

Massimo Tistarelli  
Stan Z. Li  
Rama Chellappa  
*Editors*

# Handbook of Remote Biometrics

## for Surveillance and Security

 Springer

# Advances in Pattern Recognition

For further volumes:  
<http://www.springer.com/series/4205>

Massimo Tistarelli · Stan Z. Li · Rama Chellappa  
Editors

# Handbook of Remote Biometrics

for Surveillance and Security



*Editors*

Prof. Massimo Tistarelli  
Università Sassari  
Dipto. Architettura e  
Pianificazione  
Piazza Duomo, 6  
07041 Alghero SS  
Italy  
tista@uniss.it  
mtista@gmail.com

Dr. Stan Z. Li  
Chinese Academy of Science  
Institute of Automation  
Center Biometrics Research &  
Security  
Room 1227, Zhongguancun  
95  
100080 Beijing Donglu  
China, People's Republic  
szli@cbsr.ia.ac.cn  
Stan.ZQ.Li@gmail.com

Dr. Rama Chellappa  
University of Maryland  
Center for Automation  
Research  
College Park MD 20742-3275  
USA  
rama@cfar.umd.edu  
Rama@umiacs.umd.edu

ISBN 978-1-84882-384-6      e-ISBN 978-1-84882-385-3  
DOI 10.1007/978-1-84882-385-3  
Springer Dordrecht Heidelberg London New York

Advances in Pattern Recognition Series ISSN 1617-7916

British Library Cataloguing in Publication Data  
A catalogue record for this book is available from the British Library

Library of Congress Control Number: 2009926948

© Springer-Verlag London Limited 2009

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

The use of registered names, trademarks, etc., in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

*Cover design:* Franciscu Gabriele Mulas

Printed on acid-free paper

Springer is part of Springer Science+Business Media ([www.springer.com](http://www.springer.com))

## **Foreword**

The development of technologies for the identification of individuals has driven the interest and curiosity of many people. Spearheaded and inspired by the Bertillon coding system for the classification of humans based on physical measurements, scientists and engineers have been trying to invent new devices and classification systems to capture the human identity from its body measurements. One of the main limitations of the precursors of today's biometrics, which is still present in the vast majority of the existing biometric systems, has been the need to keep the device in close contact with the subject to capture the biometric measurements. This clearly limits the applicability and convenience of biometric systems. This book presents an important step in addressing this limitation by describing a number of methodologies to capture meaningful biometric information from a distance. Most materials covered in this book have been presented at the International Summer School on Biometrics which is held every year in Alghero, Italy and which has become a flagship activity of the IAPR Technical Committee on Biometrics (IAPR TC4). The last four chapters of the book are derived from some of the best presentations by the participating students of the school. The educational value of this book is also highlighted by the number of proposed exercises and questions which will help the reader to better understand the proposed topics. This book is edited by Professors Massimo Tistarelli, Rama Chellappa, and Stan Li who are well known in the biometrics community. Much care has been devoted not only to broadly cover the biometrics subjects, but also to present an in-depth analysis of advanced techniques for acquiring and processing biometric data from a distance. Given its unique features this book will be a valuable addition to the existing biometrics literature and also a useful reference to all researchers in biometrics and pattern recognition. The chapters forming this book will certainly facilitate the advancement of research in remote biometrics (or human identification from a distance).

Beijing, China

Tieniu Tan

## Preface

Biometric is a multidimensional problem which is aimed at understanding the uniqueness of human beings to facilitate recognition. In reality, the dimensionality not only comes from the data, as engineers and computer scientists would think of, but more from the intrinsic nature of the recognition process. We expect to be able to recognize people without errors (D1), at a distance of 1–20 meters (D2), in motion (D3), and even if unaware (D4). Obviously today's commercial systems cannot cope with all four dimensions, and are never error free. Nonetheless, we believe that, in order to design 4D biometric systems, more efforts are required in the direction of dimensions D2, D3, and D4. For these reasons, this book covers several aspects of biometrics from the perspective of recognizing individuals at a distance, in motion, and under a surveillance scenario.

While this trend deserves growing attention, a strong impact is expected in many applications of biometrics, including border control, surveillance in critical infrastructures, and ambient intelligence. This book presents a wide and in-depth view of the most advanced biometric technologies for recognition at a distance.

A multiview approach is presented to the reader, with each chapter being designed to cover a different biometric subject written by authors from different research institutions, and the objective of covering the subject from different perspectives. Current existing and under preparation international standards in biometrics are also presented.

Mostly biometric techniques which do not require a close contact with the user are considered.

This comprehensive, innovative, state-of-the-art volume is designed to form and inform professionals, young researchers, and graduate students about the most advanced biometric technologies.

Alghero, Italy  
College Park, MD, USA  
Beijing, China

*Massimo Tistarelli*  
*Rama Chellappa*  
*Stan Li*

# Contents

## Part I Advances in Remote Biometrics

<b>1 Biometrics at a Distance: Issues, Challenges, and Prospects .....</b>	3
Stan Z. Li, Ben Schouten, and Massimo Tistarelli	
<b>2 Iris Recognition – Beyond One Meter .....</b>	23
James R. Matey and Lauren R. Kennell	
<b>3 View Invariant Gait Recognition .....</b>	61
Richard D. Seely, Michela Goffredo, John N. Carter, and Mark S. Nixon	
<b>4 Advanced Technologies for Touchless Fingerprint Recognition .....</b>	83
Giuseppe Parziale and Yi Chen	
<b>5 Face Recognition in Humans and Machines .....</b>	111
Alice O'Toole and Massimo Tistarelli	
<b>6 Face Recognition at a Distance: System Issues .....</b>	155
Meng Ao, Dong Yi, Zhen Lei, and Stan Z. Li	
<b>7 Long-Range Facial Image Acquisition and Quality .....</b>	169
Terrance E. Boult and Walter Scheirer	
<b>8 A Review of Video-Based Face Recognition Algorithms .....</b>	193
Rama Chellappa, Manuele Bicego, and Pavan Turaga	
<b>9 3D Face Recognition: Technology and Applications .....</b>	217
Berk Gökberk, Albert Ali Salah, Neşe Alyüz, and Lale Akarun	
<b>10 Machine Learning Techniques for Biometrics .....</b>	247
Francesca Odone, Massimiliano Pontil, and Alessandro Verri	

<b>11 Multibiometric Systems: Overview, Case Studies, and Open Issues</b>	.. 273
Arun Ross and Norman Poh	
<b>12 Ethics and Policy of Biometrics</b>	..... 293
Emilio Mordini	

**Part II Selected Contributions from Students of the International Summer School on Biometrics**

<b>13 Assessment of a Footstep Biometric Verification System</b>	..... 313
Rubén Vera Rodríguez, John S.D. Mason, and Nicholas W.D. Evans	
<b>14 Keystroke Dynamics-Based Credential Hardening Systems</b>	..... 329
Nick Bartlow and Bojan Cukic	
<b>15 Detection of Singularities in Fingerprint Images Using Linear Phase Portraits</b>	..... 349
Surinder Ram, Horst Bischof, and Josef Birchbauer	
<b>16 Frequency-Based Fingerprint Recognition</b>	..... 363
Gualberto Aguilar, Gabriel Sánchez, Karina Toscano, and Héctor Pérez	
<b>Index</b>	..... 375

## Contributors

**Gualberto Aguilar** SEPI ESIME Culhuacan, Instituto Politécnico Nacional, Av. Santa Ana #1000, México D.F. gualberto@calmecac.esimecu.ipn.mx

**Lale Akarun** Computer Engineering Department of Bebek, Boğaziçi University, TR-34342, Istanbul, Turkey, akarun@boun.edu.tr

**Neşe Alyüz** Computer Engineering Department of Bebek, Boğaziçi University, TR-34342, Istanbul, Turkey, nese.alyuz@boun.edu.tr

**Meng Ao** Center for Biometrics and Security Research and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Science, Beijing 100190, China, mao@cbsr.ia.ac.cn

**Nick Bartlow** West Virginia University, Morgantown, WV, USA, nick.barlow@mail.wvu.edu

**Manuele Bicego** University of Verona, Verona, Italy, bicego@uniss.it

**Josef Birchbauer** Siemens Austria, Siemens IT Solutions and Services, Biometrics Center, Graz, Austria, josef-alois.birchbauer@siemens.com

**Horst Bischof** Institute for Computer Graphics and Vision, Technical University of Graz, Graz, Austria, bischof@icg.tugraz.at

**Terrance E. Boult** Vision and Security Technology Lab, Department of Computer Science, University of Colorado, 1420 Austin Bluffs Parkway, Colorado Springs CO 80933-7150, USA, tboult@vast.uccs.edu

**John N. Carter** University of Southampton, Southampton, UK, jnc@ecs.soton.ac.uk

**Rama Chellappa** Department of Electrical and Computer Engineering and Center for Automation Research, UMIACS University of Maryland, College Park Maryland 20742, USA, rama@umiacs.umd.edu

**Yi Chen** Digital Persona Inc., Redwood City, CA 94063, USA, yic@digitalpersona.com

**Bojan Cukic** West Virginia University, Morgantown, WV, USA,  
bojan.cukic@mail.wvu.edu

**Nicholas W.D. Evans** Institut Eurécom, 2229 route des Crêtes, 06904  
Sophia-Antipolis, France, nicholas.evans@eurecom.fr

**Michela Goffredo** University of Southampton, Southampton, UK,  
mg2@ecs.soton.ac.uk

**Berk Gökberk** Department of Electrical Engineering, Mathematics and  
Computer Science, University of Twente, Enschede, The Netherlands,  
b.gokberk@ewi.utwente.nl

**Lauren R. Kennell** Biometric Signal Processing Laboratory, Electrical &  
Computer Engineering Department US Naval Academy – Maury Hall, Annapolis  
MD 21402-5025, USA, Kennel@usna.edu

**Zhen Lei** Center for Biometrics and Security Research and National Laboratory  
of Pattern Recognition, Institute of Automation, Chinese Academy of Science,  
Beijing 100190, China, zlei@cbsr.ia.ac.cn

**Stan Z. Li** Center for Biometrics and Security Research and National Laboratory  
of Pattern Recognition, Institute of Automation, Chinese Academy of Science,  
Beijing 100190, China, szli@cbsr.ia.ac.cn

**John S.D. Mason** Swansea University, Singleton Park, Swansea, SA2 8PP, UK,  
j.s.d.mason@swansea.ac.uk

**James R. Matey** Biometric Signal Processing Laboratory, Electrical & Computer  
Engineering Department US Naval Academy – Maury Hall, Annapolis MD  
21402-5025, USA, matey@usna.edu

**Emilio Mordini** Centre for Science, Society and Citizenship, Piazza Capo di Ferro  
23, 00186 Rome, Italy, emilio.mordini@cssc.eu

**Mark S. Nixon** University of Southampton, Southampton, UK,  
msn@ecs.soton.ac.uk

**Alice O'Toole** University of Texas, Dallas TX, USA, otoole@udallas.edu

**Francesca Odone** DISI Università degli Studi di Genova, Via Dodecaneso 35,  
16146 Genova, Italy odone@disi.unige.it

**Giuseppe Parziale** iNVASIVE CODE., Barcelona, Spain,  
geppy.parziale@invasivecode.com

**Héctor Pérez** SEPI ESIME Culhuacan, Instituto Politécnico Nacional, Av. Santa  
Ana #1000, México D.F. hmpm@prodigy.net.mx

**Norman Poh** University of Surrey, Guildford, GU2 7XH, Surrey, UK,  
normanpoh@ieee.org

**Massimiliano Pontil** Department of Computer Science, University College London Malet Place London WC1E 6BT, UK m.pontil@cs.ucl.ac.uk

**Surinder Ram** Institute for Computer Graphics and Vision, Technical University of Graz, Graz, Austria, ram@icg.tugraz.at

**Rubén Vera Rodríguez** Swansea University, Singleton Park, Swansea, SA2 8PP, UK, r.vera-rodriguez.405831@swansea.ac.uk

**Arun Ross** West Virginia University, Morgantown, West Virginia, USA, arun.ross@mail.wvu.edu

**Albert Ali Salah** Center for Mathematics and Computer Science (CWI), Amsterdam, The Netherlands, a.a.salah@cwi.nl

**Gabriel Sánchez** SEPI ESIME Culhuacan, Instituto Politécnico Nacional, Av. Santa Ana #1000, México D.F. caaann@gmail.com

**Walter Scheirer** Securics, Inc, Colorado Springs, CO, USA, wjs3@vast.uccs.edu

**Ben Schouten** Fontys University, Amsterdam, The Netherlands, ben.schouten@fontys.nl

**Richard D. Seely** University of Southampton, Southampton, UK, rds06r@ecs.soton.ac.uk

**Massimo Tistarelli** Computer Vision Laboratory, University of Sassari, Italy, tista@uniss.it

**Karina Toscano** SEPI ESIME Culhuacan, Instituto Politécnico Nacional, Av. Santa Ana #1000, México D.F., toscano@calmecac.esimecu.ipn.mx

**Pavan Turaga** Department of Electrical and Computer Engineering and Center for Automation Research, UMIACS University of Maryland, College Park Maryland 20742, USA, pturaga@cfar.umd.edu

**Alessandro Verri** DISI Università degli Studi di Genova, Via Dodecaneso 35, 16146 Genova, Italy, verri@disi.unige.it

**Dong Yi** Center for Biometrics and Security Research and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Science, Beijing 100190, China, dyi@cbsr.ia.ac.cn

**Part I**  
**Advances in Remote Biometrics**

# **Chapter 1**

## **Biometrics at a Distance: Issues, Challenges, and Prospects**

**Stan Z. Li, Ben Schouten, and Massimo Tistarelli**

**Abstract** The notion of *remote biometrics* or *biometrics at a distance* is today of paramount importance to provide a secure mean for user-friendly identification and surveillance.

In 2007 the BioSecure Network of Excellence published the Biometric Research Agenda (Schouten et al. BioSecure: white paper for research in biometrics beyond BioSecure. CWI report 2008). This network identified, as one of the most urgent topics, research in distributed (intelligent) sensor networks and the transparent use of biometrics, requiring no actions from the end-user in supervised or unsupervised ways. These citizens, sometimes in ways that are yet to be understood.

This chapter introduces the most relevant issues and challenges of biometric sensing and recognition at a distance. Most of these issues are more be deeply analyzed in the subsequent chapters of this book.

Several categorizations of biometric sensing and applications are provided including face recognition and novel biometric sensors for building intelligent environments. Several factors affecting the performances of biometric systems are identified and discussed in some detail. Some typical applications and prospects are illustrated together with their impact in today's and tomorrow's society and citizens.

### **1.1 Introduction**

Biometrics, as a mean for automatic and reliable identify recognition, plays an important role in surveillance and security, such as in access control, e-passport, and watch-list surveillance. Recent years has seen significant increase in biometric recognition applications, partly due to recent technology advances, and partly due to increased demands for security and human cost-saving reasons.

Biometric recognition systems automatically verify or identify person identity present in the input images and videos using human biometric traits. Human

---

S.Z. Li (✉)

Center for Biometrics and Security Research and National Laboratory of Pattern Recognition,  
Institute of Automation, Chinese Academy of Sciences, Beijing, China  
e-mail: szli@cbsr.ia.ac.cn

biometric traits which can be used for biometric recognition include face, iris, fingerprint, palmprint, and others. The primary biometric trait for biometrics at a distance is the face because the face is most accessible and natural biometric trait for recognition at a distance. Other useful traits are iris and gait. Multimodal biometrics fuse several biometric modules to reach more reliable recognition result.

Biometric recognition is done by extracting a biometric template in query from the input device against those of target enrolled in the database. The comparison of a query against the target is performed in one of the two modes: (1) verification (or authentication) and (2) identification (or recognition). Verification is one-to-one in that the query face is compared against the claimant's face images to verify the claimed ID. This is the case of boarder control with e-passport where the ID is claimed by the claimant. Identification is one-to-many in that the query is compared against those enrolled in the database to determine the identity of the query. Another one-to-many scenario is watch-list surveillance, where only the found matches that are confident enough (above a preset threshold) should be shown to the system operator.

Most commercial face recognition products and solutions are developed for cooperative user applications, including access control, e-passport, and national registration where a user is required to cooperate with the camera to have his/her face image captured properly, in order to be granted for the access. Less restrictive scenarios are non-cooperative user applications, such as face recognition under surveillance where person identification is done without user's intentional, cooperative effort. Pilot deployments of face recognition are also under way, such as watch-list face surveillance in subways.

Homogeneity referred to the similarity between properties of input face images and those of enrollment face images. It can be homogeneous, where both input face and enrollment face are from the same type of imaging device (e.g., camera, photo scanner) in a similar environment (e.g., lighting). On the other hand, it is heterogeneous where the input face and the enrollment face are from different imaging devices (e.g., image from system's camera vs. photo scan).

Biometric applications are categorized in terms of several key factors. Challenges in face recognition will be discussed and advanced technology will be described. Solutions for reliable face recognition will be provided for typical applications such as access control, e-passport, large face database search, and watch-list surveillance.

Table 1.1 provides a categorization of biometrics with respect to type of comparison, user cooperation, and homogeneity of input and enrollment face images. From the technical viewpoint, the easiest is one-to-one face verification such as

**Table 1.1** Categorization of biometric applications

Application	Comparison	User cooperation	Enrollment image
Access control	1:1 or 1:N	Coop	Photo, video
E-passport	1:1	Coop	Photo
Large database search	1:1 or 1:N	Coop, Non-coop	Photo, video
Watch-list surveillance	1:N	Non-coop	Photo, video

used in e-passport whereas the most challenging is one-to-many face identification in watch-list surveillance.

In the reminder of this chapter, we describe technologies for biometrics at a distance (Sections 1.2 and 1.3). Challenges in face recognition are discussed. We then provide solutions for various face recognition applications such as access control, e-passport, large face database search, and watch-list surveillance (Section 1.4). Issues in privacy and management are discussed in Section 1.5.

## 1.2 Biometric Technologies

Even though an abundance of biometric sensing technologies exist, not all are equally suited for all applications. In particular, the required level of cooperation required from the user strongly constraints the applicability of these devices in some operational environments.

Today's most common biometric sensing modalities fall into one of three categories: contact, contactless, and at a distance. The difference among these three categories is the required distance between the sensor and the user for effectively sample the biometric data.

Contact devices require the user to actually touch the biometric sensor. Some typical contact sensors are fingerprint, palm, and signature. These devices can hardly be hidden and require an active cooperation of the user [1–3].

Contactless devices are all devices which do not require the user to physically be in contact with the sensor. Nonetheless, this category includes all sensors which require a short distance, generally in the order of 1 cm to 1 m, for obtaining a good sample of the biometric data. Iris capturing devices and touchless fingerprint sensors are among the most common devices of this kind. Even though the user can keep a distance from the sensor, still an active cooperation is required. Also some kind of face recognition systems can be included in this category because the user must be close to the camera to be recognized [4].

Biometric systems capable of acquiring biometric data at a distance usually do not require the active cooperation of the user. Gait recognition [5], some face recognition systems [6, 7], and the most recently developed iris recognition systems [8] fall in this category. It is worth noting that the important aspect in this class of devices is the loose requirement of the active cooperation of the user. For this reason, this class of biometric devices is very well suited to integrate identity recognition and surveillance tasks.

Multibiometric systems generally encompass a mixture of contact and contactless devices. The reason is that multimodal sensing is generally applied to reduce the false acceptance rate thus improving the recognition performances. In other cases, multiple modalities are exploited to facilitate the data acquisition for a wider population. A notable example of multibiometric system deployment is the US-Visit program which exploits face and fingerprint data to determine the identity of travelers at the US borders.

Whenever the users are not cooperative or a high throughput is desired, a multimodal system where all modalities are contactless or at a distance is preferable.

For example, a network of cameras can be used to acquire images of persons walking through a hallway and process their faces, iris, and walking dynamics at the same time and from different vantage points. Also in this case, by exploiting several modalities (namely gate, face, and iris biometrics) and fusing them either at feature, score, or decision level, can greatly improve the performances of the identification system, still allowing for a passive or non-cooperative behavior of the users.

The remainder of this book will consider both devices classified as contactless and “at a distance,” but mainly with attention to technologies which allow to capture and process biometric data for identification at a distance.

### **1.2.1 Technology Challenges**

Biometrics evaluation reports (e.g., FERET and FRGC) and other independent studies indicate that the performance of many state-of-the-art face recognition methods deteriorates with changes in lighting, pose, and other factors. We identify four types of factors that affect system performance: (1) technology, (2) environment, (3) user, and (4) user–system interaction, summarized in Table 1.2.

**Table 1.2** Factors affecting biometric system performance

Aspect	Factors
Technology	Dealing with face image quality, heterogeneous face images, and problems below
Environment	Lighting (indoor, outdoor)
User	Expression, facial hair, facial ware, aging
User–system	Pose (alignment between camera axis and facing direction), height

To achieve a desired level of reliability and accuracy, face recognition should be performed based on intrinsic properties of the face only, mainly, 3D shape and reflectance of the facial surface. Extrinsic properties, including hairstyle, expression, posture, and environmental lighting, should be minimized.

Environmental lighting may or may not be controllable depending on the operation. The ideal case would be that both input and enrollment face images are subject to the same lighting conditions, including, lighting direction and lighting intensity. The problem of uncontrolled environmental lighting is the first challenge to overcome from face-based biometric applications. Most commercial face recognition systems have a common problem that the accuracy drops when lighting conditions of input and enrollment are different.

Regarding the user factors, most existing systems allow a limited degree of variations in expression, facial hair, and facial ware. Also these properties can be controlled in cooperative user applications. For example, the user can be advised not to play exaggerated expressions, or not to wear sunglasses. However, facial aging leads to changes in intrinsic properties (shape and skin reflectance) that progress as people age. Face recognition under significant aging of facial appearance is an unsolved problem.

Regarding user–system coordination, the user is required to face the camera to enable acquisition of frontal face images, which is a reasonable restriction in

cooperative applications. Height is a difficulty in using face biometric. While fingerprint biometric is easier to cooperate by adjusting the arm and finger, the face has to be in proper height so that the image can be properly captured.

Regarding system factors, research and development efforts have mostly been spent in the development of core technologies, aimed to solve technology challenges. Long-term effort will be invested in making reliable face technology.

### 1.3 Biometric Sensing from a Distance

The NSTC Subcommittee on Biometrics published in 2006 is a vision on the future called the Biometrics Challenge [9]. While nearly all of the deployments of biometrics are government-led and concerned with national security and border control scenarios it is now apparent that the widespread availability of biometrics in everyday life will also spin out an ever-increasing number of private applications in domain beyond national security concerns. The subcommittee identified the research on sensor technology as a primary challenge for biometrics from a distance. One year later, in 2007 the BioSecure Network of Excellence [10] published the Biometric Research Agenda [11]. This network identified research in distributed (intelligent) sensor networks and the transparent use of biometrics as one of the most urgent topics, requiring no actions from the end user in supervised or unsupervised ways. These biometric technologies are likely to have a rapidly increasing impact in the life of citizens, sometimes in ways that are yet to be understood.

#### 1.3.1 State of the Art

New sensors have been introduced in smart environments and lately biometrics, capable to detect (1) physical properties like pressure and temperature, (2) motion properties, (3) contact properties, and (4) presence like RFID and IR. An overview on new sensors can be found in [12] (see Fig. 1.1).

With the availability and advances of this new sensor technology and the improved network capabilities there is a growing interest in intelligent distributed sensor



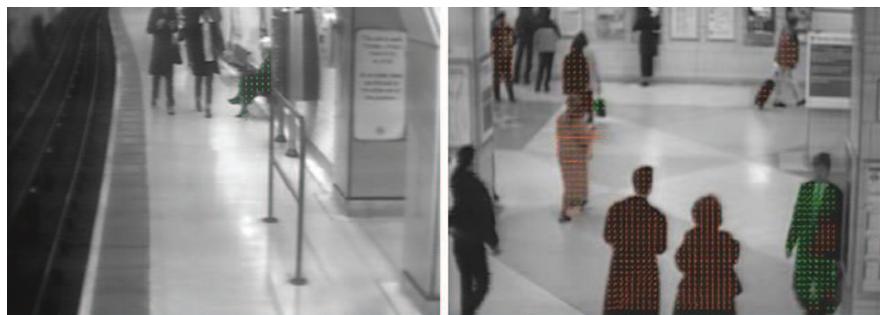
**Fig. 1.1** Some “phidget” sensors. From *left* to *right*: vibration, rotation, and humidity. See [www.phidgets.com](http://www.phidgets.com)

networks. The use of CCTV camera networks for security is an example of advanced computer vision at a distance using distributed sensor networks. According to the latest studies [13], Britain has around 4.2 million CCTV cameras – one for every 14 people in the country – and 20% of cameras globally. The applications range from (a) tracking and tracing, (b) scene modeling, (c) behavior analysis and detection of unusual behavior, to (d) the detection of specific alarm events [14]. The use of biometrics at a distance adds an important functionality to these scenarios, enabling identity management, e.g., black-list watching for shoplifters. Besides advanced security applications, future applications range from health care to multimedia and smart environments [15, 16]. Functionalities range from context awareness to identify and recognize events and actions, ubiquitous access (to information, physical locations, etc.) and natural human–computer interaction [17].

The use of camera networks for security is studied in several (funded by the European Union) research projects like PRISMATICA [18] and VITAB [19] related to technologies that will improve the effectiveness of CCTV control room operations, or that enable identification and authentication capability to improve on the protection of buildings and other commercial and public assets (see Fig. 1.2).

Other projects are targeted toward ambient intelligence. Functionalities include (among others) adaptive houses that automatically adjust lighting and temperature settings to achieve optimal user satisfaction at minimal energy costs [20, 21], smart meeting rooms that provide latecomers with a summary of the arguments so far [22], and smart beds that unobtrusively monitor the sleep pattern, heart, and breathing rate of seniors.

Apart from authenticating the user, research focuses on the recognition of behavior or actions of a subject or group of subjects. Examples can be found in [22] (AMIGO), Cognitive Home Companions [23] (IST COGNIRON), Context Aware Vision for Surveillance and Monitoring [24] (IST CAVIAR), Extraction of Information for Broadcast Video [25] (IST DETECT), Multimodal Interaction and Communications [26] (IST FAME), and Multimodal Services for Meetings and Communications [27] (IST CHIL), as well as MIT’s Project Oxygen [28], MERL’s Ambient Intelligence for Better Buildings [29], and Georgia Tech Aware Home [30].



**Fig. 1.2** Incident detection at the London subway. Courtesy of Prismaticica [18]

### 1.3.2 Performance

Advanced sensor systems can be used to explore the correlations of the biometric traits over time and place (spatio-temporal correlations) in most cases using audio and video. State-of-the-art methods in tracking humans in camera networks can be grouped according to single-camera or multicamera usage [31]. In multiview approaches, several methods are deployed using color information, blob information, or occupancy maps. When the scene is not crowded, simple background-foreground separation in combination with color features can do the detection and tracking of the humans [32, 33]. In more crowded environments, Haritaoglu et al. use vertical projection of the blob to help segment a big blob onto multiple humans [34]. Blob information and trajectory prediction based on Kalman filtering for occluded objects is used in [35].

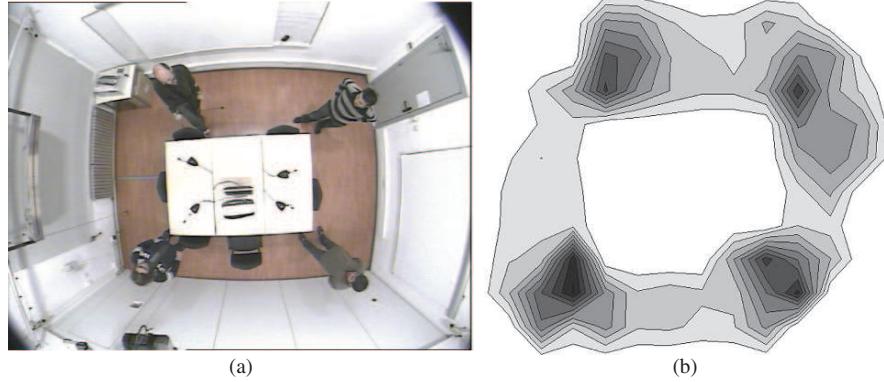
Zhou [6] applies *still to video* as well as *video to video* comparison in order to recognize humans from video footings. In [36] tracked subjects are identified during the process using face recognition and speech recognition. Additionally, color features are used to support the tracking process and to robustify the biometric recognition.

In general, biometrical recognition techniques at a distance are less robust as a simple consequence of occlusions, movements of objects and subjects, lightning, noise or simple smaller templates. However, enabling more modalities, fusion and cross correlations in time and space improve the performance of such systems. Several strategies can be applied to improve the performance of biometrics at a distance, which will be discussed in the next sections.

#### 1.3.2.1 Fusion and Multibiometrics

Multibiometrics, consolidating the evidence by multiple biometric sources can be used in parallel or serial and at different levels of the authentication process [37]. In parallel architectures, the evidence acquired from multiple sources is simultaneously processed in order to authenticate an identity [1]. A very promising candidate for solving the task of fusion is using a Bayesian network (BN) or an extension to it [38]. Using BN, one can cope with uncertain inputs. Both prior knowledge, in the form of beliefs and statistics could be incorporated in a BN (or its superset). Also Kalman filters and HMM approaches, used widely and successfully for coping with noisy inputs and tracking, respectively, could be expressed in terms of BNs. An important issue is the evaluation of the statistical correlation of the inputs [23, 36]. (Fig. 1.3)

The quality of biometric samples used by multimodal biometric experts to produce matching scores has a significant impact on their fusion. This quality can be the resultant of many factors like noise, lightning conditions, background and distance [39, 40]. The quality of biometric samples used by multimodal biometric experts to produce matching scores has a significant impact on their fusion. In [41] the problem of quality-controlled fusion of multiple biometric experts is used; they focus on the fusion problem in a scenario where biometric trait quality expressed



**Fig. 1.3** Probabilistic localization of subjects using motion and color [36]. Darker colors indicate higher probability

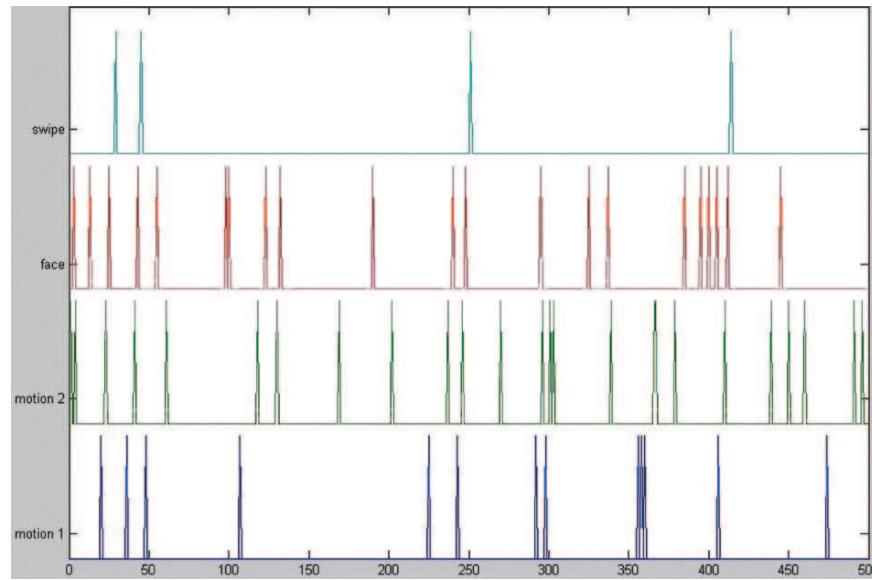
in terms of quality measures can be coarsely quantized and developed a decision fusion methodology based on fixed rules (sum and product rules). Other ways of quality-based fusion can be found in [38, 42, 43]

In the study of Jain et al. [44], the aim was to examine whether using soft biometrics, i.e., easily measurable personal characteristics, such as weight and fat percentage, can improve the performance of biometrics in verification type applications. Fusing fingerprint biometrics with soft biometrics, in this case body weight measurements, decreased the total error rate (TER) from 3.9 to 1.5% in an experiment with 62 test subjects. The result showed that simple physiological measurements can be used to support biometric recognition. Furthermore, soft biometrics are unobtrusive, there is no risk of identity theft, the perception of the big-brother effect is small, the equipment needed is low cost, and the methods are easy to understand. The use of soft biometrics may be found in non-security, convenience type cases, such as domestic applications. In certain environments with a small number of subjects, like the home environment for instance, these features are robust enough to perform the recognition process or at least capable of determining partial identity classifications like gender and age, enhancing forms of anonymous identity management (e.g., alcohol control).

Soft biometrics traits can also be used to support the authentication process. In [45] the color of the shirt a subject is wearing is identified with the face and used for authentication when face recognition is not available or robust enough.

### 1.3.2.2 Spatio-temporal Correlation Between Sensors

With the availability of sensors and their reducing cost and sophistication increasing rapidly, it has become viable to deploy sensors in a networked environment used for observation. Within an observed environment it is likely that comparing the data streams emanating from sensor network will uncover meaningful association. Such associations might in fact have a significant practical value as they can contribute



**Fig. 1.4** Patterns created by a subject using four different modalities: face, motion, and RFID [45]

to the robustness of identification process. Finding temporal patterns in multiple streams of sensor data is essential for automatic analysis of human behaviors and habits in the ambient environment [17]. Four methods can be distinguished to detect temporal patterns in sensor data streams: (a) T patterns based on searching recurring temporal patterns [46], (b) clustering time series generated by low-resolution sensors using HMM [47], (c) eigenbehaviors based on the most prominent eigenfactors of an activity matrix [48], and (d) the use of the LZW compression algorithm as a pattern extractor [49]. (Fig. 1.4)

## 1.4 Applications and Deployment of Biometrics

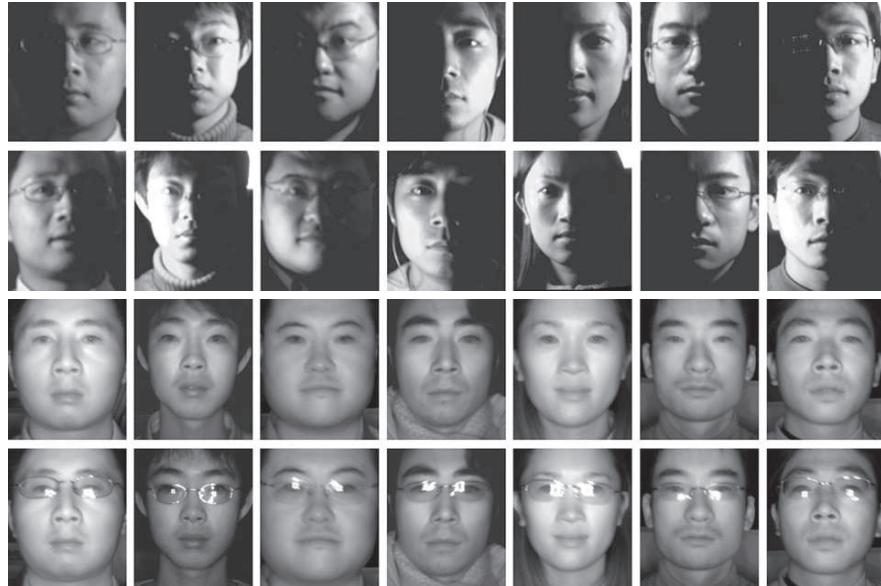
Progress has been made recently in face recognition using visible light (VL) face images. Several companies today provide solutions for personal identification under different environmental conditions. The Center for Biometrics and Security Research at the Chinese Academy of Science developed several advanced solutions for personal identification and verification from faces. The developed technology can reliably solve the problem of comparing face images taken from conventional CCTV cameras to photos and photo scans. This technology has been successfully applied for identity verification in Beijing Olympics ticket checking, where every ticket is associated with the real identify of the ticket holder, watch-list identification in CCTV surveillance in subways, in an effort to deter terrorists.

### 1.4.1 Near Infrared (NIR) Face Recognition

Most existing methods have been attempting to perform face recognition with images captured using conventional cameras working in visible light spectrum. A CCD or CMOS camera is used to capture a color or black/white face image. In such systems, lighting variation is the primary problem to deal with.

Several solutions have been investigated. Most effort is to develop face recognition algorithms that are stable under variable lighting. The effectiveness varies significantly from system to system. Another technique is to use 3D (in many cases, 2.5D) data obtained from a laser range scanner or 3D vision method. The disadvantages are the increased cost and slowed speed as well as the artifacts due to speculation. More importantly, it is shown that the 3D method may not necessarily produce better recognition results: recognition performances achieved by using a single 2D image and by a single 3D image are similar.

The use of near infrared (NIR) imaging brings a new dimension for applications of invisible lights for face detection and recognition [50]. It not only provides appropriate active frontal lighting but also minimizes lightings from other sources. Figure 1.5 compares NIR face images with visible light (VL) face images. The VL images contain large performance-deteriorating lighting changes whereas the NIR are good for face recognition. The fixed lighting direction much simplifies the problem of face recognition. The technology has then been successfully used



**Fig. 1.5** Comparison of visible light and near infrared images captured in the same environmental lighting conditions (*top*). Set of face images acquired at visible light (*bottom*). Set of face images acquired at near infrared lighting

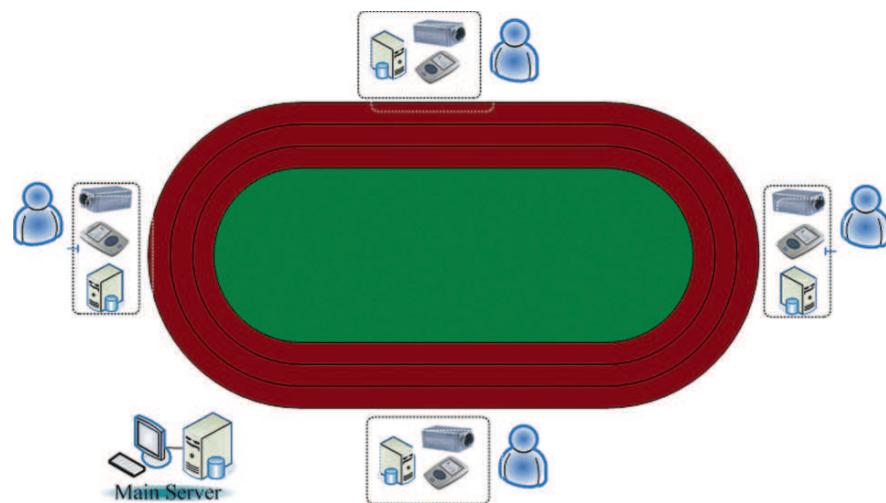
for access control and time attendance in enterprises, bank, law departments, and military as well as for self-service biometric immigration control at Shenzhen–Hong Kong border.

### 1.4.2 Access Control Systems

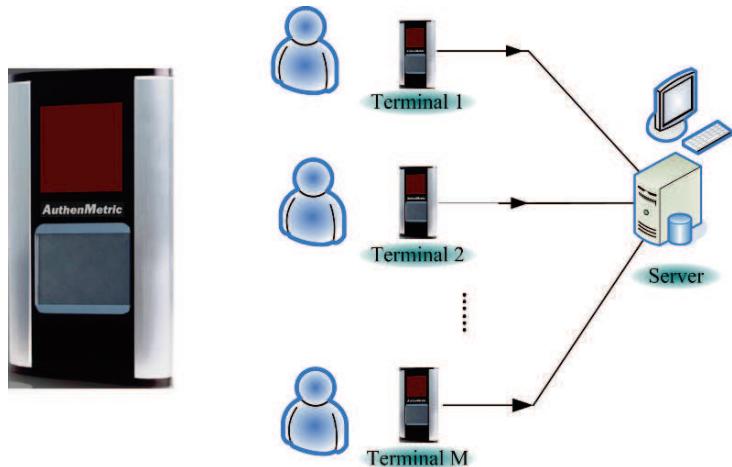
Access control is a typical example of many cooperative user applications. It can operate in either one-to-one or one-to-many mode. The one-to-one mode is generally associated with an ID card or keyboard ID input.

Access control *with* ID card uses one-to-one comparison for user verification. The VL solution has to be used if VL face images, either from photo or online camera, are required for the enrollment. The NIR-based solution gives better accuracy and speed than VL-based one if enrollment with online NIR images is allowed [50]. Figure 1.6 shows a solution for audience verification for access control at sport stadium. The terminal at each gate consists of an RFID ticket reader, a camera, and a PC processor. It uses one-to-one verification, by comparing the face image from a CCTV camera with the registered photo scan image.

Access control *without* ID card uses one-to-many comparison for user identification. A highly accurate solution is needed for one-to-many comparison. VL-based solution has not been reliable enough for that. The NIR-based solution is particularly suited for one-to-many access control in terms of accuracy. The following shows a solution for NIR-based access control, without the need for ID card. (Fig. 1.7)



**Fig. 1.6** Access control system to monitor the audience at a sport stadium



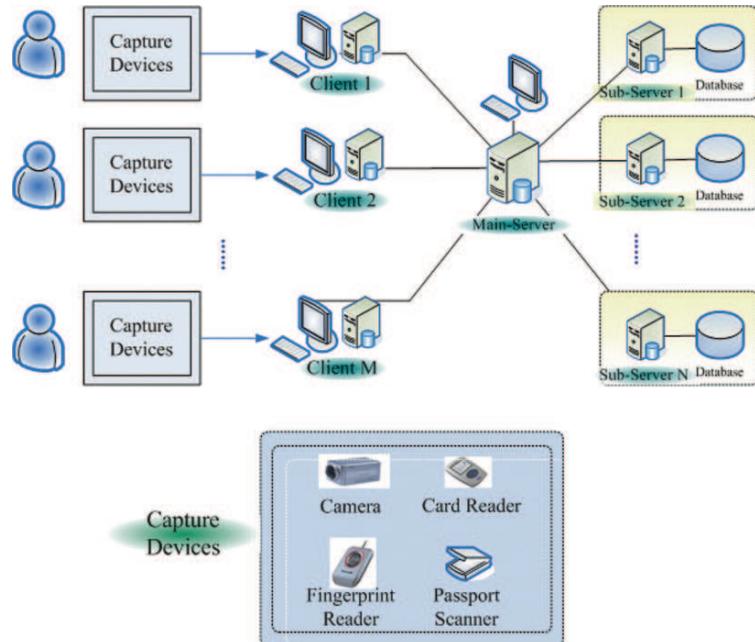
**Fig. 1.7** (left) NIR face terminal and (right) solution for one-to-many identification in access control

### 1.4.3 E-passports

The e-passport application is a special case of access control with ID card. The main difference between e-passport and access control with card is in the ID card. Special requirements are imposed on e-passport systems in RFID card reader, passport OCR scanner, and enrollment face images, by the related ICAO and ISO standards [51]. The RFID card used in an e-passport is a mass-storage card. Like other smartcards, the passport book design calls for an embedded contactless chip that is able to hold digital signature data to ensure the integrity of the passport and the biometric data. The biometric features of the user are stored in the card in the format of BioAPI. For privacy concerns, the format for storing fingerprints on a document shall be ISO/IEC standard compliant [52]. However, existing practice is to display photographs on travelers' identity documents, there should not be any privacy issues with using face images. The storage format for face images follows reference [53]. Figure 1.8 shows solutions for e-passport biometric authentication of seafarers. Australia's Smartgate and the China's Shenzhen–Hong Kong and Zhuhai–Macau border control are pioneer deployments of similar applications.

### 1.4.4 Watch-List Identification

Watch-list identification is an “open-set” recognition task, in that not all individuals in the surveillance video are in the watch list. The presence of irrelevant individuals should not bring alerts as far as possible (minimum false alarm). This has so far been most challenging task for practical face recognition. Such a system should satisfy the following design principles:

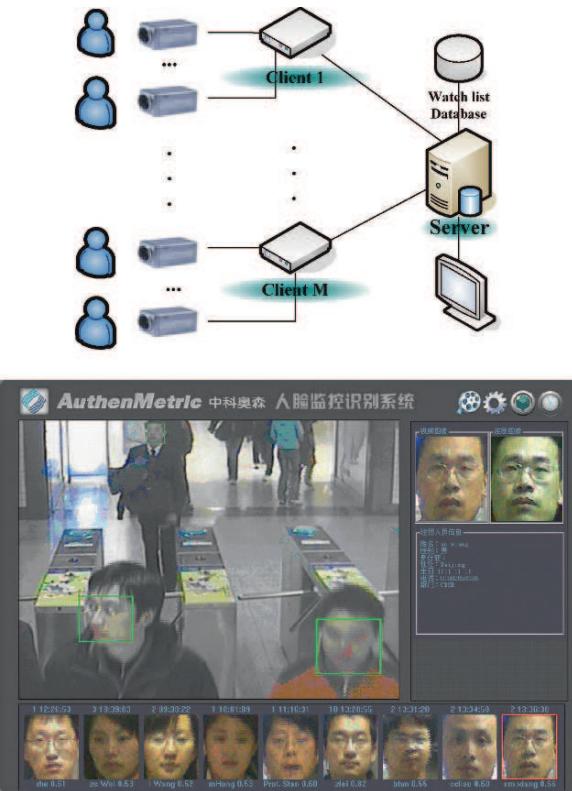


**Fig. 1.8** Deployment of the e-passport for multibiometric authentication at airport

- **NIR vs. VL.** Since the NIR-based solution is limited to a short range (smaller than 2 m), we have to resort to a VL-based solution.
- **Image and face quality.** Video images should be sharp and in good contrast. Sufficient image resolution is required since faces in surveillance video often occupy small areas in the frame. We recommend 50 pixels as the minimum inter-pupil spacing to achieve good identification result though the system can work with as small as 30 pixels.
- **Accuracy and reliability of face engine.** Surveillance cameras are often fixed at ceilings and people are non-cooperative to the video capturing. The system should be tolerant to head pose, for example, of up to  $+/- 15^\circ$  deviation from front pose. Other difficulties include partial face occlusions, beards and hairstyle changes, wearing glasses (except dark sunglasses) as well as lighting changes.
- **System speed.** For identifying people in walk at a distance, such a system should be able to detect at least 30 faces and recognize 10 faces per second.

The solution shown in Figure 1.9 supports the client/server architecture. The server station maintains the watch-list database, receives facial feature templates from the clients, performs the face identification, and provides GUI to the operator to examine the results. The server station can support several client stations. Each client station can process 1–8 video channels depending on the processor. It captures video frames, detects faces in each frame, saves frames with faces, extract facial templates, and send them to the server station.

**Fig. 1.9** Topology and user interface for a watch-list surveillance and identification system



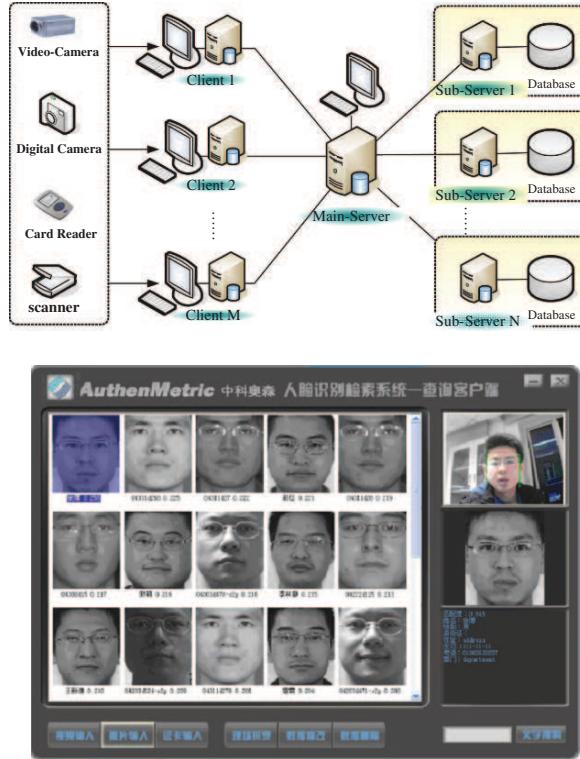
If a match is identified with sufficient confidence, an alert will rise at the server station and the client station will send the corresponding to the server station for further examination. The operator can manage watch-list database and navigate information of alerts by time and video source. See also the graphic interface below.

The enrollment of watch-list data can be imported in batch mode or one by one. The enrollment face images may come from a still face photo, a photo scan, or a connected video stream. The operator can perform query search and update in the database. The system support PC camera and CCTV cameras with frame grabber via DirectShow. The operator can configure each input camera type and video frame size, and set areas of interest in each video frame.

#### 1.4.5 Large Face Database Search

Large face database (such as that for national registration data or black-list data) search applications can operate in either one-to-one verification or one-to-many identification mode. They involve large face database of, e.g., more than a million individuals. This requires that the system should have an efficient data import

**Fig. 1.10** Topology and user interface for a large-scale face database search system



module. For one-to-many search, in particular, the response time would be a key performance measure in addition to the accuracy measure.

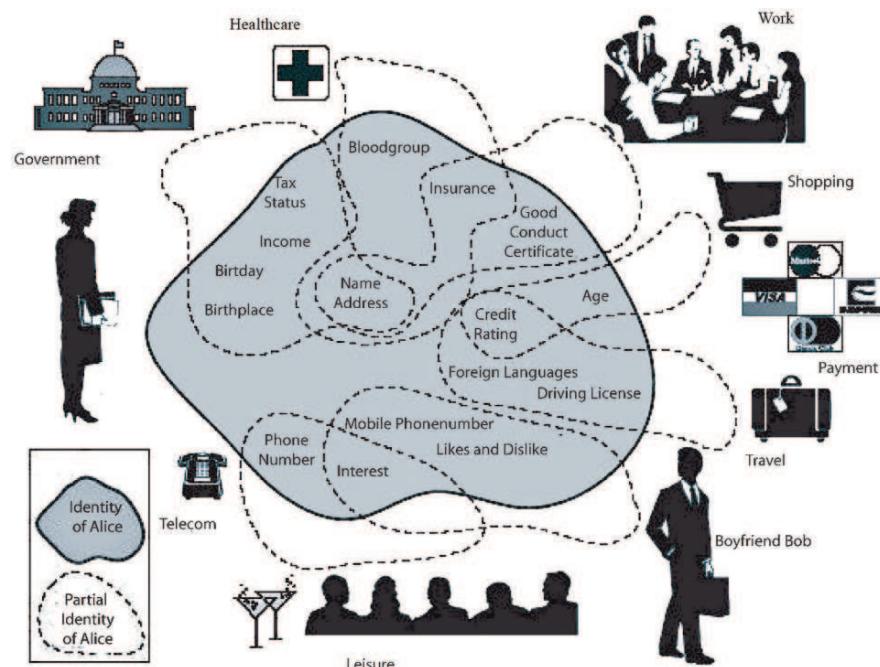
Because large databases are involved and the inquiry can come from several terminals (clients), the multi-tier architecture is used. Figure 1.10 shows our solution and the graphical user interface (GUI) of a search result. The system consists of one main server, N sub-servers, and M clients. A large database is distributed among the sub-servers. Each sub-server can host, say, a face database of a million people. A client generates the facial feature template, and sends the face image to the main server and a search request thereby. The main server then issues the search command to the sub-servers to retrieve relevant information.

## 1.5 Privacy and Identity Management in Distributed Intelligent Systems

Research in distributed (intelligent) sensor networks and the transparent use of biometrics (requiring no actions from the end user) will enable new applications in smart environments and security (a.o). To find the right balance between privacy, security, and convenience will be essential [54]. What these applications all have

in common is identity management. In classical identity management biometric systems authenticate only one (federal) identity. In modern identity management systems different profiles of identity (partial identities) exist and we foresee biometric algorithms will enable partial identity classifications like gender and age, which allow more anonymous ways of biometric authentication. In a future scenario, where users are enabled to communicate partial identities, users share different (identity) data according to different user scenarios (Fig. 1.11).

The use of biometrics at a distance will allow more transparent identity management. Given the direction of development we have to recognize that certain levels of privacy can only be maintained via technical means. We foresee a future situation where people will carry certain identity tokens (e.g., in a handheld phone, an identity card, or possibly an implanted chip) constituting partial identities by which humans present themselves enabling them to communicate with their environment through different applications [55]. The use of biometrics may subsequently be required by the environment to reach higher levels, in a mutual authentication process. One way of leaving control in the hand of individuals is to introduce negotiation into the authentication process. This may contribute to keeping a reasonable balance of power between citizens. The need for confidentiality, privacy, confidence, and trust in the integrity of exchanged information is ever greater. All these ask for privacy



**Fig. 1.11** Different relations define partial identities. The picture is taken from the tutorial of PRIME, Privacy and Identity Management for Europe, which explains privacy and identity management aspects. It is available via <http://www.prime-project.eu/>

fallback scenarios, openness, independent certification, and ways of communication for possible verification of stored data by users.

With the evolving biometric technologies toward biometrics at a distant offering transparent biometrics, new privacy issues will arise. The non-repudiation aspect of biometrics can be a bigger treat to privacy than biometrics itself. For instance, at a border crossing or in a smart environment a fresh measurement of your biometrics may be matched against the one stored in your identity token (assuming verification) without leaving a trace in the system. However, in most cases, central storage of date including time stamps will result in advanced forms of tracking and tracing.

A clearer separation between the authentication and non-repudiation roles of biometrics is necessary. When there is only authentication, a biometric check can yield a “ticket” or “credential” that allows access to further services which may actually be anonymous. This is much more privacy friendly, and may lead to easier acceptance. It makes sense to require (via regulations) that non-repudiation use of biometrics is clearly indicated to the users.

In recent papers solutions have been proposed to prevent templates to be stolen or “leaked” using template encryption. In [56] a biometric template is encrypted and a secret key is generated with so-called helper data, making it very difficult to restore the original signal. Problems are the robustness to small variations in the template. Binding this template to the application, e.g., with use of helper data or binding the template to the data making would make biometrics better conformed to the EU privacy regulations.

## References

1. D. Maltoni, D. Maio, A.K. Jain, S. Prabhakar, *Handbook of Fingerprint Recognition*, 2nd edn., Springer, New York, 2008.
2. D.D. Zhang, *Palmpprint Authentication*, Springer, New York, 2004.
3. J. Wayman, A. Jain, D. Maltoni, D. Maio, *Biometric Systems*, Springer, New York, 2004.
4. H. Wechsler, *Reliable Face Recognition Methods: System Design, Implementation and Evaluation*, Springer, New York, 2007.
5. M.S. Nixon, T.N. Tan, R. Chellappa, *Human Identification Based on Gait*, Springer, New York, 2005
6. R. Cehllappa, S.K. Zhou, Face tracking and recognition from video, in *Handbook of Face Recognition*, edited by A.K. Jain and S.Z. Li, pp. 169–192, Springer, New York, 2004.
7. S.K. Zhou, R. Chellappa, W. Zhao, *Unconstrained Face Recognition*, Springer, New York, 2006.
8. J.R. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. LoIacono, S. Mangru, M. Tinker, T. Zappia, W.Y. Zhao, Iris on the Move<sup>TM</sup>: Acquisition of Images for Iris Recognition in Less Constrained Environments, Proc. IEEE. 94(11):1936–1947, November 2006.
9. National Science and Technology Council, Subcommittee on Biometrics (Duane Blackburn co-chair). The National Biometrics Challenge. August 2007. <http://www.biometrics.gov/Documents/biochallengedoc.pdf>
10. The BioSecure Network of Excellence, <http://www.biosecure.info>
11. B.A.M. Schouten, F. Deravi, C. García-Mateo, M. Tistarelli, M. Snijder, M. Meints, J. Dittmann. BioSecure: white paper for research in biometrics beyond BioSecure. CWI report 2008, PNA-R0803, ISSN 1386-3711. 2008.

