TextonBoost : Joint Appearance, Shape
and Context Modeling for Multi-class Object
Recognition and Segmentation

Jamie Shotton2,J o h nW i n n1, Carsten Rother1, and Antonio Criminisi1
1Microsoft Research Ltd., Cambridge, UK
{jwinn, carrot, antcrim }@microsoft.com
2Department of Engineering,
University of Cambridge
jdjs2@cam.ac.uk

Abstract. This paper proposes a new approach to learning a discriminative model of object classes, incorporating appearance, shape and context information e■ciently. The learned model is used for automaticvisual recognition and semantic segmentation of photographs. Our discriminative model exploits novel features, based on textons, which jointly model shape and texture. Unary classi■cation and feature selection is achieved using shared boosting to give an e■cient classi■er which can be applied to a large number of classes. Accurate image segmentation is achieved by incorporating these classi■ers in a conditional random ■eld.E■cient training of the model on very large datasets is achieved by exploiting both random feature selection and piecewise training methods. High classi■cation and segmentation accuracy are demonstrated on three di■erent databases: i) our own 21-object class database of photographs of real objects viewed under general lighting conditions, poses and viewpoints, ii) the 7-class Corel subset and iii) the 7-class Sowerby database used in [1]. The proposed algorithm gives competitive results both for highly textured (e.g. grass, trees), highly structured (e.g. cars, faces, bikes, aeroplanes) and articulated objects (e.g. body, cow).

1

****

Weakly Supervised Learning of Part-Based
Spatial Models for Visual Object Recognition

David J. Crandall and Daniel P. Huttenlocher
Cornell University, Ithaca, NY 14850, USA
{crandall, dph }@cs.cornell.edu

Abstract. In this paper we investigate a new method of learning part-based models for visual object recognition, from training data that onlyprovides information about class membership (and not object locationor con■guration). This method learns both a model of local part appearance and a model of the spatial relations between those parts. In contrast, other work using such a weakly supervised learning paradigmhas not considered the problem of simultaneously learning appearance and spatial models. Some of these methods use a “bag” model where only part appearance is considered whereas other methods learn spatial models but only given the output of a particular feature detector. Previous techniques for learning both part appearance and spatial relations have instead used a highly supervised learning process that providessubstantial information about object part location. We show that our weakly supervised technique produces better results than these previous highly supervised methods. Moreover, we investigate the degree to whichboth richer spatial models and richer appearance models are helpful in improving recognition performance. Our results show that while both spatial and appearance information can be useful, the e■ect on performance depends substantially on the particular object class and on the di■culty of the test dataset.

1

****

Hyperfeatures – Multilevel Local Coding
for Visual Recognition

Ankur Agarwal and Bill Triggs
GRA VIR-INRIA-CNRS, 655 Avenue de l’Europe, Montbonnot 38330, France

{Ankur.Agarwal, Bill.Triggs }@inrialpes.fr
http://www.inrialpes.fr/lear/people/ {agarwal, triggs }
Abstract. Histograms of local appearance descrip tors are a popular representa-
tion for visual recognition. Th ey are highly discrimin ant and have good resist
ance
to local occlusions and to geometric and photometric variations, but they are no
t
able to exploit spatial co-occurrence statistics at scales larger than their loc
al input
patches. We present a new multilevel visual representation, 'hyperfeatures', tha
t
is designed to remedy this. The starting point is the familiar notion that to de
tect
object parts, in practice it often suf■ces to detect co-occurrences of more loca
l
object fragments – a process that can be formalized as comparison (e.g. vector
quantization) of image patches against a codebook of known fragments, followed
by local aggregation of the resulting codebook membership vectors to detect co-
occurrences. This process converts local collections of image descriptor vectors
into somewhat less local histogram vectors – higher-level but spatially coarser
descriptors. We observe that as the output is again a local descriptor vector, t
he
process can be iterated, and that doing so captures and codes ever larger assem-
blies of object parts and increasingl y abstract or 'semantic' image properties.
We formulate the hyperfeatures model and study its performance under several
different image coding methods including clustering based Vector Quantization,Ga
ussian Mixtures, and combinations of these with Latent Dirichlet Allocation.
We ■nd that the resulting high-level features provide improved performance in
several object image and texture image classi■cation tasks.
1
****
Riemannian Manifold Learning for Nonlinear
Dimensionality Reduction
Tony Lin1,■, Hongbin Zha1, and Sang Uk Lee2
1National Laboratory on Machine Perception,
Peking University, Beijing 100871, China
{lintong, zha }@cis.pku.edu.cn
2School of Electrical Engineering,
Seoul National University, Seoul 151-742, Korea
sanguk@ipl.snu.ac.kr
Abstract. In recent years, nonlinear dimensionality reduction (NLDR)
techniques have attracted much attention in visual perception and many
other areas of science. We propose an e■cient algorithm called Rie-
mannian manifold learning (RML). A Riemannian manifold can be con-
structed in the form of a simplicial complex, and thus its intrinsic
dimension can be reliably estimated. Then the NLDR problem is solved
by constructing Riemannian normal coordinates (RNC). Experimental
results demonstrate that our algorithm can learn the data's intrinsic
geometric structure, yielding uniformly distributed and well organized
low-dimensional embedding data.
1
****
Controlling Sparseness in
Non-negative Tensor Factorization
Matthias Heiler and Christoph Schn¨ orr
Computer Vision, Graphics, and Pattern Recognition Group,
Department of Mathematics and Computer Science,
University of Mannheim, 68131 Mannheim, Germany
{heiler, schnoerr }@uni-mannheim.de
Abstract. Non-negative tensor factorization (NTF) has recently been

proposed as sparse and e■cient image representation (Welling and Weber, Patt. Rec. Let., 2001) . Until now, sparsity of the tensor factorization has been empirically observed in many cases, but there was no systematic way to control it. In this work, we show that a sparsity measure recently proposed for non-negative matrix factorization (Hoyer, J. Mach. Learn. Res., 2004) applies to NTF and allows precise control over sparseness of the resulting factorization. We devise an algorithm based on sequential conic programming and show improved performance over classical NTF codes on arti■cial and on real-world data sets.
1
****
Conditional Infomax Learning: An Integrated Framework for Feature Extraction and Fusion
Dahua Lin1a n dX i a o o uT a n g1,2
1Dept. of Information Engineering,
The Chinese University of Hong Kong, Hong Kong, China
dhlin4@ie.cuhk.edu.hk
2Microsoft Research Asia, Beijing, China
xitang@microsoft.com
Abstract. The paper introduces a new framework for feature learning in classi■cation motivated by information theory. We ■rst systematically study the information structure and present a novel perspective revealingthe two key factors in information utilization: class-relevance and redundancy. We derive a new information decomposition model where a novel concept called class-relevant redundancy is introduced. Subsequently a new algorithm called Conditional Informative Feature Extraction is formulated, which maximizes the joint class-relevant information by explicitly reducing the class-relevant re dundancies among features. To address the computational di■culties in information-based optimization, we incorporate Parzen window estimation into the discrete approximation of the objective function and propose a Local Active Region method whichsubstantially increases the optimization e■ciency. To e■ectively utilize the extracted feature set, we propose a Bayesian MAP formulation for feature fusion, which uni■es Laplacian Sparse Prior and MultivariateLogistic Regression to learn a fusion rule with good generalization capability. Realizing the ine■ciency caused by separate treatment of the extraction stage and the fusion stage, we further develop an improved design of the framework to coordinate the two stages by introducing a feedback from the fusion stage to the extraction stage, which signi■cantly enhances the learning e■ciency. The results of the comparative experiments show remarkable improvements achieved by our framework.
1
****
Degen Generalized Cylinders and Their Properties
Liangliang Cao1, Jianzhuang Liu1, and Xiaoou Tang1,2
1Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, China
{llcao, jzliu, xtang }@ie.cuhk.edu.hk
2Microsoft Research Asia, Beijing, China
xitang@microsoft.com
Abstract. Generalized cylinder (GC) has played an important role in computer vision since it was introduced in the 1970s. While studying GC models in human visual perception of shapes from contours, Marr assumed that GC’s limbs are planar curves. Later, Koenderink and Ponce pointed out that this assumptiond oes not hold in general by giving some examples. In this paper, we show that straight homogeneous generalized cylinders (SHGCs) and tori (a kind of curved GCs) have planar limbs when viewed from points on speci■c straight lines. This property leads us to the de■nition and investigation of a new class of GCs, with the help of the surface model proposed by Degen for geometric modeling. We call them Degen generalized cylinders (DGCs), which include SHGCs, tori, quadrics,cyc

