton, & M. Castrillon (2009). A study for the self similarity smile detection. In J.e.a. Fierrez (Ed.), *Biometric ID Management and Multimodal Communication*, Volume 5707 of *Lecture Notes on Computer Science*, pp. 97–104. Springer.
- Freire, D., M. Castrillon, & O. Deniz-Suarez (2009). Smile detection using local binary patterns and support vector machines. In *VISSAPP(1)'09*, pp. 398–401.
- Girard, J., J. Cohn, M. Mahoor, S. Mavadati, Z. Hammal, & D. Rosenwald (2014). Nonverbal social withdrawal in depression: Evidence from manual and automatic analyses. *Image Vision Comput.* 32(10), 641–647.
- Hurley, C. & M. Frank (2011). Executing facial control during deception situations. *Journal of Non verbal Behavior* 35(2), 119–131.
- Kawulok, M. & J. Szymanek (2012, March). Precise multi-level face detector for advanced analysis of facial images. *Image Processing, IET* 6(2), 95–103.
- Kohler, C., E. Martin, N. Stolar, F. Barrett, R. Verma, C. Brensing, W. Bilker, R. Gur, & R. Gur (2008, October). Static posed and evoked facial expressions of emotions in schizophrenia. *Schizophrenia Research* 105(1–3), 49–60.

- Lanitis, A., C. Taylor, & T. Cootes (1997, Jul). Automatic interpretation and coding of face images using flexible models. *IEEE Trans Pattern Anal Mach Intell* 19(7), 743–756.
- Mase, K. (1990). An application of optical flow—extraction of facial expression. In *IAPR Workshop on Machine Vision Applications*, pp. 195–198.
- Nurzynska, K. & B. Smolka (2014). Optimal classification method for smiling vs. neutral facial display recognition. *Journal of Medical Information and Technologies* 23, 87–94.
- Ojala, T., M. Pietikäinen, & D. Harwood (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition* 29(1), 51–59.
- Ojala, T., M. Pietikäinen, & T. Mäenpää (2001). A generalized local binary pattern operator for multire solution gray scale and rotation invariant texture classification. In *Proc. 2nd Intern. Conf. on Advances in Pattern Recognition*, London, UK, pp. 397–406. Springer.
- Pantic, M. & L. Rothkrantz (2000). Automatic analysis of facial expressions: The state of the art. *IEEE Trans Pattern Anal Mach Intell* 22, 1424–1445.
- Pietikäinen, M., G. Zhao, A. Hadid, & T. Ahonen (2011). Computer vision using local binary patterns. In *Computational Imaging and Vision*, Volume 40, pp. 13–9. Springer.
- Reed, L., M. Sayette, & J. Cohn (2007). Impact of depression on response to comedy: A dynamic facial coding analysis. *Journal of Abnormal Psychology* 116(4), 804–809.
- Scherer, S., G. Stratou, M. Mahmoud, J. Boberg, J. Gratch, A. Rizzo, & L.-P. Morency (2013, April). Automatic behavior descriptors for psychological disorder analysis. In *IEEE Conf. on Automatic Face and Gesture Recognition*, Shanghai, China.
- Shan, C., S. Gong, & P. McOwan (2009, May). Facial expression recognition based on local binary patterns: A comprehensive study. *Image Vision Comput.* 27(6), 803–816.
- Shimada, K., T. Matsukawa, Y. Noguchi, & T. Kurita (2010). Appearance-based smile intensity estimation by cascaded support vector machines. In R. Koch and F. Huang (Eds.), *ACCV Workshops (1)*, Volume 6468 of *Lecture Notes in Computer Science*, pp. 277–286. Springer.
- Souto, T., A. Baptista, D. Tavares, C. Queirós, & A. Marques (2013). Facial emotional recognition in schizophrenia: preliminary results of the virtual reality program for facial emotional recognition. *Revista de Psiquiatria Clinica* 40(4), 129–134.
- Szwoch, M. (2013, June). Feedb: A multimodal database of facial expressions and emotions. In *6th Intern. Conf. Human System Interaction*, pp. 524–531.
- Szwoch, M. (2014). On facial expressions and emotions RGB-D database. In S. Kozielski, D. Mrozek, P. Kasprowski, B. Maysiak-Mrozek, and D. Kostrzewa (Eds.), *Beyond Databases, Architectures, and Structures, Volume 424 of Communications in Computer and Information Science*, pp. 384–394. Springer.
- Tsai, J., N. Pole, R. Levenson, & R. Munoz (2003, February). The effects of depression on the emotional responses of Spanish-speaking Latinas. *Cultur Divers Ethnic Minor Psychol.* 9(1), 49–63.
- Whitehill, J., G. Littlewort, I. Fasel, M. Bartlett, & J. Movellan (2009, Nov). Toward practical smile detection. *IEEE Trans Pattern Anal Mach Intell* 31(11), 2106–2111.
- Wu, Q., X. Shen, & X. Fu (2011). The machine knows what you are hiding: An automatic micro-expression recognition system. In S. DMello, A. Graesser, B. Schuller, and J.-C. Martin (Eds.), *Affective Computing and Intelligent Interaction*, Volume 6975 of *Lecture Notes in Computer Science*, pp. 152–162. Springer.

This page intentionally left blank

**Computational Vision and Medical Image Processing.** **VipIMAGE 2015** contains invited lectures and full papers presented at VIPIMAGE 2015 - V ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing (Tenerife, Canary Islands, Spain, 19-21 October, 2015). International contributions from 19 countries provide a comprehensive coverage of the current state-of-the-art in the fields of:

- 3D Vision;
- Computational Bio- Imaging and Visualization;
- Computational Vision;
- Computer Aided Diagnosis, Surgery, Therapy and Treatment;
- Data Interpolation, Registration, Acquisition and Compression;
- Industrial Inspection;
- Image Enhancement;
- Image Processing and Analysis;
- Image Segmentation;
- Medical Imaging;
- Medical Rehabilitation;
- Physics of Medical Imaging;
- Shape Reconstruction;
- Signal Processing;
- Simulation and Modelling;
- Software Development for Image Processing and Analysis;
- Telemedicine Systems and their Applications;
- Tracking and Analysis of Movement and Deformation;
- Virtual Reality.

**Computational Vision and Medical Image Processing.** **VipIMAGE 2015** will be useful to academics, researchers and professionals in Biomechanics, Biomedical Engineering, Computational Vision (image processing and analysis), Computer Sciences, Computational Mechanics, Signal Processing, Medicine and Rehabilitation.

Sponsored by



Fundação para a Ciência e a Tecnologia  
MINISTÉRIO DA CIÉNCIA, TECNOLOGIA E INNSINO SUPERIOR



FEUP FACULDADE DE ENGENHARIA  
UNIVERSIDADE DO PORTO



CRC Press  
Taylor & Francis Group  
an informa business  
[www.crcpress.com](http://www.crcpress.com)

6000 Broken Sound Parkway, NW  
Suite 300, Boca Raton, FL 33487  
Schipholweg 107C  
2316 XC Leiden, NL  
2 Park Square, Milton Park  
Abingdon, Oxon OX1 4 4RN, UK

ISBN 978-1-138-02926-2



9 781138 029262  
an informa business