12. D.J. Cook, S.K. Das, Smart Environments: Technology, Protocols and Applications]. Wiley-Interscience, Hoboken, USA, 2005.
13. Liberty CCTV, 2008 <http://www.liberty-human-rights.org.uk/issues/3-privacy/32-cctv/index.shtml>
14. H.M. Dee, S.A. Velastin, How Close Are We to Solving the Problem of automated visual surveillance? A review of Real-World Surveillance, Scientific Progress and Evaluative Mechanisms. Springer-Verlag, New York, 2007.
15. M. Friedewald, O. Da Costa, Y. Punie, P. Alahuhta, S. Heinonen, Perspectives of Ambient Intelligence in the Home Environment, Telematics and Informatics 22, pp. 221–238, Elsevier Publishing, Amsterdam, 2004.
16. E. Aarts, E. Diederiks, Ambient Lifestyle, From Concept to Experience. Bis Publishers, Amsterdam, 2006.
17. E. Pauwels, A.A. Salah, R. Tavenard, Sensor Networks for Ambient Intelligence, Workshop on Multimedia Signal Processing (MMSP), Crete, October 2007.
18. S.A. Velastin, B.A. Boghossian, B.P.L. Lo, J. Sun, M.A. Vicencio-Silva, PRISMATICA: towards ambient intelligence in public transport environments. IEEE Trans. Syst. Man Cybern. Part A 35(1):164–182, 2005.
19. Video-based Threat Assessment and Biometrics Network (ViTAB). <http://dirc web.king.ac.uk/vitab/>
20. M.C. Mozer, Lessons from an adaptive home, in Smart Environments: Technologies, Protocols, and Applications, edited by D.J. Cook and S.K. Das, Wiley Series on Parallel and Distributed Computing, 2005, pp. 273–294, Wiley, New York.
21. D. Cook, M. Youngblood, E. Heierman, K. Gopalratnam, S. Rao, A. Litvin, F. Khawaja, “Mavhom: An agent-based smart home,” in Proc. IEEE PerCom, 2003, pp. 521–524.
22. AMIGO – Ambient Intelligence for the Networked Home Environment. [Online]. Available: <http://www.hitechprojects.com/euprojects/amigo/index.htm>
23. COGNIRON: The Cognitive Robot Companion. [Online]. Available: <http://www.cogniron.org/Home.php>.
24. CAVIAR – Context Aware Vision using Image-based Active Recognition. [Online]. Available: <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>
25. D. Hall, F. Pelisson, O. Riff, J.L. Crowley, brand identification using gaussian derivative histograms, Mach Vision Appl, 16(1):41–46, 2004.
26. A. Waibel, H. Steusloff, R. Stiefelhagen, CHIL – computers in the human interaction loop, NIST ICASSP Meeting Recognition Workshop, Montreal, Canada, 2004.
27. CHIL – Computers in the Human Interaction Loop. [Online]. Available: <http://chil.server.de/servlet/is/101/>
28. MIT Project Oxygen: Pervasive Human-Centered Computing. [Online]. Available: <http://oxygen.csail.mit.edu/Overview.html>.
29. Mitsubishi Electric Research Laboratories: Ambient Intelligence for Better Buildings. [Online]. Available: <http://www.merl.com/projects/ulrs/>
30. C. Kidd, R. Orr, G. Abowd, C. Atkeson, I. Essa, B. MacIntyre, E. Mynatt, T. Starner, W. Newstetter, “The aware home: A living laboratory for ubiquitous computing research,” Proceedings of CoBuild 99, 1999.
31. F. Fleuret, J. Berclaz, R. Lengagne, P. Fua, Multi-camera people tracking with a probabilistic occupancy map, IEEE: Trans. Pattern Anal. Mach. Intell., 30(2):267–282, February 2008.
32. J. Kang, I. Cohen, G. Medioni, Tracking people in crowded scenes across multiple cameras, Asian Conference on Computer Vision (2004).
33. A. Mittal, L. Davis, M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene, Int J Comput Vision 51(3):189–203 (2003).
34. S. Haritaoglu, D. Harwood, L. Davis, W4: Real-time surveillance of people and their activities, IEEE Trans. Pattern Anal Mach Intell. 22(8):809–830, 2000.
35. J. Black, T. Ellis, P. Rosin, Multi-view image surveillance and tracking, IEEE Workshop on Motion and Video Computing (2002).

36. R. Morros, A.A. Salah, B. Schouten, C.S. Perales, J.L. Serrano, O. Ambekar, C. Kayalar, C. Keskin, L. Akarun. Multimodal identification and localization of users in a smart environment, to appear Journal on Multimodal Interfaces 2008.
37. A.A. Ross, K. Nandakumar, A.K. Jain, Handbook of Biometrics. Springer Verlag, New York, ISBN 978-0-387-22296-7. 2006.
38. D.E. Maurer, J.P. Baker, Fusing multimodal biometrics with quality estimates via a Bayesian belief network, *Pattern Recogn.*, 41(3), 821–832, 2007.
39. E. Tabassi, C. Wilson, C. Watson, NIST Fingerprint Image Quality, *NIST Res. Rep. NISTIR7151*, 2004.
40. The Ministry of the Interior and Kingdom Relations, the Netherlands. 2b or not to 2b. <http://www.minbzk.nl/contents/pages/43760/evaluatierapport1.pdf>. 2005.
41. O. Fatakusi, J. Kittler, N. Poh, Quality controlled multimodal fusion of biometric experts in progress in pattern recognition, *Image Analysis and Applications*, pp. 881–890, Springer, Berlin. Lect. Notes Comp. Sci., 4756, 2008.
42. N. Poh, G. Heusch, J. Kittler. On combination of face authentication experts by a mixture of quality dependent fusion qualifiers. In LNCS 4472, *Multiple Classifiers Systems (MCS)*, Prague, 2007, pp. 344–356.
43. J. Kittler, N. Poh, O. Fatakusi, K. Messer, K. Kryszczuk, J. Richiardi, A. Drygajlo, Quality dependent fusion of intramodal and multimodal biometric experts. In Proc. of SPIE Defense and Security Symposium, Workshop on Biometric Technology for Human Identification, 2007, vol. 6539.
44. A.K. Jain, S.C. Dass, K. Nandakumar, Soft biometric traits for personal recognition systems,” in Proceedings of International Conference on Biometric Authentication (ICBA), LNCS 3072, pp. 731–738, 2004.
45. O. Ambekar, E. Pauwels, B. Schouten, Binding low level features to support opportunistic person identification, 28th Symposium on Information Theory, Enschede, The Netherlands, May 24–25, 2007.
46. M.S. Magnusson, “Discovering hidden time patterns in behavior: T-patterns and their detection.” *Behav Res Methods Instrum Comput*, 32(1), 93–110, February 2000.
47. C. Wren, D. Minnen, S. Rao, Similarity-based analysis for large networks of ultra-low resolution sensors, *Pattern Recogn.*, 39, 1918–1931, 2006.
48. N. Eagle, A. Pentland, Eigenbehaviors: Identifying structure in routine, in Proc. of Ubicomp’06, 2006.
49. D.J. Cook, Prediction algorithms for smart environments, in *Smart Environments: Technologies, Protocols, and Applications*, edited by D.J. Cook and S.K. Das, Wiley Series on Parallel and Distributed Computing, 2005, pp. 175–192, Wiley, New York.
50. S.Z. Li, et al., Illumination invariant face recognition using near-infrared images. *IEEE-T-PAMI* (Special issue on Biometrics: Progress and Directions), 29(4):627–39, April 2007.
51. ISO/IEC 7501-1 Identification cards – Machine readable travel documents – Part 1: Machine readable passports
52. ISO/IEC 19794-2 Information Technology – Biometric Data Interchange Formats – Part 2: Finger Minutiae Data.
53. ISO/IEC 19794-5 Information Technology – Biometric Data Interchange Formats – Part 5: Face Image Data.
54. M. Rejman-Greene, Privacy issues in the application of biometrics: A european perspective in biometric systems, In *Biometric Systems*, edited by J. Wayman, A. Jain, D. Maltoni and D. Maio, Springer, New York, 2004.
55. B. Schouten, B. Jacobs, Biometrics and their use in e-Passports, *Image and Vision Computing (IMAVIS)*, Special Issue on Multimodal Biometrics, Elsevier Publishers, 2008. To appear.
56. J.P. Linnartz, P. Tuyls, New shielding functions to enhance privacy and prevent misuse of biometric templates, in Proc. AVBPA 2003, pp. 393–402, LNCS 2688, Springer-Verlag, Berlin, Heidelberg, 2003.

# **Chapter 2**

## **Iris Recognition – Beyond One Meter**

**James R. Matey and Lauren R. Kennell**

**Abstract** Iris recognition is, arguably, the most robust form of biometric identification. It has been deployed in large-scale systems that have been very effective. The systems deployed to date make use of iris cameras that require significant user cooperation; that in turn imposes significant constraints on the deployment scenarios that are practical.

There are many applications in which it would be useful to undertake iris recognition at distances greater than those provided by conventional iris recognition systems. This chapter reviews iris recognition methods and provides a framework for understanding the issues involved in capturing images for iris recognition at distances of a meter or more.

This chapter is intended to be a self-contained tutorial, but the reader will be referred to recent reviews and papers for additional detail. A small set of exercises is provided at the end of the chapter. This chapter is based on the tutorial originally presented by Matey at the International Summer School on Biometrics held in Alghero in 2007 [1].

Mention of any product in this chapter is for illustrative and tutorial purposes only and does not constitute an endorsement of the product by the authors, the US Naval Academy or the US Government.

### **2.1 Background**

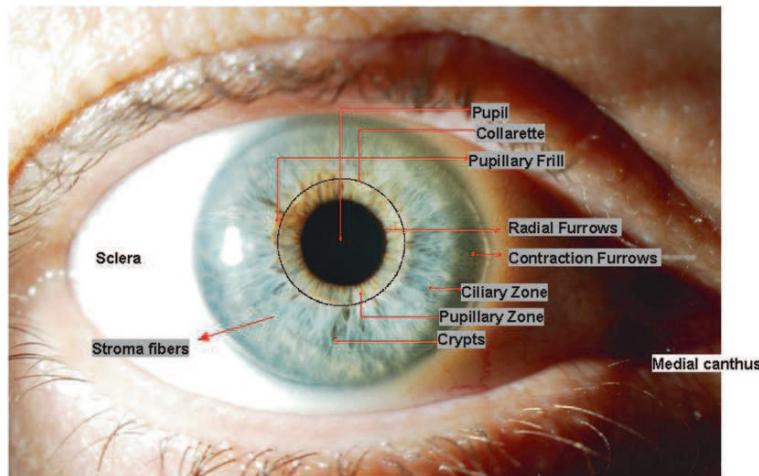
#### **2.1.1 What Is the Iris?**

The term iris has numerous meanings, including

1. in botany, flowers of the genus Iris
2. in classical mythology, the goddess of the rainbow
3. in optics, an adjustable aperture that regulates the amount of light admitted to an optical system

---

J.R. Matey (✉)  
Biometric Signal Processing Laboratory, Electrical & Computer Engineering Department,  
US Naval Academy – Maury Hall, Annapolis, MD 21402-5025, USA  
e-mail: matey@usna.edu



**Fig. 2.1** Color image of an iris taken with an ophthalmologic camera

4. in anatomy, the colored portion of the eye surrounding the pupil and adjusting its size

The last of these is the relevant one for iris recognition. In Fig. 2.1, we see an example iris with key visible features annotated

**medial canthus:** the angle formed between the upper and lower eyelid where they meet near the center of the face

**sclera:** the white of the eye

**pupil:** the opening at the center of the pupil through which light is admitted to the eye

**pupillary zone:** inner region of iris whose edge forms the boundary of the pupil; the sphincter muscles that closes the pupil resides here

**ciliary zone:** region of iris from pupillary zone to the ciliary body; the dilator muscles that open the pupil reside here

**stroma fibers:** the pigmented fibro vascular tissue that makes up most of the visible iris

**crypts, furrows:** two types of inhomogeneities in the distribution of stroma fibers

**collarette:** region that divides the pupillary zone from the ciliary zone

### 2.1.2 Why Is Iris a Good Biometric?

Any good biometric needs to have

Large inter-class variability – large differences between individuals

Small intra-class variability – small differences between samples taken from a particular individual

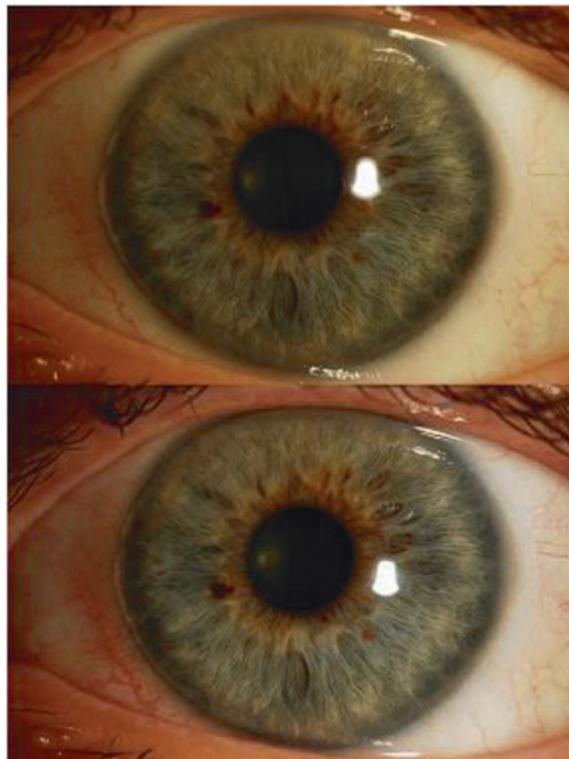
Stability over time – related to small intra-class variability

Relative ease of collection

For iris, the last of these is obvious. The iris is accessible – it can be easily seen and imaged.

Iris has a large inter-class variability. The detailed structure and distribution of the stroma fibers comes about from processes during gestation that have sensitive dependence on initial conditions [2]. The processes are akin to tearing a sheet of paper. Two sheets taken in succession from a ream of paper and subjected to a careful attempt to tear them in exactly the same way will tear differently. Experiments have shown that the details of the iris are at least as distinct as fingerprints in automated biometric identification systems [3]. Biometric templates from the most widely used algorithm have approximately 250 degrees of freedom [4] in the context of a binomial model for the imposter distribution.

Iris patterns have small intra-class variability and appear stable over time. This can be seen in the two images of Fig. 2.2; these are two pictures of the same iris, taken 6 months apart. It is difficult to visually find any significant difference between the two. Analysis using iris recognition algorithms supports this observation and there are anecdotal examples that support it over much larger periods



**Fig. 2.2** Two pictures of the same iris, taken 6 months apart using an ophthalmological camera (need permission from Springer-Verlag)

(John Daugman, private communication). Daugman's confirmation of the identity of the National Geographic "Afghan girl" is one example [5, 6]. In addition, the iris is protected by anatomical structures and reflexes that reduce the likelihood of traumatic injury – and thereby contribute to its stability.

Flynn and his colleagues at Notre Dame are collecting a data set that is large enough and spans a large enough time period that we may soon be able to test the intra-class variability more extensively than has been possible in the past.

### **2.1.3 Iris Pathologies**

For almost every biometric, some segment of the population will be unable to present a sufficiently good sample of their biometric to be enrolled or will have some trauma or disease that changes their biometric so that they no longer match a previous enrollment.

- Fingers can be lost in accidents
- Fingerprints can be changed by scarring and abrasion
- Voices can be changed by illness
- Faces can be changed by surgery, age, or accident

For iris, there are a number of pathological conditions, in addition to trauma in various forms that need to be considered in the design of any large-scale deployment. These include:

**Aniridia:** the absence of an iris. This is a congenital defect that is typically bilateral, affecting both eyes. It is linked to other congenital defects that shorten life span. The absence of an iris precludes iris recognition. Prevalence  $\sim 1:4 \times 10^4$  to  $1: 10^5$

**Albinism:** lack of pigment in skin, hair, and iris. The lack of pigment reduces contrast between the iris and the sclera and can have an adverse impact on the segmentation of the iris that we discuss below. There are several variants on this disorder. Prevalence in the general population is of the order of  $5:10^5$ ; however, some populations have higher rates.

**Essential iris atrophy:** is one of a group of iridocorneal endothelial syndromes that include Chandler syndrome and Cogan-Reese (iris nevus) syndrome. These syndromes normally present unilaterally. The key point for iris recognition is that large segments of the iris can essentially disappear or change over time. The authors were unable to find reliable estimates of prevalence for this condition.

**Tumors:** Tumors can grow anywhere in the body, including the iris. As it grows, a tumor will displace tissue and change the patterns in the iris. The authors were unable to find reliable estimates of prevalence for this condition.

More information about aniridia and albinism can be found at the US National Institutes of Health Web site [7]. The images in Fig. 2.3 illustrate these pathologies.



**Fig. 2.3** Several examples of the wide range of iris pathologies. From *top to bottom* and *left to right*: albinism, tumor, aniridia, essential iris atrophy, band keratopathy. The albinism image is courtesy of Dr. Dirk Werdermann, Ochsenfurt, Germany; the tumor image is courtesy of Prof. Bertil Damato, Ocular Oncology Service, St. Paul's Eye Unit, Royal Liverpool University Hospital, Liverpool. The lower three images are courtesy of Prof. Dr. Georg Michelson, Verlag Online Journal of Ophthalmology

#### 2.1.4 Iridology

Iridology is a technique from alternative medicine; its proponents claim to be able to diagnose a wide variety of ailments based on an examination of the iris [8]. Iridology is uniformly rejected by mainstream medicine; in double-blind tests it has been ineffective as a diagnostic tool [9]. The notion that the iris experiences significant changes in response to the health of other organs is in direct conflict with the basic tenet of iris recognition – iris patterns are essentially constant with time.

### 2.1.5 A Brief History of Iris Recognition

Bertillon (1886) “... minute drawing of the areola and denticulation of the human iris . . .” might be useful for human identification.  
 Burch (1936) suggests that iris patterns can be used for identification.  
*Never Say Never* (1983) iris recognition used as a plot element  
 Flom/Saphir (1985) patents the basic idea of iris recognition  
 Daugman (1994) patents dominant algorithm for iris recognition  
 First commercial products (1995)  
 UAE Expellee Border Control System (2001), first large-scale deployment and still (2008) the largest integrated system  
*Minority Report* (2002) iris recognition used as a plot element  
 Expiration of Flom patent (2005–2007) drives expansion of research into alternative algorithms  
 Iris on the Move<sup>TM</sup> (2005) iris recognition in less-constrained environments.  
 Expiration of Daugman patent (2011)

The notion that the complex patterns in the human iris can be used for personal identification goes back to the time of Bertillon [10], but it was not until 1994 that advances in computer hardware and automated pattern recognition technology provided the tools that enabled John Daugman to invent a practical method for iris recognition [11]. Daugman’s algorithm is the dominant algorithm; minor variations of it are used in almost all commercially significant deployments of iris recognition.

One could write a small book on the history of iris recognition, particularly if the book included a discussion of all of the algorithms that have been proposed, all of the iris acquisition devices that have been developed, and all of the legal issues that have been raised. Bowyer [12], Vatsa [13], and Thornton [14] have all conducted critical surveys of iris recognition algorithms that we will not attempt to replicate here. Table 2.1 presents a non-exhaustive list of iris recognition devices, and Table 2.2 presents a non-exhaustive list of iris recognition algorithms.

For the purpose of this paper, we will discuss two particularly interesting developments in the history of iris recognition:

United Arab Emirates Expellee Program  
 Confirmation of the identity of the Afghan girl

### 2.1.6 Some Important Definitions

There are some important terms that we will need in our discussion. Though most of these terms are defined elsewhere in this book, we have collected the following terms here as an aid to the reader (Table 2.3). These terms are generic, used for all biometrics. In the case of iris recognition, subjects normally have two eyes and each of those eyes is an independent biometric.

**Table 2.1** Non-exhaustive list of iris recognition devices, year introduced based on press accounts, and/or vendor histories

Vendor	Model	Year introduced
IrisScan	System 2000	1995
OKI	IrisPass®-S	1998
LG	2200	1999
Sensar	R1	1999
Iridian	Authenticam™	2000
Panasonic	BM-ET-100 – Authenticam™	2001
LG	3000	2001
OKI	IrisPass®-WG	2002
Panasonic	BM-ET-500	2002
IrisGuard	H-100	2003
Securimetrics	Pier™ 2.2	2003
Securimetrics	Pier™ 2.3	2003
OKI	IrisPass-H	2004
Panasonic	BM-ET-300	2004
LG	4000	2006
OKI	IrisPass®-M	2005
Sarnoff Corporation	Iris On the Move™	2005
Securimetrics	HIIDE™	2005
IriTech	Neoris 2000	2006
Jiris	JCP1000	2006
Panasonic	BM-ET-330	2006
Global Rainmakers	Hbox™	2007
Panasonic	BM-ET-200	2007
IrisGuard	AD-100	2008

### 2.1.7 United Arab Emirates Expellee Program

The United Arab Emirates (UAE) instituted their National Expellees Tracking and Border Control System (NETBCS) in August 2001. The purpose of the system is to prevent illegal immigration. Illegal immigrants who are caught trying to enter the UAE through air/sea/land ports are returned to their point of origin. In the past, these individuals would frequently try again with a new set of false documents (Table 2.4).

Under the iris recognition-based NETBCS, expellees iris patterns are enrolled in a national expellee database before they are expelled. The iris patterns of incoming travelers are checked in real time against that database – this reveals if an individual was previously expelled and should not be allowed to enter. Table 2.5 shows statistics for the system as of summer 2007. To the author's knowledge this is the largest integrated deployment of iris recognition in the world (there are other larger deployments that are scattered across multiple databases, e.g., social services in Andhra Pradesh, India).

The UAE states that, to date, they have not had a match that was not confirmed by other means. This is an impressive (and believable) result. As we shall see later, it is almost certainly due to efforts to generate the highest quality iris images possible so that the authentic distribution for the system is as narrow as possible, thus enabling

**Table 2.2** Non-exhaustive list of iris recognition algorithms

- 
- 1992 Daugman, J.G., "High Confidence Personal Identification by Rapid Video Analysis of Iris Texture," Proc. of the IEEE Int. Carnahan Conf. Security Technology, pp. 50–60, October 1992.
- 1993 Daugman, J.G., "High Confidence Visual Recognition of Persons by a Test of Statistical Independence," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 11, pp. 1148–1161, November 1993.
- 1994 Wildes, R.P., Asmuth, J.C., Green, G.L., Hsu, S.C., Kolczynski, R.J., Matey, J.R., and McBride, S.E., "A System for Automated Iris Recognition," Proc. of the Second IEEE Workshop on Applications of Computer Vision, pp. 121–128, December 1994.
- 1996 Wildes, R.P., Asmuth, J.C., Green, G.L., Hsu, S.C., Kolczynski, R.J., Matey, J.R., and McBride, S.E., "A Machine Vision System for Iris Recognition," Machine Visual and Application. 1996, 9: 1–8.
- 1998 Boles, W.W. and Boashash, B., "A Human Identification Technique Using Images of the Iris and Wavelet Transform," IEEE Transactions on Signal Processing, 46(4):1185–1188, April 1998.
- 2000 Zhu, Y., Tan, T., and Wan, Y., "Biometric Personal Identification Based on Iris Patterns," Proc. of the 15th Intl. Conference on Pattern Recognition, vol. 2, pp. 801–804, September 2000.
- 2001 El-Bakry, H.M., "Fast Iris Detection for Personal Identification Using Modular Neural Networks," Proc. of the 2001 IEEE Int Symposium on Circuits and Systems, vol. 2, pp. 581–584, May 2001.
- 2001 Lim, S., Lee, K., Byeon, O., and Kim, T., "Efficient Iris Recognition Through Improvement of Feature Vector and Classifier," ETRI Journal, 23(2):61–70, June 2001.
- 2002 Ma, L., Wang, Y., and Tan, T., "Iris Recognition Using Circular Symmetric Filters," Proc. of the 16th Intl. Conference on Pattern Recognition, vol. 2, pp. 414–417, Aug. 2002.
- 2002 Sanchez-Avila, C., Sanchez-Reillo, R., and de Martin-Roche, D., "Iris-Based Biometric Recognition Using Dyadic Wavelet Transform," IEEE Aerospace and Electronic Systems Magazine, 17(10):3–6, Oct. 2002.
- 2002 William, L., Chekima, A., Fan, L.C., and Dargham, J.A., "Iris Recognition Using Self-organizing Neural Network," 2002 Student Conference on Research and Development, pp. 169–172, July 2002.
- 2003 Ma, L., Tan, T., Wang, Y., and Zhang, D., "Personal Identification Based on Iris Texture Analysis," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 25, no. 12, pp. 1519–1533.
- 2004 Sun, Z., Tan, T., and Wang, Y., "Robust Encoding of Local Ordinal Measures: A General Framework of Iris Recognition," Proc. of the ECCV 2004 Intl. Workshop on Biometric Authentication, pp. 270–282.
- 2005 Monro, D.M. and Zhang, Z., "An Effective Human Iris Code with Low Complexity," Proc. of the 2005 IEEE International Conference on Image Processing, vol. 3, pp. III-277–280, September 2005.
- 2005 Miyazawa, K., Ito, K., Aoki, T., Kobayashi, K., and Nakajima, H., "An Efficient Iris Recognition Algorithm Using Phase-Based Image Matching," Proc. of the 2005 IEEE Int. Conference on Image Processing, pp. II-49–52, September 2005.
- 2006 Liu, C. and Xie, M. "Iris Recognition Based on DLDA," Proc. of the 18th Intl. Conference on Pattern Recognition, vol. 4, pp. 489–492, September 2006.
- 2007 Daugman, J., "New Methods in Iris Recognition," IEEE Trans. on Systems, Man, and Cybernetics, vol. 37, no. 5, pp. 1167–1175.
- 2007 Geng, J., Li, Y., and Chian, T., "SIFT Based Iris Feature Extraction and Matching," Proc. SPIE, vol. 6753, pp. 67532F.
- 2008 Hao, F., Daugman, J., and Zielinski, P., "A Fast Search Algorithm for a Large Fuzzy Database," IEEE Trans. on Information Forensics and Security, vol. 3, no. 2, pp. 203–212, June 2008.
-

**Table 2.3** Biometric terms

Biometric template	An object that summarizes and encapsulates biometric information in a uniform way to enable easy comparison of biometrics collected at different times or with different subjects. Today, these are frequently digital constructions; fingerprint cards are an example from an earlier era.
Match score	A measure of similarity (dissimilarity) between two biometric templates.
Authentic distribution	The probability distribution of match score for a biometric system, for templates compared against templates from the same subject.
Imposter distribution	The probability distribution of match score for a biometric system, for templates compared against templates from different subjects.
Match threshold	The match score above (below) which a match is declared.

**Table 2.4** Biometric failure modes

False-match	Templates A and B are from different persons, but match nonetheless.
False-non-match	Templates A and B are from the same person, but fail to match.
Failure-to-enroll	Despite good faith efforts on part of subject and operator, the system fails to generate a usable template.
Failure-to-acquire	The system fails to generate a template for someone who would not generate a failure-to-enroll.

**Table 2.5** Statistics for the UAE expellee tracking and border control system (as of 2007)

Uptime (days):	1,430
Database size (templates):	1,131,625
Number of nationalities in database:	>160
Persons searched against database:	11,201,651
Average searches per day:	> 12,000
Persons caught:	140,162
Average caught per day:	130–150
Search speed, turnaround time (s):	< 3
Total cross comparisons:	7,961,671,626,107
Daily cross comparisons:	12,894,647,310

an aggressive match threshold (substantially lower than the nominal 0.33 fractional Hamming distance value quoted in many papers).

### 2.1.8 Confirmation of the Identity of the Afghan Girl

The cover of the June 1985 issue of National Geographic was a stunning picture of a 13-year-old Afghan refugee girl, Sharbat Gula. The picture caught the attention of people all over the world. At the time, the picture was anonymous; the girl's identity was not known. Some 18 years later the photographer, Steve McCurry, joined a National Geographic expedition to Afghanistan to find the "Afghan girl." They



**Fig. 2.4** Two pictures of Sharbat Gula. The June 1985 image is on the *left*, the picture from 18 years later is on the *right* (used with the permission of the photographer, Steve McCurry)

believed they had succeeded. However, after 18 years in a war-ravaged country, Gula's appearance had changed, see Fig. 2.4. To confirm her identity, they turned to John Daugman. Working from the original high-resolution film, Daugman was able to extract iris images from the pictures, compute iris templates, and compare the templates. Daugman explained his results as follows [15]:

When I ran the search engine (the matching algorithm) on these IrisCodes, I got a Hamming Distance of 0.24 for her left eye, and 0.31 for her right eye. As may be seen from the histogram that arises when DIFFERENT irises are compared by their IrisCodes [16], these measured Hamming Distances are so far out on the distribution tail that it is statistically almost impossible for different irises to show so little dissimilarity. The mathematical odds against such an event are 6 million to one for her right eye, and 10-to-the-15th-power to one for her left eye. National Geographic accepted and published this conclusion in a second cover issue featuring Sharbat Gula, 18 years after the first, and the Society launched their "Afghan Girl's Fund" to assist the education of Muslim girls in cultures that discourage or prohibit female education.

This example is important because it made use of images that were not collected specifically for iris recognition; it is likely the first instance of the use of iris recognition in a forensic (perhaps forensic-like) context.

At the time Daugman did his analysis, the imposter distribution for Hamming distances less than 0.25 had not been experimentally confirmed; his results were based on extrapolations of the evidence available at the time. Since that time the imposter distributions have been experimentally tested by Daugman [17] and the tails of the distributions are a bit "fatter" than expected on theoretical grounds. The best estimates now for the probability of a Hamming distance of 0.24 between two different irises are closer to 1:1013 rather than the original estimate of 1:1015 – a bit higher, but still quite convincing.

The probability estimates for this case may be revised at some point in the future. The imposter distributions on which the Afghan girl results are based come from

a database of near infrared (NIR) images acquired using specialized iris recognition cameras. The images in this example are visible light images acquired using standard photographic equipment. It is possible that the imposter distribution for such visible light images may differ from that for NIR images – particularly in the tail, which requires a large database to probe. At present, there is simply not enough data to know what the tails look like for visible imagery; this is a topic of ongoing research. When more data is available, the probability estimates may be revised again.

### 2.1.9 A Generic Process for Iris Recognition

All iris recognition systems proposed or deployed to date can be fit into a four step model:

- Acquisition
- Iris localization/segmentation
- Template generation
- Template matching

In the acquisition step, we acquire an image of the iris of sufficient quality to enable the succeeding steps. The primary image quality metrics are resolution, signal/noise ratio, contrast and illumination wavelength. In the iris localization and segmentation step we identify where the iris and its borders with the pupil and the sclera are located. In many cases, the iris is then remapped into a pseudo-polar coordinate system to ease subsequent computations and to take into account variations in pupil size. Once the iris location is known, we can extract features from it and assemble those features to generate a template. Templates can then be stored for later matching against templates derived from new images.

## 2.2 Acquisition – Photons to Pixels

From quantum mechanics we know that light has both wave-like and particle-like properties. In modeling of iris recognition systems, we make use of both the wave and particle models for light propagation and illumination. The wave model is important for optical effects such as diffraction. The particle model is important in understanding the sensor. The particle of light is called the photon and we will use that term throughout the following discussion.

We will also refer to iris pixels and sensor pixels. In this context the iris pixel is the smallest feature of interest on the iris and a sensor pixel is the smallest feature that can be resolved in an image that is focused onto the sensor.

In the calculations that follow, the dimensions and specifications are not those of any particular system. They have been chosen to make the math easy.

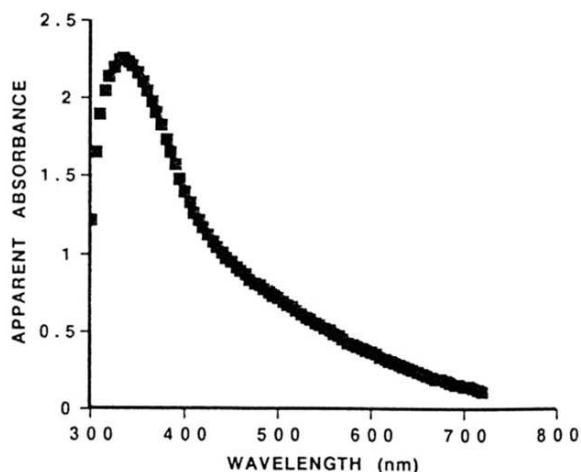
### 2.2.1 Photon Sources

One categorization of illumination sources for iris recognition reflects the degree of control that the iris recognition system has over the source: ambient or active. Ambient sources are the sources that are there in the absence of the iris recognition system. Ambient sources include daylight and office lighting. Active illumination is illumination provided by and under the direct control of the iris recognition system.

The lack of control of ambient sources – office workers can turn lights on/off, sunlight varies dramatically during the course of a day – makes it very difficult to rely on ambient light for iris recognition. Essentially all commercial iris recognition system use active illumination in some form. The lack of control of ambient light sources contributes to difficulties in acquiring good images – the variation in ambient light can change the quality of the images even in the presence of active illumination. Most iris recognition systems are designed to reject ambient light to avoid the difficulties associated with its variation. We will discuss that issue in more detail when we get to photon sensors.

Current iris recognition systems operate with illumination in the 720–900 nm range. This choice is driven by the absorption spectrum of melanin and the absorption spectrum of silicon. Melanin is the organic pigment that predominates in the coloring of human irises. It is strongly absorbing in the visible, but much less so in the near infrared (NIR) – 800–900 nm, see Fig. 2.5. People with dark eyes have more melanin than those with light eyes. Dark and light irises are more similar in their light reflecting and absorbing properties in the NIR than they are in the visible because the absorption of the melanin drops off in the NIR. Hence, the structural details of a dark iris show much better contrast in the NIR than in the visible. The melanin spectrum drives us to longer wavelengths.

Silicon is the material of choice for commercial image sensors. Silicon has a band gap of  $\sim 1.12$  eV; this corresponds to a light of wavelength of about 1100 nm.



**Fig. 2.5** Spectrum of melanin – courtesy of John Daugman

As the energy of the photon decreases with increasing wavelength ( $E = h/\lambda$  where  $E$  is the energy,  $h$  is Planck's constant and  $\lambda$  is the wavelength) toward the band gap value, the photons are less likely to interact with silicon and the detector becomes less sensitive, until, for energies less than the band gap, the photons do not have enough energy to excite a silicon sensor – and the sensor is effectively blind to such photons. The silicon absorption spectrum drives us toward shorter wavelengths.

The 720–900 nm regime is a compromise between melanin and silicon.

It is possible to build sensors using other materials (e.g., germanium) that operate at longer wavelengths than silicon. At longer wavelengths we need to consider the absorption length of the iris tissues. Human tissue is more transparent in the NIR than it is in the visible. This can be illustrated by shining an incandescent flashlight through your finger tip. Though the light is strongly scattered, a significant portion of the longer wavelengths gets through – and we see that as red.

If we move to a sufficiently long wavelength, we lose contrast and definition because the photon penetration depth in the iris tissue is larger than the size of the structures of interest and we do not get sufficient contrast to image the structures; at still longer wavelengths, water absorption cuts in and the fluids residing in aqueous humor and in the interstices of the iris structure reduce contrast.

At still longer wavelengths – thermal wavelengths – the tissue becomes self-luminous and the contrast mechanism is then tied to temperature variations. Since the iris is essentially uniform in temperature, there is no contrast at thermal wavelengths.

So, by design or not, commercial systems in the 720–900 nm regime appear to be operating close to the optimum wavelength for providing contrast and structure definition in the iris, consistent with signal/noise issues in the sensor.

There are a variety of potential sources for photons in this regime, see Table 2.6.

The majority of commercial systems make use of LED illumination. LEDs are small, bright, efficient, and have relatively narrow bandwidth, so most of the photons are at the wavelength of interest. They are also relatively easy to control – a simple constant current supply will yield a relatively constant photon output that can be easily varied or pulsed.

Incandescent, fluorescent, and arc lamps have been used in some laboratory experiments, but have not made their way into any commercial products, at least to the authors' knowledge.

**Table 2.6** Light sources

Type	Bandwidth (typical)	Pulsed/continuous
Light emitting diode (LED)	50 nm	Pulsed or continuous
Super luminescent diode (SLD)	50 nm	Pulsed or continuous
Laser	< 1 nm	Pulsed or continuous
Incandescent	Broadband > 1000 nm	Continuous
Fluorescent	Broadband with peaks	Continuous
Arc	Broadband with peaks	Continuous
Flash	Broadband with peaks	Pulsed

Flash lamps, with optical filters to control the bandwidth of the photon flux have been used in commercial system, notably the OKI IrisPass-M<sup>TM</sup>. The technology for control of flash lamps is well developed and if you can afford to throw away the portion of the spectrum that is not useful, they can be very effective.

Lasers and SLDs are interesting light sources because they have intrinsically higher radiance (watts/cm<sup>2</sup>-sr) than the other sources. Hence, they can deliver higher irradiance (watts/cm<sup>2</sup>) to a target for an optical system of fixed size. They both have safety issues that are more severe than the other light sources. In addition, laser illumination generates speckle, which tends to degrade the effective resolution of images. To date, there are no commercial products using lasers or SLDs as light sources.

### **2.2.2 Photon Transport**

The photons need to get from the illumination source to the iris and back to the sensor. Photons get lost on both traverses. One source of loss is absorption in the air, though for most applications this can be ignored as a small effect. However, the presence of fog or other mists or particulates can induce a great deal of light scattering, as anyone who has driven a car at night in the fog can attest.

A more important consideration is that the photons spread out on the outbound path and after scattering from the iris, they spread again on the inbound path. For the case of a point source (a reasonable model for an LED), the irradiance will fall off as  $1/r^2$  where  $r$  is the distance from the source. The photons that impinge on the iris will be scattered and each point on the iris will act as a new point source. The light coming back will be subject to that same factor of  $1/r^2$ . The irradiance at the sensor lens will therefore fall off as  $1/r^4$ , assuming that the sensor and the illumination source are co-located.

It is possible to add optical components to the illuminator that will collect and focus the light from the source; this can reduce the  $1/r^2$  effect on the outbound traverse. However, the  $1/r^2$  effect on the inbound traverse cannot be changed.

Another important effect is the reflectance or albedo of the iris. The iris albedo in the NIR is of the order of 10–15%, as measured by the author on a small population.

Let us consider the case of a bare, 1 watt LED at 1 m from an iris. Assuming that the LED radiates into  $2\pi$  steradians, the irradiance at the iris will be

$$E_{\text{iris}} = \frac{1 \text{ watt}}{2\pi r^2} = 0.159 \frac{\text{watt}}{\text{meter}^2} = 15.9 \frac{\text{microwatts}}{\text{cm}^2}$$

The resolution we require, as specified in the ISO standard for iris images [18] is of the order of  $100 \mu\text{m} = 0.01 \text{ cm}$ . Hence each pixel on the iris will capture of the order of  $P_{\text{iris}} = (0.01 \text{ cm}) E_{\text{iris}} = 1.59 \text{ nW}$ .

Assuming an albedo of 10%, each pixel will radiate back 0.159 nW toward the sensor and optics.

### 2.2.3 Sensor and Optics

The sensor and optics need to capture as much of that optical power as possible and convert it into a signal for subsequent processing. The power scattered by each pixel on the iris is scattered into approximately  $2\pi$  steradians, so the irradiance at the lens of the sensor will be

$$E_{\text{lens}} = \frac{P_{\text{iris}}}{2\pi r^2} = 2.5 \frac{\text{femtowatts}}{\text{cm}^2} \quad (2.1)$$

where *femto* is the prefix for  $10^{-15}$ . Assuming a lens 5 cm in radius (a good size lens), that is 100% efficient in conveying light, and that the pixels in the sensor have been matched in size to the image of the iris pixels that are imaged onto the sensor, we would then get  $\sim 200$  fW on each sensor pixel.

As noted earlier, the band gap in silicon is  $\sim 1.12$  eV =  $1.6 \times 10^{-19}$  joules. Hence, we have the energy to potentially create  $200 \times 10^{-15} / 1.6 \times 10^{-19} = 1.2 \times 10^6$  electrons per second at each pixel. At a frame rate of 30 fps, the integration time is  $\sim 30$  ms, so we have a potential of about 36,000 electrons in each pixel for each frame.

The potential is not achieved. Practical sensors have quantum efficiencies – the fraction of photons that get turned into electrons – of rather less than 1. In the NIR, 10% is respectable. Hence we may only get 3600 electrons/pixel/frame.

We assumed that the sensor pixel was matched to the iris pixel. In practice, we do not get much choice in the size of sensor pixels – we take what the camera manufacturer gives us. With the sensor pixel set and the iris pixel set, the lens must be chosen to match those pixels to each other. The magnification of the lens must therefore be the ratio of the sensor pixel size to the iris pixel size. If the sensor pixel is 10  $\mu\text{m}$ , the magnification is  $M = 10/100 = 0.1$ .

Let  $q$  equal the sensor to lens distance,  $p$  equal the lens to iris distance, and  $f$  equal the lens focal length. From simple geometrical optics we know that  $M = q/p$ . We also know that

$$\frac{1}{f} = \frac{1}{p} + \frac{1}{q} = \frac{1}{p} + \frac{1}{Mp} \quad (2.2)$$

and therefore

$$f = \frac{Mp}{M+1} \quad (2.3)$$

The size of the iris pixel, the sensor pixel, and the distance between the iris and the lens sets the focal length of the lens. At 1 m, with 10 micron sensor pixels and 100 micron iris pixels, the focal length of the lens needs to be  $\sim 0.1$  m = 100 mm.

We have assumed earlier that the lens aperture is 5 cm in radius; with a focal length of 100 mm, the *F#* for the lens would then be  $\sim 1.0$ . Practical lenses have *F#*'s above 2. Hence, we would need to use a lens with at most a radius of  $\sim 2.5$  cm. This would give a fourfold reduction in the light collected by the lens, reducing the signal to  $\sim 900$  electrons.

### 2.2.4 Signal to Noise Ratio

A simple model for sensor noise has two terms: read noise and shot noise. Read noise is the noise introduced by the electronics associated with the sensor each time a sensor pixel is read. Ten electrons is a good read noise level.

Shot noise is associated with the randomness of the arrival and absorption of the photons by the sensor. Shot noise is modeled as a binomial process for which the standard deviation is just the square root of the average. Hence, the signal/noise ratio is just the square root of the number of electrons/pixel/frame.

For a signal of 900 electrons the shot noise would be 30 electrons – rather more than the read noise and the overall signal/noise ratio would be approximately 30 – ignoring the read noise contribution. In dB terms the signal/noise ratio would be  $20 \log_{10}(30) \sim 30$  dB.

The ISO specification recommends 40 dB corresponding to a signal/noise ratio of 100. Since we are in a shot noise limited regime, increasing the signal/noise ratio by a factor of 3.3 requires an increase in signal of the order of  $3.32 \sim 10$ . We could get this by increasing the number of LEDs from 1 to 10 or by increasing the quantum efficiency of the camera from 10 to 100%. Increasing the LED count is simpler and less expensive.

If we put all of these concepts into a single model for signal/noise, we get

$$S/N \approx \frac{D}{d_i} \cdot \Delta x_i \cdot \sqrt{QE \cdot \alpha_{iris} \cdot \beta \cdot E_{iris} \cdot \Delta t_s} \cdot \sqrt{\frac{\lambda}{2hc}} \quad (2.4)$$

where

$D$  Lens diameter

$d_i$  Distance from lens to iris

$\Delta x_i$  Pixel size on the iris

QE Quantum efficiency of the sensor, including fill factor

$\alpha_{iris}$  Albedo of the iris

$\beta$  Throughput of lens

$E_{iris}$  Irradiance at the iris

$\Delta t_s$  Shutter time of the sensor

$\lambda$  Wavelength of light

$h$  Planck's LANCK'S constant

$c$  Speed of light

### 2.2.5 Safety

Note that the signal/noise equation is written in terms of the irradiance ( $\text{W/cm}^2$ ) at the iris rather than the power of the illuminator. The rationale for that choice lies in safety considerations. For an indestructible, inanimate object, we can arbitrarily increase the signal/noise of an image by simply increasing the irradiance. Eyes are, unfortunately, all too destructible. The ACGIH® [19] provides guidelines for

industrial hygiene professionals in the form of threshold limit values [20], TLVs®. TLVs are levels of exposure that the ACGIH believes that nearly all people can be repeatedly exposed to without adverse health effects.

The TLV for incoherent (non-laser) ocular, NIR irradiance is  $10 \text{ mW/cm}^2$  for long (1000 s) exposures. This TLV is concerned with damage to the cornea and lens of the eye.

The TLV for laser NIR ocular irradiance in the NIR is of the order of  $2 \text{ mW/cm}^2$ . This TLV is concerned with damage to the retina from focusing of the irradiance at the pupil to a spot on the retina.

There is an additional TLV for radiance ( $\text{W/sr}\cdot\text{cm}^2$ ) that is concerned with damage to the retina from non-laser sources. Note the term radiance rather than irradiance. Radiance is the power per steradians per unit area of source; it corresponds closely to what we perceive as brightness. The radiance TLV is too complicated to discuss here. At the time of this writing (2008) the authors are not aware of any LEDs that have a radiance in the near IR that would exceed TLVs – this is consistent with results from the ICNIRP [21] from 10 years ago. However, the state of the art is advancing. Anyone planning to undertake research in this area should consult with their radiation safety officer or health physicist before subjecting subjects to radiances or irradiances that are above those encountered in a normal office setting.

Determining the irradiance or radiance requires the use of calibrated light meters designed for the purpose such as the International Light IL-1400 [22]. Anyone planning to conduct research in this area should obtain appropriate instrumentation with the advice and assistance of their safety officer or health physicist.

### 2.2.6 Image Quality

In the preceding section, we made brief reference to the ISO standard for iris images. Let us briefly review that standard. The primary properties for the standard are shown in Table 2.7. We used two of these in our calculations above – the resolution and signal/noise ratio. We now address dynamic range and contrast.

The dynamic range is relatively simple to implement. Essentially, all frame grabbers will provide a solid 8 bits.

The contrast issue is more complicated. For the scleral boundary we need to reconsider the spectra of human tissue. In the visible, the contrast between the sclera and iris is very pronounced, particularly for dark irises. In the NIR, melanin is less absorbing and tissue is more transparent, leading to a reduction in the contrast

**Table 2.7** Primary properties for the ISO iris image standard

Property	Value
Resolution	8.3 pixels/mm $\sim 120 \mu\text{m}/\text{pixel}$
Dynamic range	8 bits, 256 gray levels
Signal/noise	40 dB S/N $\sim 100$ , 7 of the 8 bits are noise-free
Contrast	Pupil/iris $\sim 50$ levels of separation Iris/sclera $\sim 90$ levels of separation

between the iris and sclera. In some iris cameras, the illumination is provided by two sets of LEDs – one set around 750–800 nm and another at 850–900 nm. The former set provides better definition of the scleral boundary; the latter provides better definition of the iris structure.

For the pupil boundary, we need to be cognizant of the NIR equivalent of the “red eye” effect. If the light source is close to the line between the pupil and the sensor (the optic axis of the system), the light impinging on the pupil will be focused down to a small point on the retina (for the case of a point source of light). Lenses work both ways. That small point will be imaged back through the lens of the eye and be returned toward the source. If the sensor and source are close together the sensor will see the pupil as a bright circle rather than a dark one. In visible light photography we see this as the “red eye” effect – red because the retinal tissue is red. In NIR, where we have no color, it is a “bright pupil” effect.

It is important to note that the ISO standard was developed with the Daugman algorithm as the backdrop. The proponents of other algorithms have pointed out, correctly, that alternative algorithms might well work better with an alternate standard. However, the Daugman algorithm is the dominant algorithm and the standard is what it is – at least for now. As other algorithms are more thoroughly tested, we will develop sufficient understanding of their relationships to image quality to enable a next generation of standards [23]. At this writing (2008) the relationship between image quality and the Daugman algorithm, let alone the alternative algorithms, was still a topic for research.

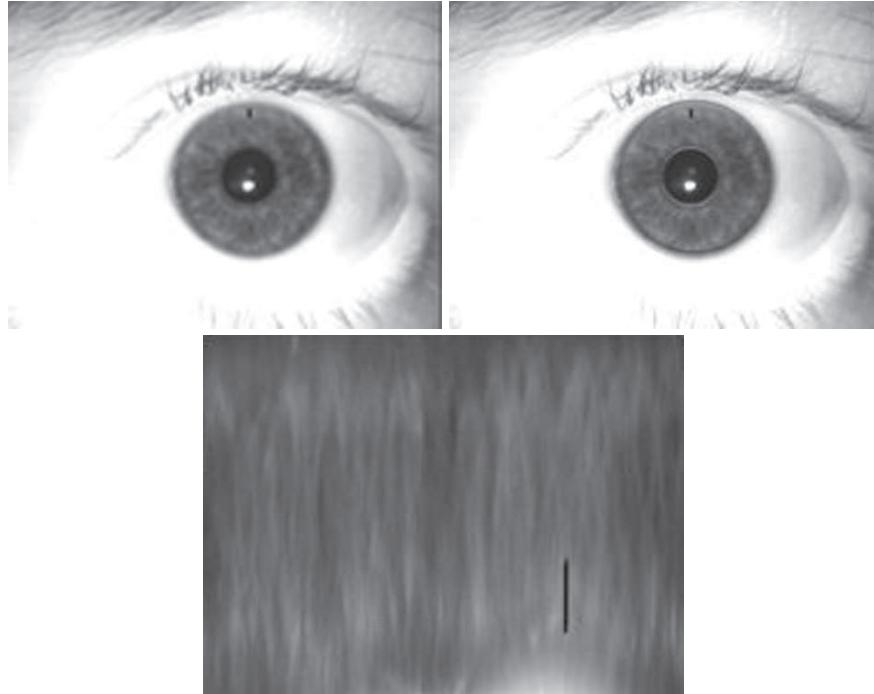
### **2.3 Segmentation, Template Generation, and Matching: Pixels to Identity**

As we can see from Table 2.2, there are many algorithms to choose from. There are several recent and excellent reviews of the state of iris recognition algorithms: Bowker [12], Vatsa [13], and Thornton [14]; in addition, the Tan group has presented local ordinal measures as a unifying principle for many of the algorithms [24].

The most thoroughly tested algorithm is the version of Daugman’s algorithm known as iris2pi. For purpose of this paper, we will briefly review iris2pi. Readers interested in alternative algorithms should consult the reviews noted and the original papers cited therein.

The first step is to identify the boundaries between the pupil and iris and the iris and sclera. In iris2pi this is done using an integro-differential operator discussed in Daugman’s 2004 paper [25]. The internal details of the segmentation are not revealed in the paper. This is not surprising. Segmentation is by far the hardest part of the template generation process and the internal details of segmentation for a number of algorithms are held as trade secrets.

Once the boundaries are established, the iris is unwrapped into a pseudo-polar coordinate system as seen in Fig. 2.6. The top edge of the image is the pupil boundary, the lower edge is the scleral boundary, and angle progresses zero to  $2\pi$  from left to right.



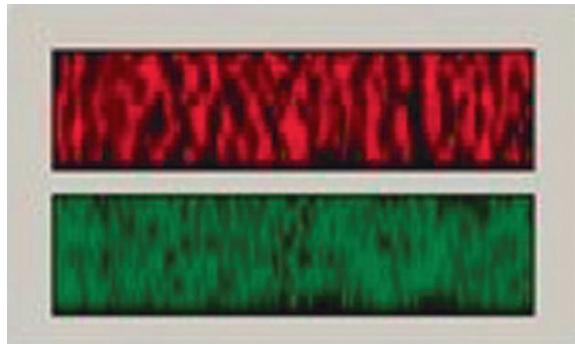
**Fig. 2.6** *Top left:* a good quality iris image with a defect introduced at 12:00. *Top right:* same image showing segmentation of pupil and iris. *Bottom:* the iris of the above image in normalized, pseudo-polar coordinates. The *upper edge* is the pupil, the *lower edge* is the sclera. The horizontal axis is angle, 0 to  $2\pi$  with zero at 03:00

An array of  $128 \times 8$  locations in angle and radius, respectively, are then selected. At each of these locations, a sine-like and a cosine-like Gabor wavelet is multiplied (in the dot product sense) against the normalized image. These sample the amount of sine-like and cosine-like signal in the bandwidth of the Gabor wavelet and over the spatial extent of the wavelet. The ratio of the two samples gives the tangent of the local phase angle. The local phase angle is then digitized to two bits and the bits are assembled into an array of  $256 \times 8$  bits. A mask array shadows the phase bit array. As the phase bits are computed, the corresponding mask bits are only set if the computation of the phase bits meets criteria for validity – these criteria include signal/noise and the presence of occlusions and specularities.

Figure 2.7 is a graphical representation of an iris template. The upper portion shows the phase bits and the lower portion the mask bits for the image of Fig. 2.6.

Two templates are compared by computing the fractional Hamming distance between the phase bit arrays of the two templates. For each bit in the phase array of the first template, we pick up the corresponding mask bit of its template, and the corresponding phase and mask bits of the second template. If either of the mask bits is not set, we do not have valid data for a comparison, so we do nothing and move to the next bit. If both mask bits are set, we increment a compared bits counter and

**Fig. 2.7** Graphical representation of a template. The *red bits* are the phase bits, the *green bits* are the mask bits. Angle runs horizontally, radius runs vertically



then compare the phase bits. If the bits differ, we then increment a bits differ counter, otherwise we do nothing more and move on to the next bit. When we have worked our way through all the bits, we have a count of the number of bits compared and the number of bits that differed. The ratio of the bits differed to the bits compared is the fractional Hamming distance.

Daugman has demonstrated that the phase bits are random, following a binomial distribution with approximately 250 degrees of freedom. For two unrelated iris images, 50% of the bits will disagree. For the binomial distribution of the phase bits, if less than  $\sim 33\%$  of the bits disagree, there is approximately a 1:106 chance that the level of agreement came about by chance. The alternative is that the templates came from a common source – the same eye.

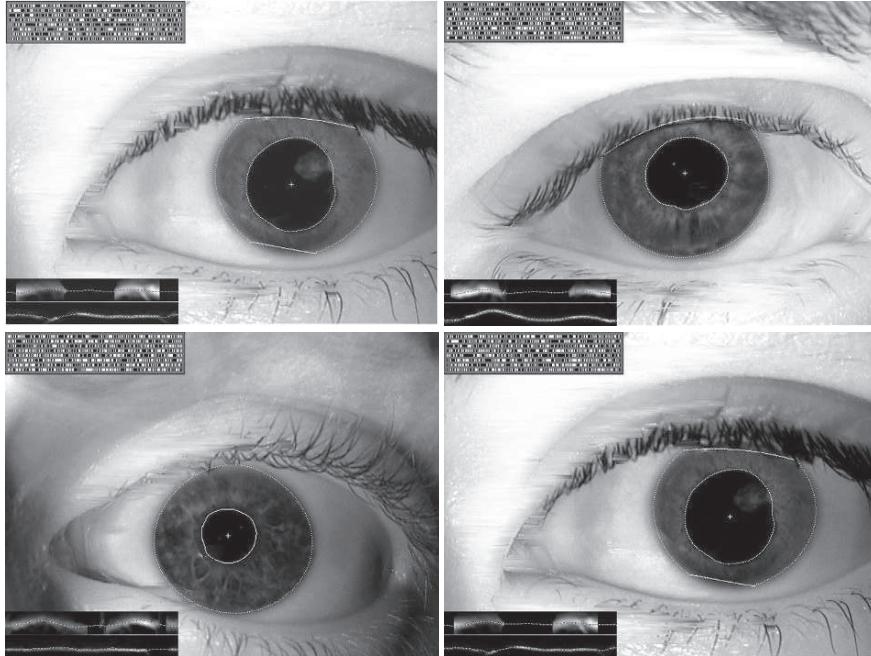
We left out an important issue – the first iris image might be rotated with respect to the second. To take this into account, fractional Hamming distance can be computed as a function of barrel shift between the two templates over a range of angles; the lowest fractional Hamming distance is then reported as a match score. This procedure has impact on the probability distributions of the match scores; the distribution of this comparison is an extreme value distribution rather than a binomial – the reader should consult the Daugman 2004 paper for details.

As noted earlier, segmentation is likely the hardest part of the iris recognition algorithm, particularly, when the iris or pupil is not circular. Daugman has developed techniques using more flexible contours to accommodate non-circularity [26]. Figure 2.8 presents four examples.

## 2.4 Issues for Iris Recognition at a Distance

The two most critical issues for any iris recognition system are resolution and signal/noise ratio. These become particularly challenging for iris recognition at a distance. These are important because they directly affect three crucial measures of system usability and performance:

- Capture volume
- Residence Time
- Sensitivity to subject motion



**Fig. 2.8** Four examples of segmentation on non-circular pupils; the pupils are also not concentric with the iris. The insets in the *top left* are the phase bits of the template. The insets at the *bottom* are diagnostics for the boundary fits. Courtesy of John Daugman

The issue of off-axis gaze is also exacerbated with iris recognition at a distance systems, because it is more difficult to get the subject to look directly at the camera when the camera is far away. We have already covered the subject of signal/noise – Eq. (2.5) is the most important result of that discussion. We now turn to the issues just raised.

#### 2.4.1 Capture Volume and Residence Time

The capture volume is the 3D volume in which the subject eye must be placed to enable the system to capture an image of sufficient quality for iris recognition. The transverse dimensions of this volume are set by the desired on-target resolution and the number of pixels in the sensor. For 100 micron resolution, a  $1024 \times 1024$  sensor will give a field of view of approximately  $10 \times 10$  cm. The depth of the capture volume is set by the depth of field of the lens system. Depth of field depends on the *F#* of the lens and is given by:

$$DOF \approx \frac{2Fc f^3 d}{f^4 - F^2 c^2 f^2} \quad (2.5)$$

where  $F$ ,  $f$ , and  $d$  are the F-number, focal length, and subject to lens distance;  $c$  is the diameter of the maximum acceptable circle of confusion; the maximum acceptable

circle of confusion is largest circle that a point on the object can map into in the image and still be regarded as in focus. In this formulation,  $c$  is expressed in object dimensions rather than image dimensions. This equation is valid for moderate to large ( $d \gg f$ ,  $d \ll \infty$ ) subject distances where  $f^2 > Fcd$ . The formula ignores diffraction limits. For a 1 m standoff, 100  $\mu\text{m}$  resolution,  $F\# = 11$ , and focal length of 100 mm, the depth of field is  $\sim 2$  cm. If  $c$  is small enough that the  $F^2c^2f^2$  term can be ignored, the equation reduces to

$$DOF \approx 2FMc = 2FM\Delta x_i \quad (2.6)$$

where  $M$  is the magnification of the lens system and we see that the depth of field depends primarily on the  $F\#$  of the lens, the magnification of the lens system and the desired, on-target, resolution.