lides, and more other GCs into one model. Our rigorous discussion is based on projective geometry and homogeneous coordinates. We present some invari-ant properties of DGCs that reveal the relations among the planar limbs, axes, a ndcontours of DGCs. These properties are useful for recovering DGC descriptions from image contours as well as for some other tasks in computer vision.
1
****
Geodesics Between 3D Closed Curves Using
Path-Straightening
Eric Klassen1and Anuj Srivastava2
1Department of Mathematics, Florida State University, Tallahassee, FL 32306
2Department of Statistics, Florida State University, Tallahassee, FL 32306
Abstract. In order to analyze shapes of continuous curves in R3,w e parameterize them by arc-length and represent them as curves on a unit two-sphere. We identify the subset denoting the closed curves, and study its di■erential geometry. To compute geodesics between any two such curves, we connect them with an arbitrary path, and then iteratively straighten this path using the gradient of an energy associated with thispath. T he limiting path of this path-straightening approach is a geodesic.
Next, we consider the shape space of these curves by removing shape-preserving transformations such as rotation and re-parametrization. To construct a geodesic in this shape space, we construct the shortestgeodesic betw een the all possible transformations of the two end shapes;
this is accomplished using an iterative procedure. We provide step-by-step descriptions of all the procedures, and demonstrate them with sim-ple examp les.
1
****
Robust Homography Estimation from Planar
C o n t o u r sB a s e do nC o n v e x i t y
Alberto Ruiz1,P e d r oE .L ´ opez de Teruel2, and Lorenzo Fern´ andez2
1Dept. Inform´ atica y Sistemas, University of Murcia
2Dept. Tecnolog´ ■a e Ingenier´ ■a de Computadores, University of Murcia
aruiz@um.es, {pedroe, lfmaimo }@ditec.um.es
Abstract. We propose a homography estimation method from the contours of planar regions. Standard projective invariants such as cross ratios or canonical frames based on hot points obtained from local dif-ferential properties are extremely unstable in real images su■ering from pixelization, thresholding artifacts, and other noise sources. We explorealterna tive constructions based on global convexity properties of the con-tour such as discrete tangents and concavities. We show that a projec-tive frame can be robustly extracted from arbitrary shapes with at leastone appr eciable concavity. Algorithmic complexity and stability are the-oretically discussed and experimentally evaluated in a number of real applications including projective shape matching, alignment and pose estimation. We conclude that the procedure is computationally e■cient and notably robust given the ill-conditioned nature of the problem.
1
****
Detecting Instances of Shape Classes That Exhibit
Variable Structure
Vassilis Athitsos1, Jingbin Wang2, Stan Sclaroff2,a n dM a r g r i tB e t k e2
1Siemens Corporate Research, Princeton, NJ 08540, USA
2Computer Science Department, Boston University, Boston, MA 02215, USA
Abstract. This paper proposes a method for detecting shapes of variable struc-ture in images with clutter. The term “variable structure” means that some shape parts can be repeated an arbitrary number of times, some parts can be optional, and some parts can have several alternative appearances. The particular variatio n
of the shape structure that occurs in a given image is not known a priori. Ex-is

ting computer vision methods, including deformable model methods, were not designed to detect shapes of variable structure; they may only be used to detect shapes that can be decomposed into a ■xed, a priori known, number of parts. The proposed method can handle both variations in shape structure and variations int he appearance of individual shape parts. A new class of shape models is intro- duced, called Hidden State Shape Models, that can naturally represent shapes of variable structure. A detection algorithm is described that ■nds instances of su ch
shapes in images with large amounts of clutter by ■nding globally optimal cor- respondences between image features and shape models. Experiments with real images demonstrate that our method can localize plant branches that consist ofan a priori unknown number of leaves and can detect hands more accurately than a hand detector based on the chamfer distance.
1
****
Direct Solutions for Computing Cylinders from
Minimal Sets of 3D Points
Christian Beder and Wolfgang F¨ orstner
Institute for Photogrammetry,
Bonn University, Germany
{beder, wf }@ipb.uni-bonn.de
Abstract. E■cient direct solutions for the determination of a cylinder from points are presented. The solutions range from the well known di- rect solution of a quadric to the minimal solution of a cylinder with ■ve points. In contrast to the approach of G. Roth and M. D. Levine (1990), who used polynomial bases for representing the geometric entities, we use algebraic constraints on the quadric representing the cylinder. Thesolutions for six to eight points directly determine all the cylinder pa- rameters in one step: (1) The eight-point-solution, similar to the esti- mation of the fundamental matrix, requires to solve for the roots of a3rd-order- polynomial. (2) The seven-point-solution, similar to the six- point-solution for the relative orientation by J. Philip (1996), yields a linear equation system. (3) The six-point-solution, similar to the ■ve- point-solution for the relative orientation by D. Nister (2003), yields a ten-by-ten eigenvalue problem. The new minimal ■ve-point-solution ■rst determines the direction and then the position and the radius of thecylinder. Th e search for the zeros of the resulting 6th order polynomials is e■ciently realized using 2D-Bernstein polynomials. Also direct solu- tions for the special cases with the axes of the cylinder parallel to acoordinat e plane or axis are given. The method is used to ■nd cylinders in range data of an industrial site.
1
****
Estimation of Multiple Periodic
Motions from Video
Alexia Briassouli and Narendra Ahuja
Beckman Insitute, University of Illinois, Urbana-Champaign,
405 N Matthews, Urbana, IL, 61801
{briassou, ahuja }@vision.ai.uiuc.edu
Abstract. The analysis of periodic or repetitive motions is useful in many applications, both in the natural and the man-made world. An important example is the recognition of human and animal activities.Existing met hods for the analysis of periodic motions ■rst extract motion trajectories, e.g. via correlation, or feature point matching. We present a new approach, which takes advantage of both the frequency and spatialinformation of the video. The 2D spatial Fourier transform is applied to each frame, and time-frequency distributions are then used to estimate the time-varying object motions. Thus, multiple periodic trajectories areextract ed and their periods are estimated. The period information is ■nally used to segment the periodically moving objects. Unlike existing