Residence time is the time for which the subject eye must remain in the capture volume. This must be at least as large as the frame acquisition time for the sensor. The residence time is coupled to the signal/noise through the sensor shutter time in equation (2.5).

### 2.4.2 Image Resolution and Aberrations

In our earlier discussion of resolution, we only dealt with the sensor pixel size and the magnification of the lens. We neglected four factors that significantly limit resolution, particularly at large distances. These factors are lens aberrations, diffraction, motion blur, and atmospheric seeing. All of these effects are well known to our colleagues in astronomy.

The simple lens equation

$$\frac{1}{f} = \frac{1}{p} + \frac{1}{q} \quad (2.7)$$

is a good model for image formation, but it is only correct in the limit of paraxial rays – rays that are close to the optic axis of the system and which are near parallel to the optic axis. Non-paraxial rays do not focus at the same point as the paraxial rays, and the focus point varies with the location of the source of the rays. The image defects that these effects cause are called aberrations. For our purposes, the four most important aberrations are

- spherical aberration
- coma
- astigmatism
- curvature of field

Spherical aberration is a direct consequence of using spherical lens surfaces. The derivation of the simple lens equation for spherical surfaces assumes that the rays are paraxial. For an example of spherical aberration, consider rays of light from a point source that is on axis and that is so far away that the rays are essentially parallel.

A perfect lens would focus all the rays impinging on it to a point. A real lens, subject to spherical aberration will focus the rays near the optic axis to the focal point of the lens, but will focus rays further from the optic axis to a different point. Hence, the image will be a circle of finite size, rather than a point. The resolution of such a system is lower than that of a perfect system.

Coma results from variation in the effective focal length of the lens over its entrance pupil. For an example of coma, consider rays of light from a point source that is off axis and that is so far away that the rays are essentially parallel. A perfect lens would focus all the rays impinging on it to a point that is off axis. A real lens, subject to coma will focus the rays impinging on one edge of the lens to one point and those from the other edge of the lens to a different point. In astronomy, this effect causes stars to have a comet-like tail, hence the name. For a system subject to coma, the image will be of finite size, rather than a point. The resolution of such a system is lower than that of a perfect system.

Astigmatism results from variation in the effective focal length of the lens for rays which propagate in different planes.

All of these aberrations can be overcome using more complicated lens surfaces and/or combinations of lenses. However, a lens system that is perfectly corrected for aberrations will only be perfectly corrected at a single  $(p, q)$  and, for practical materials, at a single wavelength.

For practical materials, the index of refraction varies with wavelength and hence the focal length also varies with wavelength. This leads to chromatic aberrations. Since iris recognition is generally carried out with narrow band illumination, we will not pursue this form of aberration further, except to point out that a lens that focuses well in the visible may not focus as well in the near IR and that anti-reflection coatings that are optimized for the visible will necessarily be suboptimal in the near IR.

Optical systems based on reflection (mirrors) rather than refraction (lenses) do not depend on index of refraction and as a result do not have chromatic aberrations.

There are no simple equations for the effect of aberrations on resolution. The most straightforward way of ascertaining the effect and optimizing a system to minimize the effect is through the use of an optical ray tracing program such as ZEMAX [27].

Direct measurements of the optical transfer function of a system can also be useful. These measurements can be largely automated through the use of software packages such as IMATEST [28].

For a given focal length, aberrations tend to be larger for larger diameter (faster, lower F#) lenses.

### **2.4.3 Diffraction Limit**

Diffraction is a fundamental physical phenomenon that results from the interference of light waves. Optical interference is analogous to the interference of ripples on a pond. If we throw two stones in a pond the circular ripples spreading out from each

entry point will eventually intersect. When two peaks or two troughs coincide, we have constructive interference – the amplitude is higher than that of either of the ripples. When a peak and trough coincide they tend to cancel; we have destructive interference.

The light rays from one edge of a lens will interfere with those from the other edge of the lens. This interference, diffraction, sets a fundamental limit to the resolving power of a lens which is given by

$$\Delta x = \frac{\lambda d}{D} = \frac{\lambda d F}{f} \quad (2.8)$$

where  $\Delta x$  is the resolution,  $\lambda$  is the light wavelength,  $D$  is the lens diameter and  $d$  is the lens to object distance,  $f$  is the focal length of the lens, and  $F$  is the  $F$ -number of the lens.

For a 1 m standoff, 850 nm light, and a resolution of 100  $\mu\text{m}$ , the lens diameter is 8.5 mm. From our earlier calculation at 1 m standoff, the focal length of the lens was 100 mm. Hence the diffraction limited  $F\#$  is  $100/8.5 \sim 12$ . This diameter is well below the diameter needed for signal/noise considerations.

The diffraction limits and signal/noise considerations drive us toward smaller  $F\#$ 's. Depth of field and aberrations drive us toward larger  $F\#$ 's. In addition, there are practical size considerations of how low the  $F\#$  can be, particularly for long focal lengths. These considerations need to be balanced in the design of any system for distances greater than 1 m. There is no simple formula.

#### 2.4.4 Motion Blur

Motion blur occurs when the object (subject) moves significantly during the exposure time of the camera. This motion can result from physical motion of the object or motion of the camera. Let us consider what this means for longitudinal and transverse motion of the object and pointing stability of the camera.

A rough measure of motion blur is the amount of motion, in pixels, that occurs during the acquisition of an image. The resolution of a system will be reduced if that motion is of the order of 1 pixel or more. Longitudinal velocity,  $v_L$  causes blur by changing the effective magnification. The magnitude of the change is approximately the ratio of the distance moved to the camera standoff. For iris images with 100 pixels iris diameter, the magnification change should be below  $\sim 0.01$ . Let  $t$  be the shutter time of the camera and  $d$  be the camera, the constraint is then  $0.01 > v_L t/d$ . For a 10 ms shutter and a 10 m standoff, the longitudinal velocity must be less than 10 m/s – about as fast as the world record ( $\sim 10$  s) in the 100 m dash. At this velocity, the problem is not the motion blur, rather it is the length of time that the subject is resident in the capture volume.

Transverse velocity is a more significant issue. Motion perpendicular to the camera subject axis simply moves the pixels at the iris as much as the motion. For a

transverse velocity  $v_T$ , a shutter time  $t$ , and a subject resolution,  $\Delta x$  (m/pixel), we require  $v_T t < \Delta x$ . For 100 pixels/cm and a shutter time of 10 ms,  $v_T < 0.01$  m/s – much smaller than the longitudinal velocity limit.

Angular jitter of the camera translates directly into transverse velocity. Angular velocity  $\varpi$  (radians/s) corresponds to a transverse velocity of the subject of  $\varpi d$ , where  $d$  is the camera to subject distance. A maximum transverse velocity of 0.01 m/s at a distance of 1 m corresponds to an angular velocity of 0.01 radians/s. At 10 m, it is 0.001 radians/s. To put this in perspective, the minute hand on a clock moves at  $\sim 0.002$  radians/s.

We can improve the motion blur performance of a system by restricting subject motion, by reducing angular jitter of the camera, or by decreasing the shutter time. Reducing the shutter time has an adverse impact on signal/noise as can be seen in Eq. 2.5.

#### **2.4.5 Atmospheric Seeing**

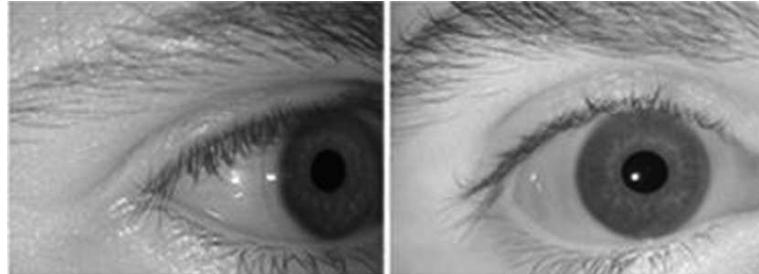
When you look down a road on a hot day, the cars approaching you appear to shimmer; when you look up at the stars at night, they seem to twinkle. Both of these are the result of fluctuations in the density of the air between you and the object. The density fluctuations cause fluctuations in the index of refraction of the air and those fluctuations cause fluctuations in the paths of the light rays traveling from object to sensor.

Dealing with these fluctuations is a hard problem. So hard that astronomers have resorted to putting telescopes in remote locations atop mountains and in orbit aboard satellites to avoid the problem. Roggeman [29] and Welsh have presented an excellent discussion of the issue. AOptix was founded by a group of astronomers familiar with these problems and is attempting to bring that expertise to bear on the issue of acquisition of iris images at a distance. We will discuss their efforts later.

As a practical matter, atmospheric seeing in an indoor environment, with reasonable precautions to avoid streams of hot or cold air impinging on the subject and camera, does not seem to be a problem at 1–5 m. Beyond 10 m it begins to be noticeable.

#### **2.4.6 Off-Axis Gaze**

Current commercially available iris recognition algorithms typically require on-axis (or nearly on-axis) iris images – the subject is looking in the direction of the camera, and usually standing a short distance away. As the subject moves further from the camera, the human factors engineering required to entice the subject to look directly at the camera becomes more difficult. In general, for all distances, near or far, it would be helpful to eliminate the requirement of “staring” at the camera. As



**Fig. 2.9** Off-axis image (*left*) and on-axis image (*right*) from the US Naval Academy database

user interfaces become more forgiving, it is increasingly important to be able to recognize off-axis images, such as the one in Fig. 2.9 from the US Naval Academy database [30].

Comparison of images from arbitrary angles generally requires additional pre-processing steps: ascertaining the gaze angles in each image, and transforming at least one of the images to correct for the deviated gaze relative to the other image, or some functionally equivalent process. Off-axis recognition is a relatively new undertaking within the iris recognition community; publications addressing off-axis segmentation/recognition began in earnest around 2006. As a partial list we mention the modifications of Daugman's algorithm for off-axis images [26] and several other approaches to off-axis segmentation and/or recognition by Sung [31], Zuo [32], Ross [33], and Schuckers [34].

In the literature to date, the most common approach to off-axis recognition is to start with segmentation, which is often accomplished by modifying on-axis techniques. On-axis methods typically exploit the roughly circular shape of the pupil/iris. The pupil (inner) boundary and limbic (outer) boundary may be detected as nearly concentric circles, and in certain algorithms the result is refined to account for irregularly shaped boundaries. This can be extended to the off-axis case, except that the circles must be generalized to something like ellipses. Common techniques to find circles/ellipses include edge detection from large gradients, active contours, and various combinations of thresholding and binary morphology (often employed to find the pupil). By whatever means off-axis segmentation is accomplished, the iris boundaries can provide major/minor axis lengths and orientation from which the horizontal and/or vertical gaze angle(s) can be computed. The off-axis image may then be rotated back to a corresponding on-axis image to generate the iris template. Alternatively, iris templates may be generated from the segmented off-axis image, in which case the pixel sampling step is used to account for the elliptical shape.

Many challenges remain in performing off-axis (or more generally, non-cooperative) recognition reliably. For segmentation, the problems that plague on-axis/cooperative cases are exacerbated in the off-axis/non-cooperative situations. The problems include inconsistent shapes of the pupil disk and iris ring; eyelid, and eyelash obscuration; specular reflections (which are more likely to intersect the pupil or limbic boundary in off-axis images); motion blur; and poor focus. For

the template generation process, considerations about the 3D shape of the iris and the effects of cornea refraction have only recently begun to be explored [30, 35]. It will not always be helpful or necessary to invoke sophisticated models of the eye. The simpler models may do just as well in many cases. But as a general principle, when it comes to the more challenging real-world iris recognition situations, even small potential improvements in eye modeling are worth investigating.

## 2.5 Current State of the Art for Iris at a Distance

At this time (summer 2008) there are no deployed commercial systems operating beyond  $\sim 1$  m. However, there are a number of prototypes that have been demonstrated and some of these are approaching commercial deployments.

The first demonstration of iris recognition at a distance was likely part of the human ID at a distance effort sponsored by DARPA and carried out by a research team at Sarnoff. In 2005, Fancourt [36] published a report on this system; it acquired iris images at 10 m using a custom designed telescope; the work had been previously described in an unpublished government report prepared in 2002. This system required the subject to be positioned in a chin rest and the camera and illumination system were built on a moderately large optical table. The system demonstrated feasibility of iris recognition at a distance, if not practicality.

In 2005, Sarnoff demonstrated its Iris on the Move<sup>TM</sup> portal system at the Biometrics Consortium Conference (Fig. 2.10). The system was built for the US Government and was described in detail by Matey [37, 38]. It was capable of acquiring iris images at distances of approximately 2–3 m with the subject walking at a normal pace. Several copies of the system were constructed for the US Government; one of those has been used to collect data for the Multi-Biometric Grand Challenge(MBGC) [39]. That data is available by request to the MBGC.



**Fig. 2.10** An Iris On the Move<sup>TM</sup> portal system.  
Picture courtesy of Dr. James Bergen of Sarnoff Corporation

**Fig. 2.11** A telescope-based iris at a distance system. The illumination system is mounted on a optical rail below the telescope. Picture courtesy of Dr. James Bergen of Sarnoff Corporation. The co-location of the illumination and sensor is an important feature



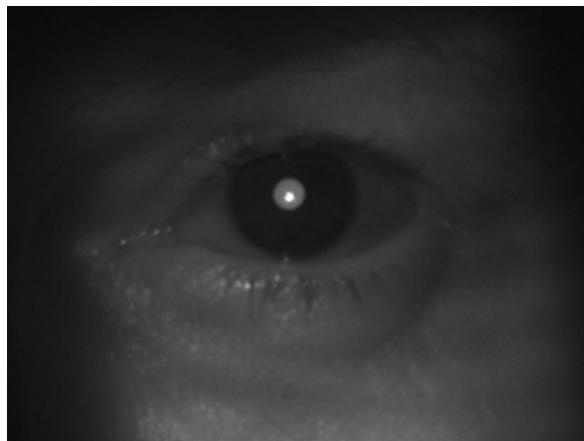
The long distance record for a system with co-located sensor and illumination was held, at least for a time, by another system that Sarnoff built for the US Government, see Fig. 2.11. This iris at a distance system is based on a Meade LX200-R F/10 8 inch reflecting telescope and a long distance illuminator consisting of an 850 nm LED focused on target to produce irradiance of approximately  $1 \text{ mW/cm}^2$ . To date, there has not been a publication describing the details of this device, but government sponsors of the research have indicated that the system generated useful images at 15+ m (Fig. 2.12).

In addition to these systems, AOptix [40], Honeywell [41], Global Rainmakers [42], and Retica [43] have all demonstrated systems working beyond 1 m – and there are likely other systems in development that are not known to the public due to the proprietary nature of commercial research in this area.

AOptix™ has incorporated adaptive optics techniques into their system, using closed-loop control to automate the subject finding and iris image acquisition process, see Fig. 2.13.

The system in Fig. 2.13 operates with a standoff distance of 1.5–2.5 m with a roughly trapezoidal capture volume that is ~1 m tall and 0.75 m wide at a 2 m standoff, for a capture volume of approximately  $0.75 \text{ m}^3$ . The system expects the subject to be standing still in that volume. AOptix demonstrated a prototype of their

**Fig. 2.12** An image captured using the system in Fig. 2.11. This was extracted from a larger image. There are approximately 200 pixels across the iris in this image captured at a distance beyond 15 m. Hamming distances below 0.25 can be routinely obtained with cooperative subjects and a commercial implementation of the Daugman algorithm



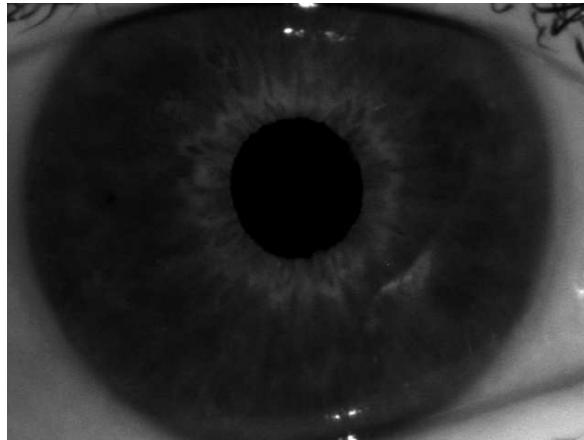
system at the 2007 Biometrics Consortium. An image from that prototype system is shown in Fig. 2.14.

AOptix™ has also worked on longer standoff systems. In June 2007, they demonstrated a system utilizing onboard coaxial imaging/illumination and adaptive optics to provide automated iris image capture at up to 18 m. The optical assembly can be seen in Fig. 2.15. An example image from that system can be seen in Fig. 2.16.



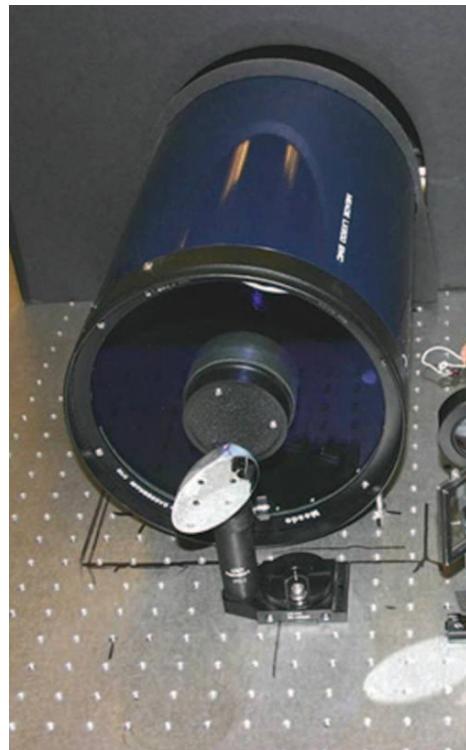
**Fig. 2.13** AOptix – waiting on a final picture. Image courtesy of Joey Pritikin of AOptix™

**Fig. 2.14** An example of the image produced by the prototype AOptix™ system. This image is 640×480; the iris is nearly 600 pixels wide. Image courtesy of Joey Pritikin of AOptix™. The image is cropped to the standard 640×480 size from a larger image

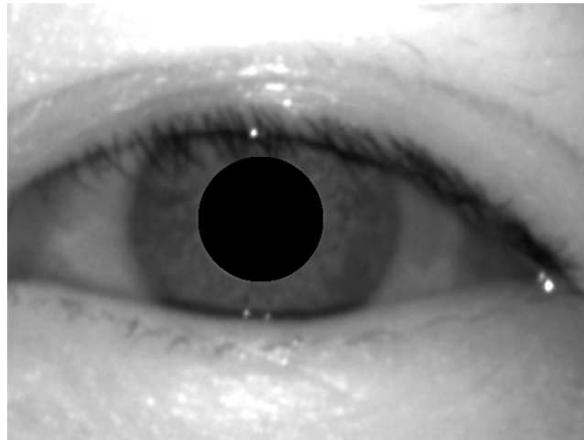


Honeywell developed their CFAIRS (combined face and iris recognition system) under contract to the US government. Their system is designed to acquire face and iris images from multiple subjects at distances from 1 to over 4 m. One of the prototype units can be seen in Fig. 2.17. An example of the iris images captured by CFAIRS can be seen in Fig. 2.18.

**Fig. 2.15** The 15 m system developed by AOptix. Image courtesy of Joey Pritikin of AOptix™



**Fig. 2.16** An image captured by the system in Fig. 2.15 at 15 m. The image is 640×480. The iris is a little less than 300 pixels wide. Image courtesy of Joey Pritikin of AOOptix™

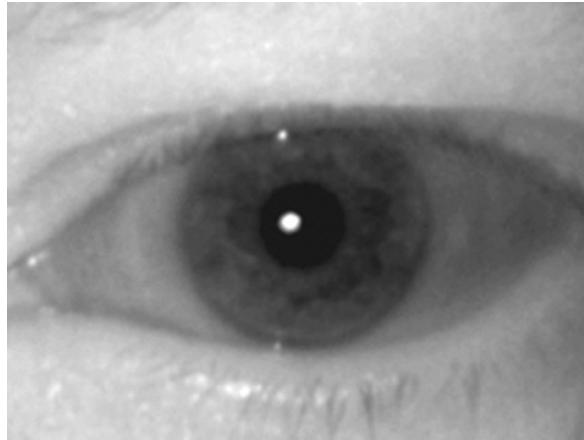


The HBOX™ from Global Rainmakers Inc. provides capability of capturing subjects in motion, similar to the capability of the Iris on the Move™ portal. The HBOX™ uses different package format, as seen in Fig. 2.19 to provide a relatively compact footprint and a straight user path with image capture at  $\sim 1.5$  m. The HBOX™ was demonstrated at the 2007 Biometrics Consortium. A still more compact version of the system, Hbox-Mini, designed for walk-up/drive-up applications was demonstrated at ISC West 2008.

**Fig. 2.17** Combined Face and Iris Recognition System (CFAIRS) built by Honeywell for face and iris acquisition and recognition at ranges from 1.5 to beyond 4 m. The device stands a bit less than 2 m high. Picture courtesy of Rand Whillock of Honeywell



**Fig. 2.18** An example iris image acquired by the system in Fig. 2.17 at 3 m. Picture courtesy of Rand Whillock of Honeywell



Bashir [44] published a description of the Retica Eagle-Eyes<sup>TM</sup> system with data that demonstrates iris recognition at 3.5–4.5 m. Retica states in their online literature that their latest Eagle-Eyes<sup>TM</sup> system provides reliable capture at 50 m. To the authors' knowledge, that claim has not yet been publicly demonstrated – though that could change before this book goes to press.

**Fig. 2.19** The HBOX<sup>TM</sup> over the door system. Picture courtesy of Keith Hanna of Global Rainmakers Inc. An affiliate of the Hoyos Group



**Table 2.8** Summary of current state of the art

< 1 m	Commercial systems currently available
1–5 m	Commercial systems likely available in the near future
5–10 m	Laboratory demonstrations
> 10 m	At present, an area for research

The current (summer 2008) state of the art, in our opinion, is summarized below. This could be radically changed by a single public demonstration (Table 2.8).

**Acknowledgments** Our thanks to Prof. Massimo Tistarelli for conducting the summer school at which Dr. Matey presented materials upon which this chapter is based and for editorial advice and assistance in the preparation of this chapter.

Our thanks to Joey Pritikin of AOptix, to Keith Hanna of Global Rainmakers/Hoyos Group, to Rand Whillock of Honeywell, and to Jim Bergen of Sarnoff for providing information regarding their systems.

The images that illustrate iris pathologies were provided by colleagues in ophthalmology and related fields. Our thanks to

- Dr. Dirk Werdermann, Ochsenfurt, Germany
- Prof. Bertil Damato, Ocular Oncology Service, St. Paul's Eye Unit Royal Liverpool University Hospital, Liverpool, for the iris tumor image.
- Prof. Dr. Georg Michelson, Verlag Online Journal of Ophthalmology

Our thanks to Steve McCurry and his staff for permission to use two of his outstanding images of Sharbat Gula.

Prof. Matey extends his thanks to colleagues at the Sarnoff Corporation for their collaborations over the past 10 years that made his summer school presentation and the resulting chapter possible.

Finally, our thanks to Prof. John Daugman for countless interesting and informative discussions, for several of the images in this chapter and for a helpful critique of the pre-publication manuscript.

## Proposed Questions and Exercises

### *Sensitive Dependence on Initial Conditions*

Take several sheets of paper from a fresh ream of paper and tear the sheets one at a time, trying to subject them to exactly the same tearing conditions. Compare the tears – can you match up half of one sheet with half of another?

### *Intra- and Inter-class Variability*

Download an iris image database. Compare the images intra- and inter- class.

Have a classmate select two images each from 10 subjects and provide them to you without identifying information. Match the images by hand. Estimate the probability of a false-match. Compare your results with the known (to your classmate) truth.

Get a copy of Steve McCurry's book Portraits (Phaidon Press, 1999). Compare the images on the front and back cover by hand. Make an estimate of the probability that the two images are of the same person.

### ***Image Collection and Time Variation***

On a table or workbench – camera (e.g., Canon A550) + chin rest + remove IR filter if possible. Get a visible block/IR pass filter AND a visible pass/IR block from Edmund Optics. 100 W incandescent lamp at  $\sim 1$  m. Camera at 1 m. 100 pixels across the iris or more. Full resolution, raw images. Run camera at fixed exposure and fixed aperture and note the exposure and aperture for each eye. Make pictures today, tomorrow, next week and next month. Compare the images. Are they the same? You want to be able to see eyelashes clearly delineated and to see significant details in the iris – you may need to saturate the image in the regions outside the iris.

### ***Segmentation***

Write a computer program that displays an iris image from a file and that allows you to mark the edges of the pupil/iris and iris/sclera boundaries with the mouse. Using the marked locations, fit suitable contours to the pupil/iris and iris/sclera boundaries. Extract the iris and save it to a file (in rectilinear or pseudo-polar coordinates) for later processing. Compute the average and standard deviation of the pixel values in the iris. Save the boundary information and the statistics to a text file.

### ***Image Metrics***

Acquire images of both of your eyes and those of several colleagues using a digital camera with the flash turned off – in the visible. Make sure that all of the images are collected under the same lighting conditions. Make a picture of a piece of matte (not glossy) white paper overlaid with a piece of matte black paper so that you have an image that is roughly half black and half white. Assume that the black paper has albedo zero and the white paper has albedo 1. Measure the albedo of the irises in the images you collected by comparison with your black/white standard. Compute the average and standard deviations of the pixels in the iris of each eye. Compute the normalized standard deviation – standard deviation divided by average. Plot the results as a function of eye color. Repeat with IR light. Taking the normalized standard deviation as a contrast measure, how does the contrast vary with eye color in the visible ? in the IR?

Are your results consistent with the literature ?

### Iris Algorithms

Obtain an iris recognition algorithm – commercial – university/open source. Compare the images from previous exercises using the algorithm. Construct an imposter distribution and an authentics distribution. Are your distributions consistent with those published for the algorithm you chose?

### Safety

Obtain a current copy of the ACGIH TLVs and BEIs. From the tables and formulae, plot the non-laser TLVs for radiance and irradiance for 850 nm photons as a function of exposure time. Plot the laser irradiance TLV for 850 nm photons as a function of exposure time. How would you insure that your calculations are correct? What steps would you take to verify the safety of an experiment based on such calculations ? The eyeball you save may be your own.

How are the results modified for the case of repetitively pulsed sources?

### References

1. 4th Summer School for Advanced Studies on Biometrics for Secure Authentication. <http://biometrics.uniss.it/>
2. Chaterjee, S. and Yilmaz, M., “Chaos, Fractals and Statistics” Statistical Science. 1992, 7: 49–121.
3. Mansfield, T., Kelly, G., Chandler, D., and Kane, J., “Biometric Product Testing Final Report”, CESG Contract X92A/4009309, Centre for Mathematics & Scientific Computing, National Physical Laboratory, Queen’s Road, Teddington, Middlesex TW11 0LW.
4. Daugman, J., “Recognizing Persons by Their Iris Patterns” in *Biometrics: Personal Identification in Networked Society*, A. Jain, R. Bolle, & S. Pankanti, eds. (Springer, New York, 2006) ISBN 0-387-28539-3.
5. Newman, C., “A Life Revealed” in *National Geographic Magazine*, April 2002.
6. Daugman, J., <http://www.cl.cam.ac.uk/~jgd1000/afghan.html>.
7. <http://www.genetests.org>
8. Jensen, B., *Iridology Simplified 5th ed.* (Bernard Jensen International, Escondido, CA, 1980) ISBN 096083608-X
9. Simon, A., Worthen, D.M., and Mitas, J.A., 2nd. “An Evaluation of Iridology”. JAMA. 1979, September 8;242(13):1385–1389.
10. Bertillon, A., “La couleur de L’Iris”. Annales de Demographie Internationale. 1886, 7: 226–246.
11. Daugman, J., “Biometric Personal Identification System Based on Iris Analysis.” U.S. Patent No. 5,291,560 issued 1 March 1994.
12. Bowyer, K.W., Hollingsworth, K., and Flynn, P.J., “Image understanding for iris biometrics: A survey”, Computer Vision and Image Understanding. 2008, 110:281–307.
13. Vatsa, M., Singh, R., and Gupta, P., “Comparison of iris recognition algorithms,” Intelligent Sensing and Information Processing, 2004. Proceedings of International Conference on, pp. 354–358, 2004.
14. Thornton, J., Savvides, M., and Kumar, B.V.K., “An Evaluation of Iris Pattern Representations,” Biometrics: Theory, Applications, and Systems, 2007. BTAS 2007. First IEEE International Conference on, pp. 1–6, 27–29 September 2007.

15. "How the Afghan Girl was Identified by Her Iris Patterns", <http://www.cl.cam.ac.uk/~jgd1000/afghan.html>
16. IrisCode™ is the name for the iris template generated by a particular group of Daugman algorithm variants.
17. Daugman, J. "Probing the Uniqueness and Randomness of IrisCodes: Results from 200 billion iris pair comparisons." Proceedings of the IEEE, vol. 94, no. 11, 2006, pp. 1927–1935.
18. ISO/IEC JTC 1/SC 37 N 504, "Biometric Data Interchange Formats – Part 6: Iris Image Data".
19. American Conference of Governmental Industrial Hygienists.
20. ACGIH, 2008 TLVs and BEIs, ISBN 978-1-882417-79-7. Available at [www.acgih.org](http://www.acgih.org).
21. Diemer, S., "Safety Aspects for Light Emitting Diodes (LEDs) and Laser Pointers: Current Positions of the ICNIRP", Proceedings of the International Laser Safety Conference, 1999.
22. International Light, 10 Technology Drive, · Peabody, MA 01960. [www.intl-lighttech.com](http://www.intl-lighttech.com).
23. Grother, P., Iris Exchange Evaluation (IREX) Evaluation 2008, Published online at <http://iris.nist.gov/irex/index.html>
24. Sun, Z., Tan, T., and Wang, Y., "Robust encoding of local ordinal measures: A general framework of iris recognition", in: Proc. BioAW Workshop, 2004, pp. 270–282.
25. Daugman, J., "How Iris Recognition Works", IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, January 2004.
26. Daugman, J., "New Methods in Iris Recognition," Systems, Man, and Cybernetics, Part B, IEEE Transactions on , vol. 37, no. 5, pp. 1167–1175, October 2007.
27. ZEMAX Development Corporation, 3001 112th Avenue NE, Suite 202, Bellevue, WA 98004-8017. [www.zemax.com](http://www.zemax.com)
28. Imatest LLC, 3478 16th Circle, Boulder, CO 80304. [www.imatest.com](http://www.imatest.com)
29. Roggemann, M.C. and Welsh, B.M., Imaging Through Turbulence (CRC Press, Boca Raton, 1996), ISBN 0-8493-3787-9.
30. Kennell, L., Broussard, R.P., Ives, R.W., and Matey, J.R. "Preprocessing of Off-Axis Iris Images for Recognition", SPIE Europe Security & Defense Conference, Cardiff, Wales, UK, September 2008.
31. Sung, E., Chen, X., Zhu, J., and Yang, J., "Towards Non-cooperative Iris Recognition Systems," Proceedings of the 2002 Seventh IEEE International Conference on Control, Automation, Robotics, and Vision, pp. 990–995, December 2002.
32. Zuo, J., Kalka, N.D., and Schmid, N.A., "A Robust Iris Segmentation Procedure for Unconstrained Subject Presentation," Special Session on Research at the Biometric Consortium Conference, 2006 Biometrics Symposium.
33. Ross, A. and Shah, S., "Segmenting Non-ideal Irises Using Geodesic Active Contours," Special Session on Research at the Biometric Consortium Conference, 2006 Biometrics Symposium.
34. Schuckers, S., Schmid, N.A., Abhyankar, A., Dorairaj, V., Boyce, C.K., and Hornak, L.A., "On Techniques for Angle Compensation in Non-ideal Iris Recognition," IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 37, no. 5, October 2007.
35. Price, J.R., Gee, T.F., Paquit, V., and Tobin, K.W., Jr., "On the Efficacy of Correcting for Refractive Effects in Iris Recognition," IEEE Conference on Computer Vision and Pattern Recognition, June 2007.
36. Fancourt, C., Bogoni, L., Hanna, K., Guo, Y., Wildes, R., Takahashi, N., and Jain, U. "Iris Recognition at a Distance", Proc. 5th Int. Conf. on Audio and Video-Based Biometric Person Authentication, 2005.
37. Matey, J.R., Naroditsky, O., Hanna, K., Koleczynski, R., LoIacono, D., Mangru, S., Tinker M., Zappia, T., and Zhao, W.Y., "Iris on the Move™: Acquisition of Images for Iris Recognition in Less Constrained Environments". Proc. IEEE. 94(11), pp. 1936–1947, November 2006.
38. Matey, J.R., Ackerman, D., Bergen, J., and Tinker, M., "Iris Recognition in Less Constrained Environments", in Advances in Biometrics, Sensors, Algorithms & Systems, N. Ratha and V. Govindaraju, eds. (Springer-Verlag, London, 2008) ISBN 978-1-84628-920-0.
39. Dr. Jonathon Phillips, MBGC Program Manager, National Institute of Standards and Technology (NIST), 100 Bureau Drive, Stop 8940, Gaithersburg, MD 20899-8940. [mbgc@nist.gov](mailto:mbgc@nist.gov).

40. Northcott, M.J. and Graves, J.E. “Iris Imaging Using Reflection from the Eye” US Patent Application 20080002863, January 3, 2008.
41. Geterman, G., Jacobsen, V., Jelinek, J., Phinney, T., Jamza, R., Ahrens, T., Kilgore, G., Whillock, R., and Bedros, S. “Combined Face and Iris Recognition System”, US Patent Application 20080075334, March 27, 2008.
42. Global Rainmakers, affiliate of Hoyos Group, 10 E 53rd St, 33rd Floor, New York, NY 10022. [www.hoyosgroup.com](http://www.hoyosgroup.com)
43. Retica Systems, Inc., 201 Jones Road, Third Floor, West Waltham, MA 02451. [www.retica.com](http://www.retica.com)
44. Bashir, F., Casaverde, P., Usher, D., and Friedman, M., “Eagle-Eye: A System for Iris Recognition at a Distance,” 2008 IEEE Conference on Technologies for Homeland Security, pp. 426–431, 12–13 May 2008.

# **Chapter 3**

## **View Invariant Gait Recognition**

**Richard D. Seely, Michela Goffredo, John N. Carter, and Mark S. Nixon**

**Abstract** Recognition by gait is of particular interest since it is the biometric that is available at the lowest resolution, or when other biometrics are (intentionally) obscured. Gait as a biometric has now shown increasing recognition capability. There are many approaches and these show that recognition can achieve excellent performance on current large databases. The majority of these approaches are planar 2D, largely since the early large databases featured subjects walking in a plane normal to the camera view. To extend deployment capability, we need viewpoint invariant gait biometrics. We describe approaches where viewpoint invariance is achieved by 3D approaches or in 2D. In the first group, the identification relies on parameters extracted from the 3D body deformation during walking. These methods use several video cameras and the 3D reconstruction is achieved after a camera calibration process. On the other hand, the 2D gait biometric approaches use a single camera, usually positioned perpendicular to the subject's walking direction. Because in real surveillance scenarios a system that operates in an unconstrained environment is necessary, many of the recent gait analysis approaches are orientated toward view-invariant gait recognition.

### **3.1 Introduction**

Much research has been done into identifying subjects by how they walk from 2D video data [32]; with publications dating as far back as 1994 [33]. Gait has even found forensic use, securing the conviction of a bank robber based on his gait [25].

Typically gait analysis techniques can be described as either model based or appearance based. Model-based analysis usually involves fitting a model representing various aspects of the human anatomy to the video data then extracting and analyzing its parameters. Appearance-based analysis often involves the analysis of a subject's silhouette shape and how it varies over time, or analysis can be carried

---

R.D. Seely  
University of Southampton, Southampton, UK  
e-mail: rds06r@ecs.soton.ac.uk

out in a more direct fashion considering the statistical distribution of pixels in the silhouette and how it varies over time.

The DARPA HumanID project spurred the creation of several large gait data sets [37, 44], each having over 100 subjects and a range of covariates such as different surfaces and footwear. Researchers were able to achieve extremely promising recognition rates on both data sets [41, 50], providing further proof that gait is a valid and usable biometric for identification purposes.

Most of the current approaches to gait biometrics process silhouettes which are obtained from images derived when a subject walks in a plane normal to the camera view. This derives a recognition metric which incorporates body shape and motion, but the process is largely dependent on the viewpoint. There is a much smaller selection of model-based approaches and these have some small angle invariance. There have been studies which have been aimed to improve viewpoint invariance [6, 23] using geometry considerations. In this chapter, we describe 2D view invariant approaches which use geometry considerations to derive a viewpoint invariant signature; we also describe developments in 3D approaches which are the complementary approach having implicit viewpoint independence since arbitrary views can be synthesized from the 3D human figure.

## 3.2 2D View Invariant Gait Recognition

### 3.2.1 Overview

Markerless 2D view independent gait identification is a recent research area and the approaches found in literature can be broadly divided into pose-based and pose-free methods. The pose-based approaches aim at synthesizing the lateral view of the human body from any other arbitrary views [14, 20, 22]. On the other hand, the pose-free ones extract some gait parameters which are independent of the human pose [3, 8, 17, 19, 53, 54].

### 3.2.2 Pose-Based Methods

The methods which generate the lateral view from data acquired at different arbitrary views are the most recent approaches to 2D viewpoint independent gait recognition. This choice is justified by the fact that the lateral view has proven the recognition capability in a great number of approaches [6, 16, 18, 32, 43, 51, 58, 59].

The biometrics research group of the University of Southampton has focused attention on 2D view invariant gait recognition from 1999 [7] where a trajectory-invariant gait signature was presented. The method of Carter and Nixon corrects the variations in gait data by knowing the walking trajectory and modeling the thigh as a simple pendulum.

The approach was then reformulated by Spencer and Carter [47] to provide a pose invariant biometric signature which did not require knowledge of the subject's trajectory. Results on synthesized data showed that simple pose correction for geometric targets generalizes well for objects on the optical axis.

More recently, these techniques have been refined for the analysis on walking subjects and applied with success on a larger number of video sequences acquired at six different point views [14, 46]. View independent gait analysis aims at synthesizing the lateral view of a walking subject without camera calibration starting from the successful results on gait biometrics from a lateral view [5]. The dynamic parameters used for biometrics in [5] are the frequential characteristics of the angles that the upper and lower legs form with the vertical axes. Therefore, the viewpoint independent method aims to synthetize the projection of the principal joints (hips, knees, ankles) on a lateral plane from their positions in the image plane.

The method is based on three main assumptions [31, 56]:

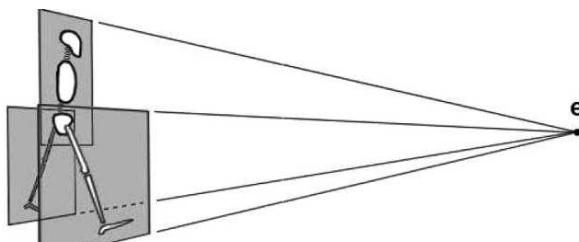
- the nature of human gait is cyclic;
- the distances between the bone joints are invariant during the execution of the movement;
- the articulated leg motion is approximately planar, since almost all of the perceived motion is contained within a single limb swing plane.

Considering a subject walking along a straight line, the multiple periods of linear gait motion appear analogous to a single period viewed from many cameras related by linear translation. Following this rationale, the positions of the points of interest, i.e., the leg joints, lie in an auto-epipolar configuration consistent with the imaged motion direction. The epipole is thus estimated by computing the intersection of the set of lines formed by linking the corresponding points of interest in each phase of the gait cycle (Fig. 3.1). Let  $\mathbf{j}_i^\ell$  be the set of joints positions for each leg  $\ell = \{1, 2\}$  at the  $i$ th frame in the image reference system. The relationship between  $\mathbf{j}_i^\ell$  and the corresponding positions in the worldspace is

$$\mathbf{j}_i^\ell \times \mathbf{P}_i \cdot \mathbf{J}^\ell = 0 \quad (3.1)$$

where

$$\mathbf{P}_i = [\mathbf{R}_e^T \ -i\mathbf{e}_0] \quad (3.2)$$



**Fig. 3.1** Epipolar configuration of the leg joints  
(source: Spencer and Carter [46])

and  $\mathbf{R}_e^T$  is the rotation matrix for aligning the epipolar vector  $\mathbf{e}_0$  with the horizontal axis X. Then,

$$\mathbf{j}_i^\ell = [\mathbf{R}_e^T \ -i\mathbf{e}_0] \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H}_v^{-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & \mathbf{H}_v \end{pmatrix} = \mathbf{H} \cdot \mathbf{j}^\ell \quad (3.3)$$

having expressed the limb plane transformation matrix with  $\mathbf{H}_v$  so that the two cross section plane lines are parallel with the vertical axis Y, centred and normalized with respect to Y and Z axes. By assuming the lengths of the articulated limbs  $\mathbf{D}_\ell^2 = \Delta\mathbf{j}_i^{\ell T} \Delta\mathbf{j}_i^\ell$  are constant over all the frames, the pose difference vectors for the limb segments at two consecutive frames,  $\Delta\mathbf{j}_i^\ell$  and  $\Delta\mathbf{j}_{i+1}^\ell$ , are related by

$$\Delta\mathbf{j}_i^{\ell T} \cdot \mathbf{H}^T \cdot \mathbf{H} \cdot \Delta\mathbf{j}_i^\ell = \Delta\mathbf{j}_{i+1}^{\ell T} \cdot \mathbf{H}^T \cdot \mathbf{H} \cdot \Delta\mathbf{j}_{i+1}^\ell \quad (3.4)$$

After recovering the fronto-parallel structure of subject gait (a more detailed description can be found in [14]) the representation of the leg joints function  $[\mathbf{J}_x^\ell(t), \mathbf{J}_y^\ell(t)]$  is found by fitting a modified Fourier series to the data with fixed fundamental frequency  $f_0$  and period  $T$ . In accordance with [11, 57], in fact, the hip rotation can be modeled by a simple pendulum

$$\mathbf{J}_x^\ell(t) = v_x t + \sum_{k=1}^n A_k \cos \left( 2\pi k f_0 \left( t + \frac{(\ell-1)T}{2} \right) + \phi_k \right) + \mathbf{J}_{x0}^\ell \quad (3.5)$$

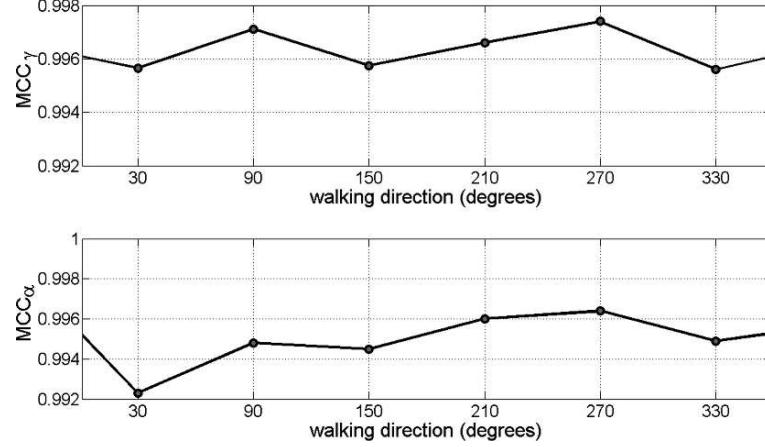
analogously for  $\mathbf{J}_y^\ell(t)$ . Considering  $n = 5$  as the sufficient number of harmonics for describing the human walking [10], the projection of the leg joints on the 3D anterior-posterior plane can be expressed in the following way

$$\check{\mathbf{J}}^\ell(t) = [h_1 \ h_2 \ h_3] g \left( t + \frac{(\ell-1)T}{2} : f_0, \mathbf{D}_\ell, v_x, v_y, F \right) \quad (3.6)$$

where  $g(t)$  is the bilateral Fourier series function with coefficients  $F$  and  $h$  being the values of the inverse normalization transform matrix. Equation (3.6) results from an optimized procedure where the coordinate positions of limb points are computed and fitted to a linear velocity model with horizontal and vertical velocities equal to  $V_x$  and  $V_y$ .

Therefore, starting from a video sequence from a single camera and without calibration, the method estimates the gait parameters projected on the lateral plane and their alignment makes them independent of the point view and allows their use for gait identification.

The first experimental tests of the method have been oriented toward the evaluation of its accuracy and therefore it has been applied on reliable and known limb trajectories extracted with reflective markers on the lower limbs. Thirty video sequences along six different camera views have been analyzed and the mean correlation coefficient (MCC) along the  $i$  ( $i = 1, \dots, 6$ ) directions has been achieved

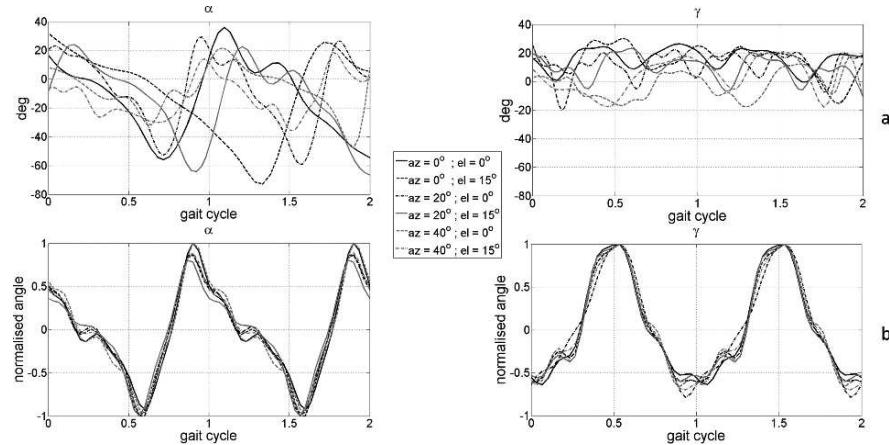


**Fig. 3.2** Tests with reflective markers: mean correlation coefficient along different walking directions (source: Goffredo et al. [14])

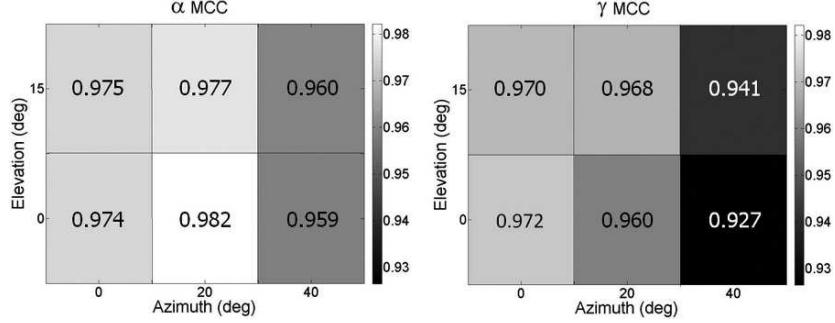
in order to quantify the angle trends matching after the view synthesis. Figure 3.2 shows the MCC for the thigh ( $\gamma$ ) and shin ( $\alpha$ ) angles along different walking directions.

Furthermore, in [14] the method sensitivity with respect to the image resolution, camera frame and SNR of Gaussian noise added to the joints trajectories, has been extracted and the minimum MCC obtained was 0.8.

More recently, the new method for viewpoint independent gait analysis has been applied on video sequences without any markers attached to the subjects [13]. A markerless gait estimation method has been designed and the lower limbs' pose has been extracted over time. Therefore, the lateral view synthesizer has been tested in a



**Fig. 3.3** Markerless tests: hip ( $\gamma$ ) and knee ( $\alpha$ ) angles in the different camera positions: (a) unprocessed; (b) corrected (source: Goffredo et al. [13])

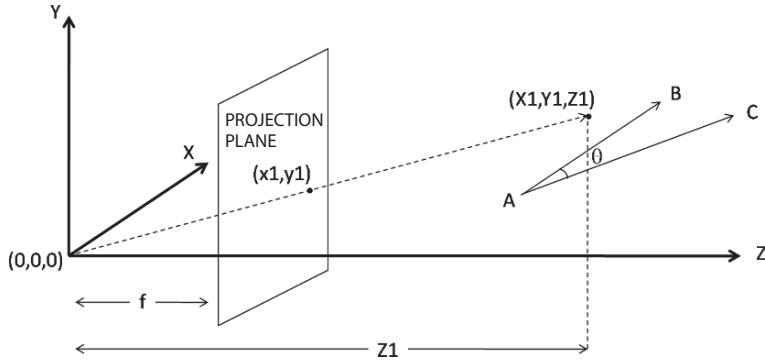


**Fig. 3.4** Markerless tests: mean correlation coefficients (MCC) with respect to the different camera positions (source: Goffredo et al. [13])

less-constrained environment. Figure 3.3(a) shows an example of the variations of  $\gamma$  and  $\alpha$  during two gait cycles for the six different camera positions. Predictably, the angles' trends are influenced by the subject's pose respect and they cannot be used directly for biometric identification. The angles after the application of the point view correction algorithm are shown in Fig. 3.3(b). The slight variations between the resulting traces in Fig. 3.3(b) are consistent with intra-subject variation between the video acquisitions. Figure 3.4 shows the variation of the MCCs with respect to the camera azimuths and elevations. The results, with a mean value of 0.919, are consistent with the value of MCC obtained using reflective markers [14]. Therefore, the correlation values obtained with the angles' trends appear particularly encouraging for its application in the wider context of gait biometrics.

In 2003, the group at the University of Maryland developed a gait recognition algorithm showing that if a person is far enough from a single camera, it is possible to synthesize the lateral view from any other arbitrary view by knowing the camera parameters [20].

Considering a subject walking with a translational velocity  $\mathbf{V} = [v_x, 0, v_z]^T$  along a straight line which subtends an angle  $\theta$  with the image plane (AC in Fig. 3.5),



**Fig. 3.5** Imaging geometry (source: Kale et al. [20])

if the distance  $Z_0$  of the person from the camera is much larger than the width  $\Delta Z$  of the person, then it is reasonable to approximate the actual 3D object as being represented by a planar one. Therefore, the angle  $\theta$  in the 3D world can be accurately estimated in two ways: (1) by using the perspective projection matrix; and (2) by using the optical flow-based SfM equations. Both methods have been proposed in [20] and, by knowing the camera calibration parameters, the lateral view has been synthesized.

In the perspective projection approach, the 3D line

$$Z = \tan(\theta) X + Z_0 \quad Y = k \quad (3.7)$$

that is at a height  $k$  from the ground plane and parallel to it corresponds to a straight line in the image plane

$$y = \frac{kf}{Z_0} - k \frac{\tan(\theta)}{Z_0} x \quad (3.8)$$

where  $f$  is the focal length of the camera and

$$x = f \frac{X}{Z_0 + \tan(\theta) X} \quad y = f \frac{Y}{Z_0 + \tan(\theta) X} \quad (3.9)$$

Therefore, if the slope of the line in the image plane is  $\alpha$ , the angle  $\theta$  can be estimated with

$$\tan(\theta) = \frac{1}{K} \tan(\alpha) \quad (3.10)$$

where  $K$  and  $f$  are obtained from the camera calibration procedure.

Conversely, the optical flow-based approach estimates the angle  $\theta$  with

$$\cot(\theta) = \frac{c(x, y) - \cot(\alpha(x, y))}{m(y, f)} \quad (3.11)$$

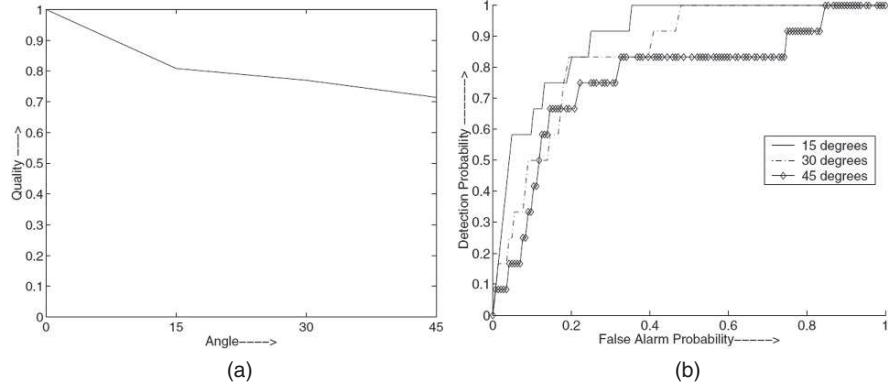
where  $c(x, y)$  and  $m(y, f)$  can be obtained from calibration data and

$$\cot(\alpha(x, y)) = \frac{p(x, y)}{q(x, y)} \quad (3.12)$$

where  $p(x, y)$  and  $q(x, y)$  represent the horizontal and vertical velocity fields of a point  $(x, y)$  in the image plane.

Therefore, any point  $(x_\alpha, y_\alpha)$  on the subject's image, walking at an angle  $\alpha$  to the image plane, can be projected on the lateral view:

$$x_0 = f \frac{x_\alpha \cos(\alpha) - f \sin(\alpha)}{-x_\alpha \sin(\alpha) + f \cos(\alpha)} \quad y_0 = f \frac{y_\alpha}{-x_\alpha \sin(\alpha) + f \cos(\alpha)} \quad (3.13)$$



**Fig. 3.6** (a) Quality degradation of the synthesized images as a function of angle; (b) ROC curves (source: Kale et al. [20])

After a camera calibration procedure for the parameters  $f$ ,  $K$ ,  $c$  and  $m$ , the method has been tested on 12 people walking along straight lines at different values of  $\theta = 0, 12, 30, 45$  and  $60^\circ$ . Figure 3.6(a) shows the quality degradation of the synthesized images as a function of angle  $\theta$ .

In order to study the performance of gait recognition on the synthesized images and keeping in view the limited quantity of training data, the DTW algorithm [22] has been used for gait recognition. Considering a gallery of people walking at lateral view, the video sequences where people walks at arbitrary angles  $\theta$  have been chosen as probes and the receiver operating characteristic (ROC) has been computed for each  $\theta$  (Fig. 3.6 b).

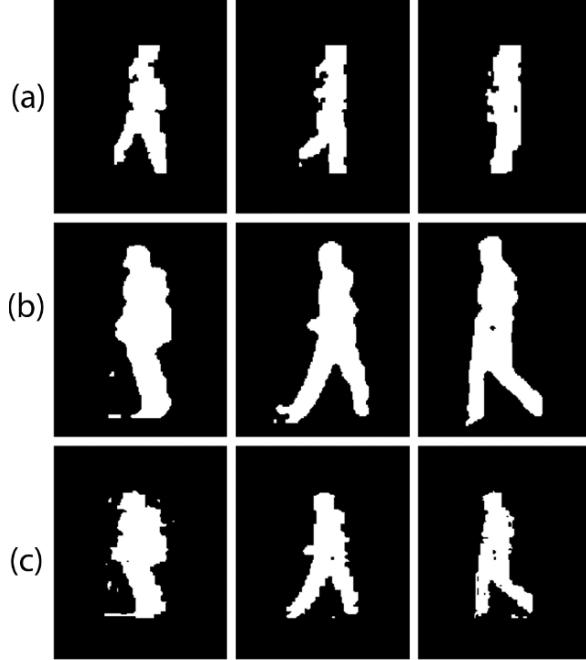
More recently, further experimental tests have been done by the same group of the University of Maryland [21] and the recognition performance has been extracted on two publicly available gait databases: the NIST database [40] and the CMU database [15]. Some of the results of the lateral view synthesis are shown in Fig. 3.8. The gait recognition is depicted in Fig. 3.7 where a comparison with the non-normalized images has been included for both the NIST and CMU databases.

### 3.2.3 Pose-Free Methods

While the pose-based approaches aim at synthesizing the lateral view of the human body from any other arbitrary views, the pose-free ones extract some gait parameters which are independent of the human pose.

One of the first approaches has been presented by Johnson and Bobick [19], who developed a multi-view gait recognition method using static body parameters. The technique does not analyse the entire dynamics of gait, but uses the action of walking to extract some identifying properties of the individual behavior. Human body limbs are first labeled by analysing the binary silhouette of the subject. Subsequently, the head, pelvis and feet positions are extracted following the geometrical proportions of human body segments. Moreover, two data compensations have been considered:

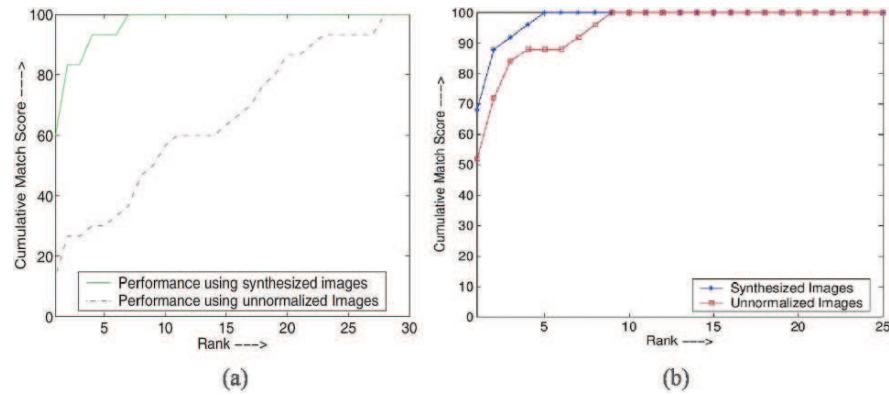
**Fig. 3.7** Sample images from the NIST database: (a) gallery images of person walking parallel to the camera; (b) space un-normalized images of person walking at 33° to the camera; (c) space synthesized images for (b) (source: Kale et al. [21])



the depth compensation and the camera view factor. The depth compensation is a conversion from pixels to centimetres via an hyperbola function dependent on  $y_b$ , the vertical location of the subject's feet:

$$CF(y_b) = \frac{A}{B - y_b} \quad (3.14)$$

where  $A$  is the vertical distance between the ground and focal point times the focal length and  $B$  is the vertical component of the optical centre.  $A$  and  $B$  are estimated



**Fig. 3.8** Recognition results on (a) NIST database; (b) CMU database (source: Kale et al. [21])

by fitting the conversion factor CF to some known subject locations by knowing the subject's height in centimetres.

After body labeling and depth compensation, a 4D walk vector is computed as

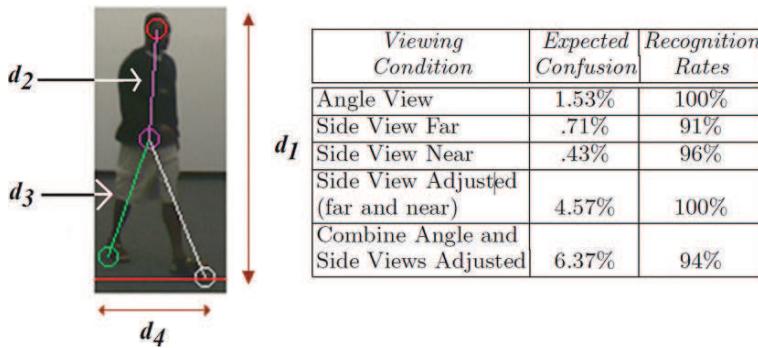
- $d_1$ : height of the bounding box around the silhouette;
- $d_2$ : distance (L2 norm) between the head and pelvis locations;
- $d_3$ : maximum value of the distance between the pelvis and left foot location, and the distance between the pelvis and right foot location;
- $d_4$ : distance between the left and right foot.

These parameters are measured only in the frame of maximum feet spread during the walking action in order to avoid self-occlusions.

The second data compensation aims at estimating four multiplicative factors to make  $d_1$ ,  $d_2$ ,  $d_3$ ,  $d_4$  independent of the camera view. For this purpose, ground truth motion capture data had been collected and analysed.

Experimental tests regarded 18 subjects walking in front of a single camera positioned at 45 and 90° with respect to the walking direction. The side-view data were captured at two different depths, 3.9 and 8.3 m from the camera. The results had been reported by using an expected confusion metric in order to predict how the subject's parameters filter the identity in a large population. Figure 3.9 shows the extracted body parameters and the recognition rates at angle view, near-side view and far-side view. The first three rows represent the results obtained from the data without any compensation applied: the obtained recognition rate is higher than 91%. By considering the appropriate scale factor based on the viewing condition, the percentage for the side view goes to 100%. Including both the scale and the view-angle adjustments, the result on all the camera positions is 94% and the confusions rates are on the order of 6%.

While Johnson and Bobick proposed a gait analysis based on some static parameters, in 2002 BenAbdelkader et al. proposed a different approach for 2D view



**Fig. 3.9** Results of the multi-view gait recognition method using the static body parameters  $d_1$ ,  $d_2$ ,  $d_3$ ,  $d_4$  (source: Johnson and Bobick [19])

independent gait recognition where the moving person is detected and tracked and an image template corresponding to the person's motion blob is extracted in each frame [3].

Subsequently, a self-similarity plot from the obtained sequence of templates has been computed in the following way:

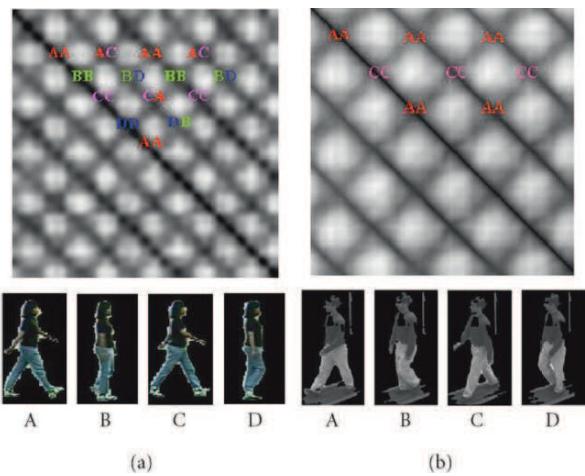
$$S(t_1, t_2) = \min_{|dx, dy| < r} \sum_{(x, y) \in B_{t_1}} |O_{t_1}(x + dx, y + dy) - O_{t_2}(x, y)| \quad (3.15)$$

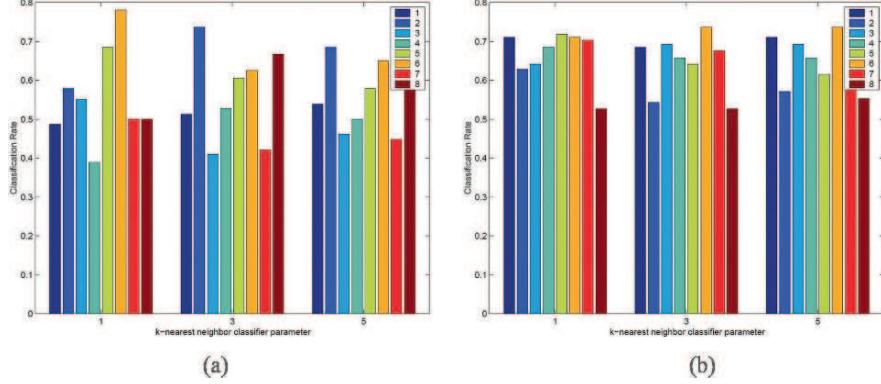
where  $N$  is the number of frames,  $1 \leq t_1, t_2 \leq N$ ,  $B_{t_1}$  is the bounding box of the person blob in frame  $t_1$ ,  $r$  is a small search radius, and  $O_{t_1}, O_{t_2}, \dots, O_{t_N}$  are the scaled and normalized image templates.

The similarity plot is a projection of the dynamics of the walking person that preserves the frequency and phase of the gait. Because gait consists of periodic contiguous steps, the similarity plot can be tiled into contiguous rectangular blocks, termed units of self-similarity (USS), each of which consists of the person's self-similarity over two periods of gait (Fig. 3.10). For recognition, the method uses principal component analysis to reduce the dimensionality of the USSs' and the  $k$ -nearest neighbour rule for classification.

Experimental tests on outdoor sequences of 44 people with four sequences of each taken on two different days achieve a classification rate of 77%. It is also tested on indoor sequences of seven people walking on a treadmill, taken from eight different viewpoints (from 0 to 120°) and on seven different days. A classification rate of 78% is obtained for near-fronto-parallel views, and 65% on average over all views. Figure 3.11 shows the classification rates for Dataset 3 for the eight viewpoints both with absolute correlation of binary silhouettes and with normalized cross-correlation of foreground images. The method appears robust to tracking and segmentation errors and to variation in clothing and background. It is also invariant to small changes in camera viewpoint and walking speed.

**Fig. 3.10** Units of self-similarity for (a) fronto-parallel sequence; (b) a non-fronto-parallel sequence. Similarity values are linearly scaled to the gray scale intensity range [0, 255] for visualization. The local minima of each SSP correspond to combinations of key poses of gait A, B, C, D (source: BenAbdelkader et al. [3])



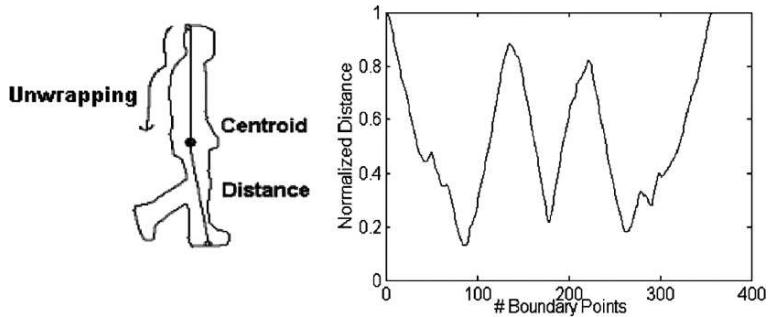


**Fig. 3.11** Classification rates for Dataset 3 for the eight viewpoints with  $k = 1; 3; 5$  and using (a) absolute correlation of binary silhouettes; (b) normalized cross-correlation of foreground images (source: BenAbdelkader et al. [3])

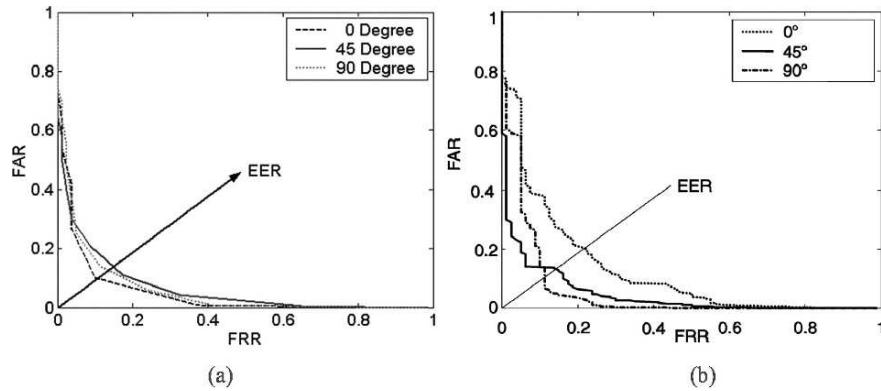
A statistical shape analysis is the solution proposed by Wang et al. [54], where the temporal changes of the detected silhouettes are represented by a 2D vector, composed of the edge points distance to the centroid position over time (Fig. 3.12). The gait signature is then obtained via the Procrustes shape analysis [29]. A supervised pattern classification technique based on the full Procrustes distance measure has been adopted and the method has been tested on the CASIA dataset-A gait database [1] (240 sequences from 20 different subjects walking at three viewing angles in an outdoor environment). Figure 3.13(a) shows the ROC curve, where the EERs (equal error rates) are about 8, 12 and 14% for 0, 90 and 45° views, respectively.

The same authors also applied the principal component analysis on the 2D silhouette representation [53]. Tests on the same database used in [54] are shown in Fig. 3.13(b).

More recently, Hong et al. [17] introduced a new representation for human gait recognition, called mass vector, defined as the number of pixels with a nonzero value



**Fig. 3.12** The 2D silhouette representation (source: Wang et al. [53])



**Fig. 3.13** ROC curves obtained with (a) Procrustes distance measure; (b) Principal component analysis (source: Wang et al. [53, 54])

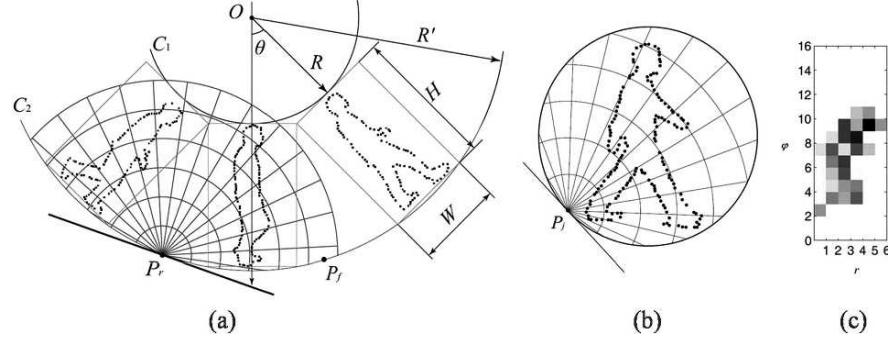
in a given row of the silhouette. Sequences of temporally ordered mass vectors have been used for gait representation and a dynamic time-warping (DTW) [23] approach for the recognition phase.

Experimental tests of the NLPR gait database have given CCR higher than 95% as reported in Table 3.1 where a comparison with the method of Kale et al. has been shown. A polar representation of a walking subject has been proposed by Chen and Gao [8] in 2007, where a 2D polar gait is obtained by tiling one period gait subsequence in a 2D polar plane along a ring (Fig. 3.14a). The gait characteristic is then achieved by a combination of appearance models of individual silhouette and contextual silhouettes in the polar plane.

For the individual frame image  $X_n$ , point sets  $\{P_j\}$  sampled from a reference circle  $C$  are employed as control points. Silhouette description relative to the control point  $P_j$  is then provided by means of shape descriptors. Orderly concatenating the shape descriptors originating from  $\{P_j\}$ , a silhouette appearance model at the  $n$ th frame, denoted as  $h_n$ , has been obtained. Figure 3.14 (b and c) shows the silhouette descriptor with angle  $\varphi = 32^\circ$  and radius  $r = 6$ . Therefore, the gait appearance model is defined as a combination of histograms and is represented to be invariant to translation, rotation and scale by means of a shape descriptor and gait images

**Table 3.1** CCR with respect to the viewpoints (source: Hong et al. [17])

View	Method	Rank				
		1 (%)	2 (%)	3 (%)	4 (%)	5 (%)
$0^\circ$	Kale et al. [22]	82.5	85	87.5	90	92.5
	Hong et al. [17]	96.25	98.75	98.75	100	100
$45^\circ$	Kale et al. [22]	92.5	95	97.5	97.5	97.5
	Hong et al. [17]	96.25	98.75	98.75	98.75	98.75
$90^\circ$	Kale et al. [22]	77.5	82.5	83.75	86.25	88.75
	Hong et al. [17]	88.75	90	90	91.25	95



**Fig. 3.14** (a) Periodic gait images plane and histogram bins for the contextual silhouettes; (b) histogram relative to the control point  $P_j$ ; (c) diagram of the polar histogram bins (source: Chen and Gao [8])

plane. Jeffrey divergence [39] and DTW have been employed for measuring the similarity of gait appearance models between test and reference sequences.

Experimental results on CASIA database [1] demonstrate that the algorithm presents a total CCR of 85.42% and a comparison with other methods [27, 36, 54] demonstrated a better performance of the approach proposed by Chen and Gao only for the lateral view.

### 3.3 3D Gait Recognition

#### 3.3.1 Introduction

An alternative to using a 2D viewpoint invariant model is to utilize a full 3D analysis technique, as this will be inherently viewpoint independent. There are several different approaches to this; a 3D model can be applied to the video data collected from a single camera, multiple cameras for improved accuracy, or to 3D volumetric data reconstructed from multiple cameras. In this section, we discuss several different 3D gait analysis methods proposed by the research community.

Angle	0 °	-45 °	45 °	-90 °	90 °
Sample					
$\lambda$	0.034	0.485	-0.592	-1	1
CCR(%)	92.5	95	85	75	65

**Fig. 3.15** CCR with respect to the view angles (source: Chen and Gao [8])

Bhanu and Han proposed a novel model-based technique[4], utilizing a 3D model with 33 degrees of freedom. Several assumptions were made in order to simplify matters; the camera was stationary, subjects walked in a constant direction and the swing of the limbs was parallel to the direction of walking. The model was fitted to silhouette data using a genetic algorithm based on the least squares technique. Recognition performance was evaluated on the stationary and kinematic features separately and combined.

The task of fitting a 3D model to a human subject in a single monocular viewpoint can prove to be quite a difficult task, as there are many degrees of freedom compared to the number of constraints created by a silhouette from a single viewpoint. Even so, it is a very popular research area with many excellent techniques proposed [12, 28, 52]. The use of multiple viewpoints greatly eases the task of fitting a model to a subject, as the set of possible configurations for the model is greatly reduced as the model is simultaneously constrained by multiple silhouettes.

There is a wide variety of data sets containing human motion from multiple viewpoints available to the research community, the most popular one among the gait community is that from Carnegie Mellon University; the Motion of Body (MoBo) database [15]. The data set was recorded indoors using a treadmill; this allowed them to record subjects walking and running at varying gradients and also requires less space. Six cameras were positioned around the subject, and the video data was captured simultaneously. The database contained 25 subjects, each having a total of 24 samples.

Orrite-Uruñuela et al. devised a technique for fitting a 3D skeletal model to multi-view video data from the CMU database [34]. This involved fitting point-distribution models to the silhouette from each frame, which were similar to the active shape model [9]. The skeletal model is then derived from the set of point-distribution models.

A similar method of fitting a 3D model was proposed by Zhao et al. [60], where multiple views were used to improve model fitting performance. A skeletal model is initially fitted to the first frame in a sequence, with the position, orientation, body geometry and joint angles being manually chosen. Tracking was then performed on the subsequent frames to find the variation in the model's parameters, which could then be used for recognition.

Using video data from multiple viewpoints, it is possible to create a 3D volumetric reconstruction of the subject. One of the most popular methods of doing this is the Visual Hull [26]; the silhouettes from each view are re-projected from the viewpoints, and the intersection of the projections gives the Visual Hull; the Visual Hull provides the maximal shape achievable in the bounds of the silhouettes. This means that the derived shape is often larger than the original, another artifact often experienced is that concave areas are always filled in. More sophisticated techniques consider the correspondence of colour information between viewpoints, in order to provide more realistic reconstructions of the original object [45].

One of the earliest uses of 3D volumetric reconstruction for gait analysis was that by Shakhnarovich et al. [42]. A small data set was created, comprised of 3D visual hull data for 12 subjects walking in arbitrary directions through the target

area, with the number of samples per subject varying between two and eight. The visual hulls were constructed using silhouettes generated from four video cameras. The trajectory of the subject was found, and 2D canonical view silhouettes were synthesized by placing a virtual camera into the volumetric space. Two-dimensional gait analysis techniques were then used to analyse the synthesized silhouettes. This approach allowed the use of 2D gait analysis techniques in a scenario where the orientation of the subjects is unknown.

Another way of deriving 3D information from multiple views is to use a stereo depth reconstruction technique, where the distance of a point from a pair of cameras can be found from the point's disparity. Stereovision is used by Urtasun and Fua to aid fitting a 3D model to a subject [48]. A sophisticated deformable model [38] was used for representing the shape of the human subjects and a motion model describing the deformation of the shape model was produced from data collected using a marker-based computer vision system to capture four subjects walking. The shape and motion models were then fitted to 3D data collected from a multi-view stereovision system using the least squares method.

### ***3.3.2 The University of Southampton Multi-biometric Tunnel***

Up until now, very little research has been carried out into the use of true 3D gait analysis techniques on volumetric data in comparison to 2D gait analysis techniques. The most likely reason for this is the difficulty in acquiring volumetric (3D) data compared to 2D video data. Also the storage capacity required to hold such volumetric data is typically much larger than video data, which in the past made handling and analysing such data sets a very large undertaking. With the ever-increasing capacity of storage devices and the growth in processing power, it is likely that the area of 3D gait analysis will start to receive much more attention from the research community.

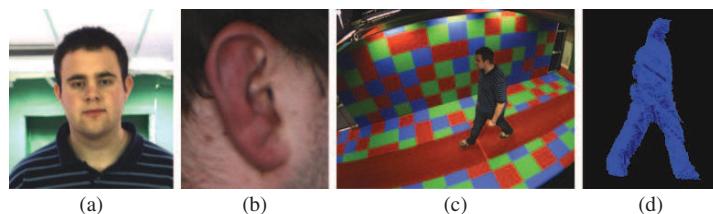
There is some debate as to the size of the database necessary for evaluation of biometrics [49]. Current approaches for gait biometrics can achieve a 100% correct classification rate on databases of around 100 people [50]. In order to provide an estimate of performance capability across a nation's population, we need a database larger than the currently available ones. Work has started at the University of Southampton creating a multi-biometric data set using the University's Multi-Biometric Tunnel [30]. The Multi-Biometric Tunnel is a unique research facility, which facilitates the capture of biometric data from a large number of subjects in a fast and efficient manner. The system has a high level of automation, using break-beam sensors at the start and end of the tunnel to control the capture of data. Each participant is given an ID card, which is scanned at the beginning of the data capture process to associate the captured data with the participant. The tunnel area is painted with bright saturated colours in a non-repeating pattern, as shown in Fig. 3.16(a). This simplifies the task of camera calibration and also makes it easier to segment the subjects from the background, resulting in high-quality silhouettes.



**Fig. 3.16** The University of Southampton Multi-Biometric Tunnel; (a) inside the tunnel; (b) placement of gait cameras

The data set will contain several of the most common non-contact biometrics; face, ear and most importantly multiple viewpoint 3D gait, as shown in Fig. 3.17. The subject's face is captured using a UXGA ( $1600 \times 1200$ ) resolution IEEE1394 video camera at a rate of 7.5 frames per second. The ear is captured by another UXGA video camera assisted by a pair of photographic strobe units. The subject is captured by eight VGA ( $640 \times 480$ ) resolution IEEE1394 cameras placed around the perimeter of the tunnel, as shown in Fig. 3.16(b); the cameras acquire video at a rate of 30 frames per second. The captured 2D video data and silhouettes will be available along with 3D visual hull reconstructions. Initial tests on a data set of 10 people gave a 100% correct classification rate when synthesizing canonical view silhouettes and calculating the average silhouette.

A simple proof of concept experiment was recently conducted using the Multi-Biometric Tunnel to establish whether it is possible to measure a change in an individual's gait when they are concealing an object using the data captured by the Multi-Biometric Tunnel. Studies conducted by researchers in the medical field show that a subject's gait varies when additional load is carried [2, 24, 35, 55]. Sarkar et al. also demonstrate that the carrying of a briefcase has an effect on a subject's gait [41]. In order to ensure that the analysis only considers the subject's gait and not any change in shape, the participants were asked to wear a harness with eight half-liter water bottles attached to it and tests were performed with the bottles empty and full creating a change in weight of 4 kg without changing the subject's shape. Four participants were used in the experiment, each providing eight samples with and without the concealed weight. A simple classifier was created, which projects



**Fig. 3.17** Example data from Multi-Biometric Tunnel; (a) face; (b) ear; (c) 2D gait; (d) 3d gait

the 3D data into a 2D canonical view, which is then used to calculate the average silhouette. Analysis of variance and canonical analysis were then used to improve the discriminatory ability of the classifier. A leave one out strategy was used with the data set for testing the performance of the classifier. A correct classification rate of 73.4% was achieved demonstrating that it is possible to detect concealed weight on a person by the change in his/her gait, even when using a very simple gait analysis technique.

### 3.4 Conclusions

To deploy gait biometrics in unconstrained views, it is necessary to relieve viewpoint dependence in recognition. To achieve this, we can incorporate geometrical considerations, and the inherent periodicity of gait, and thus achieve a technique which can correct for the effects of viewpoint. In a limited study, this has been shown to achieve viewpoint-independent recognition from image data derived from a single camera view. Conversely, 3D data acquisition requires multiple cameras and then synthesizes a solid model of the walking subject. This can achieve viewpoint independence since arbitrary views can be synthesized. A selection of approaches has been developed for these purposes and in limited evaluations has been shown to achieve recognition capability. Now that computational power is abundant and sufficient for the processing of multi-view data acquired at video rate, we anticipate more development in 3D recognition capability. Equally, as surveillance technology continues to increase, the new techniques are ready for deployment of 2D viewpoint invariant approaches.

## References

1. Casia gait database. online (2006). <http://www.sinobiometrics.com>
2. Attwells, R., Birrell, S., Hooper, R., Mansfield, N.: Influence of carrying heavy loads on soldiers' posture, movements and gait. *Ergonomics* **49**(14), 1527–1537(11) (2006). doi:10.1080/00140130600757237. <http://www.ingentaconnect.com/content/tandf/terg/2006/00000049/00000014/art00007>
3. BenAbdelkader, C., Davis, L.S., Cutler, R.: Motion-based recognition of people in eigengait space. In: FGR, pp. 267–274 (2002)
4. Bhanu, B., Han, J.: Human recognition on combining kinematic and stationary features. In: Proceedings of Audio- and Video-Based Biometric Person Authentication, *Lecture Notes in Computer Science*, vol. 2688, pp. 600–608. Springer-Verlag, New York (2003)
5. Bouchrifa, I., Nixon, M.S.: Model-based feature extraction for gait analysis and recognition. In: Mirage: Computer Vision / Computer Graphics Collaboration Techniques and Applications, vol. 4418, pp. 150–160 (2007)
6. Boyd, J.E.: Synchronization of oscillations for machine perception of gaits. *Comput. Vis. Image Underst.* **96**(1), 35–59 (2004)
7. Carter, J.N., Nixon, M.S.: On measuring gait signatures which are invariant to their trajectory. *Measurement Contrl.* **32**(9), 265–269 (1999)
8. Chen, S., Gao, Y.: An invariant appearance model for gait recognition. *Multimedia and Expo, 2007 IEEE International Conference on* pp. 1375–1378 (2–5 July 2007)

9. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models—their training and application. *Comput. Vis. Image Underst.* **61**(1), 38–59 (1995)
10. Cunado, D., Nixon, M.S., Carter, J.N.: Automatic gait recognition via model-based evidence gathering. In: L. O’Gorman, S. Shellhammer (eds.) *Proceedings AutoID ’99: IEEE Workshop on Identification Advanced Technologies*, pp. 27–30. IEEE (1999)
11. Cunado, D., Nixon, M.S., Carter, J.N.: Automatic extraction and description of human gait models for recognition purposes. *Comput. Vis. Image Underst.* **90**(1), 1–41 (2003)
12. Fua, P.: Markerless 3d human motion capture from images. In: S.Z. Li (ed.) July 2009. *Encyclopedia of Biometrics*. (ISBN 978-0-387-73003-5), Springer
13. Goffredo, M., Seely, R.D., Carter, J.N., Nixon, M.S.: Tech. Rep. n.1456, University of Southampton (2007)
14. Goffredo, M., Spencer, N., Pearce, D., Carter, J.N., Nixon, M.S.: Human perambulation as a self calibrating biometric. In: S.K. Zhou, W. Zhao, X. Tang, S. Gong (eds.) *AMFG, Lecture Notes in Computer Science*, vol. 4778, pp. 139–153. Springer, New York (2007)
15. Gross, R., Shi, J.: The cmu motion of body (mobo) database. Tech. Rep. CMU-RI-TR-01-18, Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213 (2001)
16. Hayfron-Acquah, J.B., Nixon, M.S., Carter, J.N.: Automatic gait recognition by symmetry analysis. *Pattern Recognit. Lett.* **24**(13), 2175–2183 (2003)
17. Hong, S., Lee, H., Nizami, I., Kim, E.: A new gait representation for human identification: Mass vector. *Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference on* pp. 669–673 (23–25 May 2007)
18. Huang, P., Harris, C., Nixon, M.: Recognising humans by gait via parametric canonical space. *J. Artif. Intelli. Eng.* **13**(4), 359–366 (1999)
19. Johnson, A.Y., Bobick, A.F.: A multi-view method for gait recognition using static body parameters. In: AVBPA ’01: Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication, pp. 301–311. Springer-Verlag, London, UK (2001)
20. Kale, A., Chowdhury, A., Chellappa, R.: Towards a view invariant gait recognition algorithm. *Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance, 2003.* pp. 143–150 (21–22 July 2003)
21. Kale, A., Roychowdhury, A., Chellappa, R.: Fusion of gait and face for human identification. *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP ’04). IEEE International Conference on* **5**, V-901–4 vol.5 (17–21 May 2004)
22. Kale, A.A., Cuntoor, N.P., Yegnanarayana, B., Rajagopalan, A.N., Chellappa, R.: Gait analysis for human identification. In: AVBPA, pp. 706–714 (2003)
23. Keogh, E., Ratanamahatana, C.A.: Exact indexing of dynamic time warping. *Knowl. Inf. Syst.* **7**(3), 358–386 (2005)
24. Knapik, J., Harman, E., Reynolds, K.: Load carriage using packs: A review of physiological, biomechanical and medical aspects. *Appl. Ergon.* **27**(3), 207–216 (1996). <http://www.sciencedirect.com/science/article/B6V1W-3Y0RS9P-9/2/acc0b9e0629a8cf5950fa55b4e5fc2f0>
25. Larsen, P.K., Simonsen, E.B., Lynnerup, N.: Gait analysis in forensic medicine. In: J.A. Beraldin, F. Remondino, M.R. Shortis (eds.) *Videometrics IX*, vol. 6491, p. 64910M. SPIE (2007)
26. Laurentini, A.: The visual hull concept for silhouette-based image understanding. *IEEE Trans. Patt. Anal. Mach. Intell.* **16**(2), 150–162 (1994)
27. Lee, L., Grimson, W.: Gait analysis for recognition and classification. *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on* pp. 148–155 (20-21 May 2002)
28. Liu, H., Chellappa, R.: Markerless monocular tracking of articulated human motion. In: *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 1, pp. 693–696 (2007)
29. Mardia, K., Jupp, P.: *Directional Statistics*. Wiley, New York (2000)

30. Middleton, L., Wagg, D.K., Bazin, A.I., Carter, J.N., Nixon, M.S.: A smart environment for biometric capture. In: IEEE International Conference on Automation Science and Engineering, pp. 57–62 (2006)
31. Murray, M.P., Drought, A.B., Kory, R.C.: Walking patterns of normal men. *J. Bone Joint Surg.* **46**, 335 (1964)
32. Nixon, M.S., Carter, J.N.: Automatic recognition by gait. *Proc. IEEE* **94**(11), 2013–2024 (2006)
33. Niyogi, S.A., Adelson, E.H.: Analyzing and recognizing walking figures in xyt. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 469–474 (1994)
34. Orrite-Uruñuela, C., del Rincón, J.M., Herrero-Jaraba, J.E., Rogez, G.: 2d silhouette and 3d skeletal models for human detection and tracking. In: Proceedings of the 17th International Conference on Pattern Recognition, vol. 4, pp. 244–247 (2004)
35. Pascoe, D.D., Pascoe, D.E., Wang, Y., Shim, D.M., Kim, C.K.: Influence of carrying book bags on gait cycle and posture of youths. *Ergonomics* **40**(6), 631–640(10) (1997). <http://www.ingentaconnect.com/content/tandf/terg/1997/00000040/00000006/art00003>
36. Phillips, P., Sarkar, S., Robledo, I., Grother, P., Bowyer, K.: Baseline results for the challenge problem of humanoid using gait analysis. Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on pp. 130–135 (20–21 May 2002)
37. Phillips, P.J., Sarkar, S., Robledo, I., Grother, P., Bowyer, K.W.: The gait identification challenge problem: data sets and baseline algorithm. In: Proceedings of The 16th International Conference on Pattern Recognition, vol. 1, pp. 385–388 (2002)
38. Plankers, R., Fua, P.: Articulated soft objects for video-based body modeling. In: Proceedings. Eighth IEEE International Conference on Computer Vision, vol. 1, pp. 394–401 (2001)
39. Puzicha, J., Buhmann, J., Rubner, Y., Tomasi, C.: Empirical evaluation of dissimilarity measures for color and texture. Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on vol. 2, pp. 1165–1172 (1999)
40. Sarkar, S., Phillips, P., Liu, Z., Vega, I., Grother, P., Bowyer, K.: The humanoid gait challenge problem: data sets, performance, and analysis. *Trans. Patt. Anal. Mach. Intell.* **27**(2), 162–177 (Feb. 2005)
41. Sarkar, S., Phillips, P.J., Liu, Z., Vega, I.R., Grother, P., Bowyer, K.W.: The humanoid gait challenge problem: data sets, performance, and analysis. *IEEE Trans. Patt. Anal. Mach. Intell.* **27**(2), 162–177 (2005)
42. Shakhnarovich, G., Lee, L., Darrell, T.: Integrated face and gait recognition from multiple views. In: Computer Vision and Pattern Recognition, Proceedings of the 2001 IEEE Computer Society Conference on, vol. 1, pp. 439–446 (2001)
43. Shutler, J., Nixon, M.S.: Zernike velocity moments for sequence-based description of moving features. *Image Vision Comput.* **24**(4), 343–356 (2006)
44. Shutler, J.D., Grant, M.G., Nixon, M.S., Carter, J.N.: On a large sequence-based human gait database. In: Proceedings of Fourth International Conference on Recent Advances in Soft Computing, pp. 66–72 (2002)
45. Slabaugh, G.G., Culbertson, W.B., Malzbender, T., Stevens, M.R., Schafer, R.W.: Methods for volumetric reconstruction of visual scenes. *Int. J. Comput. Vision* **57**(3), 179–199 (2004)
46. Spencer, N., Carter, J.: Towards pose invariant gait reconstruction. Image Processing, 2005. ICIP 2005. IEEE International Conference on **3**, III–261–4 (11–14 Sept. 2005)
47. Spencer, N.M., Carter, J.N.: Viewpoint invariance in automatic gait recognition. In: Proceedings of Third IEEE Workshop on Automatic Identification Advanced Technologies, AutoID'02, pp. 1–6 (2002)
48. Urtasun, R., Fua, P.: 3d tracking for gait characterization and recognition. In: Proceedings. Sixth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 17–22 (2004)
49. Veres, G., Nixon, M., Carter, J.: Is enough enough? what is sufficiency in biometric data? In: International Conference on Image Analysis and Recognition, vol. 4142, pp. 262–273. Springer Lecture Notes (2006). <http://eprints.ecs.soton.ac.uk/12832/>

50. Veres, G.V., Gordon, L., Carter, J.N., Nixon, M.S.: What image information is important in silhouette-based gait recognition? In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 776–782 (2004)
51. Wagg, D.K., Nixon, M.S.: Automated markerless extraction of walking people using deformable contour models. *Comput. Animation Virtual Worlds* **15**(3–4), 399–406 (2004)
52. Wang, L., Hu, W., Tan, T.: Recent developments in human motion analysis. *Patt. Recogn.* **36**(3), 585–601 (2003)
53. Wang, L., Tan, T., Hu, W., Ning, H.: Automatic gait recognition based on statistical shape analysis. *Image Process., IEEE Trans.* **12**(9), 1120–1131 (Sept. 2003)
54. Wang, L., Tan, T., Ning, H., Hu, W.: Silhouette analysis-based gait recognition for human identification. *IEEE Trans. Patt. Anal. Mach. Intell.* **25**(12), 1505–1518 (2003)
55. Wang, Y., Pascoe, D.D., Weimar, W.: Evaluation of book backpack load during walking. *Ergonomics* **44**(9), 858–869 (2001)
56. Winter, D.A.: *Biomechanics and Motor Control of Human Movement*. Wiley, New York (2004)
57. Yam, C., Nixon, M.S., Carter, J.N.: Extended model-based automatic gait recognition of walking and running. In: Proceedings of 3rd Int. Conf. on Audio- and Video-Based Biometric Person Authentication, AVBPA 2001, pp. 278–283 (2001)
58. Zhang, R., Vogler, C., Metaxas, D.: Human gait recognition at sagittal plane. *Image Vision Comput.* **25**(3), 321–330 (2007)
59. Zhang, Z., Troje, N.F.: View-independent person identification from human gait. *Neurocomputing* **69**(1–3), 250–256 (2005)
60. Zhao, G., Liu, G., Li, H., Pietikäinen, M.: 3d gait recognition using multiple cameras. In: Proceedings of the Seventh IEEE International Conference on Automatic Face and Gesture Recognition (FG '06), pp. 529–534. IEEE Computer Society, Los Alamitos, CA, USA (2006)

# **Chapter 4**

## **Advanced Technologies for Touchless Fingerprint Recognition**

**Giuseppe Parziale and Yi Chen**

**Abstract** A fingerprint capture consists of touching or rolling a finger onto a rigid sensing surface. During this act, the elastic skin of the finger deforms. The quantity and direction of the pressure applied by the user, the skin conditions, and the projection of an irregular 3D object (the finger) onto a 2D flat plane introduce distortions, noise, and inconsistencies on the captured fingerprint image. Due to these negative effects, the representation of the same fingerprint changes every time the finger is placed on the sensor platen, increasing the complexity of the fingerprint matching and representing a negative influence on the system performance. Recently, a new approach to capture fingerprints has been proposed. This approach, referred to as *touchless* or *contactless* fingerprinting, tries to overcome the above-cited problems. Because of the lack of contact between the finger and any rigid surface, the skin does not deform during the capture and the repeatability of the measure is quiet ensured. However, this technology introduces new challenges. Finger positioning, illumination, image contrast adjustment, data format compatibility, and user convenience are key in the design and development of touchless fingerprint systems. In addition, vulnerability to spoofing attacks of some touchless fingerprint systems must be addressed.

### **4.1 Introduction**

Historically, fingerprints were collected by applying ink on a finger and pressing it against a paper card (e.g., a tenprint card). This is known as the “ink-on-paper” capture method. Today, most civilian and criminal fingerprint systems accept live-scan digital images by directly sensing the finger surface with an electronic fingerprint scanner. Both the live-scan and the “ink-on-paper” methods require a person to press his/her finger against a flat rigid surface, also called touch-based sensing.

---

G. Parziale (✉)  
iNVASIVE CODE., Barcelona, Spain  
e-mail: geppy.parziale@invasivecode.com

Large databases have been gathered and well-established comparison techniques have been developed for this type of capture approach. Forensic experts are also exclusively trained to work with touch-based fingerprints, including latents and tenprints. As a result, it is essential that any forensic or government application utilizing fingerprint data for identification be compatible with the touch-based legacy data and algorithms.

During the capture with a live-scan device, dirt, sweat, and moisture commonly present on the finger skin are transferred onto the glass platen in correspondence of each ridge. Thus, a *latent print* remains impressed on the glass surface [1]. Using special mechanical and/or chemical tools [2], it is possible to copy the residual fingerprint and create a latex (or gelatin or other materials) replica of it that can be used to try to grant the access to the system, increasing its vulnerability [3]. This is a very sensitive problem, especially for unattended or unsupervised fingerprint systems [4].

As the demands for fingerprint identification increase, the touch-based sensing technology began facing a number of challenges. For example, the touch-based sensing is a tedious, time-consuming, and expensive process for which collecting a tenprint card could take up to 15 min. This is not acceptable in most government applications, such as US-VISIT, where long wait for fingerprint acquisition would cause significant delay for each of the  $\sim 40$  million travelers using the system every year [5].

Recently, new fingerprint devices have been proposed. They have been designed to overcome the above limitations of the legacy technology. Among the different proposed technologies, touchless fingerprinting represents a very interesting solution to the skin deformation and the latent print problems.

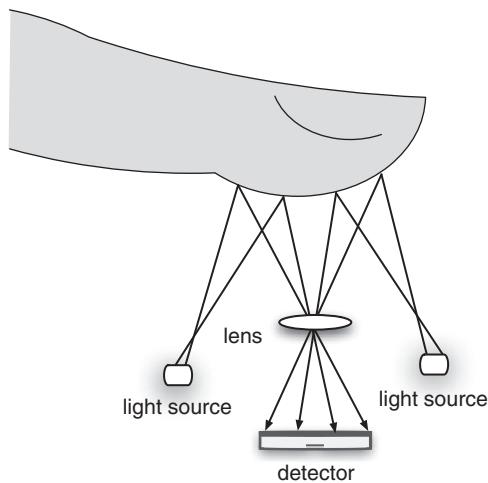
However, since the nature of the touchless fingerprint images is different than the images obtained with legacy live-scan devices, new methods of image quality check, analysis, enhancement, and protection must be implemented to provide additional flexibility for specific applications and customers. Besides, the compatibility with the legacy systems must be proven to avoid the recollection of the already existing fingerprint databases.

In this chapter, an overview of this technology is provided. Sections 4.2 and 4.3 provide an overview of the basic touchless technology and the touchless technology combined with multi-vision systems. Sensor interoperability is discussed in Section 4.4. Parametric and non-parametric methods to project a 3D fingerprint to a 2D plane are discussed in Section 4.5. A proposed method to enhance a touchless fingerprint image and an estimation of its image quality are discussed in Sections 4.6 and 4.7, respectively. The problem of vulnerability of touchless devices is reported in Section 4.8. Finally, concluding remarks are presented in Section 4.9.

## 4.2 Touchless Finger Imaging

Touchless or contactless finger imaging is essentially a remote sensing technology used for the capture of the ridge-valley pattern with no contact between the skin of the finger and the sensing area [6]. The lack of contact drastically reduces the problems intrinsic to the contact-based technology.

**Fig. 4.1** The principle of the touchless capture with light sources in front of the fingerprint



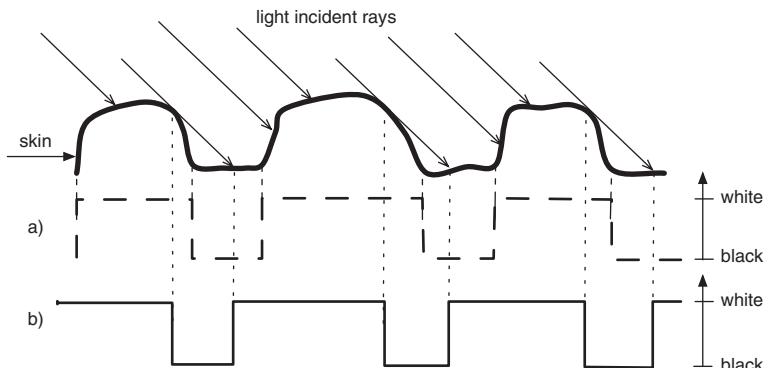
The approaches to capture a fingerprint based on touchless technology can be grouped in two main families: reflection-based touchless finger imaging (RTFI) and transmission-based touchless finger imaging (TTFI).

Figure 4.1 highlights the basic capture principle of the RTFI approach. Light sources are placed in front of the fingerprint to illuminate it. To obtain an image with sufficient contrast to well distinguish between ridges and valleys, it is necessary that

- the finger skin absorbs only a small portion of the incident light, while the majority of it (*albedo*) is reflected back to the detector; and
- the quantity of light absorbed by the valleys is different from the quantity absorbed by the ridges.

Detectors, illuminators, and lenses must be designed so that the contrast of ridge-valley pattern is sufficiently high to allow the extraction of the important features for the fingerprint matching. Parameters such as the depth-of-focus (DOF) and field-of-view (FOV) of the camera, the irradiation power and the frequency of the light sources, the shutter speed of the detectors must be correctly chosen during the design of the device to obtain the optimal contrast and facilitate the acquisition of fingerprints with a very dry or a very wet skin.

Light sources must be placed as close as possible to the detector, so that the light rays are as much as possible parallel to the optical path of the camera and the majority of the light is reflected back to the sensor chip. This reduces the *shadowing effect* caused by the 3D ridge-valley structure. If the light sources are placed far away from the normal to the finger surface, each ridge generates a shadow that is projected onto the neighbor ridge, modifying the apparent profile of the intermediate valley (Fig. 4.2). Hence, when ridges are extracted using common fingerprint algorithms, a set of pixels representing the valleys belong in reality to the adjacent ridge. The overall effect is a small shift (it can be up to a ridge width) of the whole ridge-valley structure in the opposite direction with respect to the direction from which the light rays come from.



**Fig. 4.2** Incorrect light incidence. The profile (a) provides the correct representation of the skin surface. The profile (b) highlights what happens if the light source is not perpendicular to the skin surface: the representation of the ridge–valley structure results incorrect

The wavelength, the intensity, the position, and the incident angle of the light are very important parameters to obtain the optimal contrast and the correct representation of the fingerprint. Long wavelength rays including infrared tend to penetrate the skin, absorbed by the epidermis, and the ridge–valley pattern results of less clarity.

Elli [7] measured the light reflected on the skin using a high-resolution, high-accuracy spectrograph under precisely calibrated illumination conditions. The experiments show that the human skin reflectance is influenced by both melanin and hemoglobin. The ratio of hemoglobin absorption is lower around 500 nm and higher around 545 and 575 nm. Since it is desirable to observe only the surface of the finger and reduce the effects of hemoglobin in order to obtain a high-contrast fingerprint image, the wavelength at lower hemoglobin absorption must be chosen for the light source of a touchless fingerprint sensor. Moreover, a common CCD or CMOS detector has high sensitivity still around 500 nm. Considering both the skin reflectance of the finger and the spectral sensitiveness of each camera, the final wavelength of the light for a touchless fingerprint device can be determined [8]. The above experiments bring to the conclusion that the best contrast is provided by blue light. Moreover, another reason to employ blue light is that blue is complementary to yellow and red, which are the dominant colors of the finger skin.

The advantage of the touchless approach with respect to the traditional frustrated total internal reflection (FTIR) one is that the ridge–valley pattern results fully imaged, i.e., valleys are no more part of the image background, but they bring additional information, which could be used to extract new features useful for the matching. This information is absent from the FTIR technology where the valleys belong to the image background. However, it has to be taken into account that this additional information reduces drastically the contrast of the whole ridge–valley pattern. This is a very critical point for touchless technology and the optimization of the contrast is a very complex task. Standard fingerprint algorithms cannot be directly used on this kind of images, because they are designed for FTIR devices

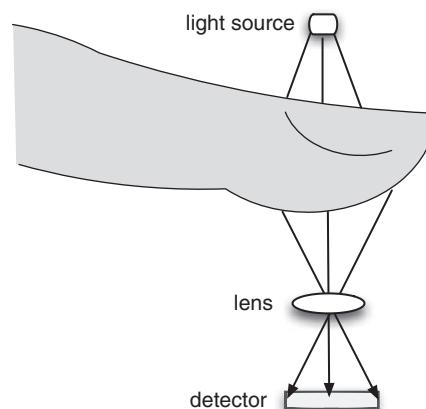
that generate much higher contrast images [9, 10]. Moreover, since the finger is a 3D object projected onto a 2D camera from a far point, the prospective effect viewed from the camera increases the apparent frequency of the ridge–valley pattern and reduces the geometrical resolution from the fingerprint center toward the side until ridges and valleys become undistinguishable. This reduces the size of the useful fingerprint area that can be correctly processed. Hence, dedicated algorithms are needed to enhance the frequency-varying ridge–valley structure with an increase of the overall computational load.

The camera lenses must be designed with special characteristics too. Since the distance between the camera and each small portions of the fingerprint is not constant due to the finger curvature and since the finger cannot be always placed exactly in the same position, the DOF and the FOV of the camera must be large enough to allow the user to place the finger with a certain degree of freedom and ensure that the fingerprint is always focused along the entire surface. However, large DOF and FOV require the use of a more complex optical system (more lenses), increasing the optical distortions and obviously the lens costs.

These optical features in addition to the above-mentioned detector characteristics make the device more expensive than a comparable touch-based device. A trade-off between the costs and the optical performance must be found. For example, to reduce the costs, some manufacturers prefer to provide a finger support, losing the advantage of avoiding to touch something.

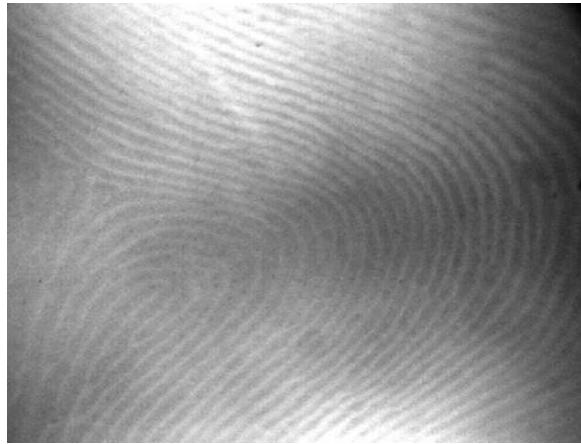
Recently, Mitsubishi Electric Corporation has proposed a new touchless capture approach, based on the transmission of the light through the finger [11]. We refer to this approach as to transmission-based touchless finger imaging (TTFT). This time, the light sources are placed in such a way that they illuminate the nail side (see Fig. 4.3). Red light with a wavelength of  $\lambda = 660$  nm is used, because it has high transmittance ratio to the skin tissue. Hence, the light penetrating the finger is collected by the detector placed in front of the fingerprint.

When the light wavelength is more than 660 nm, capillary vessels are strongly visible in fingerprint images, because of the high absorption ratio to the hemoglobin



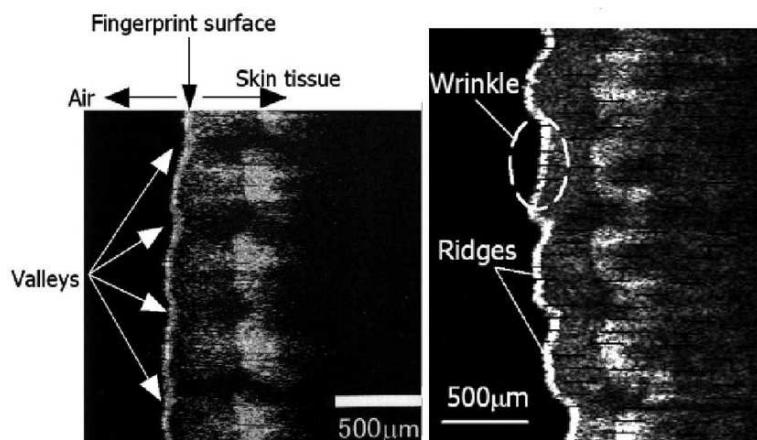
**Fig. 4.3** Touchless principle proposed by Mitsubishi Electric Corporation

**Fig. 4.4** Example of a fingerprint obtained with the Mitsubishi touchless device



of the blood. On the contrary, when the light wavelength is less than 660 nm, the brightness is insufficient to acquire images because of the high absorption ratio of the skin tissues. According to [12], using the suggested approach, it is possible to capture interior information of the finger that can be used to reduce the negative effects of the skin conditions on the final image, as it happens in the contact-based technology. Figure 4.4 represents an example of fingerprint obtained with this approach.

The principle on which the Mitsubishi approach is based comes from some studies done using the optical coherence tomography (OCT). Figure 4.5 shows a cross-sectional image of a fingerprint obtained with this technique. The bright portions of this image represent areas of low optical transmittance. A layer exists in



**Fig. 4.5** Cross section of a fingerprint obtained by optical coherence tomography. On the *left-hand side*, valleys and the corresponding high transmittance skin tissue are highlighted. On the *right-hand side* a wrinkle is shown

which optical transmittance of skin tissue corresponding to fingerprint ridges tends to be smaller than that of the fingerprint valleys. This characteristic is maintained even if the ridges are damaged. Thus, there is a skin layer whose optical characteristics correspond to the convex-concave pattern of the fingerprint surface without being affected by any concavity or convexity caused, for example, by wrinkles (right-hand side of Fig. 4.5). Detecting the optical characteristics of this internal layer enables the detection of the same pattern as that of the fingerprint without being affected by the status of the finger surface.

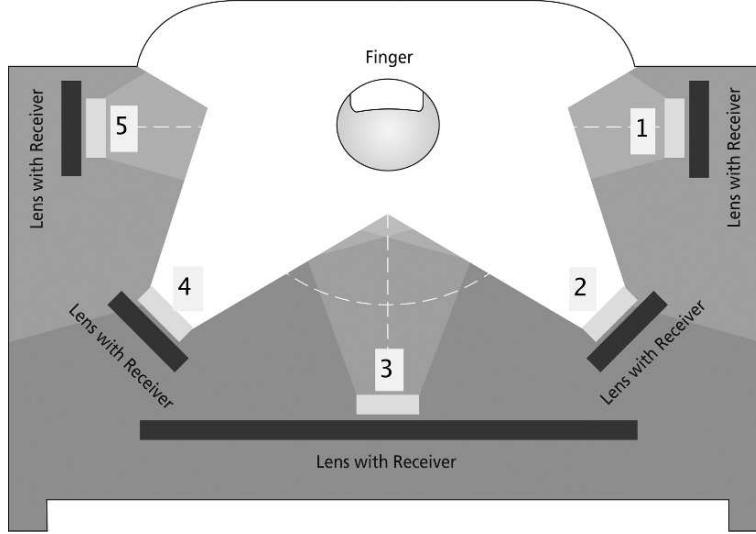
### 4.3 The 3D Touchless Fingerprinting

Touchless fingerprint devices are already available in the market, but they did not generate the sufficient interest to allow their widespread use, in spite of their advantages with respect to the legacy devices. The main reason has to be found in the relative higher costs of these sensors compared to the flat (or *dab*) touch-based ones. Besides, the curvature of the finger represents an obvious limitation for these expensive devices. In fact, the useful fingerprint area captured by a touchless device results smaller than the area acquired by a touch-based device, because the finger curvature increases the apparent frequency of the ridge-valley pattern making ridges and valleys indistinguishable on the fingerprint extremities, where the above-mentioned shadow effect also contributes to change the real ridge shape.

To improve the accuracy of fingerprint identification or verification systems, new touchless devices using more than one camera or more than one capture view have been recently proposed. These devices combine the touchless technology with a multi-vision system. In such a way, it is possible

- to acquire rolled-equivalent fingerprints with a lower failure-to-acquire error and a faster capture procedure than the traditional methods;
- to obtain the 3D representation of a finger.

Figure 4.6 highlights a schematic view of a device developed by TBS [12]. The device is a cluster of five cameras located on a semicircle and pointing to its center, where the finger has to be placed during the capture. Details of this device are reported in [14]. It contains a set of green LED arrays also located on the semicircle. During a capture, the finger is placed on a special support to avoid trembling that could create motion blur on the final image. The portion of the finger that has to be captured does not touch any surface. Moreover, the finger has to be placed in a correct position so that it is completely contained in the field-of-views of the five cameras at the same time. A real-time algorithm helps the user during the finger placement. Once the finger is in the correct position, the user receives a “Don’t move” request from the device and the capture starts automatically. During an acquisition, each LED array is set to a specific light intensity and the five cameras capture synchronously a picture of the finger. The acquired five views (Fig. 4.7) are then combined together to obtain a 3D reconstruction and then, the rolled-equivalent fingerprint.



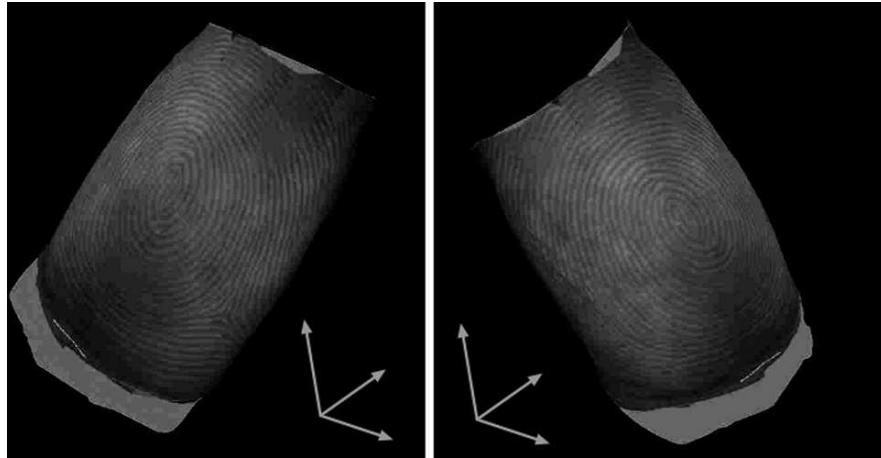
**Fig. 4.6** The schematic view of the multi-camera device developed by TBS

The 3D reconstruction procedure is based on stereovision and photogrammetry algorithms [15]. The exact position and orientation of each camera (*camera calibration*) with respect to a given reference system are computed off-line, using a 3D target on which points with known positions are marked [16–18].

The position of the middle camera (camera 3 in Fig. 4.6) has been chosen so that it could capture the central portion of the fingerprint, where the *core* and the *delta* are usually located. Then, the other cameras have been placed so that their field-of-views partially overlap. In this way, the images contain a common set of pixels (homologous pixels) representing the same portion of the skin. To compute the position of each pixel in the 3D space (3D reconstruction), the correspondences between two image pixels must be solved (*image matching*). This is done computing the cross-correlation between each adjacent image pair. Before that, the distortions generated by the mapping of a 3D object (the finger) onto the 2D image plane have to be minimized. This reduces errors and inconsistencies in finding the correspondences between the two neighbor image pair. Using shape-from-silhouette algorithms, it is possible to estimate the finger volume. Then, each image is unwrapped from the 3D model to a 2D plane obtaining the corresponding *ortho-images*, which are used to search for homologous pixels in the image acquired by each adjacent camera pair.