methods, our approach estimates multiple periodicities simultaneously,it is robu
st to deviations from strictly periodic motion, and estimates
periodicities superposed on translations. Experiments with synthetic and
real sequences display the capabilities and limitations of this approach.
Supplementary material is provided, showing the video sequences used
in the experiments.
1
****
Robust Multi-body Motion Tracking Using Commute
Time Clustering
Huaijun Qiu and Edwin R. Hancock
Department of Computer Science, University of York
York, YO10 5DD, UK
Abstract. The presence of noise renders the classical factorization method al-
most impractical for real-world multi-body motion tracking problems. The main
problem stems from the effect of noise on the shape interaction matrix, which
looses its block-diagonal structure and as a result the assignment of elements
to objects becomes dif■cult. The aim in this paper is to overcome this problemus
ing graph-spectral embedding and the k-means algorithm. To this end we de-
velop a representation based on the commute time between nodes on a graph. The
commute time (i.e. the expected time taken for a random walk to travel between
two nodes and return) can be computed from the Laplacian spectrum using the
discrete Green’s function, and is an important property of the random walk on
a graph. The commute time is a more robust measure of the proximity of data
than the raw proximity matrix. Our embedding procedure preserves commute
time, and is closely akin to kernel PCA, the Laplacian eigenmap and the diffu-
sion map. We illustrate the results both on the synthetic image sequences and re
alworld video sequences, and compare our results with several alternative method
s.
1
****
A Tuned Eigenspace Technique for Articulated
Motion Recognition
M. Masudur Rahman and Antonio Robles-Kelly
National ICT Australia■, RSISE Bldg. 115, ANU, ACT 0200, Australia
Masud.Rahman@ {rsise.anu.edu.au, nicta.com.au }
Antonio.Robles-Kelly@ {anu.edu.au, nicta.com.au }
Abstract. In this paper, we introduce a tuned eigenspace technique so as to clas
-
sify human motion. The method presented here overcomes those problems related
to articulated motion and dress texture effects by learning various human motion
s
in terms of their sequential postures in an eigenspace. In order to cope with th
e
variability inherent to articulated motion, we propose a method to tune the set
of sequential eigenspaces. Once the learnt tuned eigenspaces are at hand, the
recognition task then becomes a nearest-neighbor search over the eigenspaces.
We show how our tuned eigenspace method can be used for purposes of real-
world and synthetic pose recognition. We al so discuss and overcome the problem
related to clothing texture that occurs in real-world data, and propose a back-
ground subtraction method to employ the method in out-door environment. We
provide results on synthetic imagery for a number of human poses and illustrate
the utility of the method for the purposes of human motion recognition.
1
****
Real-Time Non-rigid Shape Recovery Via Active
Appearance Models for Augmented Reality
Jianke Zhu, Steven C.H. Hoi, and Michael R. Lyu
Department of Computer Science & Engineering,
Chinese University of Hong Kong,

Shatin, Hong Kong
{jkzhu, chhoi, lyu }@cse.cuhk.edu.hk
Abstract. One main challenge in Augmented Reality (AR) applica-
tions is to keep track of video objects with their movement, orientation,
size, and position accurately. This poses a challenging task to recover
non-rigid shape and global pose in real-time AR applications. This pa-per propos
es a novel two-stage scheme for online non-rigid shape recovery
toward AR applications using Active Appearance Models (AAMs). First,
we construct 3D shape models from AAMs o■ine, which do not involveprocessing of
the 3D scan data. Based on the computed 3D shape models,
we propose an e■cient online algorithm to estimate both 3D pose and
non-rigid shape parameters via local bundle adjustment for building up
point correspondences. Our approach, without manual intervention, can
recover the 3D non-rigid shape e■ectively from either real-time video
sequences or single image. The recovered 3D pose parameters can beused for AR re
gistrations. Furthermore, the facial feature can be tracked
simultaneously, which is critical for many face related applications. We
evaluate our algorithms on several video sequences. Promising experi-
mental results demonstrate our proposed scheme is e■ective and signi■-
cant for real-time AR applications.
1
****
A Fluid Motion Estimator for Schlieren
Image Velocimetry
Elise Arnaud1, Etienne M´ emin2,R o b e r t oS o s a3, and Guillermo Artana3
1Disi, Universit` a di Genova, 16146 Genova, Italy
arnaud@disi.unige.it
2IRISA, Universit´ e de Rennes 1, 35 042 Rennes Cedex, France
memin@irisa.fr
3Facultad de Ingenier´ ■a, Universitad de Buenos Aires
Buenos Aires 1412, Argentina
{rsosa, gartana }@fi.uba.ar
Abstract. In this paper, we address the problem of estimating the mo-
tion of ■uid ■ows that are visualized through a Schlieren system. Such
a system is well known in ■uid mechanics as it enables the visualization
of unseeded ■ows. As the resulting images exhibit very low photomet-
ric contrasts, classical motion estimation methods based on the bright-
ness consistency assumption (correlation-based approaches, optical ■owmethods) a
re completely ine■cient. This work aims at proposing a sound
energy based estimator dedicated to these particular images. The energy
function to be minimized is composed of (a)a novel data term describing
the fact that the observed luminance is linked to the gradient of the ■uiddensit
y and (b)a speci■c div curl regulariza tion term. The relevance of
our estimator is demonstrated on real-world sequences.
1
****
Bilateral Filtering-Based Optical Flow Estimation
with Occlusion Detection
Jiangjian Xiao, Hui Cheng, Harpreet Sawhney,
Cen Rao, and Michael Isnardi
Sarnoff Corporation
{jxiao, hcheng, hsawhney, crao, misnardi }@sarnoff.com
Abstract. Using the variational approaches to estimate optical ■ow between two
frames, the ■ow discontinuities between different motion ■elds are usually not
distinguished even when an anisotropic diffusion operator is applied. In this pa
-
per, we propose a multi-cue driven adaptive bilateral ■lter to regularize the ■o
w
computation, which is able to achieve the smoothly varied optical ■ow ■eld withh
ighly desirable motion discontinuities. First, we separate the traditional one-s