**Fig. 4.7** An example of five views of the same fingerprint acquired with the multi-camera device developed by TBS



**Fig. 4.8** Two views of a 3D fingerprint reconstructed using the stereo-vision approach

Once the pixel correspondences have been resolved, the third dimension of every image pixel is obtained using the camera geometry [19]. In Fig. 4.8, an example of the 3D reconstruction is highlighted.

Flashscan3D [20] proposed a 3D touchless device for the simultaneous capture of the 10 fingerprints of both hands based on structured light illumination (SLI) and multiple cameras. Structured light is the projection of a light pattern (plane, grid, or more complex shape) at a known angle onto an object. This technique can be very useful for imaging and acquiring dimensional information. The most often used light pattern is generated by fanning out a light beam into a sheet of light. When a sheet of light intersects with an object, a bright line of light can be seen on the surface of the object. By viewing this line of light from an angle, the observed distortions in the line can be translated into height variations. Scanning the object with the light constructs 3D information about the shape of the object. This is the basic principle behind depth perception for machines, or 3D machine vision.

In order to achieve significant improvements on the manner in which fingerprint images are currently acquired, SLI is used by the University of Kentucky as a mean of extracting the 3D shape of the fingerprint ridges using multiple, commodity digital cameras to acquire the fingerprint scan of all five fingers simultaneously without physical contact between the sensor and the finger. The scan process takes only 200 ms. In order to obtain a 2D rolled-equivalent fingerprint from a 3D scan, certain post-process steps are required after the acquisition. Details of this approach can be found in [21].

The advantage of this approach with respect to the TBS one is really significative. While the Surround Imager is only able to provide the 3D shape of the finger, the structured light approach of TBS provides also the 3D details of the ridge-valley structure. This has a significant impact on the vulnerability of these devices as it will be shown in the next section.

#### 4.4 Sensor Interoperability

Sensor interoperability refers to scenarios where matching is performed between fingerprint images collected from different sensor technologies (e.g., capacitive vs. optical, touch-based vs. touchless), sensor sizes, and fingerprint types (flat, rolled) at either enrollment or verification time or both. A case study [22] has suggested that sensor interoperability has significant impact on the authentication performance of a fingerprint system. In particular, when the images being matched originate from two different sensors (optical and solid state), the matching performance drastically decreases. Similar phenomenon is also observed when rolled prints are compared with flat prints. In manual fingerprint matching, forensic experts also face sensor interoperability issue as images from different sensors provide different size and quality of fingerprint data.

Note that the problem of sensor interoperability cannot be fully solved by adopting a common biometric data exchange format [23]. Such a format merely aids in the exchange of feature sets between systems/vendors [24]. Even in the ANSI/NIST common minutiae exchange format, vendors can enter proprietary fields. To address the sensor interoperability issue, we need to investigate the fundamental differences in the intrinsic characteristics of different sensing technologies and possibly develop algorithms to achieve compatibility.

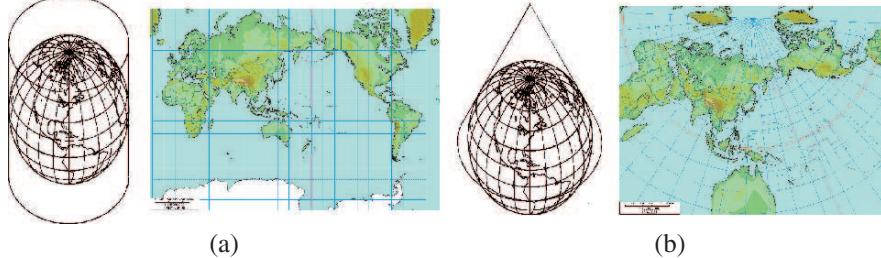
As discussed above, there are fundamental differences between the 3D touchless sensing and to 2D touch-based sensing technology. Although touchless sensing has shown great advantages over touch-based sensing, it is infeasible if touchless fingerprints are not interoperable with the touch-based legacy fingerprints, especially for forensic applications. In the next section, the interoperability issue between 3D touchless and 2D touch-based fingerprints are studied. To address the intrinsic differences between the two technologies, we have developed algorithms to convert the 3D touchless fingerprints to a rolled-equivalent 2D image basic on a “virtual rolling” concept. The resulting images are then compared with legacy rolled fingerprints in our experiments to demonstrate their interoperability.

#### 4.5 Virtual Rolling of 3D Touchless Fingerprints

The methodology to unwrap 3D objects to 2D has been extensively studied and applied in geographic information systems (GIS). One example application is map projection, which focuses on how to unwrap the globe to match with inherently flat geographic maps on paper and films [25, 26]. Other fields, including medical imaging, surface recognition, and industrial design, also involve unwrapping of 3D objects.

In general, there are two main types of unwrapping methods, parametric and non-parametric.

1. Parametric unwrapping refers to the projection of the 3D object onto a parametric model (i.e., cylindrical or conic) and the unwrapping of the model. This method often involves simple and straightforward transformations. But it also requires



**Fig. 4.9** Globe unwrapping using (a) cylindrical model; (b) conic model. Adopted from [27]

that the chosen parametric model fits the shape of the object. Otherwise, large distortions may be introduced during unwrapping.

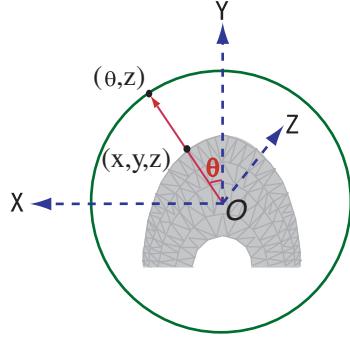
2. Non-parametric unwrapping, on the other hand, does not involve any projection on parametric models. Instead, the unwrapping is directly applied to the object to preserve local distances or angular relations. This method is often employed for irregular-shaped objects.

Figure 4.9 shows the unwrapping of the globe using two different parametric models: cylindrical and conic. Note it is not possible to unwrap the 3D sphere to a 2D plane without introducing some distortion. One can only try to minimize the distortion by using multiple models for different portions of the object to best approximate the shape locally, as shown in the figure. In the case of 3D fingerprint unwrapping, this limitation also applies because although human finger can be approximated as a cylinder or cone, distortion is still unavoidable, especially at the fingertip. In the next two sections, we will give examples of unwrapping 3D fingerprints using a parametric cylindrical model and a non-parametric method based on equidistance. We will compare the two methods and show that distortion introduced by the parametric method can be noticeably large, whereas the non-parametric method demonstrates more faithful representation of the “ground truth” of the fingerprint.

#### 4.5.1 Parametric-Based Virtual Rolling

Although human fingers vary in shape, for example, the middle finger is often more cylindrical than the thumb; it is generally true that they can be closely approximated by a cylinder. As a result, we adopt the cylindrical model for parametric unwrapping of 3D fingerprints. A simple illustration of the cylindrical-based unwrapping is to imagine projecting the fingerprint texture onto a cylinder surrounding the finger, and then unwrapping (flattening) the cylinder to obtain the 2D fingerprint. Mathematically, let the origin be positioned at the bottom of the finger, centered at the principal axis of the finger. Let  $T$  be a point on the surface of the 3D finger,  $T = (x, y, z)^T$ .

**Fig. 4.10** Parametric unwrapping using a cylindrical model (*top-down* view). Point  $(x, y, z)$  on the 3D finger is projected to  $(\theta, z)$  on the 2D plane



This 3D point is then projected (transformed) onto the cylindrical surface to obtain the corresponding 2D coordinates  $S = (\theta, z)^T$ , where  $\theta = \tan^{-1}(\frac{x}{y})$ .

A top-down view of the unwrapping model is shown in Fig. 4.10, where the Z-axis points outward from the origin. It must be noted that the finger is represented as a triangular mesh after 3D reconstruction and each vertex on a triangle would be directly projected using the above transformation. As a result, each triangle on the 3D finger would be mapped to a triangle on the cylinder, whereas points in between vertices of the triangle would be mapped using a linear interpolation.

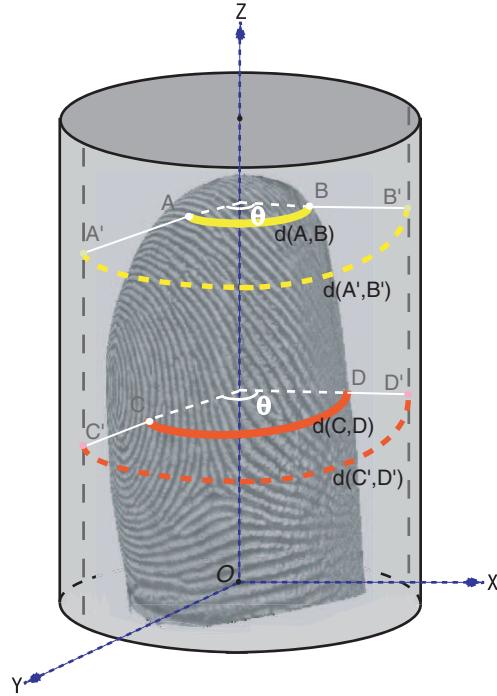
Parametric fingerprint unwrapping using the cylindrical model is efficient and straightforward, but it does not preserve the relative distances between points on the finger surface. Figure 4.11 provides a visual illustration of the problem. For example, the surface distance  $d(A, B)$  between points  $A$  and  $B$  at the fingertip is much smaller than the distance between points  $C$  and  $D$  ( $d(C, D)$ ) near the middle of the finger, but since they both correspond to the same angle  $\theta$ , the unwrapped distances  $d(A', B')$  and  $d(C', D')$  become equal. In general, each cross section of the finger, big or small, is projected into a fixed-length row in the projected 2D image. As a result, horizontal distortion is introduced as the fingerprint will be noticeably stretched, especially at the fingertip, as shown in Fig. 4.15(a).

In addition to the large stretching effects, parametric unwrapping often has limitations in preserving the size of the finger. Using the cylindrical model as an example, the mapping in the horizontal direction is based on the angle rather than the surface distance, and hence, size differences in the horizontal direction between different fingers also need to be compensated for after the unwrapping.

#### 4.5.2 Non-parametric-Based Virtual Rolling

In the non-parametric approach, an object with arbitrary shape is directly unwrapped without being projected onto a model. Instead, the objective is to preserve the geodesic distance between any two points in a local region on the object surface. This is a desirable property for our fingerprint application because the matching

**Fig. 4.11** Fingerprint unwrapping using the cylindrical model. Relative distances between points on the finger surface are not preserved after the unwrapping procedure



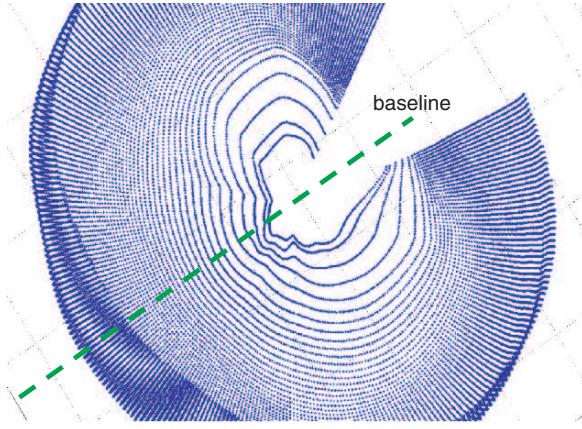
of fingerprints is often performed based on the distances between feature points. If we can preserve such distances after unwrapping, the problem of interoperability between touchless and contact-based fingerprints is then reduced to accounting for skin deformation. Since no parametric model is used, this method also guarantees that the variability in both shape and size of fingers is preserved.

The essential idea of the proposed non-parametric method is to “locally unfold” the finger surface such that both inter-point surface distances and scale are preserved to a maximum degree. More specifically, for a given 3D finger, we divide it into thin parallel slices, orthogonal to the principal axis of the finger, and unfold each slice without stretching. Because human fingers have very smooth structure, as long as each slice is sufficiently thin, the resulting unwrapped fingerprint texture will be smooth.

Figure 4.12 shows the triangular mesh representation of a 3D finger, where only vertices (no lines) of triangles are shown. Note that these vertices naturally form slices at different heights of the finger. However, distances between slices are too large to obtain a smooth unwrapping. As a result, linear interpolation is used to first extract more slices in between the vertices and create a more dense representation.

Let  $S_i$  and  $S_{i+1}$  be the given slices from the triangular mesh and  $h$  be the step size (distance between slices in the dense representation) for interpolation. Figure 4.13 gives an illustration of the procedure. Let  $S_i.P_j$ ,  $S_i.P_{j+1}$ , and  $S_{i+1}.P_k$  be the three

**Fig. 4.12** The 3D representation of the finger. Vertices of the triangular mesh are naturally divided into slices



vertices of a given triangle. The position of the interpolated point  $S_{i,1}.P_a$  is obtained as follows:

$$S_{i,1}.P_a.x = t \times S_{i+1}.P_k.x + (1 - t) \times S_i.P_j.x \quad (4.1)$$

$$S_{i,1}.P_a.y = t \times S_{i+1}.P_k.y + (1 - t) \times S_i.P_j.y \quad (4.2)$$

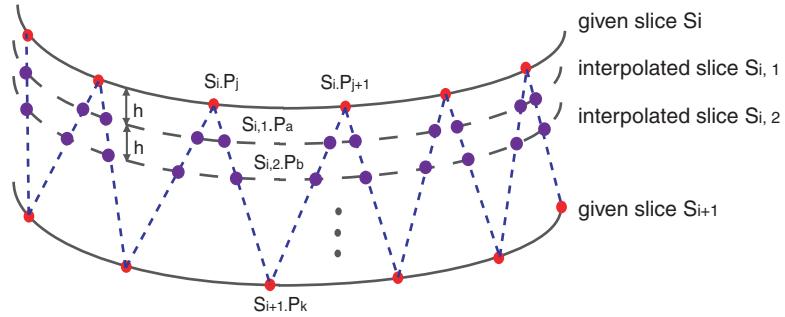
$$S_{i,1}.P_a.z = S_i.P_j.z + h, \quad (4.3)$$

where

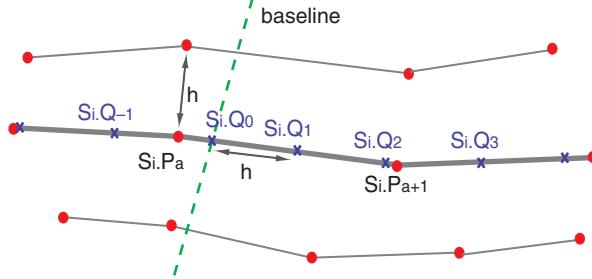
$$t = \frac{S_{i,1}.P_a.z - S_i.P_j.z}{S_i.P_j.z - S_{i+1}.P_k.z} \quad (4.4)$$

is the proportion parameter. This procedure is repeated for every step size  $h$  along the  $z$ -axis; each slice in the final dense representation corresponds to a row in the final unwrapped fingerprint image.

Once a dense representation in height has been established, we apply similar interpolation on each slice to resample points at equidistance  $h$  such that the



**Fig. 4.13** Slice interpolation. We interpolate between given slices with a step size  $h$  to obtain finer representation in the vertical direction ( $y$ -axis) for unwrapping



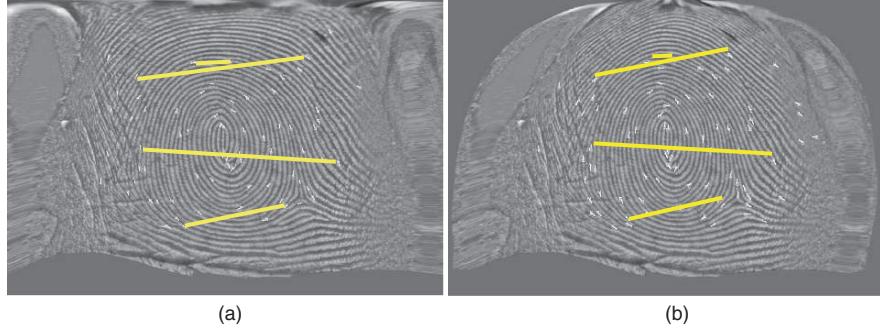
**Fig. 4.14** Equidistant resampling. We resample points on each slice with equal distance  $h$  to obtain finer representation in the horizontal direction for unwrapping. The baseline defines the central column that the fingerprint will be mapped to

neighboring points of the same slice would correspond to neighboring columns of the same row in the final unwrapped image. This step size  $h$  is set the same in the vertical direction ( $y$ -axis), to preserve the scale of the finger. A baseline to start unfolding each slice is also defined as the intersecting line (curve) between the 3D finger and a plane passing through the principal axis in the center of the finger. In other words, the direction of unwrapping is established from the center of the finger to the nail side to minimize rolling distortion. We will show that the baseline can also be defined at each nail side to simulate directional rolling from left to right or right to left. Figure 4.14 illustrates the virtual rolling procedure and the algorithm is described in detail as follows:

```

for i = 1 : n (iterate through all slices)
    for j = 1 : m (resample to the right)
        dist = ||Si.Pa+j - Si.Qj-1||;
        if (dist > h)
            t =  $\frac{h}{dist}$ ;
            Si.Qj.x = t × Si.Pa+j.x + (1 - t) × Si.Qj-1.x;
            Si.Qj.y = t × Si.Pa+j.y + (1 - t) × Si.Qj-1.y;
        else
            t =  $\frac{h-dist}{\|S_i.P_{a+j+1}-S_i.P_{a+j}\|}$ ;
            Si.Qj.x = t × Si.Pa+j+1.x + (1 - t) × Si.Pa+j.x;
            Si.Qj.y = t × Si.Pa+j+1.y + (1 - t) × Si.Pa+j.y;
    for j = 1 : l (resample to the left)
        dist = ||Si.Pa-j+1 - Si.Q-j+1||;
        if (dist > h)
            t =  $\frac{h}{dist}$ ;
            Si.Q-j.x = t × Si.Pa-j+1.x + (1 - t) × Si.Q-j+1.x;
            Si.Q-j.y = t × Si.Pa-j+1.y + (1 - t) × Si.Q-j+1.y;
        else
            t =  $\frac{h-dist}{\|S_i.P_{a-j+1}-S_i.P_{a-j}\|}$ ;
            Si.Q-j.x = t × Si.Pa-j.x + (1 - t) × Si.Pa-j+1.x;
            Si.Q-j.y = t × Si.Pa-j.y + (1 - t) × Si.Pa-j+1.y;

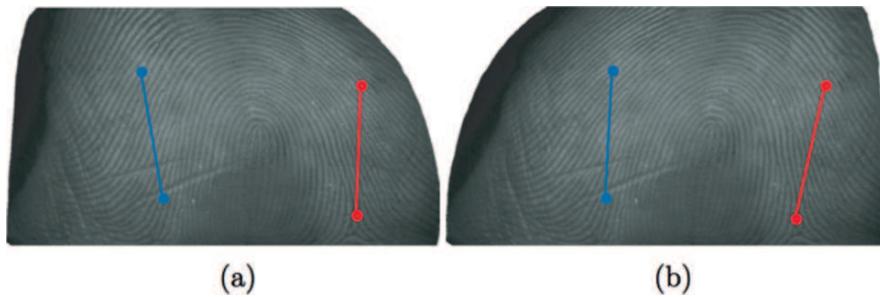
```



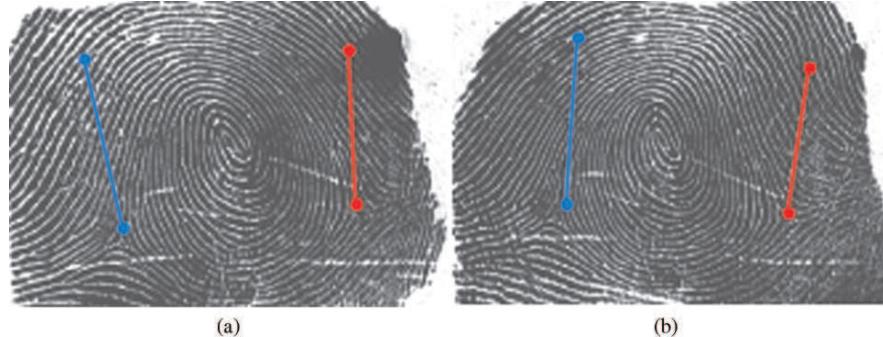
**Fig. 4.15** Unwrapping a 3D fingerprint captured with Surround Imager using **(a)** the cylindrical-based parametric method and **(b)** the proposed non-parametric method. The 2D unwrapped fingerprint in **(b)** is closer to the “ground truth” as it better retains the distances (*yellow solid lines*) between minutiae points (*white arrows*) than that in **(a)**

Figure 4.15(a) and (b) shows an example of the unwrapped touchless fingerprint image using the cylinder-based parametric and the proposed non-parametric methods. Minutiae points (shown as white arrows) are extracted using the feature extraction algorithm in [28] and distances between a few minutiae points (yellow solid lines) are shown in Fig. 4.15. These figures show that the proposed unwrapping method better preserves the inter-point distances with less distortion than the cylinder-based parametric method.

As the baseline of the unwrapping procedure changes from the center of the finger to the nail side, increasing distortion can be observed (see Fig. 4.16). These distortions are equivalent with those introduced by the legacy rolling process as the skin is pushed from one nail side to the other during rolling (see Fig. 4.17). Generally, legacy fingerprint rolling requires all fingers (from their tip to the first joint) to be rolled from “nail to nail.” In particular, thumbs should be rolled from inside to outside of the nail, while other fingers of the right hand should be rolled from left to right, and from right to left for the left hand [29]. If such protocol are to



**Fig. 4.16** Directional distortion in “virtually rolled” touchless fingerprints: **(a)** virtual rolling from *left* to *right*; **(b)** virtual rolling from *right* to *left*. Two sets of minutiae points and their relative distances are marked. Similar changes are also seen due to rolling in different directions



**Fig. 4.17** Directional distortion in legacy rolled fingerprints: **(a)** rolling from *left to right*; **(b)** rolling from *right to left*. Two sets of minutiae points and their relative distances are marked. Changes are seen due to rolling in different directions

be strictly followed, the 3D touchless fingerprints should be unwrapped accordingly to produce compatible distortions caused by rolling.

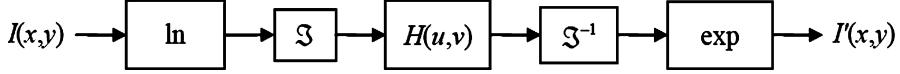
## 4.6 Enhancing Touchless Fingerprints

In touch-based fingerprinting, fingerprints are captured based on the touch/nontouch of the ridges/valleys with the sensor surface, resulting in binary-like images with ridges having gray values close to 0 and valleys close to 255. As a result, Gabor filtering is often used to enhance the image at local ridge orientation and remove noise before minutiae are extracted. In touchless finger imaging, however, fingerprints are captured as images at a distance. As a result, ridges and valleys are not separated during acquisition and the resulting touchless fingerprint images often present lower contrast (higher dynamic range) compared to the legacy fingerprint images. In this case, special enhancement is desired to increase the image contrast before Gabor filtering is applied.

We have developed an enhancement algorithm using homomorphic filters. Homomorphic filtering has been commonly used to increase the image contrast by sharpening features and flattening lighting variations in an image [30]. The image formation model assumed for homomorphic filters are often characterized by two components: illumination  $L$  and reflectance  $R$ . Both components can be combined to give the image function  $F$ :

$$I(x, y) = L(x, y) \cdot R(x, y) \quad (4.5)$$

where  $0 < L(x, y) < \infty$  and  $0 < R(x, y) < 1$ . Illumination component captures the lighting conditions present during image acquisition, while reflectance component represents the way the object reflects light, which is an intrinsic property of the



**Fig. 4.18** Homomorphic filtering procedure

object. The essential idea of homomorphic filtering is to separate the two components using log transform and use a high-pass filter to enhance reflectance while at the same time reduce the contribution of illumination. Specifically, homomorphic filtering consists of the following five steps (see Fig. 4.18):

- Use natural log to transform Eq. (4.5) from multiplicative to additive:

$$Z(x, y) = \ln[I(x, y)] = \ln[L(x, y) \cdot R(x, y)] = \ln[L(x, y)] + \ln[R(x, y)] \quad (4.6)$$

- Apply Fourier transform:

$$\mathfrak{Z}Z(x, y) = \mathfrak{Z}\ln[L(x, y)] + \mathfrak{Z}\ln[R(x, y)] \quad (4.7)$$

- High pass the  $Z(x, y)$  by means of a filter function  $H(u, v)$  in frequency domain:

$$S(u, v) = H(u, v) \cdot Z(u, v) = H(u, v) \cdot \mathfrak{Z}\ln[L(x, y)] + H(u, v) \cdot \mathfrak{Z}\ln[R(x, y)] \quad (4.8)$$

- Apply inverse Fourier transform:

$$E(x, y) = \mathfrak{Z}^{-1}S(u, v) = \mathfrak{Z}^{-1}H(u, v) \cdot \mathfrak{Z}\ln[L(x, y)] + \mathfrak{Z}^{-1}H(u, v) \cdot \mathfrak{Z}\ln[R(x, y)] \quad (4.9)$$

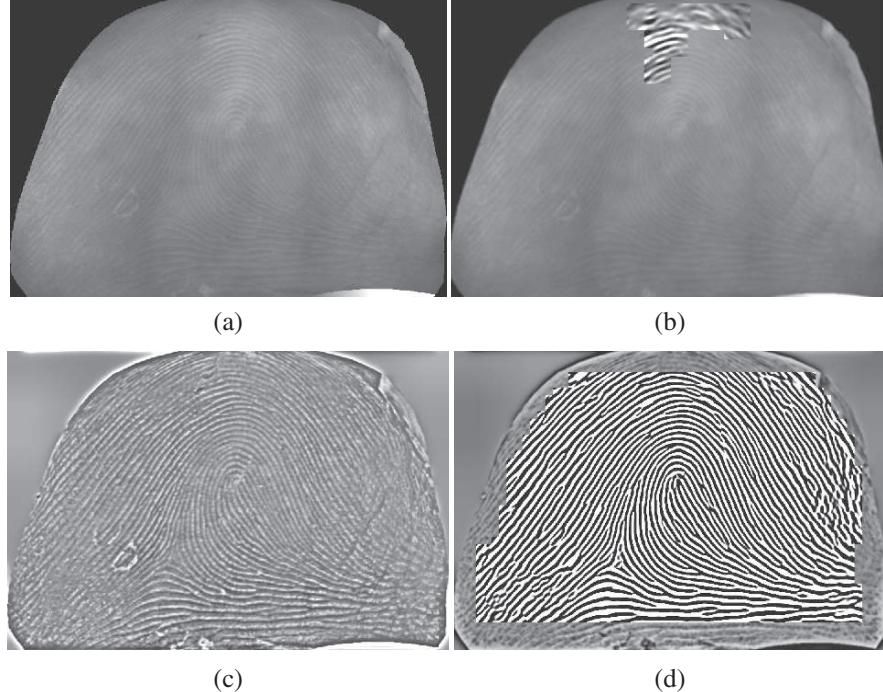
- Obtain the final enhanced image by exponential operation:

$$I'(x, y) = \exp E(x, y) \quad (4.10)$$

The high-pass filter  $H(u, v)$  used in the above transformation is defined as

$$H(u, v) = \frac{1}{1 + [\frac{D_0}{D(u, v)}]^{2n}} \quad (4.11)$$

where  $D(u, v)$  is the Euclidean distance from the origin of the centered transform,  $D_0$  and  $n$  are the cutoff frequency and order of the filter, respectively. The filter  $H(u, v)$  is designed to decrease the contribution of the low frequencies (illumination) and amplify the contribution of mid- and high frequencies (reflectance), which is the key of homomorphic filtering. Note, we also use adaptive histogram equalization to normalize the image after homomorphic filtering. Figure 4.19 shows the enhancement of an unwrapped touchless fingerprint image and the use of Gabor filtering on both the original and the enhanced images. The proposed enhancement greatly increases the image contrast and facilitates Gabor filtering to further extract the ridge/valley patterns in a touchless fingerprint image.



**Fig. 4.19** Touchless fingerprint enhancement: (a) an unwrapped touchless fingerprint image; (b) Gabor filtering of (a); (c) homomorphic enhancement of (a); (d) Gabor filtering of (c)

## 4.7 Touchless Fingerprint Image Quality

Due to imaging artifacts caused by diffused illumination, self-occlusion, and perspective distortion, the image quality of the captured touchless fingerprints vary in different local regions. This effect becomes very significant when different absorption properties of the tissues and blood vessels in the finger cause non-uniform illumination during image acquisition. As a result, it is desirable to develop a quality measure for touchless fingerprint images that takes consideration of local quality variations.

In order to estimate the quality of touchless fingerprint images, we implement a coherence-based local quality measure [31]. Coherence, estimate based on local gradients, is a measure of the strength of the dominant direction in a local region. For example, given an unwrapped touchless fingerprint  $I$ , we first partition it into a lattice of blocks of size  $b \times b$ . At each block  $B$ , let  $g_s$  denote the gradient of the gray level intensity at site  $s \in B$ , and the covariance matrix of the gradient vectors for all  $b^2$  sites in this block is given by

$$J = \frac{1}{b^2} \sum_{s \in B} g_s g_s^T \equiv \begin{bmatrix} G_x^2 & G_{xy} \\ G_{yx} & G_y^2 \end{bmatrix} \quad (4.12)$$

where  $\begin{bmatrix} G_x^2 & G_{xy} \\ G_{yx} & G_y^2 \end{bmatrix}$  is also called the Hessian matrix. This symmetric matrix is positive semidefinite with eigenvalues

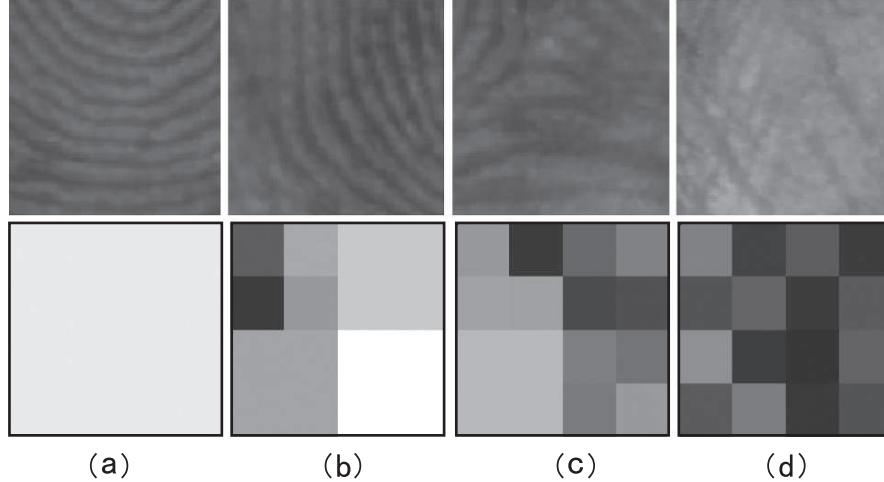
$$\begin{aligned}\lambda_1 &= \frac{1}{2}(trace(J) + \sqrt{trace^2(J) - 4 det(J)}) \\ \lambda_2 &= \frac{1}{2}(trace(J) - \sqrt{trace^2(J) - 4 det(J)})\end{aligned}\quad (4.13)$$

where  $trace(J) = G_x^2 + G_y^2$ ,  $det(J) = G_x^2 G_y^2 - G_{xy} G_{yx}$ , and  $\lambda_1 \geq \lambda_2$ . The block-wise quality is then measured by the *normalized coherence*, defined as

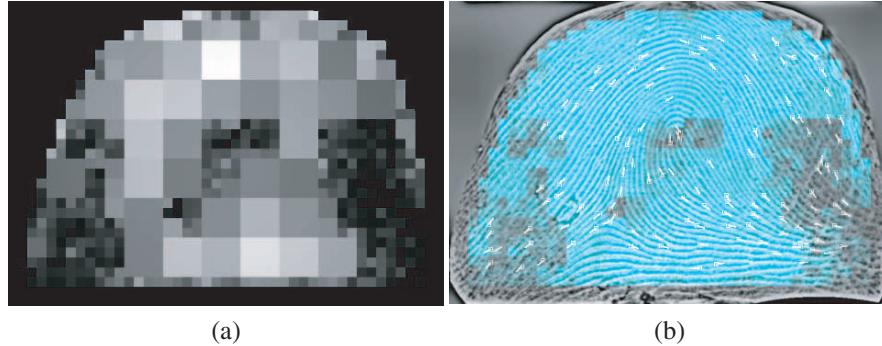
$$q = \frac{(\lambda_1 - \lambda_2)^2}{(\lambda_1 + \lambda_2)^2} = \frac{(G_x^2 - G_y^2)^2 + 4G_{xy}G_{yx}}{(G_x^2 + G_y^2)^2} \quad (4.14)$$

with  $0 \leq q \leq 1$ . This measure reflects the clarity of the local ridge–valley structure in each foreground block  $B$ . If the local structure presents a distinct orientation, then  $\lambda_1 \gg \lambda_2$ , resulting in  $q \approx 1$ . On the contrary, if the local structure lacks a clear orientation, we obtain  $\lambda_1 \approx \lambda_2$  and consequently  $q \approx 0$ . Figure 4.20 shows the resulting block-wise local quality measure.

To obtain a smooth evaluation of quality, we further extend the above quality estimation in a cascade framework by progressively investigating quality. That is, instead of estimating quality block by block, we first divide the fingerprint foreground into relatively large regions (say block size is  $4b \times 4b$ ), and estimate the quality using the coherence-based measure proposed above. If the quality value of a



**Fig. 4.20** Cascade quality assessment in local regions (80 × 80 pixels) with (a) excellent quality; (b) good quality; (c) median quality; and (d) poor quality. The *first* and *second* rows show the fingerprint regions and their corresponding cascade quality, respectively

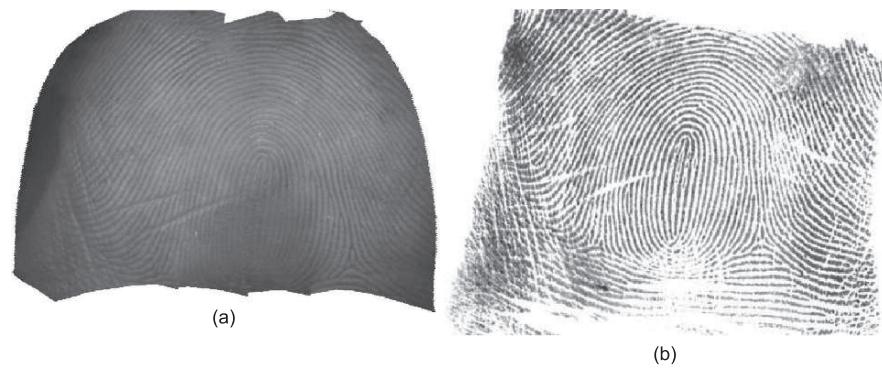


**Fig. 4.21** Touchless fingerprint quality: (a) local quality map estimated for Fig. 4.19 (a); (b) minutiae overlaid with local quality map. Minutiae extracted in poor quality regions are shown less reliable than those extracted in good quality regions

region is sufficiently high, the estimation process for this region is terminated, meaning that the whole region is of good quality. Otherwise, that region is divided into four quadrants (each with size  $2b \times 2b$ ) for more detailed estimation. This process continues until individual blocks are reached. Finally, an overall quality measure of the given fingerprint image is obtained by taking the average of the local quality measures. Figure 4.20 shows the quality estimated at four local regions ( $80 \times 80$  pixels) in Fig. 4.19 (a) and Fig. 4.21 shows the overall quality map of the image.

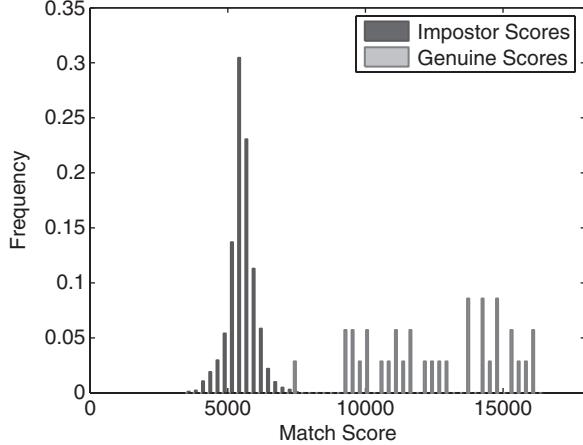
#### 4.7.1 Experiments

To demonstrate the compatibility of the unwrapped touchless fingerprints with legacy rolled images, TBS provided a small database with 38 fingers captured using a line-scan sensor (at 1000 ppi). For each finger, one 3D touchless print and one 2D ink-on-paper rolled print are obtained (see Fig. 4.22). The 3D touchless fingerprints



**Fig. 4.22** Visualizing compatibility between (a) a touchless fingerprint from line-scan sensor using the proposed non-parametric unwrapping; (b) the corresponding ink-on-paper rolled fingerprint

**Fig. 4.23** Matching touchless with ink-on-paper fingerprints. Genuine and impostor match scores are well separated with only one genuine score (7,483) below the maximum impostor score (7,725), indicating that touchless fingerprints are compatible with legacy rolled fingerprint images

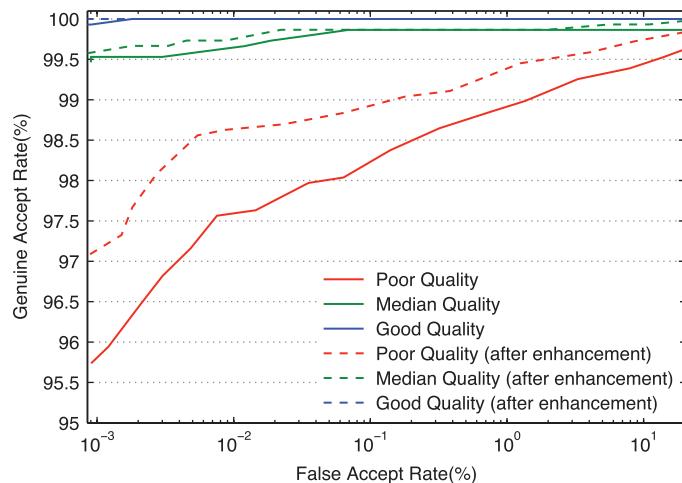
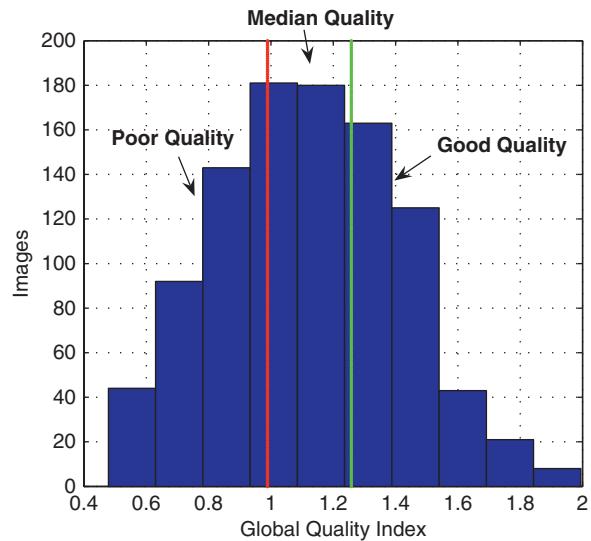


are then converted to 2D using the proposed non-parametric “virtual rolling” algorithm. The resulting 2D rolled-equivalent fingerprints are also enhanced using the proposed homomorphic filtering with parameters  $D_0 = 0.1$  and  $n = 1$ .

To evaluate the interoperability between the converted touchless and touch-based fingerprints, we use a commercial fingerprint matcher (Neurotechnologija Verifier 4.2) [32] to evaluate the cross-matching performance. Figure 4.23 shows the distributions of all the match scores. In total, there are 38 genuine scores and 2812 impostor scores, which include all within-group (ink-on-paper vs. ink-on-paper) and between-group (touchless vs. ink-on-paper) impostor scores. There is almost no overlap (only one genuine score (7483) is below the maximum impostor score (7725)) between genuine and impostor score distributions, suggesting compatibility between touchless and ink-on-paper rolled fingerprints in this small database.

To evaluate the effects of the proposed enhancement and quality evaluation algorithms, we conduct experiments on a larger touchless fingerprint database, containing 1000 touchless fingerprint images (102 fingers,  $\sim 10$  impressions per finger) at 500 ppi, captured using the TBS Surround Imager. These touchless fingerprints are converted using the proposed “virtual rolling” algorithm and the quality of each image is estimated using the proposed quality measure. Figure 4.24 shows the histogram of the estimated overall quality indices for all images in the database. We then divide the data into three equal-size quality bins, namely good, median, and poor, and perform matching in each quality bin. Figure 4.25 shows the ROC curves of matching in each quality bin as solid lines. The ROC curves degrade as image quality drops, suggesting that the proposed quality index is a good indicator of the matching performance. Next, we perform the proposed homomorphic filtering enhancement on all images and perform matching again in each quality bin. The corresponding ROC curves are shown in dashed lines in Fig. 4.25. It is shown that the proposed enhancement algorithm leads to consistently higher matching accuracy among all three image quality bins, with the most significant improvement in the poor quality bin.

**Fig. 4.24** Histogram of the estimated global quality indices. The green line and red line indicate the thresholds that divide the database into three equal-sized quality groups



**Fig. 4.25** ROC curves for the three quality-based partitions of the database before and after image enhancement

## 4.8 Vulnerability of Touchless Fingerprint Imaging Devices

Fingerprint recognition is not spoof-proof. One of the potential threats for fingerprint recognition systems is the fake finger attack. Recently, the feasibility of this attack model was studied by a few researchers [2, 33] and it has been shown that many commercial systems are vulnerable to spoof attacks by synthetic fingerprints

developed with materials such as gelatin, silicon, and latex. Due to the lack of liveness detection capability, fingerprint recognition has been limited from high-secure applications.

Like any other fingerprint technology, touchless fingerprint technology is not free from this problem. In fact, it may even make the attack easier compared to touch-based technology. As explained earlier, the multi-camera imaging approach proposed by TBS has some disadvantages compared to the structured lighting approach proposed by Flashscan 3D, as the former only captures the 2D texture of a finger (which is then mapped onto a 3D shape obtained by the stereovision), while the latter captures all details on a finger in 3D. As a result, a 2D picture or simply a drawing of a fingerprint on a matte paper presented in front of the sensor cameras is sufficient to attack the system and obtain an impostor access.

We conduct the following experiment based on a database of 10 subjects (6 males and 4 females). The images were collected from each finger of each subject in three sessions during a 2-month period using the TBS Surround Imager. Due to problems with the sensor in capturing large fingers, 4 out of the 6 male subjects were not able to provide their thumbs. As a result, a total of 276 ( $= 10 \times 10 - 4 \times 2$ ) fingerprints were collected. We randomly selected 30 fingerprints from the database as the *set for the attack*. Among them, half were rescaled and printed with a commercial laser printer at a resolution of 3000 ppi and the other half were printed and manually traced (drawn) on a semi-transparent paper. The drawing was then scanned using a flatbed scanner, re-scaled using a image editor software, and printed using a laser printer. To increase the absorption of the light and simulate more precisely the skin reflection properties, we use a paper with a similar color of the light sources of the sensor.

Next, the 30 printed fingerprint images were folded on a finger and presented to the TBS Surround Imager. As expected, the fake fingerprints are acquired without liveness being checked. Figure 4.26 illustrates the spoof attack, in which a fingerprint image was printed on a green paper and then glued on to a finger. A commercial algorithm was used to match the fake fingerprint against the authentic fingerprint image obtained from the real finger. The experiment was repeated for 20 fingerprints and the false acceptance rate was 100%.

To improve the security of touchless fingerprint systems, Diaz-Santana et al. [35] proposed a liveness detection method that takes advantage of the sweating activity of the human skin. Using high-magnification lenses and special illumination techniques, it is possible to capture the continuous perspiration activity of the pores present on the fingerprint friction ridges. The presence of this activity ensures the liveness of the finger and hence, protects against a fake fingerprint attack as described above.

To demonstrate the feasibility of this approach, experiments were also conducted using a new generation high-definition camera (Sony HDR-H1C1E), optical-fiber illuminators, and lenses with a magnification factor of 40x. The frame rate and the shutter speed of the camera were fixed to 30 frame/s and 1/250 ms, respectively. The acquired sweating pores video sequences were then down-sampled to 1 frame/s to reduce computational load. Each frame was processed to extract the



**Fig. 4.26** Process to prepare a fake fingerprint to attack the Surround Imager. A drawing is obtained from a printout of an image generated by the device. The drawing is then scaled down and folded on to a finger. More details on [34]

sweat pores using wavelets (top-hat wavelets) and traditional motion tracking techniques (optical-flow) were used to follow the presence/absence of sweat coming out from the pores present on the fingerprint friction ridges.

Note that the proposed liveness detection approach is very costly due to the high cost of the lenses, the camera, and the illuminators used. Hence, the implementation of this method would be only feasible for high-secure applications where the loss generated by a false acceptance is greater than the cost of the device. Moreover, the proposed liveness detection method is only suitable for touchless devices. In case of touch-based technology, the sweat spreads on the platen reflecting all the light coming from the LED and the perspiration activity would not be visible.

## 4.9 Summary

Touchless imaging, especially 3D touchless imaging, is a novel fingerprint sensing technology. Because of its contact-free property, the resulting fingerprint images well preserve the fingerprint ground truth and are likely to lead to more consistent and accurate fingerprint representation and matching. However, designing a touchless fingerprint imaging device is a challenging problem as it heavily relies

on the proper positioning and configuration of optics and are prone to problems that are intrinsic to imaging, e.g., diffused illumination, self-occlusion, and perspective distortion. In addition, the fingerprint images obtained using the touchless sensing technology have many different characteristics compared to legacy fingerprint images, thus are not fully compatible. In this chapter, we highlight some major approaches to touchless device design and their principles. We also propose a virtual rolling process to achieve compatibility between 3D touchless fingerprints and 2D legacy fingerprints. Further, we propose an image enhancement algorithm specifically designed for touchless fingerprints to optimize the contrast between ridges and valleys. The effectiveness of this enhancement algorithm is demonstrated by improving the matching accuracy of touchless fingerprints, especially among poor-quality images. Finally, we discuss the vulnerability and potential methods to achieve liveness detection capability of touchless sensing devices.

## References

1. E. German, *Latent Print Examination*, <http://www.onin.com/fp>, 2007.
2. Matsumoto, T., Matsumoto, H., Yamada, K. and Hoshino, S., *Impact of Artificial Gummy Fingers on Fingerprint Systems*. Proceedings SPIE, Vol. 4677, pp. 275–289, San Jose, USA, Feb 2002.
3. Bolle, R.M., Connell, J.H. and Ratha, N.K., *Biometric Perils and Patches*. Pattern Recognition, vol. 25, no. 12, pp. 2727–2738.
4. Matsumoto, T., Matsumoto, H., Yamada, K. and Hoshino, S., *Impact of Artificial Gummy Fingers on Fingerprint Systems*. Proceedings on SPIE, Vol. 4677, pp. 275–289, San Jose, USA, Feb 2002.
5. *Testimony of Jim Williams Director in US-VISIT Program, Department of Homeland Security, Before The Senate Appropriations Subcommittee on Homeland Security, January 25, 2006*, <http://appropriations.senate.gov/hearmarkups/JWTestimonyFINAL.pdf>.
6. Parziale, G., *Touchless Fingerprinting Technology*, a chapter in *Advances in Biometrics: Sensors, Systems and Algorithms*, Eds. by Nalini K. Ratha and Venu Govindaraju, Springer-Verlag Ltd, Berlin, Dec 2007.
7. Elli, A., *Understanding the Color of Human Skin*. Proceedings of the 2001 SPIE conference on Human Vision and Electronic Imaging VI, SPIE Vol. 4299, pp. 243–251.
8. Song, Y., Lee, C. and Kim, J., *A New Scheme for Touchless Fingerprint Recognition System*. Proceedings of 2004 International Symposium on Intelligent Signal Processing and Communication Systems, pp. 524–527.
9. Krzysztof M., Preda M. and Axel M., *Dynamic Threshold Using Polynomial Surface Regression with Application to The Binarization of Fingerprints*. Proceedings of SPIE on Biometric Technology for Human Identification, Orlando, USA, pp. 94–104, 2005.
10. Lee, C., Lee, S. and Kim, J., *A Study of Touchless Fingerprint Recognition System*. Springer, New York, Vol. 4109, pp. 358–365, 2006.
11. Sano, E., Maeda, T., Nakamura, T., Shikai, M., Sakata, K., Matsushita, M. and Sasakawa, K., *Fingerprint Authentication Device Based on Optical Characteristics Inside a Finger*. In Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (June 17–22, 2006). CVPRW. IEEE Computer Society, Washington, DC, 27.
12. Shiratsuki, A., Sano, E., Shikai, M., Nakashima, T., Takashima, T., Ohmi, M. and Haruna, M., *Novel Optical Fingerprint Sensor Utilizing Optical Characteristics of Skin Tissue Under Fingerprints*. International Society for Optical Engineering, Proceedings SPIE, Vol. 5686, pp. 80–87, 2006.

13. TBS Touchless Fingerprint Imaging, <http://www.tbsinc.com>.
14. Parziale, G., Diaz-Santana, E. and Hauke, R., *The Surround Imager: A Multi-Camera Touchless Device to Acquire 3D Rolled-Equivalent Fingerprints*. Proceedings of IAPR International Conference on Biometrics (ICB), pp. 244–250, Hong Kong, China.
15. Hauke, R., Parziale, G. and Paar, G., *Method and Arrangement for Optical Recording of Biometric Data*. Patent. PCT/DE2004/002026, 2004.
16. Tsai, R., *An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision*. Proceedings of IEEE Conference on computer Vision and Pattern Recognition, Florida, USA, 1086, pp. 364–374.
17. Sonka, M., Hlavac, V. and Boyle, R., *Image Processing, Analysis, and Machine Vision*. Second Edition, Brooks/Cole Publishing, USA. 1999.
18. Gruen, A. and Huang, T.A. (Eds.), *Calibration and Orientation of Cameras in Computer Vision*. Springer-Verlag, Berlin, 2001
19. Hartley, R. and Zisserman, A., *Multiple View Geometry in Computer Vision*. Cambridge University Press, UK, 2003.
20. Flashscan 3D Touchless Fingerprint Sensor, <http://www.flashscan3d.com/>.
21. Fatehpuria, A., Lau, D.L., and Hassebrook, L.G., *Acquiring a 2-D Rolled Equivalent Fingerprint Image from a Non-Contact 3-D Finger Scan*. Biometric Technology for Human Identification III, edited by Patrick J. Flynn, Sharath Pankanti, SPIE Defense and Security Symposium, Orlando, Florida, Vol. 6202, pp. 62020C-1 to 62020C-8, 2006.
22. Ross, A. and Jain, A.K., *Biometric Sensor Interoperability: A Case Study in Fingerprints*, in Proceedings of ECCV International Workshop on Biometric Authentication (BioAW), Prague, Czech Republic, May 2004, vol. LNCS 3087, pp. 134–145, Springer, New York.
23. Bolle, R.M., Rathna, N.K., Senior, A. and Pankanti, S., *Minutia Template Exchange Format*, in Proc. of IEEE Workshop on Automatic Identification Advanced Technologies, 1999, pp. 74–77.
24. Podio, F.L., Dunn, J. S., Reinert, L., Tilton, C.J., O’Gorman, L., Collier, P., Jerde, B. and Wirtz, M., *Common Biometric Exchange File Format (CBEFF)*, Technical Report NISTIR 6529, NIST, January 2001.
25. Snyder, J. P., *Flattening the Earth: Two Thousand Years of Map Projections*, The University of Chicago Press, Chicago, 1993.
26. Yang, O., Tobler, W., Snyder, J. and Yang, Q. H., *Map Projection Transformation*, Taylor and Francis, Abinsdon, 2000.
27. From FLand: map projection methods, <http://gpscycling.net/fland/map/projection.html>
28. Jain, A.K., Hong, L. and Bolle, R., *On-line Fingerprint Verification*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 4, pp. 302–314, April 1997.
29. *Tips for Rolling Fingerprints*, <http://apps.mentoring.org/safetynet/fingertips.pdf>.
30. Gonzalez R.C. and Woods, R.E., *Digital Image Processing*, Prentice Hall, Upper Saddle River, NJ, 2002.
31. Chen, Y., Dass, S. C. and Jain, A. K., *Fingerprint Quality Indices for Predicting Authentication Performance*, in Proc. International Conference on Audio- and Video-Based Biometric Person Authentication, pp. 160–170, 2005.
32. *Neurotechnologija Verifinger SDK*, <http://www.neurotechnologija.com/vfsdk.html>.
33. Putte, T. and Keuning, J., *Biometrical Fingerprint Recognition: Dont Get Your Fingers Burned*. Proc. IFIP TC8/WG8.8, 4th Working Conf. Smart Card Research and Adv. App. pp. 289–303, 2000.
34. iFingerSys, <http://www.ifingersys.com>
35. Diaz-Santana, E. and Parziale, G., *Liveness Detection Method*. EP1872719, 2006.

# **Chapter 5**

## **Face Recognition in Humans and Machines**

**Alice O'Toole and Massimo Tistarelli**

**Abstract** The study of human face recognition by psychologists and neuroscientists has run parallel to the development of automatic face recognition technologies by computer scientists and engineers. In both cases, there are analogous steps of data acquisition, image processing, and the formation of representations that can support the complex and diverse tasks we accomplish with faces. These processes can be understood and compared in the context of their neural and computational implementations. In this chapter, we present the essential elements of face recognition by humans and machines, taking a perspective that spans psychological, neural, and computational approaches. From the human side, we overview the methods and techniques used in the neurobiology of face recognition, the underlying neural architecture of the system, the role of visual attention, and the nature of the representations that emerges. From the computational side, we discuss face recognition technologies and the strategies they use to overcome challenges to robust operation over viewing parameters. Finally, we conclude the chapter with a look at some recent studies that compare human and machine performances at face recognition.

### **5.1 Introduction**

Face recognition is a fundamental skill that humans acquire early in life and which remains an integral part of our perceptual and social abilities throughout our life span. Faces provide information about the identity of people, about their membership in broad demographic categories of humans (including sex, race, and age), and about their current emotional state. Humans perceive this information effortlessly and apply it to the ever-changing demands of cognitive and social interactions. Face recognition is now, also, within the capabilities of machines. It is clear that machines are becoming increasingly accurate at the task of face recognition. In this chapter we provide a survey of the essential elements of face recognition by man and machine. We look first at what is known about how humans recognize faces

---

A. O'Toole (✉)  
University of Texas, Dallas, TX, USA  
e-mail: otoole@utdallas.edu

from both a psychological and a neural perspective. Next, we turn our attention to face recognition technologies and to the challenges they must overcome to be useful in applications. Finally, we discuss what is known about the performance of face recognition algorithms relative to humans.

## 5.2 Human Face Perception and Recognition

Face recognition by “biological systems” has been studied for decades using techniques from psychology and neuroscience. Psychologists have examined the factors that affect human perception and memory for faces beginning from the most basic of viewing parameters to more complex factors useful for understanding the “memorability” of individual faces. Perceptual illusions and distortions of faces using surprisingly simple methodologies have also given insight into the nature of human representations of faces. The factors that affect human memory for faces can also tell us about the quality and flexibility of face representations. From a neuroscience perspective, specialized brain areas and neural mechanisms have been posited for faces based on neuropsychological case studies of prosopagnosia [1], electrophysiological recordings in monkey cortex, and more recently on data from functional neuroimaging studies [2]. Combined, techniques from neuroscience have indicated a number of brain regions that activate selectively when people look at faces.

In this part of the chapter, we begin first with an overview of technologies and methods available to study the neurobiology of human face perception and recognition (Section 5.2.1). These methods contribute data from distinctly different perspectives and must be integrated to understand the complete set of neural and psychological processes involved in human face recognition. Second, we discuss what is known about the neural architectures for the complex types of facial analyses carried out by the human visual system (Section 5.2.2). As we will see, these brain areas are organized according to the tasks they carry out, rather than by image analysis issues that seem more computationally intuitive. In the third section, we differentiate the foveal and peripheral processings of faces (Section 5.2.3). Fourth, we consider the role of visual attention in face processing (Section 5.2.4). Fifth, we will take stock of the combination of factors presented to this point to begin to consider how faces are represented in the human visual system (Section 5.2.5). An important part of this is to take into account the relevance of face motion and its role in perception and memory (Section 5.2.6).

At the outset, it is worth noting that the information in human faces serves a number of distinct, ecologically necessary tasks. Faces provide information for the recognition of individuals, for visually based categorizations (e.g., by sex, race, and age), and for various kinds of adaptive social interaction (e.g., perceiving facial expression, intent, and direction of gaze). An overriding theme of both the psychological and neural literatures in recent years is that there are multiple brain areas and representations involved in processing faces. Concomitantly, numerous findings connect across the psychological and neural literatures enabling insights that would be unavailable from either discipline alone. We will focus on techniques and some

key findings that have had important implications for understanding the nature of human face representations.

### **5.2.1 Technologies and Methods in the Neurobiology of Face Perception**

Several techniques and technologies from neuroscience have provided a wealth of information about the nature of human representations of faces. These include: (a) neuropsychological case studies of patients with brain injury; (b) neurophysiological studies in monkeys that provided the first evidence of neurons that respond selectively to faces; (c) and functional neuroimaging studies that have offered insight into the network of brain areas important for face processing. In each case, we will present the method and discuss some classic findings that have changed the kinds of questions we can ask about the structure of face processing in the brain.

#### **5.2.1.1 Neuropsychological Case Studies**

To date, one of the most compelling demonstrations of a dedicated neural face processing system in the human brain is based on neuropsychological case studies of prosopagnosia, a selective deficit in recognizing faces following brain injury [1]. Prosopagnosia occurs with no associated deficits in object recognition and no general deficits in intelligence, memory, or cognitive function [3]. Prosopagnosics are aware of their difficulties in recognizing faces and rely prominently on the coding of peripheral visual cues like clothing or on non-visual cues such as voices to recognize familiar individuals.

Studies that have examined brain damage in prosopagnosia have revealed bilateral lesions in the inferior temporal (IT) cortex [4, 5] and occasionally unilateral lesions in the right hemisphere of IT [6, 7]. As we shall see, the data from these case studies are largely consistent with the results from other more sophisticated functional neuroimaging technologies suggesting IT cortex as a primary brain region selective for face recognition.

Neuropsychological case studies have also reported cases where expression processing in faces is impaired, though this is less common given the redundancy of these vital emotion-related processing systems cortically and subcortically. Bruce and Young [8] were the first to propose a comprehensive psychological model of the human face processing system. In that model, they proposed functional independence between the neural processing of facial identity and facial expression based largely on neuropsychological evidence in the form of *double dissociations*. Double dissociations occur when multiple neuropsychological cases exist to demonstrate the loss of one function with the sparing of a second related function, and vice versa. In the expression–identity case, patients with prosopagnosia often retain the ability to recognize facial expressions [9–13]. Concomitantly, some brain-damaged patients with expression perception deficits retain the ability to recognize faces [10, 14–18]. Although a strict separation of these systems has been questioned

recently [19], there is still ample reason to believe that identity and expression are processed in parallel but neural sub-systems with different functions.

Combined, these case studies illustrate the role of dedicated brain areas in face processing that operate independently of more general purpose object recognition functions. These studies comprise the first basis of the claim that inferior temporal brain regions are important for face processing. Combined with an understanding of the pattern of brain damage, these kinds of case studies also provided the first evidence for right hemisphere dominance in face processing—a finding that is now well accepted based on converging evidence in functional neuroimaging.

### 5.2.1.2 Neurophysiology

Much of what is known about the organization of visual processing in the brain comes from painstaking work by neurophysiologists who have recorded from cells in along neural pathways involved in vision.

These pathways connect the two retinae, via the optic nerve, to the lateral geniculate nucleus (LGN) of the thalamus, the visual areas in the occipital cortex (V1, V2, V3, and V4) and to higher level visual areas in inferotemporal (IT) cortex, mediotemporal cortex (MT or V5), and superior temporal sulcus (STS). The recording of activity in a single neuron in animal brains (usually monkeys) is the primary methodology used. This involves the insertion of an electrode into a single neuron (or in the vicinity of a small group of neurons) while simultaneously presenting a visual stimulus. Neurons are characterized by their receptive fields, which suggest their functions. A *receptive field* of a neuron defines the range of stimuli to which the neuron responds (i.e., alters its firing rate).

At low levels of visual processing, neurons are selective for features such as line orientation [20], motion direction [21], and color [22].

At higher levels of visual processing, Gross [23] reported that neurons are selectively responsive to faces and hands in IT cortex. These cells respond over a broad range of stimulus positions and orientations, but are remarkably selective to faces or hands. Gross noted in a 1998 review of this early work that the findings were met with a rather skeptical reaction and thus stood alone for nearly a decade following the publication of their now classic paper [24]. Since that time, single unit recording studies in macaques have isolated neurons selectively responsive to either facial identity or facial expression (e.g., [25–27]). The response profiles of these neurons mirror the kind of double dissociation found in neuropsychological case studies.

Expression selective cells were found in more superior areas of the temporal sulcus, while identity selective cells were found in more inferior areas [25]. Moreover, the superior temporal sulcus (STS) also contains neurons that respond to other biological motion-related aspects of faces such as eye gaze and pose [27].

### 5.2.1.3 Functional Neuroimaging Methods

Functional neuroimaging methods have been used extensively in the last decade to probe the responsiveness of various brain regions to faces. These methods include, but are not limited to, positron emission tomography (PET), electro-encephalography

(EEG), and functional magnetic resonance imaging (fMRI). Although these techniques operate on different technical principles, they share the goal of detecting the neural activation changes that result when a person is engaged in a cognitive/perceptual task or when a person perceives an external stimulus (e.g., face, sound). PET works via the injection of a radioactive isotope directly into the blood stream. Neural activity is labeled as it decays. The method has excellent spatial resolution, but poor temporal resolution. EEG works via the recording of surface electrical potentials from neural activity and has excellent temporal resolution, but poor spatial resolution. fMRI works via a labeling of bold oxygen level changes (i.e., the BOLD signal). These changes occur most prominently at the sites characterized by high levels of neural activity. This method has excellent spatial resolution. In recent applications, this resolution has been on the order of 1 mm cubic voxels. Temporal resolution in fMRI depends on the size of the brain region imaged, but in most applications is takes about 2 s for the acquisition of a full brain volume. Most recent functional neuroimaging studies of face perception have relied on fMRI because it represents a good compromise between spatial and temporal resolutions. As we will see in the next section, functional neuroimaging studies, combined with the other methods we have presented, point to several brain areas involved in face processing. We will describe these in terms of their functions and representations in the section on the neural architecture of the face system (Section 5.2.2).

Before proceeding to questions of neural areas and architectures, it is worth noting there are three standard methods for linking brain areas with their relevance for individual tasks using functional neuroimaging data. The first is a *preference method* in which one sets up a stimulus comparison (e.g., faces versus objects) to demonstrate which brain regions “prefer” or respond more strongly to one stimulus (e.g., faces) versus the other (e.g., objects). The second type of method for linking regions to tasks makes use of the common finding that the brain response to a stimulus decreases with repeated presentations of the “same” stimulus. There are several related techniques in this category including repetition priming or suppression (e.g., [28]) and functional magnetic resonance adaptation (fMR-A) [29]. In these repetition-based techniques, the habituation or adaptation of the neural response is used to leverage information about the dimensions of the stimulus that a brain area is coding. Strong repetition suppression to a stimulus is taken as an indication that a brain area is selective for the stimulus. When a stimulus feature changes, if the fMRI response recovers, one can infer that the region is sensitive to the feature dimension in question. For example, if a region is selective to face identity, the fMRI signal attenuates with repeated images of a particular face. If the viewpoint of the image is changed and the signal fails rebound, one can infer that the area codes identity at a level independent of viewpoint. If, on the other hand, the response rebounds, then one can infer that the region codes something about the viewpoint of the image.

Finally, a third method that is becoming increasingly popular relies on the use of simple pattern-based classifiers for gauging the discriminability of brain activation profiles to different stimuli or tasks [30, 31]. The idea is to train a classifier that can discriminate brain response patterns by experimental condition. This approach is sometimes referred to as a “brain-reading” approach because it allows a researcher

to “look at” a pattern of brain activity and to determine the likelihood that a person is experiencing a particular perception. An advantage of the method is that it is possible also to assess the importance of different regions to the task by “ablating them” (i.e., deleting them from the classifier input). If classification performance declines, then one can conclude that the brain regions have information useful for the task.

### **5.2.2 Neural Architectures for Face Perception and Recognition**

We begin with the brain areas selective for faces and then discuss the functions associated with these areas. The functions will be discussed in the context of a proposed network of these areas that provides a coherent framework for understanding their functions [32].

We will concentrate on the *core system* that consists of the fusiform face area (FFA), posterior superior temporal sulcus (pSTS), and the occipital face area (OFA). All three of these regions respond more strongly to faces than to objects. Combined, the areas are involved in high-level visual processing of faces for recognition/identification and for interfacing with social cognitive systems that process facial expressions and head movements.

*Fusiform face area (FFA).* Using a preference method in fMRI, Kanwisher et al. [2] reported on a brain area they called the *fusiform face area* (FFA) located in the fusiform gyrus of ventral temporal cortex (VT) at the base of the brain. This small contiguous brain area was significantly more responsive when participants viewed faces than when they viewed other objects. Moreover, the FFA retains its preference for faces across stimulus manipulations that eliminate alternative lower-level visual explanations of its selectivity. Kanwisher and colleagues proposed the area as a specialized module for face processing. There is a general agreement in the literature that the FFA shows a strong and robust preference for faces. The analysis of brain activity patterns in response to faces and other objects in the FFA, however, has led to an active debate on the nature of the neural representations that underlie these responses (for a review, cf. [31]). In particular, researchers have debated the degree to which the FFA contains a modular (encapsulated) representation of faces or a distributed representation that shares neural resources with object codes. The modular account assumes that the FFA is a special purpose area that encodes only faces. The distributed representation assumes that topographical organization of high-level visual features can contribute to the coding of objects and faces. A third hypothesis suggests that the FFA may be specialized for expert, within-category processing of objects or faces [33].

The primary challenge to the modular account of face representations in the FFA came from one of the first studies to use a pattern-based classification approach to the analysis of fMRI data [34]. Haxby et al. [34] posited a distributed account of face and object representations based on the results of applying a classifier algorithm to categorize brain scans from an fMRI experiment. In the experiment, participants viewed eight categories of objects (faces, houses, cats, bottles, scissors, shoes,

chairs, and scrambled control stimuli). The results of the classifier indicated that the patterns of brain responses to object categories were highly discriminable. To look at the distributed versus modular representation of faces and objects, the classifier was re-applied to different subsets of voxels. Haxby et al. [34] found that the voxels maximally activated in response to particular categories could be deleted from the classifier input with only minor cost to classification accuracy. In particular, deleting voxels that were the most responsive to faces did not strongly affect the ability of the classifier to discriminate faces from other categories of objects. This finding supports a distributed account of neural representation, because it suggests that the regions of brain that respond maximally to particular categories of objects are not required to accurately categorize the brain patterns by object category.

Spiridon and Kanwisher [35] countered this argument, again using a pattern-based classification approach, but with a different operational definition of modular and distributed. They found that voxels giving their maximal response to a particular object (e.g., faces or houses) were only minimally able to classify other objects (e.g., chairs and shoes). They concluded in favor of a modular organization of ventral temporal cortex. The papers supporting modular versus distributed models of representation launched a flood studies refining the methods used and applying more sophisticated classifiers to the problem (e.g., [36–39]). These studies opened up an active dialogue on the use of pattern-based classifiers for functional neuroimaging analysis.

The literature to date generally supports the idea that FFA is strongly involved in detecting faces and in processing facial identity.

*Superior temporal sulcus (STS).* In addition to the FFA, it has been clear for sometime that regions in the STS are also involved in processing faces and other body parts, particularly when they are in motion [40]. As noted in Kanwisher et al. [2], the STS region of most of the subjects they tested was activated more in response to faces than to objects. It is now known that faces *in motion* reliably activate the STS, whereas static images activate STS less reliably. The case for STS in processing faces and bodies in motion comes from a combination of electrophysiological studies in monkeys and functional neuroimaging studies in humans. These studies have been reviewed in detail elsewhere [32, 40]. For present purposes, there is direct evidence for the responsiveness of STS to changes in the direction of eye gaze [41], movements of the head [42], and expression [27]. More generally, the perception of biological motion (including motion of the whole body, hand, eyes, and mouth) consistently activates the STS.

*Occipital Face Area (OFA).* This brain region has been considered a kind of “entry point” for the face processing system, where high-level visual features of faces are coded. Recent work using fMRI-A and repetition suppression techniques is supportive of the sensitivity of this region to any image-based change in faces, including viewpoint and other structural changes in a face. There is little evidence for specificity of the area to coding faces as individuals.

The FFA, STS, and OFA constitute the core of the neural processing of faces. Building on functional neuroimaging data in humans and electrophysiological data from non-human primates, Haxby et al. [32] proposed a neural system that associated

these regions with tasks. This model emphasizes a distinction between representing the invariant versus changeable aspects of the face. The model is built from the core brain areas and four additional areas that extend the system for specific tasks. They posit FFA as the area responsible for representing the invariant facial information useful for identifying faces and for processing any features and configurations needed to categorize faces (e.g., by sex, race, and age). Haxby et al. propose the STS as the region responsible for processing the changeable aspects or movements of faces. This is important for detecting gaze information, head orientation, facial speech, and expression. The OFA, third component of the core system, is the lateral inferior occipital gyrus, which abuts both the lateral fusiform and STS regions and is proposed as a precursor region to both the fusiform and STS.

The distributed model also extends the inferotemporal system to the anterior temporal area, which is involved in the retrieval of personal identity, name, and biographical information. The STS system is extended to include brain areas involved in spatially directed attention for gaze and orientation detection, pre-lexical speech perception from lip movements, and the perception of emotion from expression. The STS extender system underscores its importance in processing social information from faces.

In summary, following this highly influential theory of neural face processing [32], converging research in neuroscience points to two main face processing systems. One system codes the invariant aspects of a face, including features that are useful for categorization and memory for individual faces. From a psychological perspective, a representation that is useful for recognition is one that has parceled out irrelevant variations in the appearance of a face from changes in viewpoint, illumination, etc. The second neural system appears to code the changeable aspects of faces, including how faces move to communicate social information through expressions, facial speech, and attention cues such as gaze. This latter system processes the movements of faces, perhaps somewhat independent of their identity. Again from a psychological perspective, a representation that is useful for social processing should minimize the encoding of static features in favor of fast online processing of subtle changes in the face. This representation might have little connection to memory systems, as the processing of particular facial movements is needed in the here and now to react to transitory signals.

### **5.2.3 Foveal and Peripheral Processes**

To achieve any visual task, including face recognition, humans are able to purposefully control the flow of input data limiting the amount of information gathered from the sensory system [43–45]. This is needed to reduce the space and computation time required to process the incoming information. The anatomy of the early stages of the human visual system is a clear example: despite the formidable acuity in the fovea centralis (1 min of arc) and the wide field of view (about  $140 \times 200$  degrees of solid angle), the optic nerve is composed of only  $10^6$  nerve fibers. The space-variant distribution of the ganglion cells in the retina allows a formidable data

flow reduction. In fact, the same resolution would result in a space-invariant sensor of about  $6 \times 10^8$  pixels, thus resulting in a compression ratio of 1:600 [46]. The probability density of the spatial distribution of the ganglion cells, which conveys the signal from the retinal layers to the optic nerve and is responsible for the data compression, follows a logarithmic-polar law. The number of cells decreases from the center of the retina toward the periphery, with the maximal resolution in the fovea [47]. The same data compression can be obtained on electronic images, either by using a specially designed space-variant sensor [48] or re-sampling a standard image according to the log-polar transform [45, 46]. The analytical formulation of the log-polar mapping describes the mapping that occurs between the retina (retinal plane  $(\rho, \theta)$ ) and the visual cortex (log-polar or cortical plane  $(\xi, \eta)$ ).

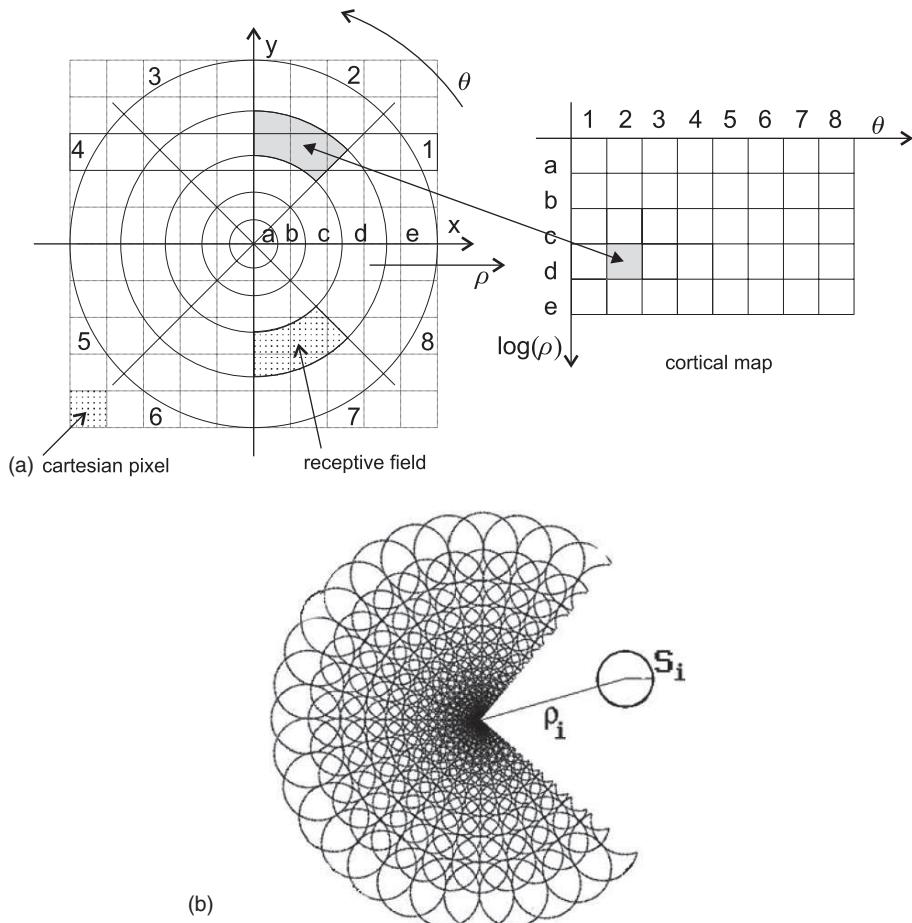
The derived logarithmic-polar law, taking into account the linear increment in size of the receptive fields, from the central region (fovea) toward the periphery is given by

$$\begin{cases} x = \rho \cos \theta \\ y = \rho \sin \theta \end{cases} \quad \begin{cases} \eta = q \theta \\ \xi = \ln_a \frac{\rho}{\rho_0} \end{cases} \quad (5.1)$$

where  $a$  defines the amount of overlap among neighboring receptive fields,  $\rho_0$  is the radius of the innermost circle,  $\frac{1}{q}$  is the minimum angular resolution of the log-polar layout, and  $(\rho, \theta)$  are the polar coordinates of an image point. A model of this space-variant topology is sketched in Fig. 5.1.

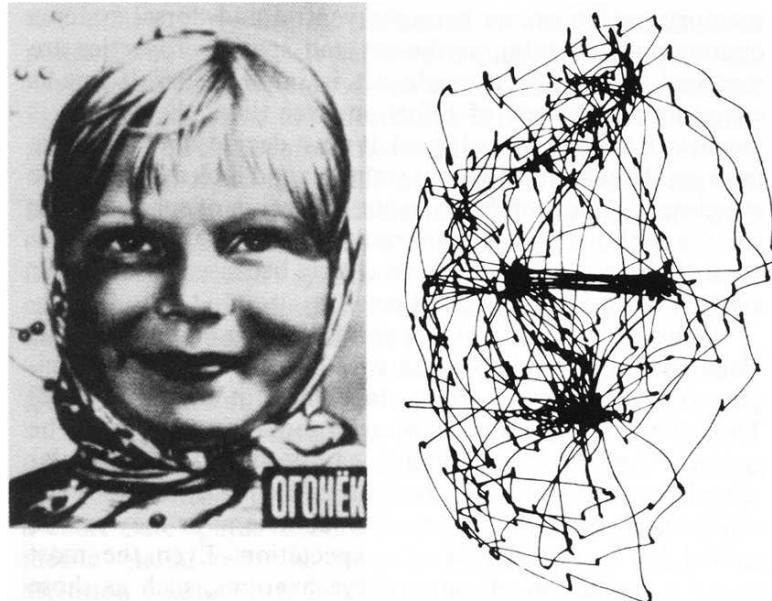
Other models for space-variant image geometries have been proposed, like the truncated pyramid [49], the reciprocal wedge transform (RWT) [50], and the complex logarithmic mapping (CLM) [51]. Several implementations of space-variant imaging have been developed: space-variant sensors [48], custom-designed image re-sampling hardware [52], and special software routines [45, 53]. Given the high processing power of current computing hardware, image re-mapping can be performed at frame rate without the need of special computing hardware and also allows the use of conventional, low cost cameras.

Multiple redundant mechanisms that exist in human vision are applied to cope with limited viewing, such as distance of the object or its distance from the fovea within the field of view. On the other hand, the scanpath performed by the eyes is greatly influenced by the task and may be the case that the performance in recognition (or in the face representation) is influenced by factors such as the capability to collect multiple fixations (Fig. 5.2). This is quite well expressed in [54]. Henderson et al. reported the results of a visual perception experiment performed on a population of eight subjects. Participants viewed 20 faces for 10 sec each during a learning session and then were tested on those faces and 20 new faces in a recognition session. Two learning conditions were compared: in the free viewing learning condition, participants were allowed to move their eyes freely during learning; in the restricted viewing learning condition, participants were required to maintain fixation in the center of each face during learning, and any deviation from that position caused the display to be replaced by a mask. Scores in the recognition session were compared as a function of learning condition. In addition, eye movement patterns



**Fig. 5.1** (a) Log-polar sampling for Cartesian image re-mapping and (b) discrete log-polar model

were compared from learning to recognition and during recognition as a function of whether eye movements were free or restricted during learning. Quoting [54] the conclusions drawn were the following: “In a free viewing learning condition, participants were allowed to move their eyes naturally as they learned a set of new faces. In a restricted viewing learning condition, participants remained fixated in a single central location as they learned the new faces. Recognition of the learned faces was then tested following the two learning conditions. Eye movements were recorded during the free viewing learning condition, as well as during recognition. The recognition results showed a clear deficit following the restricted viewing condition, compared with the free viewing condition, demonstrating that eye movements play a functional role during human face learning. Furthermore, the features selected for fixation during recognition were similar following free viewing and restricted



**Fig. 5.2** Schema of the saccades performed by the human visual system analyzing an unfamiliar face (reprinted from [55])

viewing learning, suggesting that the eye movements generated during recognition are not simply a recapitulation of those produced during learning.”

Therefore, the face perception and recognition processes involve both foveal and peripheral visions, with the former providing a better description of the face. Whenever the space and time constraints allow to collect multiple fixations of the face, they are used to provide a richer description of the subject. If the subject is too far away or there is not sufficient time to perform a series of saccades to move the fovea over several facial landmarks, the visual system limits the processing to the feature extracted from a single view of the face.

#### 5.2.4 Visual Attention and Selective Processing

A very general and yet very important perceptual mechanism in humans is visual attention [55]. This mechanism is exploited by the human perceptual system to parse the input signal in various dimensions: “signal space” (low- or high-frequency data), depth (image areas corresponding to objects close or far from the observer), motion (static or moving objects), etc. The selection is controlled through ad hoc band-limiting or focusing processes, which determine the areas of interest in the scene to which the gaze is directed [56].

In the case of face perception, both space-variant image re-sampling and the adoption of a selective attention mechanism can greatly improve the performance

of any recognition/authentication algorithm. While the log-polar mapping allows to adaptively reduce the frequency content of the input signal, more sophisticated processes are needed to discard low-information areas in the image. Visual attention in humans is also devoted to detect the most informative areas in the face to produce a compact representation for higher level cognitive processes.

Behavioral studies suggest that, in general, the most salient parts for face recognition are, in order of importance, eyes, mouth, and nose [57]. Eye-scanning studies in humans and monkeys show that eyes and hair/forehead are scanned more frequently than the nose [55, 58], while human infants focus on the eyes rather than the mouth [59]. Using eye-tracking technology to measure visual fixations, Klin [60] reported that adults with autism show abnormal patterns of attention when viewing naturalistic social scenes. These patterns include reduced attention to the eyes and increased attention to mouths, bodies, and objects. The high specialization of specific brain areas for face analysis and recognition motivates the relevance of faces for social relations. On the other hand, this further demonstrates that face understanding is not a low-level process but involves higher level functional areas in the brain.

Even though visual attention is generally focused on almost fixed facial landmarks, this does not imply that these are the only areas processed for face perception. Facial features are not simply distinctive points on the segmented face, but rather a collection of image features representing specific (and anatomically stable) areas of the face such as the eyes, eyebrows, ears, mouth, nostrils, etc. Two different kinds of landmarks can be defined:

- face-invariant landmarks, such as the eyes, the nose, the mouth, the ears, and all other elements which are typical of every face;
- face-variant landmarks, which are distinctive elements for a given subject's face [61, 62].

The face-invariant landmarks are important to distinguish faces from non-faces and constitute the basic elements to describe both familiar and unfamiliar faces. All face-variant landmarks constitute the added information, which is learned by the human visual system, to uniquely characterize a subject's face. As a consequence, attention is selectively driven to different areas of the face corresponding to the subject's specific landmarks. This hypothesis is grounded, not only on considerations related to the required information processing, but also on several observations of the eye movements while processing human faces [55, 58–60, 63]. In all reported tests, the gaze scanpaths were different according to the identity of the presented face. As a consequence, the classification of subjects based on the face appearance must be tuned to extract and process the most salient features of each subject's face.

### ***5.2.5 Representation of Faces in the Human Visual System***

Psychologists have suggested that the human face is processed as a “Gestalt,” highlighting the idea that, in perceptual terms, a face is more than the sum of its parts. More concretely, the uniqueness of an individual's face derives not from a set of

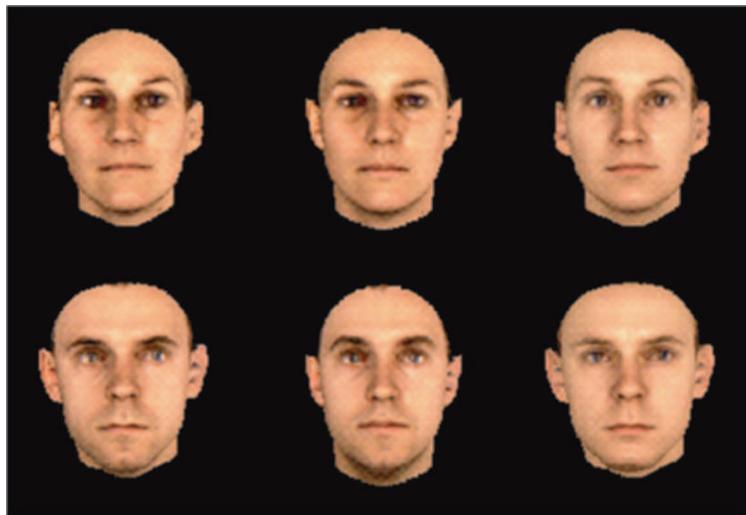


**Fig. 5.3** Thatcher illusion shows that gross distortions of configural information in faces are barely noted when the face is inverted

discrete features that vary in form, but from the complex, subtle arrangement of these features. The importance of the face configuration is perhaps best illustrated with a demonstration called the “Thatcher illusion,” so named because it was first illustrated with the face of Margaret Thatcher [64]. The face on the left side of Fig. 5.3 is an inverted version of the face on the right. The grotesqueness of face is barely noticed when the face is inverted. The Thatcher illusion has been interpreted in terms of human sensitivity to face configuration and concomitantly to the special purpose nature of face processing mechanisms. In short, the use of configural information for perceiving faces is thought to comprise an important part of human expertise for faces. Even small changes in face configuration are noted easily when the face is upright. However, as with all perceptual expertise, there are strict limits on the extent to which *general* perceptual principles apply to the analysis of faces. The Thatcher illusion illustrates that human ability to perceive configural information in faces operates in a mono-oriented fashion.

A related illustration of the lack of general purpose visual processes for face analysis is evident in older work by Yin [65], who tested recognition of upright and inverted faces, houses, and other commonly mono-oriented objects. He showed that although mono-oriented objects/scenes are more difficult to recognize inverted than upright, faces are disproportionately affected by inversion.

The limited generality of face processing can be seen also in the difficulties we have recognizing faces in the photographic negative [66]. Even highly familiar faces are challenging to recognize in the photographic negative. The paradox is that all of the contour information is retained in negative images. The fact that we recognize faces easily from line drawings suggests that shape information from faces contributes to our representation. Moreover, inaccurate shape information, as might be suggested by contrast reversal, can actually interfere with the processing of identity information from the face contours. In fact, there is evidence that human face recognition relies on both 3D shape and 2D surface reflectance (i.e., albedo or pigmentation) information. This was illustrated in a recognition experiment using faces



**Fig. 5.4** The original faces appear on the *left*. The shape-normalized faces appear in the *center* and the reflectance-normalized faces appear on the *right*

that were either *shape normalized* or *reflectance normalized* [67] (see Fig. 5.4). These normalized faces were created with a 3D morphing algorithm applied to laser scans of human heads [68]. This morphing algorithm allows for independent control of the 3D shape and reflectance of a face. The shape-normalized faces were made by morphing the reflectance of original faces onto the shape of the average face. This set of faces varied only in reflectance. The reflectance-normalized faces were made by morphing the reflectance of average face onto the shape of the original faces. This set of faces varied only in shape. In a human behavioral experiment, recognition performance for the original faces, varied in both shape and reflectance, was close to the sum of recognition performance for the shape-normalized faces and the recognition performance for the reflectance-normalized faces. This indicates that relatively complex and complete information about the shape and reflectance information in faces contributes to the human representations that support recognition [69].

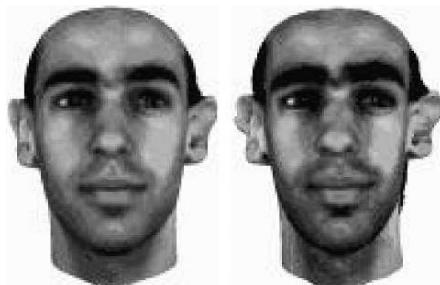
#### 5.2.5.1 The Importance of the “Average” or Prototype Face

A number of psychological findings have suggested the importance of a prototype or average face as a reference point for human face representations. The earliest of these comes from work that showed that face recognition accuracy correlates inversely with ratings of face “typicality” [70]. Faces that humans rate as “typical” are recognized less accurately than faces rated as distinctive or unusual. The relationship between typicality and recognizability has been replicated dozens of times and is one of the most reliable findings in the face perception literature. The inverse relationship between typicality and recognition accuracy is conceptualized most frequently in the context of a face space theory of representation [71]. By this

account, human face representations can be thought of metaphorically as points in a multidimensional space with the average face at the center and with “feature” axes defining the space. Typical faces are close to the center of the space, near the average, and distinctive faces are further from the average. If one adds the assumption that the density of faces declines as a function of the distance from the center of the space, then typical faces lie in a more crowded part of the space than unusual faces. Recognition difficulty comes, therefore, from the fact that typical faces are more confusable with other faces than are unusual faces.

An interesting consequence of this relationship can be seen in caricatures of faces which can be recognized more accurately than veridical faces [72]. The paradox is that a caricature is often a highly distorted version of a face and yet seems nonetheless to be a “good likeness.” In making caricatures, artists or computer graphics algorithms exaggerate the distinctive aspects of faces, *relative to some population of faces*. An example of a caricature made by the 3D morphing program of Blanz and Vetter [68] appears in Fig. 5.5. An important point of the perceptual effectiveness of caricatures is that it illustrates the importance of distinctiveness relative to an average. This “average” is likely dependent on our experience with faces and may therefore vary with that experience. It also suggests that human memory may code the uniqueness of individual faces in terms of how a face *differs* from the average, rather than in absolute terms.

A related phenomenon is the “other-race effect” for face recognition [73, 74]. This is the well-known finding that humans recognize faces of their own race more accurately than faces of other races. The other-race effect for face recognition has been established in numerous human memory studies and in meta-analyses of these studies [75]. The effect can be summed up in the oft-heard phrase, They all look alike to me. This anecdote suggests that our ability to perceive the unique identity of other-race faces is limited, relative to our ability to perceive the unique identity of faces of our own race. This phenomenon suggests the possibility that the “average” face varies as a function of a person’s race. This is not due to race per se but to the likelihood that people have more experience with faces of their own race than with faces of other races. Although humans have additional social prejudices that



**Fig. 5.5** Original face (*left*) and a caricature (*right*) made with the 3D morpher which exaggerates both shape and reflectance information [68]

impact our ability to recognize other-race faces, there is solid evidence for a perceptual component to the phenomenon [76]. Indeed, the other-race effect develops in infancy and can be measured as a narrowing of perceptual face discrimination abilities from 3 to 9 months of age [77]. Three-month-old Caucasian infants in a recent study recognized African, Middle Eastern, Chinese, and Caucasian faces; 6-month olds recognized only Chinese and Caucasian faces; and 9-month olds recognized only Caucasians [77]. This perceptual narrowing is likely to be a consequence of feature selection processes that optimize the encoding of uniqueness for the faces we see, most usually faces of our own race. Experience in infancy and early child development, therefore, may play an important role in tuning the perceptual system to the statistics of the environment in which it must operate. The cost of this optimization is a perceptual limit on the quality of representations that can be formed for stimuli that are not well described by these statistics [78].

In sum, there is evidence that human face representations include information both about the 3D shape of a face and its reflectance. Moreover, human face representations may be organized around a center prototype or average face. This makes sense given the high similarity of faces to each other, because it puts the representational power into coding *differences* between an individual face and the average rather than on the face itself.

### 5.2.6 Relevance of Motion

Faces are nearly always in motion, and the nature of motion generally conveys information useful for social communication. Rigid motions of the head can give indication of intent or attentional shift. Eye gaze has a similar function. Non-rigid movements of the face such as facial expressions and facial speech movements are likewise socially informative. The effects of these movements on face recognition have been investigated in a number of studies and have been reviewed recently [79]. The combination of findings was summarized in a model that refers to the face network model of Haxby et al. [32]. As noted previously, Haxby and colleagues posited separate systems for processing the invariant and changeable aspects of faces, with the former in IT cortex and the latter in STS. Building on the separation of the neural system into these changeable and invariant processing components and integrating behavioral results for the effects of motion on face recognition, a recent model suggests a function for the motion-based coding of identity [79].

From a behavioral point of view, O'Toole et al. [79] proposed that motion might be helpful in face recognition in one of two complementary ways.

The *dynamic identity signature hypothesis* posits that identity-specific information exists in the form of facial gestures that are unique to individuals. There is good evidence that humans can use dynamic identity signatures to recognize people they know reasonably well [80, 81]. The *representation enhancement hypothesis* posits that motion might be used in a perceptual way to build a better 3D representation.

This might make use of structure-from-motion processing. Somewhat surprisingly, there is little evidence to suggest that motion information is useful in this

perceptual way. Thus, for unfamiliar faces, seeing a face motion does not increase recognition performance. It is worth considering the possibility that motion may actually be detrimental for face recognition. This could be due to the fact that when a face is in motion, it may distract humans from processing identity information. As noted, facial motion generally carries social information and the task of processing both identity and social information is a divided attention task.

Putting together the behavioral evidence and placing it in the context of the neural system, O'Toole et al. [79] proposed the following. Information about identity is coded both in the FFA and in the STS. The FFA processes static feature and is configural for both familiar and unfamiliar faces. This is a high-resolution face recognition system that relies on input from the parvocellular processing visual stream [82]. The STS may also have a code for face identity in the form of dynamic identity signatures. The literature supports the conclusion that dynamic information contributes more to face recognition in poor viewing conditions. This might be because facial structure is a more reliable cue to recognition than the dynamic identity signature. Thus, motion information is most beneficial when viewing conditions are not optimal for extracting the facial structure. A second important conclusion is that face familiarity mediates the role of dynamic information in recognition [79]. This is due to the fact that characteristic motions and gestures occur only intermittently and so become reliable cues to identity at a slower rate than static structure information. The relative importance of motion information to recognition, and its likelihood of succeeding, will increase with a viewer's experience with the face.

### 5.3 Face Recognition Technologies

Automatic face identification and verification have been extensively studied by researchers for more than two decades. Despite the relatively low performances, as compared to other biometric modalities such as iris, it has a great potential in application due to the large acceptability and collectability. Moreover, face recognition is among the most successful applications devised in the field of computer vision and pattern recognition. Several approaches have been proposed which are based either on 2D or 3D representation of the face appearance. Also multispectral approaches have been proposed which are based on color or near infrared imaging.

In general, face recognition technologies are based on a two-step approach:

- An off-line enrollment procedure is established to build a unique template for each registered user. The procedure is based on the acquisition of a pre-defined set of face images selected from the input image stream (or a complete video), and the template is built upon a set of features extracted from the image ensemble.
- An online identification or verification procedure where a set of images are acquired and processed to extract a given set of features. From these features a face description is built to be matched against the user's template.

Regardless of the acquisition devices exploited to grab the image streams, a simple taxonomy can be based on the computational architecture applied to extract distinctive and possibly unique features for identification and to derive a template description for subsequent matching.

The two main algorithmic categories can be defined on the basis of the relation between the subject and the face model, i.e., whether the algorithm is based on a subject-centered (ego-centric) representation or on a camera-centered (ego-centric) representation. The former class of algorithms relies on a more complex model of the face, which is generally 3D or 2.5D, and it is strongly linked with the 3D structure of the face. These methods rely on a more complex procedure to extract the features and build the face model, but they have the advantage of being intrinsically pose invariant. The most popular face-centered algorithms are those based on 3D face data acquisition and on face depth maps.

The ego-centric class of algorithms strongly relies on the information content of the gray level structures of the images. Therefore, the face representation is strongly pose variant and the model is rigidly linked to the face appearance, rather than to the 3D face structure. The most popular image-centered algorithms are the holistic or subspace-based methods, the feature-based methods, and the hybrid methods. Over these elementary classes of algorithms several elaborations have been proposed. Among them, the kernel methods greatly enhanced the discrimination power of several ego-centric algorithms, while new feature analysis techniques such as the local binary pattern (LBP) representation greatly improved the speed and robustness of Gabor-filtering-based methods. The same considerations are valid for eco-centric algorithms, where new shape descriptors and 3D parametric models, including the fusion of shape information with the 2D face texture, considerably enhanced the accuracy of existing methods.

### **5.3.1 Subspace Methods**

Subspace algorithms stem from the assumption that any collection of  $M$  face images contains redundancies which can be eliminated by applying a tensor decomposition. This procedure produces a set of basis vectors representing a lower dimensional space with respect of the original image ensemble. Given the basis vectors, every face can be reconstructed in the reduced space. To facilitate the process each  $N \times N$  face image is represented as a vector obtained by aligning the image rows. The resulting  $N \times N \times M$  matrix is decomposed to obtain the non-singular basis vectors. The classification is usually performed by projecting the newly acquired face image into the low-dimensional space and computing a distance measure from all classes represented in the space.

Various criteria have been employed to determine the bases of the low-dimensional spaces. A class of algorithms define projections best representing the population but without information related to the different classes. Other approaches more explicitly address the discrimination between classes. The statistical independence in the low-dimensional feature space is also enforced to retrieve the linear

projections. The pioneer approach is the eigenfaces (PCA) approach [83, 84]. The eigenface representation is based on a linear transformation that maximizes the scatter of all projected samples by decorrelating the data. This corresponds to a singular value decomposition (SVD) applied to the image data set.

Given a set of  $M$  images  $x_i$ , each composed of  $N = n \times m$  pixels, the basis space of the data set is given by the  $M \times N$  principal components matrix  $\mathbf{U}$ . This can be computed by applying the singular value decomposition:

$$\mathbf{X} = \mathbf{U} \cdot \mathbf{D} \cdot \mathbf{V} \quad (5.2)$$

where each row of the matrix  $\mathbf{X}$  is obtained by the sequence of rows of each image and the gray level of each pixel is normalized to the image mean. The  $N \times N$  matrix  $\mathbf{D}$  has the singular values of  $\mathbf{X}$  on its main diagonal and zero elsewhere. These correspond to the eigenvalues of  $\mathbf{U}$ . The main limitation of PCA is the orthonormality constraint. The computed set of basis vectors are always orthonormal, therefore they are unable to well identify the directions of maximal variability.

The PCA approach has been recently extended to a nonlinear algorithm using kernel functions (KPCA) [85, 86].

A variation of the PCA decomposition is the non-negative matrix factorization (NMF) [87]. The main difference with PCA is that it does not allow negative elements in both the bases vectors and the weights of the linear combination. An extension of NMF that gives even more localized bases by imposing additional locality constraints is the linear discriminant analysis (LDA) [89, 90]. Belhumeur [91] originally proposed a method derived from PCA and LDA, where the PCA dimensionality reduction is followed by the Fisher's linear discriminant (FLD) optimization criterion. FLD selects the linear subspace  $\Psi$  maximizing:

$$\frac{|\Psi^T \mathbf{S}_b \Psi|}{|\Psi^T \mathbf{S}_w \Psi|} \quad (5.3)$$

where  $\mathbf{S}_b$  is the between-class scatter matrix and  $\mathbf{S}_w$  is the within-class scatter matrix:

$$\mathbf{S}_b = \sum_{i=1}^m N_i (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})^T \quad \mathbf{S}_w = \sum_{i=1}^m \sum_{\mathbf{x} \in \mathbf{X}_i} (\bar{\mathbf{x}} - \bar{\mathbf{x}}_i)(\bar{\mathbf{x}} - \bar{\mathbf{x}}_i)^T$$

while  $\mathbf{S}_b$  determines the difference among images of different subjects,  $\mathbf{S}_w$  defines the similarity among images of the same subject. In order to avoid singularities in the computation of  $\mathbf{S}_b$  and  $\mathbf{S}_w$ , a PCA decomposition is first applied to the data set  $\mathbf{X}$ . The advantage of FLD over PCA decomposition is that the FLD projection space maximized the separability of the different classes.

Direct LDA (D-LDA) algorithms were proposed [92–94] to prevent the loss of discriminatory information that occurs when a PCA decomposition is applied prior to LDA. The LDA decomposition algorithms have been generalized to the kernel versions, also called general discriminant analysis (GDA) [95] or Kernel Fisher

discriminant analysis (KFDA) [96]. In GDA/KFDA the original input space is projected, by a nonlinear mapping, to a high-dimensional feature space where different classes of faces are linearly separable [95, 97].

The discrimination power of these algorithms strongly depends on the number of samples per class which is included in the data set. As a consequence, the generalization capability of these methods heavily relies on the pose and lighting variations recorded for each subject [98, 99].

The independent component analysis (ICA) is another subspace method aimed at finding linear projections (the subspace vectors) that minimize the statistical dependence between its components [100]. ICA defines a generative model for the observed multivariate data, which is typically given as a large database of samples. In the model, the data variables are assumed to be linear mixtures of some unknown latent variables, and the mixing system is also unknown. The latent variables are assumed non-Gaussian and mutually independent. They are called the independent components of the observed data. These independent components, also called sources or factors, can be found by ICA. Like PCA, ICA yields a linear projection  $\mathbf{R}^N \rightarrow \mathbf{R}^M$  but with different properties. In practice, the ICA is aimed to decompose the input signal (the face images) into a linear combination of independent sources. The basic assumption is

$$\mathbf{x}^T = \mathbf{As}^T \quad (5.4)$$

where  $\mathbf{A}$  is the unknown *mixing matrix*. The ICA algorithms try to find the separating matrix  $\mathbf{W}$  such that:

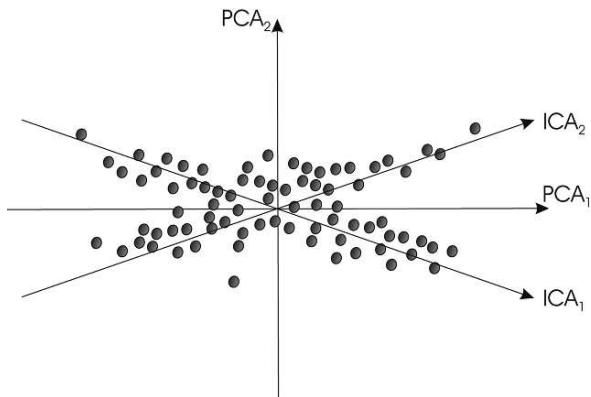
$$\mathbf{Wx}^T = \mathbf{WAs}^T \quad (5.5)$$

The ICA decomposition can be applied to the face recognition problem in two ways:

- The face images are the system variables, the independent basis images are the rows of the matrix  $\mathbf{S}$ , and the pixels are the observations. As a consequence the obtained basis vectors represent local features of the face images.
- The sources are independent coefficients (the eigenvalues) computed from an SVD applied to the data set, the basis images are the columns of the mixing matrix  $\mathbf{A}$ , and the variables are the pixels. The basis vectors represent global properties of the face images and each vector represents a variation mode of the entire face. The computed basis vectors more closely resemble the PCA decomposition.

As shown in Fig. 5.6, while the PCA basis vectors are orthonormal and aligned to the direction of maximal variability of the data, the resulting ICA basis vectors are non-orthonormal but aligned along both directions of maximal variability of the data. In order to better represent local face properties, ICA has been also applied to face representations based on the Gabor decomposition [101]. A nonlinear version of ICA using kernel methods (K-ICA) has been proposed in [102].

**Fig. 5.6** Principal components defined by the PCA or Karunen Lowe decomposition and by the independent component analysis (ICA) decomposition



All the subspace methods have been very popular to implement face recognition systems. Nonetheless they are all very sensitive to changes in the background and alignment errors. For this reason, most of the time the face is manually cropped from the image. On the other hand, as the data set is processed as a single matrix, geometrical coherence must be enforced among all face instances. Consequently all face images must be carefully aligned to a common reference frame. A small error in the face alignment can induce very large errors in the face classification.

### 5.3.2 Facial Features and Landmark-Based Methods

Attention and purposively planned fixations play a crucial role in face recognition by humans. Attentive processes, in turn, are generally guided by salient features which can be located by computing a “saliency map” of the viewed space. The same salient points can provide useful information for face recognition algorithms. In fact, not all face areas on the image convey the same amount of information. For example, the forehead and the cheeks have a much simpler structure and less distinctive patterns than the eyes or the nose.

Salient points on the face are used in registering face images, normalizing expressions, and to perform recognition based on both the geometrical distribution and the gray level pattern at landmark positions. Even though extensive studies have been reported in craniology to precisely define a rich set of face landmarks, there is no universal set of landmark points accepted for recognition. Fred Bookstein [103] defined landmarks as “points in one form for which objectively meaningful and reproducible biological counterparts exist in all the other forms of a data set.” The most frequently used landmarks on faces are the nose tip, eye and mouth corners, center of the iris, tip of the chin, the nostrils, the eyebrows, and the nasion. It is worth noting that many times in the literature discriminant areas within the face, such as the eyes and the mouth, are also referred as “facial features.” This terminology

sometimes generates an ambiguity. In fact, more often the term “feature” is used in pattern recognition to define a specific representation extracted from the gray level pattern. As an example, the eigenface vectors are also termed “features” and also the numerical representations obtained from the multi-channel Gabor filtering applied to a gray level image are called “features.” For this reason, the patterns extracted from specific, discriminant locations of the face image are here termed as “landmarks” instead of “features.”

In geometric-based methods, the distribution of landmark points on the face is used in the form of heuristic rules that involve angles, distances, and areas [104–106]. In structure-based methods, the geometry is incorporated into a complete structure model. For instance in the elastic bunch graph-matching (EBGM) approach, a graph models the relative positions of landmarks, where each node represents one point of the face and the arcs are weighted according to expected landmark distances. At each node, a set of templates are used to evaluate the local feature similarity [107]. Since the possible deformations depend on the landmark points (e.g., mouth corners deform much more than the nose tip), landmark-specific information can be incorporated into the structural model [108]. As the jointly optimized constraint set is enlarged, the system runs more frequently into convergence problems and local minima, which in turn makes a good – and often manual – initialization necessary.

Several approaches have been proposed to derive face representations from a number of components or sub-images of the face. Moghaddam proposed a component-based version of PCA where the face subspace is composed of a number of sub-spaces build from partial images of the original face images [109]. The landmark chosen were the eyes and the mouth. Tistarelli et al. proposed a method based on the extraction of facial landmarks which are re-sampled by applying a log-polar mapping [45]. The actual recognition was performed by applying a normalized cross correlation between two face representations. The value of the correlation determines the similarity between the two representations and hence the two subjects. The correlation value is then used as a score for classification. The elastic graph matching (EGM) is another popular class of face recognition techniques based on landmarks. EGM is a practical implementation of the dynamic link architecture (DLA) for object recognition [110, 111]. In EGM the reference object graph is created by overlaying a rectangular elastic sparse graph on the object image and calculating a Gabor wavelet bank response at each graph node. The values accumulated at each graph node represent one *jet*. The graph-matching process is implemented by a stochastic optimization of a cost function which takes into account both jet similarities and node deformation. A two-stage coarse-to-fine optimization procedure suffices for the minimization of such a cost function:

$$S_\phi(J^0, J^1) = \frac{\sum_j a_j^0 a_j^1 \cos(\phi_j^0 - \phi_j^1) - \bar{d}\bar{k}_j}{\sqrt{\sum_j (a_j^0)^2 \sum_j (a_j^1)^2}} \quad (5.6)$$

where  $J^0$  and  $J^1$  are the two jets sets to be compared,  $(a_j^0, \phi_j^0)$  and  $(a_j^1, \phi_j^1)$  represent the jets coefficients (amplitude and phase) for the two sets,  $\bar{d}$  is the displacement between the two sets and  $\bar{k}_j$  is the *wave vector* or the set of kernels coefficients.

A variation of the EGM is the elastic bunch graph matching (EBGM). In the bunch graph structure a set of jets are computed for every node for different instances of the same face (e.g., with mouth and eyes opened or closed) [112, 113]. In this way, the bunch graph representation can cope for several variations in the face appearance.

Another approach similar to the EGM is the morphological elastic graph matching (MEGM) [114–116]. In this case, the Gabor features are replaced by multiscale morphological features obtained through a dilation–erosion filtering of the facial image [117].

Discriminant techniques have been employed in order to enhance the recognition and verification performances of all these approaches. The use of linear discriminating techniques at the feature vectors for selecting the most discriminating features has been proposed in [113, 114, 118]. Several schemes that aim at weighting the graph nodes according to their discriminatory power have been proposed [114, 118–120]. In [118] the selection of the weighting coefficients has been based on a nonlinear function that depends on a small set of parameters. These parameters have been determined on the training set by maximizing a criterion using the simplex method. In [114] the set of node weighting coefficient was not calculated by some criterion optimization but by using the first- and second-order statistics of the node similarity values. A Bayesian approach for determining which nodes are more reliable has been used in [118]. A more sophisticated scheme for weighting the nodes of the elastic graph by constructing a modified class of support vector machines has been proposed in [119]. In this work it has also been shown that the verification performance of the EGM can be highly improved by proper node weighting strategies. The subspace of the face verification and recognition algorithms considers the entire image as a feature vector and its aim is to find projections that optimize some criterion defined over the feature vectors that correspond to different classes.

The main drawback of all the landmark-based methods is that they require the facial images to be perfectly aligned [121]. That is, all the facial images should be aligned in order to have all the fiducial points (e.g., eyes, nose, and mouth) represented at the same position inside the feature vector. For this purpose, the facial images are very often aligned manually and moreover they are anisotropically scaled. Perfect automatic alignment is in general a difficult task to be assessed. On the contrary, elastic graph matching does not require perfect alignment in order to perform well. The main drawback of the elastic graph matching is the time required for multiscale analysis of the facial image and for the matching procedure. It is commonly known that differences in illumination considering face recognition compose one of the very important issues in this aspect. How computers conceive individual's face geometry is also a problem that researchers are called to solve in order to increase the robustness and the stability of a face recognition system.

### 5.3.3 Dealing with Illumination

Face recognition (at least in 2D) is mostly based on the analysis of photometric properties of the face surface. For this reason, changes in the face reflectance which are due to changes in the illumination (either source orientation or energy) cannot be distinguished from changes due to shape deformations. In order to cope with this problem, several methods have been studied to either normalize the face illumination or to compute the illumination-independent component of the face surface. A comprehensive survey on illumination compensation methods for face recognition can be found in [122].

The proposed methodologies for illumination compensation fall in one of the three main categories:

- *Histogram-based adaptive techniques.* The face image is divided into sub-regions which are analyzed by means of adaptive histogramming techniques to normalize the intensity level of each sub-region. Several approaches are presented in [123] and [124].
- *Re-lighting techniques.* The intensity level of each image pixel can be considered as the ratio of two components: the light energy impinging the skin surface and the actual skin reflectance. The skin reflectance is the illumination-independent component which can then be computed as the ratio between the intensity and the illumination values. These methods aim at estimating the pixel-by-pixel illumination and consequently the skin reflectance.
- *Synthesis of illumination-invariant representations.* One example is the Hue component in color space, which is invariant to shadows. Another example is the face space manifold, including several instances of the same faces with varying illumination direction.

The complex nature of the human skin tissue makes it very difficult to accurately model the response of the face surface to illumination. The skin is not a simple material, but it is made of several layers of tissue having different chromatic and light reflectance properties [125]. For this reason any derived model of the skin reflectance can be only an approximation of the true surface properties and it can be only valid under certain assumptions.

The method proposed in [126] belongs to the second category and it is based on harmonic images. Stemming from the observations that the human faces share similar shape and the face surface is quasi-constant, the nine low-frequency components of the lighting are estimated from the input images. The face image is normalized by a re-lighting procedure based on the illumination ratio image. The authors propose to calibrate the input face image to the canonical illumination. The experimental results show a significant improvement in the face recognition performances.

In the paper by Gross and Brajovic [127], the luminance field is computed by means of an anisotropic diffusion derived from the minimization of the functional:

$$F(L) = \int \int_{\omega} \rho(x, y)(L(x, y) - I(x, y))^2 dx dy + \lambda \int \int_{\omega} (L_x^2 + L_y^2) dx dy \quad (5.7)$$

where the first integral constraints the luminance  $L(x, y)$  to be close to the captured pixel image intensity  $I(x, y)$ , the second integral enforces the smoothness of the recovered luminance field, controlled by the parameter  $\lambda$ , and  $\rho(x, y)$  determines the level of anisotropy of the diffusion process. The functional can be solved through the Lagrange multipliers to determine the value of  $L(x, y)$ .  $\rho(x, y)$  varies with the image luminance and it is defined as the reciprocal of the Weber contrast function:

$$\frac{\rho(i, j)}{\min(I_i, I_j)} = \frac{|I_i - I_j|}{\sqrt{I_i I_j}} \quad (5.8)$$

The approach presented in [128] falls into the third category. A face-lighting subspace is designed based on three or more training face images illuminated by non-coplanar lights. The lighting of any face image is represented as a point in this subspace. The main contribution of this paper is a very general framework for analyzing and modeling face images under varied lighting conditions. The concept of face-lighting space is introduced as well as a general face model and general face imaging model for face space modeling under varied lighting conditions. In a practical system, based on subspace analysis, illumination and pose are two problems that need to be faced concurrently.

In [129] the statistics of the derivative of the irradiance images (log) of human face is analyzed. An illumination insensitive distance measure is defined based on the min operator of the derivatives of two images. The proposed measure for recovering the reflectance component is compared with the median operator proposed by Weiss [130]. When the probes are collected under varying illuminations, the experiments of face recognition on the CMU-PIE database show that the proposed measure is much better than the correlation of image intensity and a little better than the Euclidean distance of the derivative of the log image used in [131].

### 5.3.4 Digital Representation of Human Faces

Holistic methods for face identification require a large (statistically significant) training set to build the base vectors determining the low-dimension space. The generalization capabilities of these methods have been tested to some extent but are still unclear. Up to now tests have been performed on databases with limited size. Even the FRGC database [132] only comprises few thousands subjects. Scaling up to larger databases, including hundred of thousands individuals, even if possible, would make the problem very difficult to be numerically analyzed. Managing the identity by these face representations requires to be able to discriminate each single individual through a single feature space, but this can be hardly guaranteed. The best

performing face recognition methods, based on holistic processing, under real conditions reach an equal error rate (EER) around 1%. This corresponds to 100 wrongly classified subjects over a database of 10,000 individuals or 1000 over 100,000. The template size depends on the dimensionality of the representation space, i.e., the number of basis vectors selected for the database representation. This value depends on the population of subjects, the variability of the face appearance (pose, expression, lighting, etc.), the number of classes, and the discrimination power to be achieved. Therefore, coping with many variations in the face appearance, for example, to deal with ageing, the size of the subspace and hence the representation can become indefinitely large. An advantage of feature-based approaches is the strong dependence on the input signal rather than on the population of subjects analyzed. Therefore, the subject's identity is represented exclusively from information related to data captured from each subject. The relation to the "rest of the world" is limited to the classification parameters which must be tuned for classification. The resulting face template can be very compact as it is limited to the geometric structure of the features with the associated values. This allows to cope with many variations, including ageing, without affecting the size of the representation.