tep
variational updating model into a two-step ■ltering-based updating model. Then, employing our occlusion detector, we reformulate the energy functional of op-tic al ■ow estimation by explicitly introducing an occlusion term to balance the energy loss due to the occlusion or mismatches. Furthermore, based on the two-step updating framework, a novel multi-cue driven bilateral ■lter is proposed to substitute the original anisotropic diffusion process, and it is able to adaptiv ely
control the diffusion process according to the occlusion detection, image inten-sity dissimilarity, and motion dissimilarity. After applying our approach on var i-ous video sources (movie and TV) in the presence of occlusion, motion blurring ,
non-rigid deformation, and weak textureness, we generate a spatial-coherent ■ow ■eld between each pair of input frames and detect more accurate ■ow disconti-nui ties along the motion boundaries.
1
****
Geometry and Kinematics with Uncertain Data
Christian Perwass, Christian Gebken, and Gerald Sommer
Institut für Informatik, CAU Kiel,
Christian-Albrechts-Platz 4, 24118 Kiel, Germany
{chp, chg, gs }@ks.informatik.uni-kiel.de
Abstract. In Computer Vision applications, one usually has to work with un-certain data. It is therefore important to be able to deal with uncertain geome-try and uncertain transformations in a uniform way. The Geometric Algebra of conformal space offers a unifying framework to treat not only geometric enti-tie s like points, lines, planes, circles and spheres, but also transformations like re■ection, inversion, rotation and tran slation. In this text we show how the un -
certainty of all elements of the Geometric Algebra of conformal space can beappr opriately described by covariance matrices. In particular, it will be shown that it is advantageous to represent uncertain transformations in Geometric Alge -
bra as compared to matrices. Other important results are a novel pose estimation approach, a uniform framework for geometric entity ■tting and triangulation, the testing of uncertain tangen tiality relations and the treatment of catadioptric cam-
eras with parabolic mirrors within this framework. This extends previous work by Förstner and Heuel from points, lines and planes to non-linear geometric entiti es
and transformations, while keeping the linearity of the estimation method. We give a theoretical description of our approach and show exemplary applications.
1
****
Euclidean Structure from N≥2 Parallel Circles:
Theory and Algorithms
Pierre Gurdjos1, Peter Sturm2, and Yihong Wu3
1IRIT-TCI, UPS, 118 route de Narbonne,
31062 Toulouse, cedex 9, France
Pierre.Gurdjos@irit.fr
2PERCEPTION, INRIA Rhône-Alpes,
655, avenue de l'Europe, 38330 Montbonnot, France
Peter.Sturm@inrialpes.fr
3NLPR-IA, Chinese Academy of Sciences, P.O. Box 2728,
No. 95 East Road of Zhong Guan Cun, Beijing 100080, China
yhwu@nlpr.ia.ac.cn
Abstract. Our problem is that of recovering, in one view, the 2D Eu-clidean structure, induced by the projections of Nparallel circles. This structure is a prerequisite for camera calibration and pose computation. Until now, no general method has been described for N>2. The main

contribution of this work is to state the problem in terms of a system of linear equations to solve. We give a closed-form solution as well as bundle adjustment-like re■nements, increasing the technical applicability andnumerical stability. Our theoretical approach generalizes and extends all those described in existing works for N= 2 in several respects, as we can treat simultaneously pairs of orthogonal lines and pairs of circles within a uni■ed framework. The proposed algorithm may be easily implemented,using well-known numerical algorithms. Its performance is illustrated by simulations and experiments with real images.
1
****
Overconstrained Linear Estimation of Radial Distortion and Multi-view Geometry
R. Matt Steele and Christopher Jaynes
Center for Visualization and Virtual Environments,
University of Kentucky, Lexington, Kentucky, USA
Abstract. This paper introduces a new method for simultaneous estimation of lens distortion and multi-view geometry using only pointcorrespondences. The new technique has signi■cant advantages over the current state-of-the art in that it makes more e■ective use of correspondences arising from any number of views. Multi-view geometry in thepresence of lens distortion can be expressed as a set of point correspon-dence constraints that are quadratic in the unknown distortion parameter. Previous work has demonstrated how the system can be solved e■ciently as a quadratic eigenvalue problem by operating on the normalequations of the system. Although this approach is appropriate for situations in which only a minimal set of matchpoints are available, it does not take full advantage of extra correspondences in overconstrained situations, resulting in signi■cant bias and many potential solutions. The new technique directly operates on the initial constraint equations and solves the quadratic eigenvalue problem in the case of rectangular matrices. Themethod is shown to contain signi■cantly less bias on both controlled and real-world data and, in the case of a moving camera where additional views serve to constrain the number of solutions, an accurate estimateof both geometry and distortion is achieved.
1
****
Camera Calibration with Two Arbitrary Coaxial Circles
Carlo Colombo, Dario Comanducci, and Alberto Del Bimbo
Dipartimento di Sistemi e Informatica,
Via S. Marta 3, 50139 Firenze, Italy
{colombo, comandu, delbimbo }@dsi.unifi.it
Abstract. We present an approach for camera calibration from the image of at least two circles arranged in a coaxial way. Such a geometric con■guration arises in static scenes of objects with rotational symmetry or in scenes including generic o bjects undergoing rotational motion around a ■xed axis. The approach is based on the automatic localization of a surface of revolution (SOR) in the image, and its use as a calibration artifact. The SOR can either be a real object in a static scene,or a “virtual surface” obtained by frame superposition in a rotational sequence. This provides a uni■ed framework for calibration from single images of SORs or from turntable sequences. Both the internal and ex-ternal calibration parameters (square pixels model) are obtained from two or more imaged cross sections of the SOR, whose apparent contour is also exploited to obtain a better calibration accuracy. Experimental results show that this calibration approach is accurate enough for several vision applications, encompassing 3D realistic model acquisition from single images, and desktop 3D object scanning.
1

****
Molding Face Shapes by Example
Ira Kemelmacher and Ronen Basri■
Dept. of Computer Science and Applied Math.,
The Weizmann Institute of Science,
Rehovot 76100, Israel
Abstract. Human faces are remarkably similar in global properties, in-
cluding size, aspect ratios, and locations of main features, but can vary
considerably in details across individuals, gender, race, or due to facial
expression. We propose a novel method for 3D shape recovery of a face
from a single image using a single 3D reference model of a di■erent per-
son's face. The method uses the input image as a guide to mold the
reference model to reach a desired reconstruction. Assuming Lambertian
re■ectance and rough alignment of the input image and reference model,
we seek shape, albedo, and lighting that best ■t the image while preserv-
ing the rough structure of the model. We demonstrate our method by
providing accurate reconstructions of novel faces overcoming signi■cant
di■erences in shape due to gender, race, and facial expressions.
1
****
Reconstruction of Canal Surfaces from Single
Images Under Exact Perspective
Vincenzo Caglioti and Alessandro Giusti
Politecnico di Milano
{caglioti, giusti }@elet.polimi.it
Abstract. This paper addresses the reconstruction of canal surfaces
from single images. A canal surface is obtained as the envelope of a
family of spheres of constant radius, whose center is swept along a spacecurve,
called axis. Previous studies either used approximate relationships
(quasi-invariants), or they addressed the recognition based on a geomet-
ric model. In this paper we show that, under broad conditions, canal sur-faces c
an be reconstructed from single images under exact perspective. Inparticular, ca
nal surfaces with planar axis can even be reconstructed from
a single fully-uncalibrated image. An automatic reconstruction method
has been implemented. Simulations and experimental results on real im-ages are a
lso presented.
1
****
Subspace Estimation Using Projection Based
M-Estimators over Grassmann Manifolds
Raghav Subbarao and Peter Meer
Department of Electrical and Computer Engineering,
Rutgers University, Piscataway, NJ 08854, USA
{rsubbara, meer }@caip.rutgers.edu
Abstract. We propose a solution to the problem of robust subspace es-
timation using the projection based M-estimator. The new method han-dles more ou
tliers than inliers, does not require a user de■ned scale of
the noise a■ecting the inliers, handles noncentered data and nonorthog-
onal subspaces. Other robust methods like RANSAC, use an input for
the scale, while methods for subspace segmentation, like GPCA, are not
robust. Synthetic data and three real cases of multibody factorization
show the superiority of our method, in spite of user independence.
1
****
Anastass iaAngelopoulou1,J os´eG a r c ´ ■aRodr´■guez2,andAlexandraPsarrou1
1Harrow School of Computer Science, University of Westminster,
Harrow HA1 3TP, United Kingdom
2Departamento de Tecnolog´ ■a Inform´ atica y Computaci´ on, Universidad de Alic
ante,
Apdo. 99. 03080 Alicante, Spain

Abstract. Recovering the shape of a class of objects requires estab-
lishing correct correspondences between manually or automatically an-notated lan
dmark points. In this study, we utilise a novel approach to
automatically recover the shape of hand outlines from a series of 2D train-
ing images. Automated landmark extraction is accomplished through theuse of the
self-organising model the growing neural gas (GNG) network
which is able to learn and preserve the topological relations of a given set
of input patterns without requiring a priori knowledge of the structure of
the input space. To measure the quality of the mapping throughout the
adaptation process we use the topographic product. Results are given forthe trai
ning set of hand outlines.
1
****
Towards Optimal Training
of Cascaded Detectors
S. Charles Brubaker, Matthew D. Mullin, and James M. Rehg
College of Computing and GVU Center,
Georgia Institute of Technology, Atlanta, GA 30332
{brubaker, mdmullin, rehg }@cc.gatech.edu
Abstract. Cascades of boosted ensembles have become popular in the
object detection community following their highly successful introduc-tion in th
e face detector of Viola and Jones [1]. In this paper, we explore
several aspects of this architecture that have not yet received adequate
attention: decision points of cascade stages, faster ensemble learning, and
stronger weak hypotheses. We present a novel strategy to determine the
appropriate balance between false positive and detection rates in the
individual stages of the cascade based on a probablistic model of theoverall cas
cade’s performance. To improve the training time of individ-
ual stages, we explore the use of feature ■ltering before the application
of Adaboost. Finally, we show that the use of stronger weak hypothe-ses based on
 CART can signi■cantly improve upon the standard face
detection results on the CMU-MIT data set.
1
****
Learning and Incorporating Top-Down Cues
in Image Segmentation
Xuming He, Richard S. Zemel, and Debajyoti Ray
Department of Computer Science, University of Toronto
{hexm, zemel, debray }@cs.toronto.edu
Abstract. Bottom-up approaches, which rely mainly on continuity prin-
ciples, are often insu■cient to form accurate segments in natural images. In
order to improve performance, recent methods have begun to incorporate
top-down cues, or object information, into segmentation. In this paper, we
propose an approach to utilizing category-based information in segmenta-
tion, through a formulation as an image labelling problem. Our approach
exploits bottom-up image cues to create an over-segmented representation
of an image. The segments are then merged by assigning labels that corre-
spond to the object category. The model is trained on a database of images,
and is designed to be modular: it learns a number of image contexts, whichsimpli
fy training andextendtherange of object classes and image database
size that the system can handle. The learning method estimates model pa-
rameters by maximizing a lower bound of the data likelihood. We examine
performance on three real-world image databases, and compare our system
to a standard classi■er and other conditional random ■eld approaches, as
well as a bottom-up segmentation method.
1
****
Learning to Detect Objects of Many Classes
Using Binary Classi■ers
Ramana Isukapalli1, Ahmed Elgammal2, and Russell Greiner3

1Lucent Technologies, Bell Labs Innovations, Whippany, NJ 07981, USA
2Rutgers University, New Brunswick, NJ 08854, USA
3University of Alberta, Edmonton, CA T6G 2E8, CA
Abstract. Viola and Jones [VJ] demonstrate that cascade classi■cation methods can successfully detect objects belonging to a single class, such as faces. Detect-
ing and identifying objects that belong to any of a set of “classes”, many class detection , is a much more challenging problem. We show that objects from each class can form a “cluster” in a “classi■er sp ace” and illustrate examples of such
clusters using images of real world objects. Our detection algorithm uses a “de-
cision tree classi■er” (whose internal nodes each correspond to a VJ classi■er)to propose a class label for every sub-image Wof a test image (or reject it as a negative instance). If this Wreaches a leaf of this tree, we then pass Wthrough a
subsequent VJ cascade of classi■ers, speci■c to the identi■ed class, to determine
whether Wis truly an instance of the proposed class. We perform several empir-
ical studies to compare our system for detecting objects of any of Mclasses, to the obvious approach of running a setofMlearned VJ cascade classi■ers, one for each class of objects, on the same image. We found that the detection rates are comparable, and our many-class detection system is about as fast as running asingle VJ cascade, and scales up well as the number of classes increases.
1
****
A Unifying Framework for Mutual Information
Methods for Use in Non-linear Optimisation
Nicholas Dowson and Richard Bowden
Centre for Vision Speed and Signal Processing,
University of Surrey, Guildford, GU2 7JW, UK
{n.dowson, r.bowden }@surrey.ac.uk
http://www.ee.surrey.ac.uk/personal/n.dowson
Abstract. Many variants of MI exist in the literature. These vary pri-
marily in how the joint histogram is populated. This paper places thefour main variants of MI: Standard sampling, Partial Volume Estima-
tion (PVE), In-Parzen Windowing and Post-Parzen Windowing into a single mathematical framework. Jacobians and Hessians are derived ineach case. A particular contribution is that the non-linearities implicit to standard sampling and post-Parzen windowing are explicitly dealt with. These non-linearities are a barrier to their use in optimisation. Side-by-
side comparison of the MI variants is made using eight diverse data-sets, considering computational expense and convergence. In the experiments, PVE was generally the best performer, although standard sampling of-
ten performed nearly as well (if a higher sample rate was used). Thewidely used sum of squared di■erences metric performed as well as MI unless large occlusions and non-linear intensity relationships occurred. The binaries and scripts used for testing are available online.
1
****
Random Walks, Constrained Multiple Hypothesis
Testing and Image Enhancement■
Noura Azzabou1,2, Nikos Paragios1, and Frederic Guichard2
1M A S ,E c o l eC e n t r a l ed eP a r i s ,
Grande V oie des Vignes, Chatenay-Malabry, France
noura.azzabou@certis.enpc.fr, nikos.paragios@ecp.fr
http://www.mas.ecp.fr
2DxOLabs, 3, Rue Nationale, 92100 Boulogne, France
{nazzabou, fguichard }@dxo.com
http://www.dxo.com
Abstract. Image restoration is a keen problem of low level vision. In this paper