The studies performed through fMRI experiments on face perception allow us to infer also some indications on the complexity of the representation of faces in the human brain. Leveroni et al. defined 31 areas in the human brain, mainly related to the superior temporal sulcus and the ventral striatum, which are activated during face perception. The total volume of the active areas sums to 21.2 ml. Given a total mass of 1400 ml for the brain cortex, with an approximate total of  $100 \times 10^9$  neurons, we end up to a total of about  $1.5 \times 10^9$  neurons. Counting in average 12,000 synapses per neuron in the visual cortex, the active areas during face perception sums to  $18 \times 10^{12}$  synapses. Limiting the representation to one bit of information per synapse, the information carried by the neural network is roughly equivalent to  $2.3 \times 10^{12}$  bytes. Considering a simple scenario where a subject typically learns to recognize 1,000 faces, the amount of information which can potentially be stored in such a network equals a video stream of about 70 s with a resolution of 1 Mega pixels per frame. Even though the neural architecture subduing face perception in the human brain is also able to deal with complex changes such as facial expression and ageing, still this simple comparison allows us to understand the complexity and the amount of information involved in the perception of faces and the subsequent recognition of individuals.

#### **5.3.4.1 Video Streams**

When identifying people at a distance it is possible to collect information over time. This process allows to build a rich representation than using a single snapshot. It is therefore possible to define a "dynamic template." This representation can encompass both physical and behavioral traits, thus enhancing the discrimination power of the classifier applied for identification or verification. The representation of the subject's identity can be arbitrarily rich at the cost of a large template size. Several

approaches have been proposed to generalize classical face representations based on a single view to multiple view representations. Examples of this kind can be found in [133, 134] and [135–137] where face sequences are clustered using vector quantization into different views and subsequently fed to a statistical classifier. Recently, Kruger, Zhou and Chellappa [138, 139] proposed the “video-to-video” paradigm, where the whole sequence of faces, acquired during a given time interval, is associated to a class (identity). This concept implies the temporal analysis of the video sequence with dynamical models (e.g., Bayesian models) and the “condensation” of the tracking and recognition problems. Other face recognition systems, based on the still-to-still and multiple stills-to-still paradigms, have been proposed [140–142]. However, none of them is able to effectively handle the large variability of critical parameters like pose, lighting, scale, face expression, and some kind of forgery in the subject appearance (e.g., the beard). Typically, a face recognition system is specialized on a certain type of face view (e.g., frontal views), disregarding the images that do not correspond to such a view. Therefore, a powerful pose estimation algorithm is required. In order to improve the performance and robustness, multiple classifier systems (MCSs) have been recently proposed [143]. Achermann and Bunke [144] proposed the fusion of three recognizers based on frontal and profile faces. The outcome of each expert, represented by a score, i.e., a level of confidence about the decision, is combined with simple fusion rules (majority voting, rank sum, and Bayes’ combination rules). Lucas [133, 134] used a n-tuple classifier for combining the decisions of experts based on sub-sampled images. Other interesting approaches are based on the extension of conventional, parametric classifiers to improve the “face space” representation. Among them are the extended HMMs [145], the pseudo-hierarchical HMMs [146, 147], and parametric eigenspaces [148], where the dynamic information in the video sequence is explicitly used to improve the face representation and, consequently, the discrimination power of the classifier. In [149] Lee et al. approximated face manifolds by a finite number of infinite extent subspaces and used temporal information to robustly estimate the operating part of the manifold. There are fewer methods that recognize from manifolds without the associated ordering of face images. Two algorithms worth mentioning are the mutual subspace method (MSM) of Yamaguchi et al. [150, 151] and the Kullback–Leibler divergence-based method of Shakhnarovich et al. [152]. In MSM, infinite extent linear subspaces are used to compactly characterize face sets, i.e., the manifolds that they lie on. Two sets are then compared by computing the first three principal angles between corresponding principal component analysis (PCA) subspaces [153]. The major limitation of MSM is its simplistic modeling of manifolds of face variation. Their high nonlinearity invalidates the assumption that data are well described by a linear subspace. Moreover, MSM does not have a meaningful probabilistic interpretation. The Kullback–Leibler divergence(KLD)-based method [152] is founded on information-theoretic grounds. In the proposed framework, it is assumed that  $i$ th person’s face patterns are distributed according to  $p_i(x)$ . Recognition is then performed by finding  $p_j(x)$  that best explains the set of input samples – quantified by the Kullback–Leibler divergence. The key assumption in their work, that makes divergence computation tractable, is that face patterns are normally distributed.

### 5.3.4.2 SIFT Features for Face Representation and Identification

Most systems for face identification and verification are based on a face representation which is composed of a collection of image features. For these representations to be effective the extracted features must be stable over time and properly selected. The first requirement inhibits external factors, such as changes in lighting and facial expression, from perturbing the representation. The second requirement allows to eliminate spurious features which are due to photometric phenomena not directly related to physical characteristics of the face shape.

The scale invariant feature transform (SIFT) has been proposed by Lowe [154] and proved to be very effective for general 3D object recognition. A relevant advantage of SIFT features is the invariance to image rotation, scaling, translation and partly to illumination changes, and projective transform. The SIFT features are detected through a staged filtering approach that identifies stable points in the scale-space. This is achieved by the following steps:

- select candidates for feature points by searching peaks in the scale-space from a difference of Gaussian (DoG) function;
- localize the feature points by using the measurement of their stability;
- assign orientations based on local image properties and;
- calculate the feature descriptors which represent local shape distortions and illumination changes.

After candidate locations have been found, a detailed fitting is performed to the nearby data for the location, edge response, and peak magnitude. To achieve invariance to image rotation, a consistent orientation is assigned to each feature point based on local image properties. The histogram of orientations is formed from the gradient orientation at all sample points within a circular window of a feature point. Peaks in this histogram correspond to the dominant directions of each feature point.

For illumination invariance eight orientation planes are defined. Toward this end, the gradient magnitude and the orientation are smoothed by applying a Gaussian filter and then sampled over a  $4 \times 4$  grid with eight orientation planes.

In the method proposed in [155] the face image is first photometrically normalized by using histogram equalization. The rotation, scale, and translation invariant SIFT features are extracted from the face image. The feature selection is not performed on the face representation itself, but rather the features to be retained are selected according to the matching scheme. Consequently, the features to be matched are selected on the probe and test images according to three envisaged different schemes alongside the graph construction. The three schemes, namely gallery image-based, reduced point-based, and regular grid-based match constraints, are defined. Each face is first represented with a complete graph drawn on all feature points extracted using the SIFT operator [154]. During matching, the constraint schemes are applied to find the corresponding sub-graph in the probe face image given the complete graph in the gallery image.

### Graph Matching and Feature Reduction Schemes

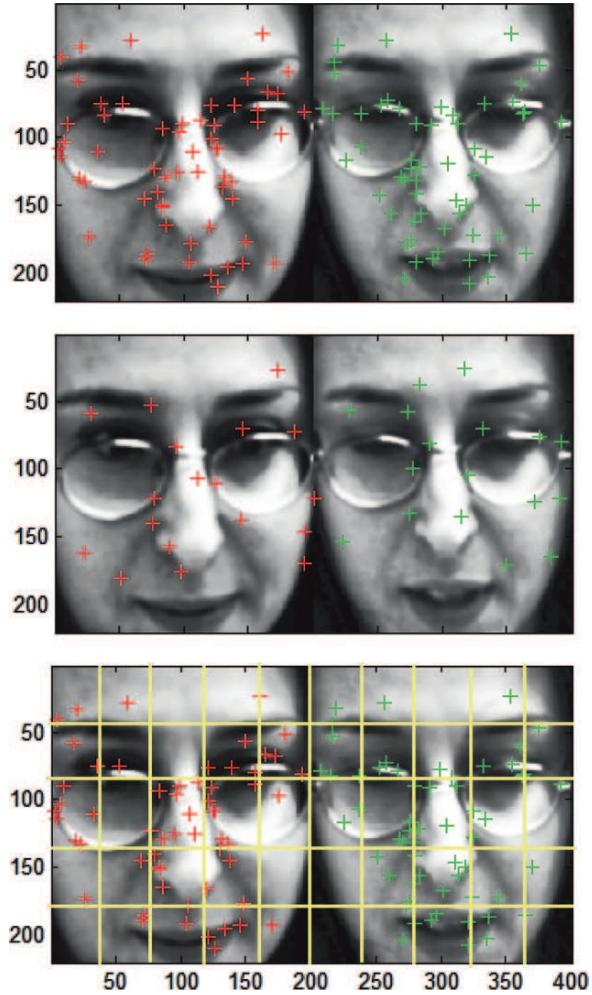
As explained before, each feature point is composed of four types of information: the spatial coordinate, key point descriptor, scale, and orientation. The key point descriptor is a 1D vector of 128 real values.

In order to select the sub-graph to be matched against the probe image, a feature reduction strategy is applied according to one of the following schemes (see Fig. 5.7):

- *Gallery image constraint.* It is assumed that matching points will be found around similar positions, i.e., fiducial points on the face image. To eliminate false matches a minimum Euclidean distance measure is computed by means of the Hausdroff metric. It may be possible that more than one point in the first image correspond to the same point in the second image. Let  $N$  = number of interest points on the first image;  $M$  = number of interest points on the second image. Whenever  $N \leq M$ , many interest points from the second image are discarded, while if  $N \geq M$ , many repetitions of the same point match in the second image. After computing all the distances, only the point with the minimum distance from the corresponding point in the second image is paired. The mean dissimilarity scores are computed for both the vertexes and the edges. A further matching index is given by the dissimilarity score between all corresponding edges. The two distances are then averaged.
- *Reduced point constraint.* After completing the previous phase, there can still be some false matches. Usually, false matches are due to multiple assignments, which exist when more than one point are assigned to a single point in the other image, or to one-way assignments. The false matches due to multiple assignments are eliminated by pairing the points with the minimum distance. The false matches due to one-way assignments are eliminated by removing the links which do not have any corresponding assignment from the other side. The dissimilarity scores on reduced points between two face images for nodes and edges is computed in the same way as for the gallery-based constraint. Lastly, the average weighted score is computed. Since the matching is done on a very small number of feature points, this graph-matching technique proved to be more efficient than the previous match constraint
- *Regular grid constraint.* In this technique, the images are divided into sub-images, using a regular grid with overlapping. The matching between a pair of two face images is done by computing distances between all pairs of corresponding sub-image graphs and finally averaging them with dissimilarity scores for a pair of sub-images. From an experimental evaluation, sub-images of dimensions 1/5 of width and height represent a good compromise between localization accuracy and robustness to registration errors. The overlapping was set to 30%. The matching score is computed as the average between the matching scores computed on the pairs of image graphs.

The proposed graph-matching technique is tested on the BANCA database. For this experiment, the matched controlled (MC) protocol is followed, where the

**Fig. 5.7** Selected feature points according to the three proposed matching strategies: (top) gallery-based, (middle) reduced point-based, and (bottom) regular grid-based match constraints



images from the first session are used for training, whereas second, third, and fourth sessions are used for testing and generating client and impostor scores. The results obtained are summarized in Table 5.1 and 5.2.

**Table 5.1** Prior EER on G1 and G2 for the two methods: “GIBMC” stands for gallery image-based match constraint, “RPBMC” stands for reduced point-based match constraint, and “RGBMC” stands for regular grid-based match constraint

	GIBMC (%)	RPBMC (%)	RGBMC (%)
Prior EER on G1	10.13	6.66	4.65
Prior EER on G2	6.92	1.92	2.56
Average	8.52	4.29	3.6

**Table 5.2** WER for the two different graph-matching techniques: “GIBMC” stands for gallery image-based match constraint, “RPBMC” stands for reduced point-based match constraint, and “RGBMC” stands for regular grid-based match constraint

	GIBMC (%)	RPBMC (%)	RGBMC (%)
WER (R = 0.1) on G1	10.24	7.09	4.07
WER (R = 0.1) on G2	6.83	2.24	3.01
WER (R = 1) on G1	10.13	6.66	4.6
WER (R = 1) on G2	6.46	1.92	2.52
WER (R = 10) on G1	10.02	6.24	4.12
WER (R = 10) on G2	6.09	1.61	2.02

## 5.4 Face Recognition at a Distance

The human face is a very distinctive landmark in the human body which is very well suited for performing recognition at a distance. Most of the visible facial features can be captured at a low-spatial resolution, and therefore it is not required to capture images at a very close distance.

In order to capture face images from a distance three processes are involved:

- segmentation of the moving body;
- detection and tracking of the head; and
- localization of the face.

All three processes require a proper sampling of the sequence both in space and time to isolate the information content which is relevant for further processing.

The detection and segmentation of the human body have been recently investigated within the context of visual surveillance. Several methods have been proposed which are based on background subtraction techniques such as Gaussian mixture models. The rationale underpinning all these methods is the definition of a proper learning mechanism which is capable of capturing the characteristic spatio-temporal frequencies of a moving body.

Further considerations and details on the methodology to be applied to capture face images are reported in chapter 7 of this book, specifically devoted to the acquisition of face images at a distance.

## 5.5 A Comparative Study of Human and Machine Performances

In this section we discuss comparisons between humans and state-of-the-art face recognition algorithms. The motivation for these comparisons is that face recognition algorithms are targeted often to applications in security and identity verifications. Most of these applications are being done currently by humans. Therefore, in addition to testing algorithms, it is important to know how well *humans* perform at

the task. Arguably, algorithms can be useful for security applications if they meet or exceed the performance of humans who are currently performing the task.

In this section, we will look at two human–algorithm comparison studies. The first is aimed simply at comparing quantitatively the performance of humans and algorithms on a difficult face recognition test [156]. The second looks at the potential for fusing human and algorithm recognition judgments with the goal of improving performance [157]. The latter looks at the qualitative accord between human and machine performances with the rationale that fusion will be helpful only insofar as there are qualitative differences in recognition strategy.

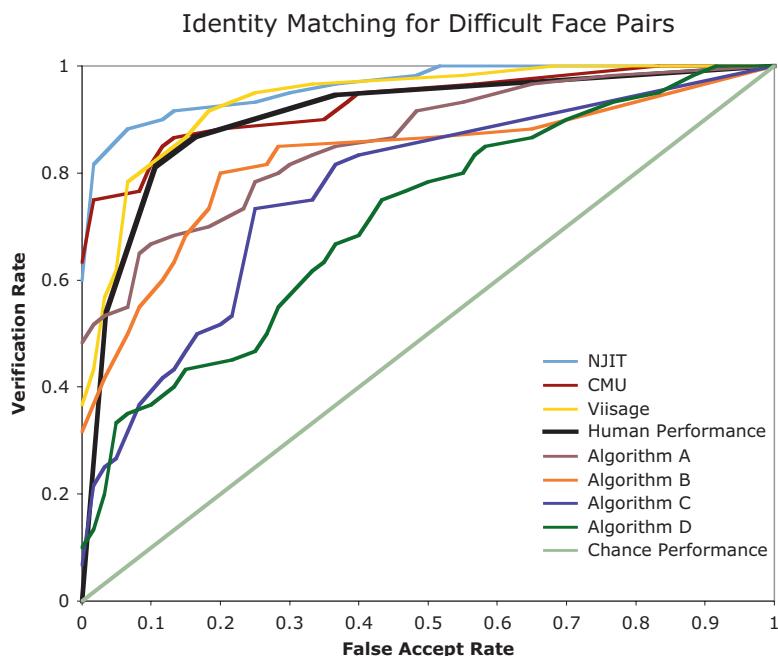
The source of the algorithms for the human–machine comparisons comes from periodic tests of automatic face recognition systems organized by the US government. The tests are aimed at spurring the development of algorithms toward higher levels of performance and are open to participants from academia and industry worldwide. The advantage of these evaluations for understanding the “state of the art” in computer-based face recognition systems is that they test many algorithms simultaneously with a standardized protocol and the same, large set of stimulus images. Here we made use of algorithm data from the Face Recognition Grand Challenge (FRGC) (2004–2006) [158]. The comparisons focus on the performance of algorithms and humans matching the identity of faces in pairs of images tested under different illumination conditions, with one image taken under controlled illumination (similar to a passport image) and the other image taken indoors with uncontrolled illumination. Of the several algorithm evaluations in the FRGC, the uncontrolled illumination identity match test yielded the poorest algorithm performance. To illustrate the difficulty with illumination, it is informative to compare the performance of algorithms matching faces under controlled conditions with algorithms matching faces between controlled and uncontrolled illuminations. In the controlled illumination experiment of the FRGC, algorithms achieved a median verification rate of 0.91 at the 0.001 false acceptance rate. The verification rate is the proportion of matched pairs correctly judged to be the same person and the false acceptance rate is the proportion of mismatched pairs judged incorrectly to be the same person. Algorithms in the uncontrolled illumination experiment achieved a median verification rate of 0.42 at a false acceptance rate of 0.001. Combined, these findings suggest that illumination variation remains a serious problem for algorithms.

### ***5.5.1 Human and Machine Performances: Quantitative Comparison***

How do state-of-the-art face recognition algorithms perform relative to humans? To answer this question, a direct comparison of human and machine performances was conducted [156]. Algorithms in the FRGC were required to match identity in all possible pairs of 16028 “target” faces (controlled illumination) and 8014 “probe” faces (uncontrolled illumination). Individual algorithm results were tallied as a matrix of roughly 128 million similarity scores between all possible pairs of images. The

similarity score matrix was used to compute the algorithms' "same person" versus "different person" judgments. Specifically, accuracy was scored using receiver operating characteristic (ROC) curves. For the human-machine comparison, 120 *difficult* face pairs and 120 *easy* face pairs were sampled using a baseline principal components analysis (PCA) algorithm. This algorithm was chosen for sampling because it is widely available and well known, but it is not state of the art. Thus, it provides a tool for screening pair difficulty level. The difficult face pairs were defined as match pairs that had similarity scores two standard deviations below the mean (i.e., two dissimilar images of the same person) and no-match pairs that had similarity scores two standard deviations above the mean (i.e., two similar images of different people). Easy face pairs were defined inversely.

In the human experiments, subjects viewed the 240 face pairs and rated the likelihood that the pairs were the same person, using a 5-point scale that ranged from "sure they are the same" to "sure they are different." Using these data, it is possible to create an ROC curve analogous to that used to score the algorithms' performance. For the machine side of the comparison, an ROC curve was computed for each algorithm using the same 240 face pairs used in the human test. The performance of humans and algorithms for the difficult face pairs appears in Fig. 5.8. For the difficult face pairs, three algorithms were more accurate than humans[159–161] and



**Fig. 5.8** The ROC curves obtained by testing 7 different computer algorithms for face identification and one obtained from the performance of humans on the set of 240 "difficult" image pairs [68]

four algorithms were less accurate. For the easy face pairs, the algorithms fared even better relative to humans with all but one algorithm surpassing human performance.

The results of the comparison, therefore, support the idea that the best face recognition algorithms perform better than humans on a challenging task. This may lead us to wonder if human face recognition skills are overrated. In answer to this question, it is important to remember that human face recognition skills are at their best for the faces of people we know well [162–164]. Indeed, when we know someone well, we can recognize them from a distance in quite poor viewing conditions. Human abilities with relatively unfamiliar faces are less impressive. In these cases, changes in viewpoint and illumination can have strong effects on human accuracy. The use of unfamiliar face matching in this study, however, is an appropriate test for estimating the accuracy of a human security guard whose job is to match and recognize unfamiliar faces.

### ***5.5.2 Fusing Human and Machine Performances: Qualitative Comparison***

Do humans and machines make similar errors? This question gets at the qualitative similarity between human and machine performances. A sensible way to address this question is to fuse the match estimates made by machines and humans. The rationale behind this relies on the assumption that if qualitative differences exist, then fusion will result in performance improvements relative to either algorithms or humans acting alone. Two experiments in algorithm and human fusion were carried out in a recent study [157]. In the first experiment, similarity scores from the seven algorithms tested in the FRGC were fused for the 120 difficult face pairs considered in the previous study [156]. In the second experiment, the human-generated estimate of similarity for these face pairs was included in the fusion.

The fusions were done using a statistical learning algorithm known as partial least squares (PLS) regression (cf., [165]). This algorithm learns a statistical mapping from a set of input variables to a set of output measures. In this case, the similarity scores for the seven algorithms on the face pairs were the input and the match status of the individual face pairs (1 or 0) was the output. The robustness of the match status predictions was verified with a cross-validation procedure in which the learning algorithm was trained with 119 pairs and tested with the left-out pair, rotating the left-out pair through 120 iterations. This produces an accuracy measure that is the proportion of correctly classified left-out face pairs.

In the first experiment, fusion of the seven algorithms together reduced the error rate to 0.06, which was approximately half of the error rate of the best-performing algorithm operating alone [160]. In the second experiment, the human similarity estimates for the matched pairs were included in the fusion. These estimates came from the average similarity rating produced by the 49 subjects in the experiment. This human-machine fusion was nearly perfect with an error rate of less than 0.005.

The results of the two fusion experiments indicate that there are sufficient qualitative differences in the performance of different algorithms to make use of fusion

for improving performance. There is also enough qualitative difference in human and machine performances for the human system to make a strong contribution to the overall accuracy.

In summary, fusing humans and machines may be a viable way to combine the strengths of both “algorithms” and to compensate for the weaknesses.

**Acknowledgments** Funding from the Technical Support Working Group (TSWG) to A. O’Toole supported the human–machine comparison work described in Section 5.5. Funding to M. Tistarelli from the European Union *Biosecure* Network of Excellence and from the Italian Ministry for Research is also acknowledged.

## Proposed Questions and Exercises

- Describe the nature of the evidence that neuropsychological double dissociation studies have provided about the human neural systems for human face recognition.
- Describe the two main components of the neural processing system for faces and their functions?
- What psychological evidence suggests that human face representations are organized around a prototype or average face?
- Design an experiment to assess the role of shape and reflectance information in the “other-race effect” for face recognition.
- In what ways does motion information contribute to face perception and recognition?
- Make a list of man–machine comparisons that would be useful to carry out in the next few years. Choose one of these and design an experiment that would make the comparison.
- What is the overall neural architecture subserving face perception?
- In your opinion, why is motion perception involved in face recognition?
- What are the computational processes involved in face recognition?
- Read the paper [166] listed in the references and explain how fMRI tests demonstrate the relevance of facial landmarks for face perception.
- Design a computational architecture for binding shape and motion processing for face recognition.
- How can a space-variant sampling of the image plane improve the feature extraction process?
- What are the advantages of a representation of faces based on global features rather than on local features?
- What data would you include in a subject-specific representation of faces?
- How can ageing effects be handled in a face template representation?
- Try to devise a system architecture to process the subject’s identity by face verification (1 to 1 recognition) and by face identification (1 to many recognition) adopting a subject-specific template. What is the added complexity in the identification process as compared to verification?

- How would you include quality measures to the proposed computational model for face processing?
- How many images should be processed for a reliable dynamic template? What parameters and variables are involved? How would you quantitatively determine the reliability of the resulting recognition system?
- What is the relevance of coupled shape and motion processing of faces for surveillance and identification from a distance?

## References

1. J. Bodamer. Die prosopagnosie. *Arch. Psychiat. Nerv.* 179:6–54, 1947.
2. N. Kanwisher, J. McDermott, and M. Chun. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17:4302–4311, 1997.
3. J. Barton. Disorders of face perception and recognition. *Neurol Clin.* 21:521–548, 2003.
4. A. Damasio. Prosopagnosia. *Trends Neurosci.* 132–135, 1985.
5. J. Meadows. The anatomical basis of prosopagnosia. *J. Neurol. Neurosurg. Psychi.*, 37:489–501, 1974.
6. E. DeRenzi. Prosopagnosia in two patients with CT scan evidence of damage confined to the right hemisphere. *Neuropsychologia*, 24:385–389, 1986.
7. T. Landis, J. Cummings, L. Christen, J. Bogen, and H-G. Imbof. Are unilateral right posterior cerebral lesions sufficient to cause prosopagnosia? Clinical and radiological findings in six additional patients. *Cortex*, 22:243–52, 1986.
8. V. Bruce and A.W. Young. Understanding face recognition. *Br. J. Psychol.* 77(3):305–327, 1986.
9. R. Bruyer, C. Laterre, X. Seron, P. Feyereisen, E. Strypstein, E. Pierrard, and D. Rectem. A case of prosopagnosia with some preserved covert remembrance of similar faces. *Brain Cogn.* 2:257–284, 1983.
10. J. Hornak, E. Rolls, and D. Wade. Face and voice expression identification in patients with emotional and behavioral changes following ventral frontal lobe damage. *Neuropsychologia*, 34:173–181, 1996.
11. F. Parry, A. Young, J. Saul, and A. Moss. Dissociable face processing impairments after brain injury. *J. Clin. Exp. Neuropsychol.* 13:545–558, 1991.
12. E. Shuttleworth, V. Syring, and N. Allen. Further observations on the nature of prosopagnosia. *Brain Cogn.* 1:307–322, 1982.
13. D. Tranel, A. Damasio, and H. Damasio. Intact recognition of facial expression, gender, and age in patients with impaired recognition of face identity. *Neurology*, 38:690–696, 1988.
14. A. Calder, A. Young, D. Rowland, D. Perrett, J. Hodges, and H. Etcoff. Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cogn. Neuropsychol.* 13:699–745, 1996.
15. G. Humphreys, N. Donnelly, and M. Riddoch. Expression is computed separately from facial identity, and it is computed separately for moving and static faces: Neuropsychological evidence. *Neuropsychologia*, 31:173–181, 1993.
16. J. Kurucz and J. Feldmar. Prosopo-affective agnosia as a symptom of cerebral organic brain disease. *J. Am. Geriatr. Soc.* 27:91–95, 1979.
17. J. Kurucz, J. Feldmar, and W. Werner. Prosopo-affective agnosia associated with chronic organic brain syndrome. *J. Am. Geriatr. Soc.* 27:225–230, 1979.
18. A. Young. Face recognition impairments. *Philos. Trans. Roy. Soc. Lond.* 335B:47–54, 1992.
19. A. Calder and A. Young. Understanding the recognition of facial identity and facial expression. *Nature Rev. Neurosci.* 6:641–651, 2005.
20. D.H. Hubel and T. Wiesel. Receptive fields, binocular interaction, and functional architecture in cat's visual cortex. *J. Physiol.* 160:106–154, 1962.

21. J. Movshon, and W. Newsome. Neural foundations of visual motion perception. *Curr. Direct. Psychol. Sci.* 1:35–39, 1992.
22. S. Zeki. Color coding in the cerebral cortex: The reaction of cells in monkey visual cortex to wavelengths and colours. *Neuroscience*, 9:741–765, 1983.
23. C. Gross, C. Rocha-Miranda, and D. Bender. Visual properties of neurons in inferotemporal cortex of monkeys. *J. Neurophysiol.* 35:96–111, 1972.
24. C. Gross. Brain, Vision, and Memory. MIT Press, Cambridge, MA, 1998.
25. M. Hasselmo, E. Rolls, and G. Baylis. Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey. *Exp. Brain Res.* 75:417–429, 1989.
26. J. Hietanen, D. Perrett, M. Oram, P. Benson, and W. Dittrich. The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex. *Exp. Brain Res.* 89:157–171, 1992.
27. D. Perrett, J. Hietanen, M. Oram, and P. Benson. Organization and function of cells responsive to faces in temporal cortex. *Phil. Trans. Roy. Soc. Lond. B Biol. Sci.* 335:23–30, 1992.
28. E. Eger, S. Schweinberger, R. Dolan, and R. Henson. Familiarity enhances invariance of face representations in human ventral visual cortex: fMRI evidence. *NeuroImage*, 26:1128–1139, 2005.
29. K. Grill-Spector, T. Kushnir, S. Edelman, G. Avidan, Y. Itzhak, and R. Malach. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, 24, 187–203, 1999.
30. J. Haynes, and G. Rees. Decoding mental states from brain activity in humans. *Nature Rev. Neurosci.* 7:523–534, 2006.
31. A.J. O'Toole, F. Jiang, H. Abdi, N. Penard, J. Dunlop, and M. Parent. Theoretical, statistical, and practical perspectives on pattern-based classification approaches to the analysis of functional neuroimaging data in ventral temporal cortex. *J. Cogn. Neurosci.* 19:1735–1752, 2007.
32. J.V. Haxby, E.A. Hoffman, and M.I. Gobbini. The distributed human neural system for face perception. *Trends Cogn. Sci.* 20(6):223–233, 2000.
33. I. Gauthier, M.J. Tarr, A.W. Anderson, P. Skudlarski, and J.C. Gore. Activation of the middle fusiform face area increases with expertise recognizing novel objects. *Nature Neurosci.* 2: 568–573, 1999.
34. J.V. Haxby, M.I. Gobbini, M.L. Furey, A. Ishai, J.L. Shouten and J.L. Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293 :2425–2430, 2001.
35. M. Spiridon and N. Kanwisher. How distributed is visual category information in human occipito-temporal cortex? An fMRI study. *Neuron*, 35:1157, 2002.
36. T. Carlson, P. Schrater, and S. He. Patterns of activity in the categorical representations of objects. *J. Cogn. Neurosci.* 15:704–717, 2003.
37. D. Cox and R. Savoy. Functional magnetic resonance imaging (fMRI) “Brain Reading”: Detecting and classifying distributed patterns of FMRI activity in human visual cortex. *NeuroImage* 19:261–270, 2003.
38. Y. Kamitani and F. Tong. Decoding the visual and subjective contents of the human brain. *Nature Neurosci.* 8:679–685, 2005.
39. A.J. O'Toole, F. Jiang, H. Abdi, and J. Haxby. Partially distributed representations of objects and faces in ventral temporal cortex. *J. Cogn. Neurosci.* 17:580–590, 2005.
40. T. Allison, A. Puce and G. McCarthy. Social perception from visual cues : Role of the STS region. *Trends Cogn. Sci.* 4:267–278, 2000.
41. E.A. Hoffman and J.V. Haxby. The distinct representations of eye gaze in and identity in the distributed human neural system for face perception. *Nature Neurosci.* 3:80–84, 2000.
42. M. Oram and D.I. Perrett. Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the Macaque monkey. *J. Neurophysiol.* 76, 109–129, 1996.
43. D.H. Ballard. Animate vision. *Artifi. Intelli.* 48:57–86, 1991.

44. Y. Aloimonos. Purposize, qualitative, active vision. *CVGIP: Image Understand.* 56(special issue on qualitative, active vision):3–129, July 1992.
45. M. Tistarelli. Active/space-variant object recognition. *Image Vision Comput.* 13(3):215–226, 1995.
46. E. L. Schwartz, D. N. Greve, and G. Bonmassar. Space-variant active vision: definition, overview and examples. *Neural Networks* 8(7/8):1297–1308, 1995.
47. C. A. Curcio, K. R. Sloan, R. E. Kalina, and A. E. Hendrickson. Human photoreceptor topography. *J. Comput. Neurol.* 292 (4) : 497–523, 1990.
48. G. Sandini, G. Metta. Retina-like sensors: Motivations, technology and applications In *Sensors and Sensing in Biology and Engineering*, T.W. Secomb, F. Barth and P. Humphrey (Eds). Springer-Verlag, Berlin, 2002.
49. P. J. Burt. Smart sensing in machine vision. In *Machine Vision: Algorithms, Architectures, and Systems*. Academic Press, New York, 1988.
50. F. Tong and Z.N. Li. The reciprocal-wedge transform for space-variant sensing. In 4th IEEE Intl. Conference on Computer Vision, pages 330–334, Berlin, 1993.
51. E. L. Schwartz. Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception. *Biol. Cyber.* 25, 181–194, 1977.
52. T.E. Fisher and R.D. Juday. A programmable video image remapper. In Proceedings of SPIE, volume 938, pages 122–128, 1988.
53. E. Grossi and M. Tistarelli. Log-polar Stereo for Anthropomorphic Robots. In Proc. of 6th European Conference on Computer Vision, pages 299–313, Springer Verlag LNCS 1842, 2000.
54. J.M. Henderson, C.C. Williams, and R.J. Falk. Eye movements are functional during face learning. *Mem. Cogn.* 33(1), 98–106, 2005.
55. A.L. Yarbus. *Eye Movements and Vision*. Plenum Press, New York, 1967.
56. Y. Yeshurun and E. L. Schwartz. Shape description with a space-variant sensor: Algorithms for scan-path, fusion and convergence over multiple scans. *IEEE Trans. PAMI, PAMI-11*:1217–1222, Nov. 1993.
57. J. Shepherd. Social factors in face recognition. In G. Davies, H. Ellis & J. Shepherd (Eds.). *Perceiving and Remembering face*, Academic Press, London, 55–79, 1981.
58. F. K. D. Nahm, A. Perret, D. Amaral, and T. D. Albright. How do monkeys look at faces?. *J. Cogn. Neurosci.* 9, 611–623, 1997.
59. M. M. Haith, T. Bergman, and M. J. Moore. Eye contact and face scanning in early infancy. *Science*, 198, 853–854, 1979.
60. A. Klin. Eye-tracking of social stimuli in adults with autism. *NICHD Collaborative Program of Excellence in Autism*, Yale University, New Haven, CT, May 2001.
61. M. Tistarelli, and E. Grossi. Active vision-based face authentication. *Image and Vision Computing: Special issue on Facial Image Analysis*, M. Tistarelli ed., vol. 18, no. 4, 299–314, 2000.
62. M. Bicego, E. Grossi, and M. Tistarelli. On finding differences between faces. in *Audio-and Video-based Biometric Person Authentication*, T. Kanade, A. Jain, and N.K. Ratha, Eds., vol. LNCS 3546, pp. 329–338. Springer, 2005.
63. C. Goren, M. Sarty, and P. Wu. Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics* 56, 544–549, 1975.
64. P. Thompson. Margaret Thatcher: A new illusion. *Perception* 9:483–484, 1980.
65. R.K. Yin. Looking at upside-down faces. *J. Exp. Psychol.* 81:141–145, 1969.
66. R.E. Galper and J. Hochberg. Recognition memory for photographs of faces. *Am. J. Psychol.* 84(3), 351–354, 1971.
67. A.J. O'Toole, T. Vetter, and V. Blanz. Two-dimensional reflectance and three-dimensional shape contributions to face recognition across viewpoint change. *Vision Res.* 39:3145–3155, 1997.
68. V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *SIGGRAPH'99 Proceedings*, ACM: Computer Society Press, 187–194, 1999.

69. P. Sinha, B.J. Balas, Y. Ostrovsky, R. Russell. Face recognition by humans: 19 results all computer vision researchers should know about. *Proceedings of the IEEE*, vol. 94, no. 11, pp 1948–1962, 2006.
70. L. Light, F. Kayra-Stuart, and S. Hollander. Recognition memory for typical and unusual faces. *J. Exp. Psychol. Human Learn. Mem.* 5:212–228, 1979.
71. T. Valentine. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quar. J. Exp. Psychol.* 43A:161–204, 1991.
72. G. Rhodes. *Superportraits: Caricatures and Recognition*. Psychology Press, Hove, UK, 1997.
73. R.S. Malpass and J. Kravitz. Recognition for faces of own and other race faces. *J. Personality Soc. Psychol.* 13:330–334, 1969.
74. J.C. Brigham and P. Barkowitz. Do “They all look alike?” The effects of race, sex, experience and attitudes on the ability to recognize faces. *J. Appl. Soc. Psychol.* 8:306–318, 1978.
75. C.A. Meissner and J.C. Brigham. Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychol Public Policy Law* 7: 3–35, 2001.
76. G. Bryant and G. Rhodes. Recognition of own-race and other-race caricatures: Implications for models of face recognition. *Vision Res.* 38:2455–2468, 1998.
77. D.J. Kelly, P.C. Quinn, A.M. Slater, A.M. Lee, L. Ge and O. Pascalis. The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychol. Sci.* 18:1084–1089.
78. P.K. Kuhl, K.A. Williams and F. Lacerdo. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255:606–608, 1992.
79. A.J. O’Toole, D. Roark, and H. Abdi. Recognition of moving faces: a psychological and neural perspective. *Trends Cogn. Sci.* 6:261–266, 2002.
80. H. Hill and A. Johnston. Categorizing sex and identity from the biological motion of faces. *Curr. Biol.* 11:880–885, 2001.
81. B. Knappmeyer, I.M. Thornton and H.H. Bulthoff. The use of facial motion and facial form during the processing of identity. *Vision Res.* 43:1921–1936, 2003.
82. W.H. Merigan. P and M pathway specialization in the macaque. In A. Valberg and B.B. Lee, editors. *From Pigments to Perception*. Plenum, New York, pp. 117–125, 1991.
83. M. Kirby and L. Sirovich, Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. Patt. Anal. Mach. Intelli.* 2 (1), 103108, Jan. 1990.
84. M. Turk and A. P. Pentland. Eigenfaces for recognition. *J. Cogn. Neurosci.* 3 (1), 7186, 1991.
85. B. Schölkopf, A. Smola, and K. R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* 10, 1299–1319, 1999.
86. L. Chengjun. Gabor-based kernel PCA with fractional power polynomial models for face recognition. *IEEE Trans. Patt. Anal. Mach. Intelli.* 26(5), 572–581, May 2004.
87. D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization, *Nature*, 401, pp. 788–791, 1999.
88. S.Z. Li, X.W. Hou, and H.J. Zhang, Learning spatially localized, parts-based representation, in CVPR, 2001, pp. 207–212.
89. H. Yu and J. Yang. A direct lda algorithm for high-dimensional data with application to face recognition. *Patt. Recogn.* 34, 2067–2070, 2001.
90. L. Juwei, K.N. Plataniotis, and A.N. Venetsanopoulos. Face recognition using lda-based algorithms. *IEEE Trans. Neural Networks* 14(1), 195–200, 2003.
91. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Patt. Anal. Mach. Intelli.* 19(7), 711–720, 1997.
92. G. Baudat and F. Anouar. Generalized discriminant analysis using a kernel approach. *Neural Comput.* 12, 2385–2404, 2000.
93. K.-R. Müller, S. Mika, G. Rätsch, K. Tsuda, and B. Scholkopf. An introduction to Kernel-based learning algorithms. *IEEE Trans. Neural Networks* 12(2), 181–201, 2001.
94. L. Juwei, K.N. Plataniotis, and A.N. Venetsanopoulos. “Face recognition using kernel direct discriminant analysis algorithms.” *IEEE Tran. Neural Networks* 14(1), 117–126, 2003.

95. S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K.-R. Müller, Fisher discriminant analysis with Kernels, Proc. IEEE Int Workshop Neural Networks for Signal Processing IX, pp. 41–48, Aug. 1999.
96. S. Mika, G. Ratsch, B. Scholkopf, A. Smola, J. Weston, and K.-R. Müller. *Invariant Feature Extraction and Classification in Kernel Spaces*, Advances in Neural Information Processing Systems 12, MIT Press, Cambridge, MA, 1999.
97. A. Martinez and A. Kak. PCA versus LDA. *IEEE Trans. Patt. Anal. Mach. Intell.* 23(2), 228–233, 2001.
98. A.K. Jain and B. Chandrasekaran. Dimensionality and sample size considerations in pattern recognition practice. in Handbook of Statistics, P. R. Krishnaiah and L. N. Kanal (Eds.) North-Holland, Amsterdam, vol. 2, pp. 835-855, 1987.
99. S.J. Raudys and A.K. Jain. Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Trans. Patt. Anal. Mach. Intell.* 13(3), 252–264, 1991.
100. M.S. Bartlett, J.R. Movellan and T.J. Sejnowski. Face recognition by independent component analysis. *IEEE Trans. Neural Networks* 13(6), 1450–1464, 2002.
101. L. Chengjun and H. Wechsler. Independent component analysis of Gabor features for face recognition. *IEEE Trans. Neural Networks* 14,(4), 919–928, July 2003.
102. F. Bach and M. Jordan. Kernel independent component analysis. *J. Mach. Learn. Res.* 3, 1–48, 2002.
103. F.L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Trans. Patt. Anal. Mach. Intelli.* 11, 567–585, 1989.
104. F.Y. Shih and C. Chuang. Automatic extraction of head and face boundaries and facial features. *Inf. Sci.* 158, 117–130, 2004.
105. K. Sobottka and I. Pitas. A fully automatic approach to facial feature detection and tracking. in J. Bigün, G. Chollet, G. Borgefors (eds.), *Audio- and Video-Based Biometric Person Authentication*, LNCS, vol.1206, pp.77-84, Springer Verlag, Berlin, 1997.
106. M. Zobel, A. Gebhard, D. Paulus, J. Denzler and H. Niemann. Robust facial feature localization by coupled features. in *4th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Grenoble, France, 2000.
107. L. Wiskott, J. M. Fellous, and C. V. der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. Patt. Anal. Mach. Intelli.* 19, 775–779, 1997.
108. Z. Xue, S.Z. Li and E.K. Teoh. Bayesian shape model for facial feature extraction and recognition. *Patt. Recogn.* 36, 2819–2833, 2003.
109. A. Pentland, B. Moghaddam and T. Starner. View-based and modular eigenspaces for face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 84–91, 21–23 June 1994, Seattle, Washington, USA.
110. J. Lange, C. V. D. Malsburg, R. P. Wurtz and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Compu.* 42(3), 300–311, March 1993.
111. L. Wiskott. Phantom faces for face analysis. *Patt. Recogn.* 30(6), 837–846, 1997.
112. B. Duc, S. Fischer, and J. Bigün. Face authentication with Gabor information on deformable graphs. *IEEE Trans. Image Process.* 8(4), 504–516, Apr. 1999.
113. C. Kotropoulos, A. Tefas, and I. Pitas. Frontal face authentication using morphological elastic graph matching. *IEEE Trans. Image Process.* 9(4), 555–560, Apr. 2000.
114. P. T. Jackway and M. Deriche. Scale-space properties of the multiscale morphological dilation-erosion. *IEEE Trans. Patt. Anal. Mach. Intelli.* 18(1), 38–51, 1996.
115. A. Tefas, C. Kotropoulos and I. Pitas. Face verification using elastic graph matching based on morphological signal decomposition. *Signal Process.* 82(6), 833–851, 2002.
116. N. Kruger. An algorithm for the learning of weights in discrimination functions using A priori constraints. *IEEE Trans. Patt. Anal. Mach. Intelli.* 19(7), 764–768, July 1997.
117. A. Tefas, C. Kotropoulos and I. Pitas. Using support vector machines to enhance the performance of elastic graph matching for frontal face authentication. *IEEE Trans. Patt. Anal. Mach. Intelli.* 23(7), 735–746, 2001.

118. K. I. Chang, K. W. Bowyer, and P. J. Flynn. Face Recognition Using 2D and 3D Facial Data. Workshop in Multimodal User Authentication, pp. 25-32, Santa Barbara, California, December. 2003.
119. Kyong I. Chang, Kevin W. Bowyer and Patrick J. Flynn. An evaluation of multi-modal 2D+3D face biometrics. *IEEE Trans. Patt. Anal. Mach. Intelli.* 27(4), 619–624, 2005.
120. C. Kotropoulos, A. Tefas and I. Pitas. Frontal face authentication using discriminating grids with morphological feature vectors. *IEEE Trans. Multimedia* 2(1), 14–26, Mar. 2000.
121. S. Arca, P. Campadelli, R. Lanzarotti. A face recognition system based on automatically determined facial fiducial points. *Patt. Recogn.* 39(3), 432–443, March 2006.
122. E.G. Llano, H. Mendez-Vazquez, J. Kittler and K. Messer. An illumination insensitive representation for face verification in the frequency domain. In Proceedings of the 18th international Conference on Pattern Recognition – Vol 01, pp. 215-218, (August 20–24, 2006), IEEE Computer Society, Washington, DC.
123. S. M. Pizer, E. P. Amburn, J. D. Austin, R. Cromartie, A. Geselowitz, T. Greer, B. M. ter Haar Romeny, J. B. Zimmerman, K. Zuiderveld. Adaptive histogram equalization and its variations. *CVGIP* 39(3):355–368, September 1987.
124. Y. Adini, Y. Moses and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. Patt. Anal. Mach. Intelli.* 19(7):721–732, 1997.
125. H. W. Jensen. Digital face cloning. In Proceedings SIGGRAPH'2003 Technical Sketch, San Diego, July 2003.
126. Laiyun Qing, Shiguang Shan, and Xilin Chen. Face relighting for face recognition under generic illumination. ICASSP 2004, pp. 733–736, 2004.
127. R. Gross and V. Brajovic. An image preprocessing algorithm for illumination invariant face recognition. Proc. of International Conference on Audio- and Video-Based Biometric Person Authentication, pp. 10–18, 2003.
128. Haitao Wang, Stan Z. Li, Yangsheng Wang, and Weiwei Zhang. Illumination modeling and Normalization for Face Recognition Proceedings of IEEE International Workshop on Analysis and Modeling of Faces and Gestures. Nice, France. 2003.
129. Laiyun Qing, Shiguang Shan, and Wen Gao. Face recognition under varying lighting based on derivates of log image. In SINOBIOMETRICS, pp. 196–204, 2004.
130. Y. Weiss. Deriving intrinsic images from image sequences. Proc. ICCV,01. Vol. II, pp. 68–75, 2001.
131. M. Turk and A. P. Pentland. Eigenfaces for recognition. *J. Cogn. Neurosci.* 3(1), 71–86, 1991.
132. P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. J. Worek, Overview of the face recognition grand challenge, In Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp.947-954, 2005.
133. S. M., Lucas. Continuous n-tuple classifier and its application to real-time face recognition. In IEE Proceedings-Vision Image and Signal Processing, vol. 145, no. 5, October 1998, p. 343.
134. S. M. Lucas, and T. K., Huang. Sequence recognition with scanning N-tuple ensembles. In Proceedings ICPR04 (III) pp 410–413.
135. B., Raytchev, and H., Murase. Unsupervised recognition of multi-viewface sequences based on pairwise clustering with attraction andrepulsion. *Comput. Vision Image Understand.* 91 (1-2), July–August 2003, pp. 22–52.
136. B. Raytchev, and H. Murase. VQ-Faces: Unsupervised face recognition from image sequences. In Proceedings ICIP02 (II), pp. 809–812.
137. B. Raytchev, and H. Murase. Unsupervised face recognition from image sequences. In Proceedings ICIP01(I), pp. 1042–1045.
138. S. Zhou, V. Krueger, and R. Chellappa. Probabilistic recognition of human faces from video. *Comput. Vision Image Understand.* 91(1-2), 214–245, July–August 2003.
139. S.K. Zhou, R. Chellappa, and B. Moghaddam. Visual Tracking and Recognition using appearance-adaptive models in particle filters. *Image Process.* 13(11) 1491–1506, November 2004.

140. Y. Li, S. Gong, and H. Liddell. Modelling faces dynamically across views and over time. In Proceedings IEEE International Conference on Computer Vision, pages 554-559, Vancouver, Canada, July 2001.
141. Y. Li, S. Gong, and H. Liddell. Support vector regression and classification based multiview face detection and recognition. In Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR'00), Grenoble, France, pp. 300– 305, 2000.
142. A.J. Howell and H. Buxton. Towards unconstrained face recognition from image sequences. In Proceeding. of the IEEE International Conference on Automatic Face and Gesture Recognition (FGR'96), Killington, VT, pp. 224–229, 1996.
143. F. Roli and J. Kittler, (Eds.) Multiple Classifier Systems. Springer Verlag, Berlin, LNCS 2364, 2002.
144. B. Achermann and H. Bunke. Combination of classifiers on the decision level for face recognition. Technical Report IAM-96-002, Institut für Informatik und angewandte Mathematik, Universität Bern, January 1996.
145. X. Liu, and T. Chen. Video-based face recognition using adaptive hidden Markov models. In Proceedings CVPR03 (I), pp. 340–345.
146. Bicego, M., Grossi, E. and Tistarelli, M.: Person authentication from video of faces: a behavioral and physiological approach using Pseudo Hierarchical Hidden Markov Models. In Proceedings Intern.l Conference on Biometric Authentication 2006, Hong Kong, China, January 2006, pp 113-120, LNCS 3832.
147. Tistarelli, M., Bicego, M. and Grossi, E.: Dynamic face recognition: From human to machine vision. Image and Vision Computing: Special issue on Multimodal Biometrics, M. Tistarelli and J. Bigun ed.s, doi:10.1016/j.imavis.2007.05.006.
148. Arandjelovic, O., Cipolla, R.: Face Recognition from Face Motion Manifolds using Robust Kernel Resistor-Average Distance. In Proceedings FaceVideo04, pp 88.
149. Lee, K.C., Ho, J., Yang, M.H., Kriegman, D.J.: Video-based face recognition using probabilistic appearance manifolds. In Proceedings CVPR03 (I), pp 313-320.
150. Yamaguchi, O., Fukui, K., Maeda, K.: Face Recognition Using Temporal Image Sequence. In Proceedings IEEE AFGR98, pp 318-323.
151. K. Fukui and O. Yamaguchi: Face recognition using multiviewpoint patterns for robot vision. In Proceedings International Symposium of Robotics Research, 2003.
152. Shakhnarovich, G., Fisher, J.W., Darrell, T.J.: Face Recognition from Long-Term Observations. In Proceedings ECCV02 (III), pp 851.
153. Raytchev, B., Murase, H.: Unsupervised Face Recognition from Image Sequences. In Proceedings ICIP01(I), pp 1042-1045.
154. D.G. Lowe.: Object recognition from local scale invariant features. Proc. of International Conference on Computer Vision, pp 1150-1157, 1999.
155. D.R. Kisku, A. Rattani, E. Grossi, M. Tistarelli. Face Identification by SIFT-based Complete Graph Topology. Proc. of 5th IEEE Workshop on Automatic Identification Advanced Technologies, pp 63-68. Alghero, Italy, June 2007.
156. A.J. O. Toole, P.J. Phillips, F. Jiang, J.J. Ayadd, N. Penard and H. Abdi.: Face recognition algorithms surpass humans matching faces across changed in illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29:1642–1646, 2007.
157. A.J. O. Toole, H. Abdi, F. Jiang, J.J. and P.J. Phillips.: Fusing face recognition algorithms and humans. *IEEE Transactions on Systems, Man and Cybernetics*, 37:1149-1155, 2007.
158. P.J. Phillips, W.T. Scruggs, A.J. O. Toole, P.J. Flynn, K.W. Bowyer, C.L.; Schott, and M. Sharpe. FiRTV 2006 and ICE 2006 Large-Scale Experimental Results. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12 Mar. 2009. IEEE Computer Society, < <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2009.59> >
159. M. Husken, B. Brauckmann, S. Gehlen, and C. von der Malsburg. Strategies and benefits of fusion of 2D and 3D face recognition. Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 3:174, 2005.

160. C. Liu.: Capitalize on dimensionality increasing techniques for improving face recognition Grand Challenge performance. *IEEE Trans. Patt. Anal. Mach. Intell.* 28: 725-737, 2006.
161. C.M Xie, M. Savvides and V Kumar.: Kernel correlation filter based redundant class-dependence feature analysis (KCFA) on FRGC2.0 Data IEEE International Workshop Analysis & Modeling Faces & Gestures, 32-43, 2005.
162. V. Bruce, Z. Henderson, C. Newman, and A.M. Burton.: Matching identities of familiar and unfamiliar faces caught on CCTV images. *J. Exp. Psychol.: Appl.* 7:207-218, 2001.
163. A.M. Burton, V. Bruce, and P.J.B. Hancock.: From pixels to people: a model of familiar face recognition. *Cogn. Sci.* 23:1-31, 1999.
164. A.M. Burton, S. Wilson, M. Cowan, and V. Bruce.: Face recognition in poor-quality video. *Psychological Science*, 10:243-248, 1999.
165. H. Abdi, D. Valentine, B. Edelman, and A.J. O'Toole.: More about the difference between men and women: evidence from linear neural networks and the principal-component approach. *Perception* 24:539-562, 1995.
166. E. H. Aylward et al.: Brain activation during face perception: evidence of a development change. *J. Cogn. Neurosci.* 17:308-319, 2005.

# **Chapter 6**

## **Face Recognition at a Distance: System Issues**

**Meng Ao, Dong Yi, Zhen Lei, and Stan Z. Li**

**Abstract** Face recognition at a distance (FRAD) is one of the most challenging forms of face recognition applications. In this chapter, we analyze issues in FRAD system design, which are not addressed in near-distance face recognition, and present effective solutions for making FRAD systems for practical deployments. Evaluation of FRAD systems is discussed.

### **6.1 Introduction**

Research and development of face recognition technologies and systems have been done extensively for decades. In terms of distance from user to the camera, face recognition systems can be categorized into near-distance (often used in cooperative applications), middle-distance, and far-distance ones. The latter cases are referred to as face recognition at a distance (FRAD).

According to the NIST's face recognition evaluation reports on FERET and FRGC tests [7] and other independent studies, the performance of many state-of-the-art face recognition methods deteriorates with changes in lighting, pose, and other factors. Those factors which can affect system performance are summarized into four types: (1) technology, (2) environment, (3) user, and (4) user–system interaction, shown in Table 6.1.

For near-distance face recognition, camera can easily capture high-resolution and stable face images, but in FRAD systems, the quality of face images become a big issue. The user–system interaction in middle to far face recognition systems are not so simple. To build a robust FRAD system, these issues should be solved: resolution, focus, interlace effect, and motion blur.

In FRAD systems, image sequence from a live video is usually used for tracking and identifying people of interest. Video-based face recognition is a great challenge in face recognition area, which attracts many researchers' attentions in recent

---

M. Ao (✉)

Center for Biometrics and Security Research and National Laboratory of Pattern Recognition,  
Institute of Automation, Chinese Academy of Science, Beijing 100190, China  
e-mail: mao@cbsr.ia.ac.cn

**Table 6.1** Performance affecting factors

Aspect	Factors
Technology	Dealing with face image quality, heterogeneous face images, and problems below
Environment	Lighting (indoor, outdoor)
User	Expression, facial hair, facial ware, aging
User-System	Pose (alignment between camera axis and facing direction), height

years [11]. McKenna et al. [6] modeled face eigenspace in video data via principal component analysis. Probabilistic vote approach is used to fuse the sequence information. Zhou et al. [12, 13] took advantage of the time and temporal information to improve the recognition performance. In [8], an active face tracking and recognition system is proposed, in which two cameras, a static and a PTZ, work cooperatively. The static camera is used to take image sequences for face tracking while the PTZ camera is used for face recognition. In this way, the system is supplied with high-quality images for face recognition since the PTZ camera can be adjusted to focus on the face to be recognized. However, the above recognition methods are initially developed to recognize one person in video sequence. Therefore, how to fuse the temporal and identity information for recognizing multi-faces in one scene is still an open problem to be studied.

This chapter is focused on issues in FRAD systems using video sequences. It is organized as follows. Section 6.2 provides an analysis of problems in FRAD systems. Section 6.3 presents solutions for making FRAD systems. Section 6.4 presents two examples of FRAD systems: the face verification system used in Beijing 2008 Olympic Games, and a system for watch-list face surveillance in subway. Finally, how to evaluate FRAD systems is discussed in Section 6.5.

## 6.2 Issues in Video-Based Face Recognition

### 6.2.1 Low Image Resolution

Low resolution is a difficult problem of face recognition at a distance; see Fig. 6.1. In this case, the view of the camera is usually wide and the proportion of the face in the whole image is small. So the facial image is always at low resolution which degenerates both the performances of face detection and the recognition engines.

While there is a long way to go to develop reliable algorithms to achieve good performance with low-resolution face images, using a high-definition camera is a current solution to this problem. However, a high-resolution image will decrease the speed of the face detection.



**Fig. 6.1** High-resolution image (*left*) and low-resolution one (*right*)

### 6.2.2 *Out of Focus*

In the application of face recognition at a distance, the distance between the face and the camera is in a spacial extant. That means in the most cases the face is out of the focus of the lens which makes the face image blur; see Fig. 6.2.



**Fig. 6.2** The face at the focus (*left*) and the face out of focus (*right*)

Although the focus is conceptually a point, physically the focus has a small extent, which is called the blur circle. This non-ideal focusing is caused by aberrations of the imaging optics. Aberrations tend to get worse as the aperture diameter increases. So using a small aperture lens can decrease the degree of the blur.

### **6.2.3 Interlace in Video Images**

Interlace refers to the methods for painting a video image on an electronic display screen by scanning or displaying each line or row of pixels. This technique uses two fields to create a frame. One field contains all the odd lines of the image, the other contains all the even lines of the image. Because each frame of interlaced video is composed of two fields that are captured at different moments in time; interlaced video frames exhibit motion artifacts if the faces are moving fast enough to be in different positions when each individual frame is captured. Interlace increases the difficulties to correctly detect and recognize the face image; see Fig. 6.3.