,
we propose a novel - assumption-free on the noise model - technique based on random walks for image enhancement. Our method explores multiple neighbors sets (or hypotheses) that can be used for pixel denoising, through a particle ■lter-
ing approach. This technique associates weights for each hypotheses according to its relevance and its contribution in the denoising process. Towards accounti
ng
for the image structure, we introduce perturbations based on local statistical p
rop-
erties of the image. In other words, particle evolution are controlled by the im
age
structure leading to a ■ltering window adapted to the image content. Promising experimental results demonstrate the potential of such an approach.
1
****
From Tensor-Driven Diffusion to Anisotropic
Wavelet Shrinkage
Martin Welk1, Joachim Weickert1, and Gabriele Steidl2
1Mathematical Image Analysis Group,
Faculty of Mathematics and Computer Science,
Saarland University, 66041 Saarbr¨ ucken, Germany
{welk, weickert }@mia.uni-saarland.de
http://www.mia.uni-saarland.de
2Faculty of Mathematics and Computer Science,
A5 University of Mannheim, 68131 Mannheim, Germany
steidl@math.uni-mannheim.de
http://kiwi.math.uni-mannheim.de
Abstract. Diffusion processes driven by anisotropic diffusion tensors are known to be well-suited for structure-preser ving denoising. However, numerical imple-mentations based on ■nite differences introduce unwanted blurring artifacts that deteriorate these favourable ■ltering properties. In this paper we introduce a n
ovel
discretisation of a fairly general class of anisotropic diffusion processes on a 2-D grid. It leads to a locally semi-analytic scheme (LSAS) that is absolutely stable, simple to implement and offers an outstanding sharpness of ■ltered im-ages. By showing that this scheme can be translated into a 2-D Haar wavelet shrinkage procedure, we establish a connection between tensor-driven diffusion and anisotropic wavelet shrinkage for the ■rst time. This result leads to couple
d
shrinkage rules that allow to perform highly anisotropic ■ltering even with the simplest wavelets.
1
****
SURF: Speeded Up Robust Features
Herbert Bay1, Tinne Tuytelaars2, and Luc Van Gool1,2
1ETH Zurich
{bay, vangool }@vision.ee.ethz.ch
2Katholieke Universiteit Leuven
{Tinne.Tuytelaars, Luc.Vangool }@esat.kuleuven.be
Abstract. In this paper, we present a novel scale- and rotation-invariant interest point detector and descriptor, coined SURF (Speeded Up Ro-bust Features
). It approximates or even outperforms previously proposed
schemes with respect to repeatability, distinctiveness, and robustness, yet can be computed and compared much faster.
This is achieved by relying on integral images for image convolutions;
by building on the strengths of the leading existing detectors and descrip-tors (in casu , using a Hessian matrix-based measure for the detector, and
a distribution-based descriptor); and by simplifying these methods to the essential. This leads to a combination of novel detection, description, andmatch

ing steps. The paper presents experimental results on a standardevaluation set, as well as on imagery obtained in the context of a real-life object recognition application. Both show SURF's strong performance.
1
****
Top-Points as Interest Points for Image Matching
B. Platel, E. Balmachnova, L.M.J. Florack■, and B.M. ter Haar Romeny
Technische Universiteit Eindhoven, P.O. Box 513,
5600 MB Eindhoven, The Netherlands
{B.Platel, E.Balmachnova, L.M.J.Florack,
B.M.terHaarRomeny }@tue.nl
Abstract. We consider the use of top-points for object retrieval. These points a re
based on scale-space and catastrophe theory, and are invariant under gray value scaling and offset as well as scale-Euclidean transformations. The differential properties and noise characteristics of these points are mathematically well un-derstood. It is possible to retrieve the exact location of a top-point from any coarse
estimation through a closed-form vector equation which only depends on local derivatives in the estimated point. All these properties make top-points highly suitable as anchor points for invariant matching schemes. By means of a set of repeatability experiments and receiver-operato r-curves we demonstrate the per-formance of top-points and differential invariant features as image descriptors.
1
****
Machine Learning for High-Speed
Corner Detection
Edward Rosten and Tom Drummond
Department of Engineering,
Cambridge University, UK
{er258, twd20 }@cam.ac.uk
Abstract. Where feature points are used in real-time frame-rate appli-cations, a high-speed feature detector is necessary. Feature detectors such as SIFT (DoG), Harris and SUSAN are good methods which yield high quality features, however they are too computationally intensive for usein real-time applications of any complexity. Here we show that machine learning can be used to derive a feature detector which can fully process live PAL video using less than 7% of the available processing time. Bycomparison neither the Harris detector (120%) nor the detection stageof SIFT (300%) can op erate at full frame rate.
Clearly a high-speed detector is of limited use if the features produced are unsuitable for downstream processing. In particular, the same scene viewed from two di■erent positions should yield features which corre-spond to the same real-world 3D locations[1]. Hence the second contri-bution of this paper is a comparison corner detectors based on this crite-rion applied to 3D scenes. This comparison supports a number of claims made elsewhere concerning existing corner detectors. Further, contraryto our ini tial expectations, we show that despite being principally con-structed for speed, our detector signi■cantly outperforms existing feature detectors according to this criterion.
1
****
Smooth Image Segmentation by Nonparametric
Bayesian Inference
Peter Orbanz and Joachim M. Buhmann
Institute of Computational Science, ETH Zurich
{porbanz, jbuhmann }@inf.ethz.ch
Abstract. A nonparametric Bayesian model for histogram clustering is proposed to automatically determine the number of segments when Markov Random Field constraints enforce smooth class assignments. Thenonparametr

ic nature of this model is implemented by a Dirichlet pro-
cess prior to control the number of clusters. The resulting posterior can
be sampled by a modi■cation of a conjugate-case sampling algorithm
for Dirichlet process mixture models. This sampling procedure estimates
segmentations as e■ciently as clustering procedures in the strictly con-
jugate case. The sampling algorithm can process both single-channel andmulti-cha
nnel image data. Experimental results are presented for real-
world synthetic aperture radar and magnetic resonance imaging data.
1
****
Shape Analysis and Fuzzy Control
for 3D Competitive Segmentation
of Brain Structures with Level Sets
Cyb`ele Ciofolo and Christian Barillot
IRISA / CNRS, Team VisAGeS,
Campus de Beaulieu, 35042 Rennes Cedex, France
{Cybele.Ciofolo, Christian.Barillot }@irisa.fr
http://www.www.irisa.fr/visages/visages-eng.html
Abstract. We propose a new method to segment 3D structures with
competitive level sets driven by a shape model and fuzzy control. To this
end, several contours evolve simultaneously toward previously de■ned
targets. The main contribution of this paper is the original introduction
of prior information provided by a shape model, which is used as an
anatomical atlas, into a fuzzy decision system. The shape information is
combined with the intensity distribution of the image and the relativeposition o
f the contours. This combination automatically determines the
directional term of the evolution equation of each level set. This leads
to a local expansion or contraction of the contours, in order to matchthe border
s of their respective targets. The shape model is produced
with a principal component analysis, and the resulting mean shape and
variations are used to estimate the target location and the fuzzy states
corresponding to the distance between the current contour and the tar-
get. By combining shape analysis and fuzzy control, we take advantage
of both approaches to improve the level set segmentation process withprior infor
mation. Experiments are shown for the 3D segmentation of
deep brain structures from MRI and a quantitative evaluation is per-
formed on a 18 volumes dataset.
1
****
Variational Motion Segmentation with Level Sets
Thomas Brox1, Andr´ es Bruhn2, and Joachim Weickert2
1CVPR Group, Department of Computer Science, University of Bonn,
R¨omerstr. 164, 53113 Bonn, Germany
brox@mia.uni-saarland.de
2Mathematical Image Analysis Group, Faculty of Mathematics and Computer Science,
Saarland University, Building 27, 66041 Saarbr¨ ucken, Germany
{bruhn, weickert }@mia.uni-saarland.de
Abstract. We suggest a variational method for the joint estimation of optic ■ow
and the segmentation of the image into regions of similar motion. It makes use
of the level set framework following the idea of motion competition, which is ex
-
tended to non-parametric motion. Moreover, we automatically determine an ap-
propriate initialization and the number of regions by means of recursive two-pha
se
splits with higher order region models. The method is further extended to the sp
a-
tiotemporal setting and the use of additional cues like the gray value or color
forthe segmentation. It need not fear a quantitative comparison to pure optic ■o
w es-
timation techniques: For the popular Yosemite sequence with clouds we obtain the

currently most accurate result. We further uncover a mistake in the ground truth
.Coarsely correcting this, we get an average angular error below 1 degree.
1
****
Ellipse Fitting with Hyperaccuracy
Kenichi Kanatani
Department of Computer Science,
Okayama University, Okayama 700-8530, Japan
kanatani@suri.it.okayama-u.ac.jp
Abstract. For ■tting an ellipse to a point sequence, ML (maximum
likelihood) has been regarded as having the highest accuracy. In this pa-per, we
demonstrate the existence of a "hyperaccurate" method which
outperforms ML. This is made possible by error analysis of ML followed
by subtraction of high-order bias terms. Since ML nearly achieves the
theoretical accuracy bound (the KCR lower bound), the resulting im-
provement is very small. Nevertheless, our analysis has theoretical sig-
ni■cance, illuminating the relationship between ML and the KCR lowerbound.
1
****
A Physically-Motivated Deformable Model
Based on Fluid Dynamics
Andrei C. Jalba and Jos B.T.M. Roerdink
Institute for Mathematics and Computing Science,
University of Groningen, P.O. Box 800,
9700 AV, Groningen, The Netherlands
{andrei, roe }@cs.rug.nl
Abstract. A novel deformable model for image segmentation and shape
recovery is presented. The model is inspired by ■uid dynamics and is
based on a ■ooding simulation similar to the watershed paradigm. Unlike
most watershed methods, our model has a continuous formulation, beingdescribed b
y two partial di■erential equations. In this model, di■erent
■uids, added by placing density (dye) sources manually or automatically,
are attracted towards the contours of the objects of interest by an imageforce.
In contrast to the watershed method, when di■erent ■uids meet
they may mix. When the topographical relief of the image is ■ooded,
the interfaces separating homogeneous ■uid regions can be traced to
yield the object contours. We demonstrate the ■exibility and potential
of our model in two experimental settings: shape recovery using manual
initializations and automated segmentation.
1
****
Video and Image Bayesian Demosaicing
with a Two Color Image Prior
Eric P. Bennett1, Matthew Uyttendaele2,C .L a w r e n c eZ i t n i c k2,
Richard Szeliski2, and Sing Bing Kang2
1The University of North Carolina at Chapel Hill, Chapel Hill, NC■
2Microsoft Research, Redmond, WA
Abstract. The demosaicing process converts single-CCD color repre-
sentations of one color channel per pixel into full per-pixel RGB. We
introduce a Bayesian technique for demosaicing Bayer color ■lter array
patterns that is based on a statistically-obtained two color per-pixel im-
age prior. By modeling all local color behavior as a linear combination
of two fully speci■ed RGB triples, we avoid color fringing artifacts while
preserving sharp edges. Our grid-less, ■oating-point pixel location archi-
tecture can process both single images and multiple images from video
within the same framework, with multiple images providing denser color
samples and therefore better color reproduction with reduced aliasing.
An initial clustering is performed to determine the underlying local two
color model surrounding each pixel. Using a product of Gaussians statis-
tical model, the underlying linear blending ratio of the two representative

colors at each pixel is estimated, while simultaneously providing noise re-
duction. Finally, we show that by sampling the image model at a ■ner
resolution than the source images during reconstruction, our continuous
demosaicing technique can super-resolve in a single step.
1
****
Generalized Multi-sensor Planning
Anurag Mittal
Dept of Computer Science and Engg■,
Indian Institute of Technology Madras,
Chennai-600036, India
Abstract. Vision systems for various tasks are increasingly being deployed. Al-
though signi■cant effort has gone into improving the algorithms for such tasks,
there has been relatively little work on determining optimal sensor con■guration
s.
This paper addresses this need. We speci■cally address and enhance the state-of-
the-art in the analysis of scenarios where there are dynamically occuring object
s
capable of occluding each other. The visibility constraints for such scenarios a
re
analyzed in a multi-camera setting. Also analy zed are other static constraints
such
as image resolution and ■eld-of-view, and algorithmic requirements such as stere
o
reconstruction, face detection and background appearance. Theoretical analysis
with the proper integration of such visibility and static constraints leads to a
 generic
framework for sensor planning, which can th en be customized for a particular ta
sk.
Our analysis can be applied to a variety of applications, especially those invol
ving
randomly occuring objects, and include surveillance and industrial automation.
Several examples illustrate the wide applicability of the approach.
1
****
Variational Shape and Re■ectance Estimation
Under Changing Light and Viewpoints
Neil Birkbeck1, Dana Cobzas1, Peter Sturm2, and Martin Jagersand1
1ComputerScience, Universityof Albe rta,Canada
{birkbeck, dana, jag }@cs.ualberta.ca
2INRIARhone-Alpes, F rance
peter.sturm@inrialpes.fr
Abstract. Fitting pa rameterized3 D shape an dgene ral re■ectance
models to 2 Dimage data ischalleng ingduet ot hehighd imens ional ityof
theproble m.Thepropose dmethodcombines t he capab ilities of class ical
andphotometrics t e reo, allo wing fo raccurate reconst ruction of bot htex-
turedandnon-text uredsurfaces. In pa rticular,wep resent a va riational
methodimplemente das a P DE-driven s urface evol ution inte rleave dw ith
re■ectance est imation. Thesurface isrepresente don an a dapt ivemesh
allowing topolog ical c hange. Toprovidetheinputdata, wehave designed
ac a p t ures e t upthat s imultaneo uslyacquires bot hviewpoint an dlight
variationwhileminimizing self-s hadowing. Ourcapt uremethodisf e a s i-
ble fo rreal-worldappl icat ion as itrequires a moderate a mount of input
data an dprocess ing t ime. In expe riments, models of people an deve ryday
objects werec a p t uredfromaf e wdozen images taken withac o n s umer
digital ca mera.Thec a p t urep rocess recove rsap hoto-cons istent model
of spat iallyvaryingLambertian an dspec ularre■ectance an dahighly
accurate geo metry.
1
****

Specularity Removal in Images and Videos:
A PDE Approach

Satya P. Mallick1, Todd Zickler2, Peter N. Belhumeur3, and David J. Kriegman1
1Computer Science and Engineering, University of California at San Diego, CA 92093
2Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138
3Computer Science, Columbia University, New York, NY 10027

Abstract. We present a uni■ed framework for separating specular and diffuse re■ection components in images and videos of textured scenes. This can be usedfor specularity removal and for independently processing, ■ltering, and recombining the two components. Beginning with a partial separation provided by an illumination-dependent color space, the challenge is to complete the separation using spatio-temporal information. This is accomplished by evolving a partial differential equation (PDE) that iteratively erodes the specular component at each pixel. A family of PDEs appropriate for d iffering image sources (still images vs. videos), differing prior information (e.g., highly vs. lightly textured scenes), or differing prior computations (e.g., optical ■ow) is introduced. In contrast to many other methods, explicit segmentation and/or manual intervention are not required .We present results on high-quality images and video acquired in the laboratory in addition to images taken from the Inte rnet. Results on the latter demonstrate robustness to low dynamic range, JPEG artifacts, and lack of knowledge of illuminant color. Empirical comparison to physical removal of specularities using polarization is provided. Finally, an application termed dichromatic editing is presented in which the diffuse and the specular components are processed independently to produce a variety of visual effects.

1

****

Carved Visual Hulls for Image-Based Modeling

Yasutaka Furukawa1and Jean Ponce1,2
1Department of Computer Science, University of Illinois at Urbana Champaign, USA
2D´epartement d'Informatique, Ecole Normale Sup´ erieure, Paris, France
{yfurukaw, ponce }@cs.uiuc.edu

Abstract. This article presents a novel method for acquiring high-quality solid models of complex 3D shapes from multiple calibrated photographs. After the purely geometric constraints associated with the silhouettes found in each image have been used to construct a coarse surface approximation in the form of a visual hull, photoconsistency constraints are enforced in three consecutive steps: (1) the rims where the surface grazes the visual hull are ■rst identi■ed through dynamic programming; (2) with the rims now ■xed, the visual hull is carved using graph cuts to globally optimize the photoconsistency of the surface and recover its main features; (3) an iterative (local) re■nement step is ■nally used to recover ■ne surface details. The proposed approach has been implemented, and experiments with six real data sets are presented, along with qualitative comparisons with several state-of-the-art image-based-modeling algorithms.

1

****

What Is the Range of Surface Reconstructions
from a Gradient Field?

Amit Agrawal1, Ramesh Raskar2, and Rama Chellappa1
1Center for Automation Research, University of Maryland,
College Park, MD, USA 20742

{aagrawal, rama }@cfar.umd.edu
2Mitsubishi Electric Research Labs (MERL),
201 Broadway, Cambridge, MA, USA 02139
raskar@merl.com
Abstract. We propose a generalized equation to represent a continuum of sur-
face reconstruction solutions of a given non-integrable gradient ■eld. We show
that common approaches such as Poisson solver and Frankot-Chellappa algo-rithm a
re special cases of this generalized equation. For a N×Npixel grid, the
subspace of all integrable gradient ■elds is of dimension N
2−1. Our frame-
work can be applied to derive a range of meaningful surface reconstructions from
this high dimensional space. The key observation is that the range of solutions is
related to the degree of anisotropy in applying weights to the gradients in the inte-
gration process. While common approaches use isotropic weights, we show thatby u
sing a progression of spatially varying anisotropic weights, we can achieve
signi■cant improvement in reconstructions. We propose (a) α
-surfaces using bi-
nary weights, where the parameter αallows trade off between smoothness and
robustness, (b) M-estimators and edge preserving regularization using continu-
ous weights and (c) Diffusion using af■ne transformation of gradients. We pro-
vide results on photometric stereo, compare with previous approaches and showtha
t anisotropic treatment discounts noise while recovering salient features in
reconstructions.
1
****
Practical Global Optimization for
Multiview Geometry
Sameer Agarwal1, Manmohan Krishna Chandraker1, Fredrik Kahl2,
David Kriegman1, and Serge Belongie1
1UniversityofCalifornia, San Diego, CA92093,USA
{sagarwal, mkchandraker, kriegman, sjb }@cs.ucsd.edu
2LundUniversity,Lund,Sweden
fredrik@maths.lth.se
Abstract. Thisp a p e rpresents a p ract ical methodfor■nding t he
provabl ygloball yopt imal sol ution to n umerousp roble msinp roject ive
geometryincludingmultiviewtriangulation, ca meraresect ioning an dho-
mographyestimation. Unlike t raditional methodswhichmayget t rappe d
inl o c a l minimaduet ot he non-convex nat ureo ft hese p roble ms, this
app roachprovides a t heoretical g uarantee of global opt imality.Thef o r-
mulation relies on recent develop ments infract ional p rogramm ing an d
thetheoryof convex underestimato rsa n dallowsauni■edframeworkf o r
minimizing t hes t a n dardL2-no rmofreproject ion e rrorswhichiso p t imal
underGaussian no ise as well as t hemorerobustL1-no rmw h ichisl e s s
sens itive to o utliers.Thee ■ c a c yof o uralgo rithm isempiricallydemon-
stratedbygoodperformance on expe riments fo rbothsynthetica n dreal
data. An open so urce MA TLABtoolbox t hatimplements t hea l g o rithm
isa l s o madea v a ilable to fac ilitate f urtherresea rch.
1
****
Perspective n-View Multibody
Structure-and-Motion Through Model Selection
Konrad Schindler, James U, and Hanzi Wang
Inst itute fo rVision S ystemsE n g inee ring,
Monas hUniversity,Clayton, 3800 VI C,Australia
{Konrad.Schindler, James.U, Hanzi.Wang }@eng.monash.edu.au
Abstract. Multi-bodystructure-an d-motion (MSaM) istheproble mto
establ ishthemultiple-v iewgeometryof an image seq uence of a 3Dscene,
wherethe scene cons ists of multiple rigidobjects moving relat ive to eac h

other.S of a r,s o l utions have been p ropose dforseve ralrestrictedsett ings, suchas onl ytwov iews, a■ne p roject ion, an dperspect ive p roject ion of linea rlymoving po ints. We g ive a sol ution fo rsequences of seve ralimages, full pe rspect ive p roject ion, an dgene ralrigidmotion. It can dealwiththe fact t hat t hes e to fc o rrespon dences c hanges ove rtime, an disrobust to outliers.Thep ropose dsolution isb a s e don Monte- Carlo sa mpling an d clustering of t wo-v iewmotions, l inking t hemthroughthes e q uence, an d model select ion to yieldtheb e s te x p l a n a t ion fo rthee n t ires e q uen ce.

1

****

Confocal Stereo

Dept. of ComputerScience, UniversityofToronto

Abstract. We p resentconfocal stereo ,an e wmethodforcomputing3D shape b ycont rolling t hef o c usa n dape rtureo fal e n s . Themethodis spec i■call ydesignedforreconst ructing scenes withh ighgeometricc o m- plex ityor■ne-scale text ure.Toa c hieve t his,weintroduce t heconfocal constancy prope rty,whichstates t hat as t hel e n sa p e rturev a ries, t he pixel intens ityof a v isible in-foc uss c e n ep o intwill va ryinas c e n e - independent way,that can be p redictedbypriorradiometricl e n sc a l i- bration. Theo n l yrequirement isthatincoming radiance withinthec o n e subten dedbythel a rgest ape rtureisn e a rlyconstant. F irst,wedevelop adeta iledlens model that facto rsoutthedistortions inhighresol ution SLRcameras (12MP o rmore)withlarge-ape rturel e n s e s(e . g . ,f 1.2). This allo wsus to asse mble an A×Fape rture-foc usimage (AFI) fo r eachpixel, t hat collects t heundistortedmeasurements ove rallAape r- tures an dFfocuss e t t ings. In t heA F I representat ion, confocal constanc y reduces to colo rcomparisons withinregions of t heA F I ,a n dleadst of o - cusmetrics that can be eval uatedsepa ratel yforeachpixel. We p ropose twosuchmetrics an dpresent initialreconst ruction results fo rcomplex scenes.

1

****