**Fig. 6.3** The image captured by CCTV camera with interlace problem

To minimize the artifacts caused by interlaced video, a process called de-interlacing can be utilized. However, this process is not perfect, and it generally results in a lower resolution, particularly in areas with objects in motion. Using a progressive scan video system is the ultimate solution of this problem.

### **6.2.4 Motion Blur**

Motion blur is a frequent phenomenon in digital image system. It may occur when the object is moving rapidly or the camera is shaking; see Fig. 6.4. To avoid the motion blur, the camera should use rapid exposures, which causes a new problem. When taking rapid exposures, the aperture stop should be increased. This makes conflict to the out-of-focus problem.

**Fig. 6.4** When the exposure of the camera (progressive scan camera) is not rapid enough, the motion blur occurs



### 6.3 Making FRAD Systems

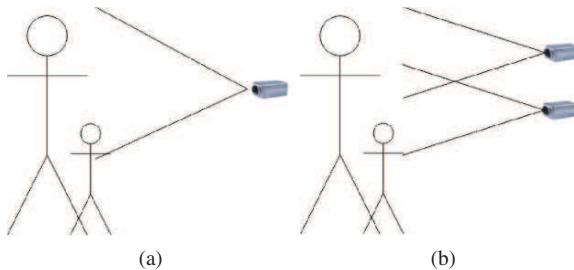
For a cooperative user system, the level of the cooperation of the users directly determine the ultimate performance. How to make the users feel natural and barrier-free is a problem of designing a face recognition system. A good design of the face recognition system could substantially not only improves the practical application performance, but also enhances the users' satisfaction with the system. The design of a cooperative user system are mainly related to the following questions: how to cover most of the user's height, how to get frontal face images, and how to capture high-quality images.

For a non-cooperative user system, there are also some hints to increasing the performance. To combine the tracking technology and recognition technology together would get a better result. The system first tracks the person's face. Then the system would get a series of images of a same person. Using these images to recognize a person is easier than just using a single image. This method receives a higher accuracy.

#### 6.3.1 Cover Most Users' Heights

For a cooperative user system, different users with different heights brings a great problem. How to make the vision of camera cover most of the users' height is an important problem of designing the system. There are usually two solutions: using

**Fig. 6.5** Cover most users' heights: (a) single camera scheme and (b) multi-camera scheme



a single large vision camera and using multiple cameras; see Fig. 6.5. Both options have their pros and cons.

Using a single large vision camera is to directly cover most of the users' height. Staple camera's image aspect ratio is fixed, 4:3 or 16:9. Here, we rotate 90° to make the camera cover a higher field of vision. At this time the proportion of the face image is decreased due to the expansion of the vision of the camera. Take a  $640 \times 480$  pixels camera as an example. If the camera is requested to cover the height of 1 m, the face image size is about 90 pixels with the eyes distance. Therefore, the scheme of using a single large vision camera always requests a high-resolution camera.

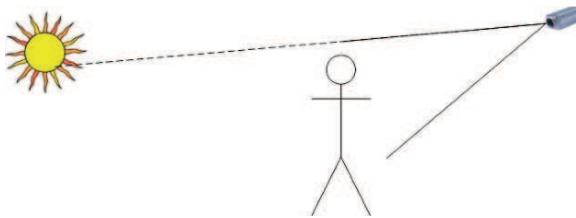
Using multiple cameras is to make each camera cover a different height. In multi-camera scheme, it is necessary to solve the problem how to use multi-images. There are usually two ways: merging the multi-images and using multi-images independently. Merging images brings a new preprocessing problem. When using multi-images independently, the multi-visions of the cameras should be overlapped in order that the face image is not cut into two images.

### 6.3.2 Capture Frontal Faces

The face algorithm gets the best results when the face images are frontal ones. How to make users to face the camera so as to capture the frontal face image is another system design problem. To achieve this purpose, we can place some devices to attract the attention of the user so that the system can capture the frontal face image. Placing a screen below the camera showing the images captured by the camera may be a good choice. The screen is able to attract the users' attention. Similar to the role of the mirror, most people would self-consciously watch the screen with their own image. As the distance between the screen and camera is close, watching the screen is nearly equal to watch the camera.

### 6.3.3 Capture High-Quality Images

The image quality is whether the clarity and exposure of the image meet the requirements. The main reason of blur is out of focus and the movement of the face and the exposure problem is mainly due to the changes of environmental light and the bright background light. To avoid out of focus, we can select a large depth of vision field lens and to avoid the motion blur, we can adjust the sensitivity of the camera and

**Fig. 6.6** To dodge the sun

the speed of the shutter. When using analog video cameras, the speed of the capture device is another problem. When using high-resolution camera, the image captured by the decoding card will be jagged fuzzy because of the movement of the objects. One solution is to make the user keep stable during the recognition process. The auto-aperture lens is the only choice to solve the exposure problem caused by the change of the environment light. However, an auto-aperture lens camera captures the image with serious exposure problem in the case of bright background light, particularly when a strong illuminate such as the sun is in the field of camera vision. In order to avoid such a case, the camera should be placed at a high place; see Fig. 6.6.

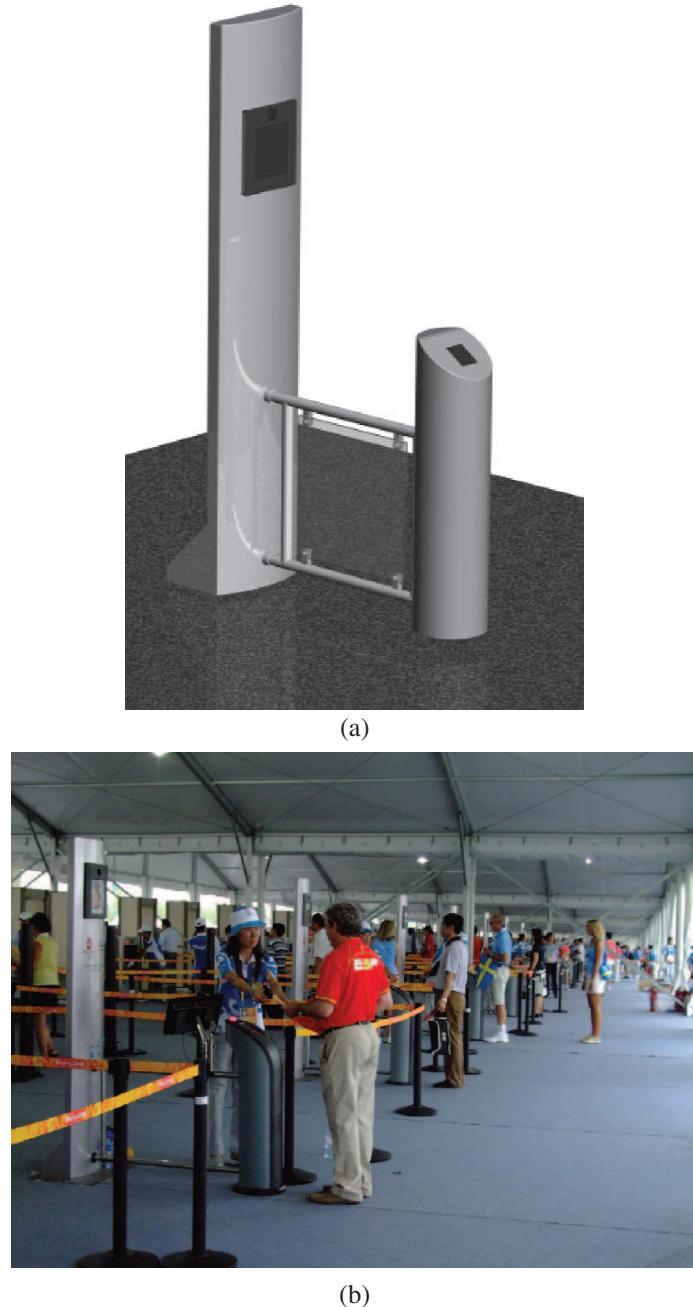
## 6.4 Examples of FRAD Systems

### 6.4.1 Face Biometric for Beijing 2008 Olympic Games

The CBSR-AuthenMetric face recognition system is developed based on the above principles and has been used as a means of biometric verification in Beijing 2008 Olympic Games. This is the first time that a biometric is used for Olympic events; see Fig. 6.7(b). This system verifies in 1:1 mode the identities of the ticket holders (expectators) on the entry to the National stadium (Bird Nest). Every ticket holder is required to submit the registration form together with a 2 inch ID/passport photo attached. The face photos are scanned into the system. Every ticket is associated with a unique ID number. When the ticket is read in, the system takes the face images and compares them with the extracted face templates for the ID. The throughput for face verification (excluding walk-in and ticket reading times) is 2 seconds per person.

The system equipment consists of the following hardware parts: a CCTV camera, a PC system, a software system, a feedback LCD, and a standing casing. An industrial design of the system (with an RFID ticket reader incorporated) is shown in Fig. 6.7(a). The body–camera distance is about 1.5 m. The system should also take care of body height between 1.45 and 2.0 m.

The software system consists of three main modules: face detection, feature template extraction, and template matching. In the first, the input image is processed by AdaBoost and multi-block local binary pattern (MB-LBP)-based face detection [3, 10]. Effective MB-LBP and Gabor features are extracted and template matching classifiers are learned using statistical learning [1, 4]. The self-quotient image (SQI) technique [9] is used to deal with the illumination change problem.



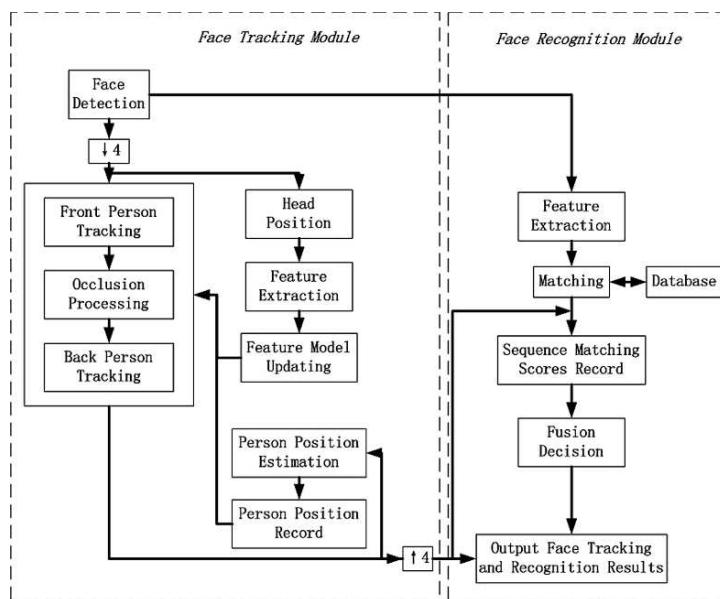
**Fig. 6.7** Face verification used in Beijing 2008 Olympic Games. **(a)** The industrial design of the system. **(b)** On-site deployments and applications

The system has to deal with several technical challenges. It works outdoors between 3 p.m. to 8 p.m., so can face toward possible sunlight shed directly into the camera. This is the first challenge to the system. The second challenge is the non-standard photo images. Although requirements are specified as using 2 in ID/passport photos, some registrants use non-ID/passport photos, small photos, or unclear photos. The photo scanning process contains flaws: some scanned photos are out of focus, and some are scans of wrong parts of registration forms. Other changes are related to the coordination with other parts of the whole security system.

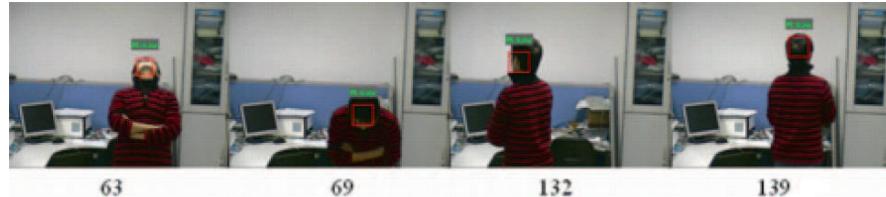
#### 6.4.2 Face Surveillance

Face surveillance is a non-cooperative user application. In such settings, the system should be able to follow the faces when the people under tracking are not facing to the camera or when the people's state is not able to be recognized. In addition, mutual occlusions may occur as multiple faces are moving and interact with one another and some faces may disappear in several frames due to total occlusion. Moreover, the quality of the video is usually low because of the low resolution and object motion. Therefore, the system should track the faces and recognize the faces with a series of face images. This combination enhances the recognizing result. So the face tracking is necessary in such a task.

Figure 6.8 shows a method for incorporating face recognition into face tracking [5]. In the face tracking module, Gaussian mixture models (GMMs) are used to



**Fig. 6.8** Combining face tracking and face recognition



**Fig. 6.9** Combining face tracking and face recognition can deal with largely rotated faces. From left to right: the face looking upward (frame 63), looking downward (frame 69), turning aside (frame 132), and turning back (frame 139)

represent the appearances of the tracked head and the upper body. Two GMMs are used to represent the appearance of each person. One model is applied to the head appearance to keep head tracking and to predict the head position in the next frame. The other is applied to that of the upper body to deal with occlusions. These two models are updated online.

The face recognition module uses an LBP and Adaboost method based on that in [2] to obtain identity matching scores for each frame. These matching scores are computed over time to obtain a score sequence. The matching scores are fused and used to help associate the tracked persons in consecutive frames, as well as to provide face recognition results. When the fused scores are very slow, the system will consider the corresponding persons have not been enrolled before. The recognition result can be shown on the tracked object; see Fig. 6.9.

This system is used in municipal subways for watch-list face surveillance. Subway scenes are often crowded and contain simultaneously moving objects including faces. Figure 6.10 shows a real scene. The cameras are fixed at the entrances and



**Fig. 6.10** Watch-list face surveillance at entrance of subways. A watch-list person is alerted by the red rectangle on the face

exists of the subway, where the people would face to the camera naturally. The system will automatically alarm when people in the watch-list appear in the field of view.

## 6.5 Evaluation of FRAD Systems

FRAD evaluations can be classified into three types: algorithm evaluation, application system evaluation, and application operational evaluation. Algorithm evaluation tests the performance using the data in a public database or a certain database for testing the accuracy of algorithms. Application system evaluation tests the face recognition system in the laboratory or a simulator environment. The face recognition system is constructed similarly to the real case. Some people test the system according to the process in real using. Application operational evaluation is to test the system in the real using. The system records the data in real using process for some time. The result of the testing is obtained by analyzing the log file of the system. These three types of system evaluations are ordered by the difficulty level increasing.

Algorithm evaluation of face recognition algorithm is a method which is used in a very wide range. In such an evaluation there are many well-known public database available such as used for FERET and FRGC. However, algorithm evaluation cannot be fully representative of the system in use as the final performance. There is a great distinction of the data between the real face recognition system and the database in personnel, the quality of the image shooting, the shooting environment, and the photography equipment used. So algorithm evaluation of the representative system can only be used for testing the performance of face recognition algorithm. Algorithm is the most crucial factor of a face recognition system performance, but not the only factor.

Application system evaluation is a mostly used method. In the simulation environment, the user tests the system in accordance with the real use of the system processes. In such a test, the simulated environment is different from the real environment in lighting condition and others. Also the users are different from the real users in experience, habit, and knowledge. As a result, the application system evaluation gives a result different from the real performance. And this result is always better than the real one.

Application operational evaluation is able to represent the real performance of the system. In real using process, the system records the data which are necessary in analysis. The result of the evaluation is given by analysis of the log file. The result of this evaluation matches the users' feeling.

Face recognition system performance does not entirely hinge on the performance of the algorithm. Only an algorithm performance testing is not enough for a face recognition system. The algorithm performance is merely one of the ultimate factors of the system performance. The different face recognition system using a same face recognition algorithm always gives different manifestations. How to increase the system performance without changing the face recognition algorithm is an important problem.

## Proposed Questions and Exercises

- What hardware and software modules are needed for a general face recognition system and a video-based FRAD system?
- What are the main issues in FRAD? In what aspects is surveillance video-based FRAD more challenging than cooperative, near-distance face recognition?
- How would you propose solutions for dealing with these challenges?
- How face detection, tracking, and matching could be combined to deal with problems in FRAD?
- How do you expect camera properties and lens would affect the performance?
- How a multiple camera system could be used to deal with problems in FRAD?
- How would a super-resolution algorithm help solving the low-resolution problem?
- Implement a Matlab algorithm for de-interlacing.
- Implement a Matlab algorithm for de-blurring.
- What are criteria for performance evaluation of a FRAD system? Why is it more difficult than that of a face recognition algorithm engine?
- Assuming you are buying a FRAD system for a watch-list FRAD application, propose a protocol to test candidate products how it meets your requirements.

## References

1. Z. Lei, R. Chu, R. He, and S. Z. Li. Face recognition by discriminant analysis with gabor tensor representation. In *Proceedings of IAPR International Conference on Biometric*, volume 4642/2007, Seoul, Korea, 2007.
2. S. Z. Li, R. Chu, S. Liao, and L. Zhang. Illumination invariant face recognition using near-infrared images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(Special issue on Biometrics: Progress and Directions), April 2007.
3. S. Z. Li and Z. Q. Zhang. FloatBoost learning and statistical face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1112–1123, September 2004.
4. S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li. Learning multi-scale block local binary patterns for face recognition. In *Proceedings of IAPR International Conference on Biometric*, volume 4642/2007, Seoul, Korea, 2007.
5. R. Liu, X. Gao, R. Chu, X. Zhu, and S. Z. Li. Tracking and recognition of multiple faces at distances. In *Proceedings of IAPR International Conference on Biometric*, volume 4642/2007, Seoul, Korea, 2007.
6. S. McKenna, S. Gong, and Y. Raja. Face recognition in dynamic scenes. In *Proceedings of British Machine Vision Conference*, pages 140–151. BMVA Press, 1997.
7. P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.
8. S. Prince, J. Elder, Y. Hou, M. Sizinstev, and E. Olevsky. Towards face recognition at a distance. *Crime and Security, 2006. The Institution of Engineering and Technology Conference on*, pages 570–575, June 2006.
9. H. Wang, S. Z. Li, and Y. Wang. Face recognition under varying lighting conditions using self quotient image. *fg*, 0:819, 2004.
10. L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li. Face detection based on multi-block lbp representation. In *Proceedings of IAPR International Conference on Biometric*, volume 4642/2007, Seoul, Korea, 2007.

11. W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, pages 399–458, 2003.
12. S. Zhou, V. Krueger, and R. Chellappa. Face recognition from video: A condensation approach. *fg*, 0:0221, 2002.
13. S. Zhou, V. Krueger, and R. Chellappa. Face recognition from video: a condensation approach. *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 221–226, May 2002.

# **Chapter 7**

## **Long-Range Facial Image Acquisition and Quality**

**Terrance E. Boult and Walter Scheirer**

**Abstract** This chapter introduces issues in long-range facial image acquisition and measures for image quality and their usage. Section 7.1 on image acquisition for face recognition discusses issues in lighting, sensor, lens, blur issues, which impact short-range biometrics but are more pronounced in long-range biometrics. Section 7.2 introduces the design of controlled experiments for long-range face and why they are needed. Section 7.3 introduces some of the weather and atmospheric effects that occur for long-range imaging, with numerous of examples. Section 7.4 addresses measurements of “system quality,” including image-quality measures and their use in prediction of face recognition algorithm. This section also introduces the concept of failure prediction and techniques for analyzing different “quality” measures. The section ends with a discussion of post-recognition “failure prediction” and its potential role as a feedback mechanism in acquisition. Each section includes a collection of open-ended questions to challenge the reader to think about the concepts more deeply. For some of the questions we answer them after they are introduced; others are left as an exercise for the reader.

### **7.1 Image Acquisition**

Before any recognition can even be attempted, the system must acquire an image of the subject with sufficient quality and resolution to detect and recognize the face. The issues examined in this section are the sensor issues in lighting, image/sensor resolution issues, the field of view, the depth of field, and effects of motion blur.

#### **7.1.1 In the Beginning: Let There Be Light**

To recognize a face one needs an image with visible features, which requires that we collect an image with sufficient light levels and quality. Understanding the impact of

---

T.E. Boult (✉)

Vision and Security Technology Lab, Department of Computer Science, University of Colorado,  
Colorado Springs, CO 80918-7150, USA  
e-mail: t.boult@vast.uccs.edu

illumination variation, or normalizing to reduce it, is by far the most well studied of the issues associated with lighting and face recognition [1, 5, 8, 10, 17, 29]. While this type of work is very important, it is more focused on algorithms and not acquisition, and hence not covered in this chapter. This section will focus on illumination aspects associated with acquisition, in particular collecting and measuring light.

When working at close range in daylight conditions, the issue of sufficient lighting is not a critical concern. However, as one starts looking at long-range face-based recognition, especially for 24-h “surveillance,” assuring sufficient light level is critical. Addressing this raises two unique issues: how to measure those light levels and what sensors to use to collect in lower light and/or long-range settings.

Long-range face needs very long focal lengths, often in the range 800–3200 mm. Combining distance with the inherent limits on optics results in high F-numbers levels. For example, the Questar Ranger 3.5, which is a portable telescope used in long-range surveillance, provides 1275–3500 mm focal lengths, but it comes at a cost of light, with the 89 mm (3.5 in.) providing F13.2 at 1175 mm and F35 at 3500 mm.<sup>1</sup> Recalling that each F-stop is a 50% loss of light, this telescope will measure intensity that is orders of magnitude smaller than that measured with a more traditional F4 lens used for close/moderate-range face recognition. This need for light is even more exasperated by the need for faster shutter speeds to avoid motion blur issues that will be described later in this chapter. Understanding the available lighting for long-range settings thus is far more important than for standard face recognition. A question then is how to report light levels for long-range experiments, especially for low-light conditions. This is important not just for scientific experimentation, but for practical concerns if one wants to determine if conditions are sufficient for a particular system to operate. The most common measure for low-light imaging is in terms of *lux*. Lux is a measure of illuminance (the accumulated light energy reaching a surface) and measures how much light is in the scene. Given the lux reaching a surface, and the bi-directional reflectance function of the material/subject, one can estimate the luminous flux (the light leaving the surface in a particular direction). Luminous flux is measured in *lumens*. One can also compute the luminous emittance, which is the luminous flux per unit area emitted by a source. Luminous emittance, like luminance, is measured in lux. Given the luminous flux one can use field of view of the lens and its F-stop to estimate the amount of light reaching the sensor from the targets. When the models are done right, this can be effective in predicting the light reaching the camera and hence the response of the sensor.

Unfortunately, to use this approach for long-range low-light imaging there are a number of difficulties. First, the reflectance function of the face varies considerably across the population. More significantly, the reflection is directional and is impacted significantly by self-shadowing, so measuring the scene irradiance with a traditional lux meter is not very effective without accounting for reflectance, shading, and shadowing which requires a detailed calculation after measurement, making it difficult to use without advanced computer models. Finally, an issue especially

---

<sup>1</sup> <http://www.company7.com/questar/surveillance/querange.html>

important for low-light settings is that even higher end handheld light sensors are only effective down to 0.01 lux. These sensors use a light-to-voltage conversion that makes them good for bright scenes. But even though their accuracy is officially rated at  $\pm 0.01$  lux, in practice, it is quite tenuous below 0.1 lux. In many of our field experiments, the available lux sensors report underflow or zero (it is too dark for them to operate). There are higher end NVIS lux meters, such as the ANV-410 and TSP-410, but these are significantly more expensive and still have the issues of not providing sufficiently directional measurements to measure light that will reach the sensor.

There is an alternative, which is to directly measure the light leaving the face in the direction of the sensor: *luminance*. The candela per square meter ( $\frac{\text{cd}}{\text{m}^2}$ ) is the SI unit of luminance; nit is a non-SI name also used for this unit. A candela is a lumen per steradian (solid angle), so a  $\frac{\text{cd}}{\text{m}^2}$  (nit) is equivalent to a  $\frac{\text{lumen}}{\text{m}^2}$  (sr), whereas a lux is a  $\frac{\text{lumen}}{\text{m}^2}$  and there is no simple conversion between lux and nits without using knowledge of the view subtended by the source (face), which varies with distance. Luminance is valuable because its quantities describe the “brightness” of the source and do not vary with distance, whereas illuminance in lux (the “light” falling on a surface) must be manipulated to estimate how much light there is to measure. Putting it another way, illuminance is a good measure to use when asking how well people or cameras can function anywhere within a dimly lit environment, but luminance is the better measure to use for how well they can view a particular target (see [14]). The question then is how to effectively measure luminance for long-range face, especially if experimenting in low-light conditions.

To address these problems and provide for a simple in-field measurement, we have adapted a different type of measurement sensor. Using a sensor originally designed for “sky quality” measurements or “sky darkness” measurements, we have a device that can operate at much lower light levels and can measure a narrow enough FOV to capture just the data of the face. The sensor being used is the SQM-L,<sup>2</sup> based on the TAOS TSL237S sensor, which is a light-to-frequency converter. The SQM-L has an added lens so that the full width half maximum (FWHM) of the sensor is  $\sim 20^\circ$ . The sensitivity to a point source  $\sim 19^\circ$  off-axis is a factor of 10 lower than on-axis and falls off faster beyond that. We will be experimenting with adding a component for further restriction of the field of view. The SQM-L sensor reports magnitudes/arcsecond, which is an astronomical unit of measurement, but which is easily converted into  $\frac{\text{cd}}{\text{m}^2}$ . If we let  $s$  be the SQM-L value reported, then luminance  $\frac{\text{cd}}{\text{m}^2} = 108,000 \times 10^{-0.4*s}$ .

We use the SQM-L for long-range face experiments by having the subject look toward the camera, so it has appropriate lighting falling on the face, and then aiming the sensor on the center of their face while holding the sensor about 18 in. away. At this range a face subtends approximately  $18^\circ$ , i.e., the sensor is measuring the light leaving the face and little else (but some care has to be used if there are distant lights behind the subject, and not shadow the face from any light sources with the

---

<sup>2</sup> <http://unihedron.com/projects/darksky/>



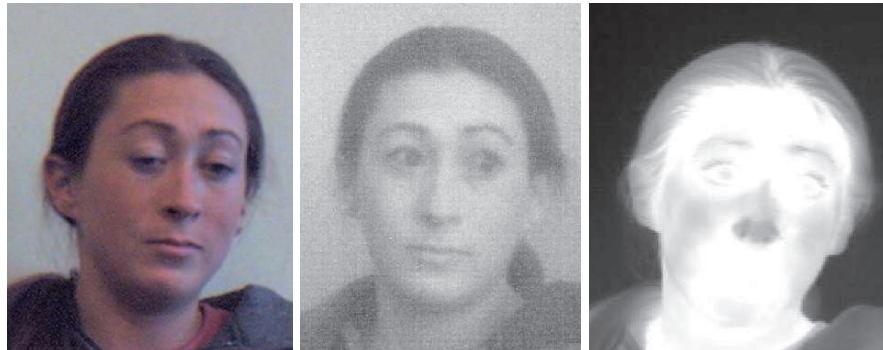
**Fig. 7.1** Example of low-light long-range EMCCD imagery. The measured scene illuminance for the left image was 0.01 lux, and illuminance was not measurable for the other two images. The measured face luminance *left to right* was 0.089, 0.0768, and 0.015 nits, respectively

hand/sensor). We call this measurement the “face luminance” and consider it the most useful overall lighting measurement for estimating performance of a long-range face system in low-light conditions. This is really a measurement of luminance but it can be converted to luminous flux using the area of a face and the solid angle subtended by a face from the target range, which is simple scaling.

For example in Fig. 7.1, we have long-range images in low-light conditions. The images were obtained at approximately 100 m with an F5.6 Sigma 300–800 mm. Capture occurs under starlight conditions 60, 90, and 120 min after sunset, with a street light 100 m off on the subject’s left. We prefer the face luminance measurement approach because it works in the low-light setting where we want to operate and it already accounts for the complex lighting/face shape interactions and is easily converted to a direct measure of the luminous flux heading in the direction of the camera. It is also very easy to “measure” in the field: hold the sensor, face the direction of the camera and push the button on the sensor and hold (maybe up to 60 s if its really dark), then read the measurements from the unit’s LEDs. These measurements are more repeatable and reliable than using simple lux-based estimations of overall illumination and then trying to convert it to lux at the sensor.

The second major issue impacting light levels for acquisition is the inherent imaging system sensitivity. This is significantly impacted by the sensor. Again, since long-range face is generally for surveillance, there is a general need to consider low-light conditions.

One approach often suggested for dealing with low-light settings is the use of infrared sensors for face recognition. For long-wave IR (8–14  $\mu\text{m}$ ), the human body is a light source and such images could be collected in total darkness. While there has been some significant progress in the area of LWIR face (see [4, 11, 23, 25]), we believe LWIR is too limited for long-range face for several reasons. First, the need for long focal length lenses and high-resolution sensors for long-range face – the combination of which are simply not available for LWIR. The resolution issues will be discussed in the next section. The second limitation of LWIR is that since



**Fig. 7.2** *Left to right* shows the same subject in a normally lit visible light camera, a low-light intensified imagery, and in LWIR thermal imagery. The intensified imagery was obtained using an American Eagle 603U which is a GenIII+ intensifier (specs are the same as PVS-14 commonly used by the US Military). The intensified image was captured by an IQ-EYE smart camera with  $1280 \times 1024$  resolution. The thermal (LWIR) sensor is an NYTEK WEB-50 Micro-Bolometer,  $8-14 \mu\text{m}$  sensor, with images captured from the analog  $640 \times 480$  video output. The visible images were captured from an IQ-EYE 1 megapixel sensor. Images having faces with 80 pixels between eyes, which is the lower end, is expected for good recognition

long-range face is usually for non-cooperative subjects, LWIR requires specialized enrollment whereas visible recognition can use standard intelligence photos.

For comparison, consider Fig. 7.2, which shows example images (close range) of three different types of sensors: a standard visible image, an intensified image, and a thermal image. This data set can be obtained, for US researchers, from the author. An interesting open research question is the development of an LWIR recognition system that can operate with visible image galleries, that is, with some initial work in the area converting thermal into visible images addressed in [6].

The alternative for low-light operation is to use some type of intensified imagery. There are a few alternatives within this group ranging from the very common tube-based intensifier optically coupled to a CCD sensor, to an intensified CCD, to the current generation of electron multiplying CCD. In our early work in low light, we used tube-based intensifiers coupled with a CDD. One disadvantage of this is the blurring induced by the micro-charge plates of the intensifier and the visible “channel” artifacts, which have also been noted by other researchers (see Fig. 1 in [22]). Our more recent work in long-range face/surveillance [24] has moved to using EMCCD technology, based on a Salvador Imaging camera using the TC 285 Chip. This provides  $1004 \times 1002$  pixel images with  $8 \mu\text{m}$  pixels and an overall quantum efficiency of 65%. This sensor can operate from full sunlight down to starlight conditions. While the full details of our long-range low-light experiments are beyond the scope of this chapter, Fig. 7.1 provides some examples of cropped face data collected with an 800-mm Sigma F5.6 lens at more than 100 m under very low light conditions. Except in the first of these photos, the naked eye camera operators could not even see the subject was there, let alone recognize them. These images show there is potential for long-range low-light face recognition using EMCCD technology.

### **7.1.2 Resolution: What Does It Mean and How Much Do We Need?**

In acquisition, presuming we have sufficient number of photons, the next most important issue is the resolution of the target. While it is quite common to hear people talk about resolution in terms of number of pixels, it is more accurate to talk about the effective resolution. One can formally define this using the modulation transfer function (MTF) of the imaging system, which can account for both blur and contrast loss. Under some simplifying assumptions one can decompose the MTF into the product of the optical (lens) MTF, the sensor geometry MTF, and the diffusion MTF [9].

The ability of a lens to resolve detail is usually determined by the quality of the lens, though some very high-end lenses and telescopes are diffraction limited. The effective aperture of the lens diffracts the light rays so a single point in space forms a diffraction pattern in the image, which is known as the Airy disk. If the system is not diffraction limited, then other lens artifacts produce patch such that different rays leaving a single scene point do not arrive at a single point in the image, giving rise to what is called the “circle of confusion,” even though it can be a far more complex shape. Ideally, the circle of confusion will be smaller than a sensor pixel.

Most MTF tables provided by lens manufacturers (see [3]) will show the MTF as a function of image position or distance from the center of the image. MTF values above 0.6 are considered satisfactory, while some lenses such as the Canon EF 400 mm f2.8 IS USM, which we use for some of our long-range experiments, have a circle of confusion of 0.035 mm and MTF values above 0.9 over the whole field of view. Even when extending with the Canon 2xII extender (making it an 800-mm F5.6 lens), the MTF is above 0.7 everywhere. In general, zoom lenses will have lower MTF because of the more complex lens designs that limit the optimization. (Note: you can buy adapters for C-mount to Canon lenses, with complete rs232-based control of lens parameters such as focal distance, aperture, and stabilization parameters. These adapters are open air but because they increase the separation to adapt the 35-mm format to C-mount they may degrade the MTF.)

It is important that when working with long-range biometric the lenses matched or overqualified for the sensor choice. Modern high-quality lenses are multiple-element multi-coated designs optimized by the manufacturers for particular sensor choices and with particular wavelengths in mind. If you can see vignetting, spatially varying blur (when focused on a flat target) or color “fringe” artifacts, find a better lens. It is also important to note that few lenses are optimized outside of the visible range, so be particularly careful in choices if working in the NIR range.

In the remainder presume that the optics are properly adapted to the sensor such that the overall MTF is not significantly limited by the optics, atmospheric, or by motion blur because if those are the limiting factors, it makes little sense to discuss sensor “resolution.” In practice, the important consideration is that the blur of the system is less than a pixel, otherwise the image can be effectively down-sampled by a factor of the blur and not loose significant information. If you are doing long-range biometrics, the minimum is to measure your effective blur, or you can waste a lot

of time working on issues which are limited by blur. In short, a large sensor/image size with a blurred image is not providing the resolution you might think.

Assuming good optics, resolution for long-range face becomes a question of ensuring enough pixels on face to support recognition, and sufficiently above the minimum needed for recognition to deal with the loss of resolution due to atmospheric turbulence. Formal models for atmospheric loss have been derived in the literature. See [27]. Diving into those models is beyond the chapter, but using such models can estimate an atmospheric blur level for long-range face and expand the resolution requirements by an equivalent amount. We have routinely expanded the 60 pixel inter-pupil distance (IPD), used for close-range face recognition, to 80 pixel IPD for our 100-m experiments.

Given the desired goal of 80 pixels between the eyes and an average physical size of 4 between 60 and 72 mm, combined with pixel size (for example, 8  $\mu\text{m}$  for visible spectrum sensors, 15  $\mu\text{m}$  for LWIR) and size of the sensor (in number of pixels), one can then estimate the focal length needed to produce an image with the necessary spatial resolution on the subject. Deriving the formula is left as an exercise for the reader. In doing so, do not forget to account for any “adaptors,” e.g., converting a 35-mm camera lens into a C-mount is a change in format and back-focal distance that impacts the effective focal length.

In reviewing Table 7.1, you should note that lenses for visible sensors up to 1000 mm are readily available and up to 3500 mm available as special order via fields “telescopes.” Most intensified CCDs are only  $640 \times 480$  and LWIR sensors are only  $320 \times 240$  (though there are exceptions for both, there are no  $1280 \times 1024$  LWIR sensors). Long-wave IR lenses up to 300 mm are available and up to 1000 mm is a special order (and massive).

**Table 7.1** Focal lengths needed to achieve 80 pixel average inter-pupil distance for different sensor sizes

Range (m)	$1280 \times 1024$	$640 \times 480$	$320 \times 240$
50	333	625	1250
100	667	1250	2500
150	1000	1875	3750
200	1333	2500	5000
250	1667	3125	6250
300	2000	3750	7500

### 7.1.3 The Working Volume: Depth of Field and Field of View

The working volume, the region where the subject is in focus and within the field of view, is clearly important for the acquisition system design. Depth of field (DOF) defines the ranges around the focus distance where subjects will be in sharp focus. DOF increases with decreasing lens aperture and decreases with focal length, so for long-range face it is a much more significant issue. The depth of field for a lens is not symmetric, with different formula for the distance in front of the focus plane and behind.

Formally we can derive these as

$$\text{Front depth of field} = \frac{d \cdot F \cdot a^2}{f^2 + d \cdot F \cdot a} \quad (7.1)$$

$$\text{Rear depth of field} = \frac{d \cdot F \cdot a^2}{f^2 - d \cdot F \cdot a} \quad (7.2)$$

where  $f$  is the focal length,  $F$  is the F number,  $d$  is the diameter of the circle of confusion, and  $a$  is the subject distance from the first principal of the lens to subject. Note that if  $f^2 < d \cdot F \cdot a$ , the rear depth of field is considered infinite.

The important things to note here is that the depth of field decreases with the square of the focal length and for long focal length lenses can be quite short. Focus is further exasperated by the fact that for long-range face the optical axis usually does not intersect a ground plane where the target will be, because they will be walking well above the ground, and thus there is no way to easily pre-focus the image. Fast auto-focus or having subjects “walk through” the DOF region are the most common choices.

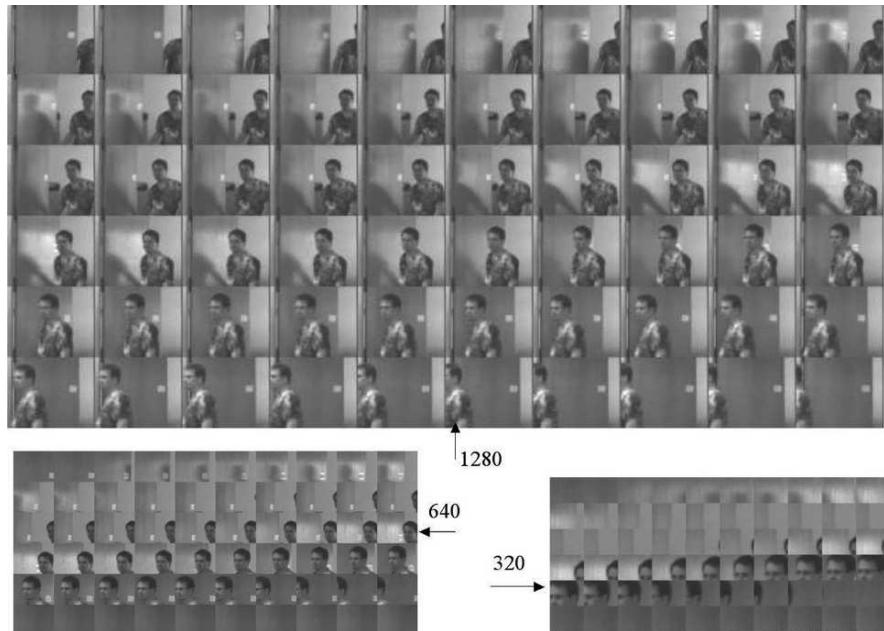
While DOF is directly impacted by distance, the FOV of a lens is not. Ignoring blurring, the field of view necessary to maintain sufficient resolution for long-range face is actually the same as that needed for near-field “non-cooperative” subjects. The increased resolution requirements to account for atmospheric blur do change it, but the change is effectively the same as requiring a larger inter-pupil distance in pixels.

The more significant difference is that there are many near-field face applications presuming cooperative subjects at effective choke points to limit subject positioning. With non-cooperative subjects, a larger field of view is needed to allow for subject movement. This is especially acute in maritime biometrics where the subjects, and the sensor, may be moving with the waves.

The FOV is defined by the combination of the sensor resolution and the focal length. Presuming a focal length just sufficient for the minimum resolution provides the maximum FOV. Again, one can easily derive, via basic geometry, the FOV, the associated pixel resolution on the sensor, and the effective physical size at the focal point of the working volume. Example figures are shown in Table 7.2, with the derivation of the formula left to the reader. In deriving the table, we presumed an I'D of 80 pixels and an overall head size of 160 pixels and assured the head is within the frame. Figure 7.3 puts that data into perspective and also shows how the FOV affects “time of target” if one is using a stationary camera aimed at a choke point.

**Table 7.2** Usable resolution and size of FOV; the maximum size can reasonably be used for face recognition. Conservative estimates are half the sizes/times shown

Sensor resolution	2048 × 1520	1280 × 1024	640 × 480	320 × 240
Usable size in pixels	1888 × 1320	1120 × 824	480 × 280	160 × 40
Usable physical FOV (in ft)	5.8' × 4.3'	3.6' × 2.6'	1.5' × 0.9'	0.5' × .1'
Allowed height variation	±25in	±15.6 in	±5.4 in	± 0.6 in



**Fig. 7.3** Example showing image sequences of a subject exiting a doorway and how the sensor resolution and FOV affect the effective number of frames where there are sufficient face data for potential recognition. Note how  $640 \times 480$  is just large enough for a head and would not capture good data for someone significantly taller or shorter. The  $320 \times 240$  sensor is relatively useless

It is not just that the larger sensor gives a larger FOV; the larger FOV translates into more frames on target as they cross the larger FOV.

#### 7.1.4 Motion Artifacts

The last significant “sensor” issue to be discussed on acquisition is motion artifacts including motion blur. In any face-based system with non-controlled subjects the issue of subject motion must be addressed. This section addresses some of those motion artifact issues.

At first one might again presume that these issues are the same for long-range face as they are for any non-cooperative subject. That is, in part, correct. However, the long focal lengths necessary for long-range face can mean that even a slight vibration in the sensor mounting can produce far more significant results. The vibrations near field imaging on a basic camera mount on the wall might produce unnoticeable interlace artifacts, but the same vibration magnified by an 800-mm lens might tear the image apart and seem as here simple: *Do not even think of using an interlace camera for long-range face recognition.*

Beyond interlace artifacts, there are two major artifacts which impact long-range face. While they also impact near-field face recognition, the fact that non-cooperative



**Fig. 7.4** Example of motion blur. Subject is moving at a walking pace toward the EMCCD camera. Images are taken at approximately 100 M from the camera at dusk. The top of the walking stride produces minimal motion blur (scene has approximately 0.04 lux, yielding face lumens of 0.115 nits)

distance subjects can be moving faster or that the long-range sensor could be on a moving platform induces greater potential for these issues to be objectionable.

The first of these issues is the well-known motion blur. It can occur because of platform motion, including vibrations, as well as because of subject motion. We have found that for long-range face with walking subjects, most of the gait cycle will have noticeable vertical motion blur with a significant reduction at the top of the stride. Figure 7.4 shows an example with both a clear face image and various images showing motion blur. These images were with an 800-mm f5.6 lens with a shutter speed of 1/30 of a second. These types of issues are further exasperated if attempts are made to use slower shutter speeds, or if the camera or subject is on a moving platform such as a ship.

A less well-known issue, which may not be obvious at first, arises with modern CMOS sensors that use a rolling shutter. Before we describe the issue, study the images in Fig. 7.5 and see if you can discern what it is. Note that the images have



**Fig. 7.5** Examples of rolling shutter artifact

a very fast integration time speed (1/10,000 of a second), so if you thought it was motion blur, think again. And it is not a depth of field issue either.

There are two primary reasons to use a rolling shutter. First it saves one transistor per cell compared to a true “snapshot” shutter, and second it allows the integration time to be almost equal to the frame rate without significant buffering or fast readout circuits. The concept is quite simple; think of it as having two pointers to sensor pixels, both “rolling down” the sensor. One pointer is for readout of data, and the other is for reset or erase operation. The time difference between a row’s erase and next read defines the effective integration time (shutter speed). Each pixel sees (and accumulates) the light for the same exposure time (from the moment the erase pointer passes it till it is read out), but that happens at different times.

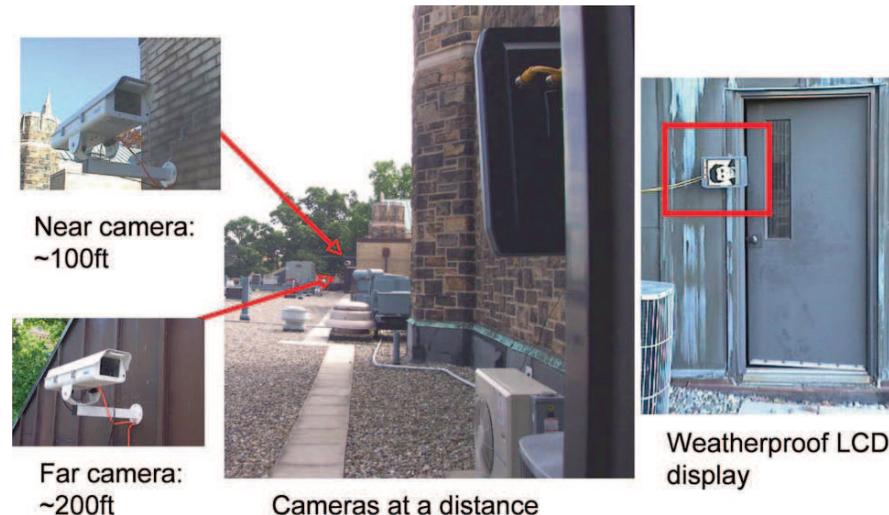
All this sounds good – a wider range of integration times at a lower cost. So what is the problem? Looking back at Fig. 7.5 again, we will give you a hint. The wall to the subject’s left is a normal doorway, it is a vertical edge and “straight.” Your cell phone camera is almost certainly a rolling shutter CMOS sensor – you can try some experiments on your own and see how significant the skew, warp, or wobble can be.

The issue for rolling shutters is that even with a short integration time, the shutter is capturing data at different times for the top and bottom of the image. In the example the camera was subjected to horizontal motion fast enough that the top of the image saw the wall in a different position than the middle or the bottom. Now ask yourself what that would be doing to your face recognition algorithm, and you will start to appreciate the issue and probably think twice about rolling shutter sensors, even if they are the cost-effective solution for getting multi-megapixel arrays.

This first section has reviewed the early and more static aspects of image acquisition for long-range face. The next section examines approaches for controlled experiments for long-range evaluation, which is a necessary precursor before we can get into the impacts of weather and atmosphere.

## 7.2 Photo-Heads: Controlled Experiments in Long-Range Face

Even after the images are acquired the atmosphere and weather impacts can be critical for long-range face acquisition. Studying them is a challenge as it is hard to collect enough data under varying conditions. To address this we designed a specialized experimental setup called Photo-heads. The setup of the initial photo-head experiment is shown in Fig. 7.6, and example images in Fig. 7.7. This “photo-head” data is unique, in that it is a well-known set of 2D images (FERET) that were displayed on a special LCD and then re-imaged from approximately 94 and 182 ft. (We are currently implementing another photo-head setup at much greater distances, with 3D animated imagery.) At these distances we needed a very long FOV lens, for which we used Phoenix 500-mm zoom lenses (for 35-mm cameras), with C-Mount adapters and Panasonic PAL cameras. The marine LCD was 800×600 resolution with 300 Nits and a special anti-reflective coating. For display the FERET face



**Fig. 7.6** The photo-head experimental setup. Two cameras are positioned at two different distances from a mounted weather-proof LCD display on a rooftop. Data capture occurred from dawn till dusk. Experiments were conducted over 2 years, capturing weather for all seasonal conditions

images were scaled up for display. As one can see from the examples in Fig. 7.7, which are all from the same subject, the FERET data have a range of inter-pupil distances, poses, and contrasts. This re-imaging model allows the system to control pose/lighting and subjects so as to provide the repeatability needed to isolate the effect of long-distance imaging and weather. As one can see, the collection produced images sufficient for identification but with the types of issues, e.g., loss of contrast and variations in size, that one would expect in a realistic long-distance collection. All experiments herein used FaceIt (V4), the commercial face recognition system from Identix. This algorithm was one of the top performers in the National Face Recognition vendor tests [16]. These tests were completed in the 2001–2004 time frame.



**Fig. 7.7** Example photo-heads: four different views of the same gallery subject taken at 100 ft. Moving *left to right*: image taken at dawn, mid-morning, early afternoon, and evening

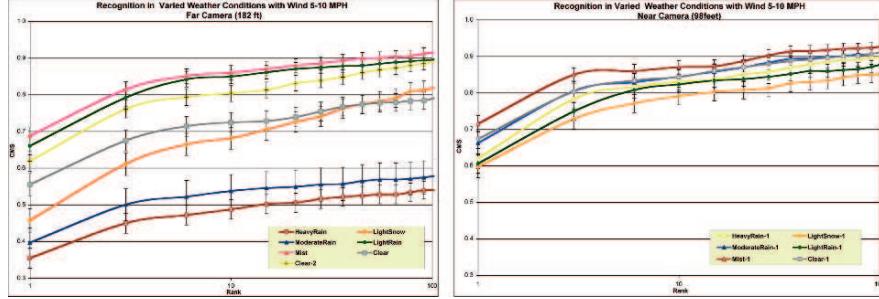
This photo-head data set is well suited to formally study the issues to be encountered in using biometrics for long-range “uncooperative” subjects in surveillance video. One of the most controlled variations is what we call “self-matching” the probe and the gallery is based on the same image, except that the probe has been subject to the long-range (re)imaging process, atmospheric disturbances, and the weather. The self-matching experiments are tightly controlled – they have exactly the same pose and subject-lighting conditions. For initial testing we used a camera at approximately 15 ft and the rank-1 self-matching performance was over 99%, showing that the re-imaging process and LCD are not a significant issue. We then moved to the real photo-head collections. We generally ran each data set, which includes 1024 images, with four images of each subject, every 15 min, with collections over 4 months. The resulting 1.5TB of photo-head data was included in the DARPA HBASE, and subsets of the data are available from the authors. With four images per subject we can use the BRR technique [13] to estimate standard errors and statistical confidence. All our graphs include such error bars, though for clarity it is often shown only for the first plot point as it usually does not vary much as we change “rank” in the CMS curves.

### 7.3 In the Middle: Atmospheric and Weather

The obvious utility of a photo-head setup is the ability to capture outdoor conditions at all time of the year. Clearly, harsh weather conditions will have a significant impact on recognition performance, but even seemingly good conditions can have unexpected impacts on recognition performance, depending on the interaction of atmospherics. Figure 7.8 shows the visual impact of weather in three different conditions captured during the photo-head collection. Note the images are rotated for display; the white on the left edge of the middle image is the snow building up on the top of the display.



**Fig. 7.8** From *left* to *right*: clear conditions, snow conditions, and rain conditions



**Fig. 7.9** CMC curves under various weather settings with self-matching

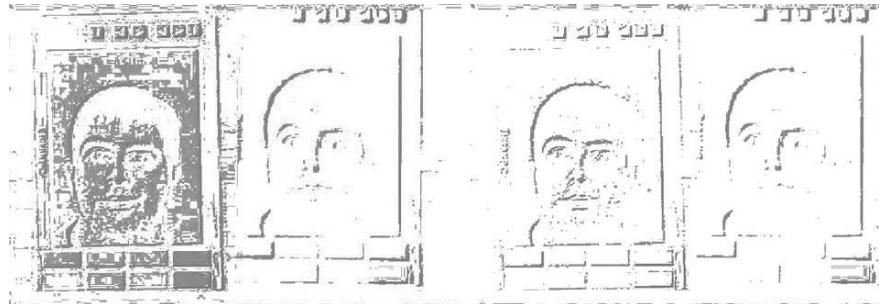
The two graphs in Fig. 7.9 show the impact of different weather conditions on face recognition. These are semi-log cumulative match curves with error bars from BRR. The curve shows the recognition rate on the vertical axis and the log “rank” used to decide correct recognition on the horizontal axis. Rank- $N$  recognition means the person was within the top  $N$  scores of the systems.

Two things should be apparent from these graphs. First, looking at rank-1 recognition (or even rank-3), off-the-shelf systems are not sufficient for these ranges even under the best of weather conditions and ideal pose/expression. Recall, these are self-matching experiments; only the imaging systems, atmospheric, and weather are stopping it from being identical images for probe and gallery.

Second, and not surprising, the far camera, at approximately 182 ft, was much more significantly impacted by the variations in weather. (The best weather rank-1 recognition at 182 ft was < 70%.) While it is not shown here, increasing wind even more significantly impacted the system, in part because at these ranges even a small deflection of the camera causes significant blur and may take the face out of the sensors field of view. (With these long FOV lenses, we needed 30" housings which increase wind loading.) These graphs are computed over more than 20,000 images, and with the “controls” of the photo-head collections we know the images are identical; thus the variations are not artifacts of individual errors, pose, or expression changes. The techniques of [15] improved performance slightly, not statistically significantly, in large part because they do not address blur or geometric distortion, only contrast and dynamic range.

There are also some initially surprising results within these curves. If you look at the far camera results, you will see that light-rain and mist are statistically better than “clear” days. Can you generate a plausible hypothesis why “clear” days were not better? We controlled for reported wind speed, so it is not that.

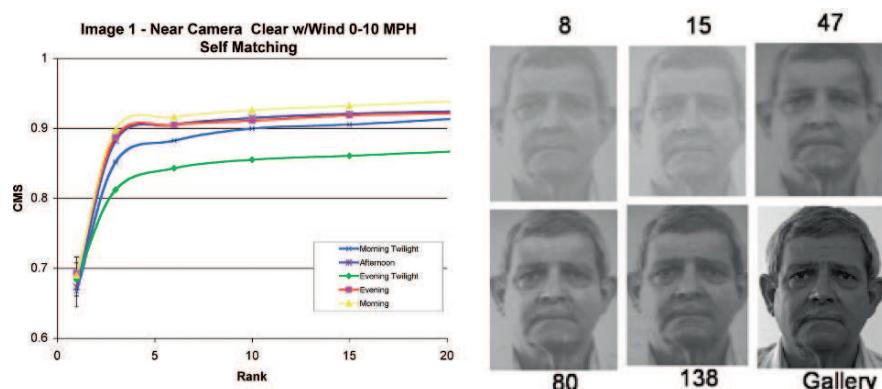
In addition to variations due to obvious weather effects, our experiments also showed that there were variations due to time of day. Atmospherics, such as thermal waves, can have a significant impact on recognition performance. Figure 7.10 shows thermal activity in different images computed from a base frame and subsequent frames from a sequence timed over an entire day. The four images (two from the far camera and two from the near) shown are from two successive captures only



**Fig. 7.10** Different images highlighting thermal activity and natural lighting changes for a sequence of frames captured several minutes apart. The first and third images are produced by the far camera at 200 ft, and the second and fourth images are produced by the near camera at 100 ft

a few minutes apart. Note the significant variation in the far camera between the two capture instances. Beyond atmospherics, rapid natural lighting changes, such as when the sun is shining down on the scene and then is quickly hidden by a passing cloud, can also impact the collection. The significant variation visible in the first image is likely due to this effect. But what about the others. You can see from the “structure” of the differences that it is not just a shifting of the image significant differences are up and to the left of edges in the upper left, but down and to the right on the lower/right part of the image. The “difference” patterns are more like localized zooms, probably caused by atmospheric lensing from thermals.

The impact of atmospherics and natural lighting changes on the far camera’s recognition rate is shown in Fig. 7.11. These differences are statistically significant. Note that to reduce the impact of pose and lighting variations, these images are using the exact same image on the display as in the recognition database; the only variations between the probe and the gallery are those caused by the imaging system.



**Fig. 7.11** On the *left*: variations over the time of day. On the *right*: recognition rank for various “quality” images

Recall that indoors at 15 ft the performance on this type of data is nearly perfect. Even with this very strong constraint, we see that at 182 ft on a clear and low-wind day, for Rank- $N$  recognition the performance of one of the best commercial algorithms of its day is below 65% with  $N < 4$  and still below 80% recognition rate even when  $N$  is 10. Again, these are averaged over hundreds of trials with 1024 images per trial, so this is not a sampling artifact.

A first guess might be that the weather impacts the raw “image quality,” which is determining the performance. We examined various measures of facial image quality and (to our surprise) many of the errors had nothing to do with human-perceived or measured image quality. While better quality images generally did do better, it was not as strongly related as one might hope. The right half of Fig. 7.11 shows some examples of the recognition rank (i.e., where the image ended up when probes are sorted by match score) for a collection of images from a “same image” experiment. Rank and image quality in this set were inversely correlated. A detailed discussion of problematic issues with quality is presented in the next section.

Our research set off to find the causes of this unexpectedly poor performance. After considerable investigation, we hypothesized that the poor performance was due in large part to error in localization of the eyes. In [18] we presented an analysis of this theory. To definitively show the cause we added registration markers within our photo-head data to allow us to transform the original eye coordinates to provide eye locations in the captured images. The graphs show the recognition performance (with error bars) for the off-the-shelf FaceIt algorithms and when used forced FaceIt to use the correct eye positions. The results on both cameras were statistically significant, and when the eyes are corrected the performance of both far and near cameras is similar. These results are, of course, highly optimistic because the data for correction are artificial calibration points and second this is self-matching, with the same image as probes and gallery so the near perfect recognition is to be expected. It is important to note that the “eye locations” being discussed are not just a question of where in the image the eyes appear but how that position related to where it should be in the image. In the “good-quality” images of Fig. 7.11, the corrected eye position is not in the middle of the eye! Atmospheric turbulence and lensing effect can distort the face image to point that to work properly the system needs to use a different eye position for its coordinate system and normalization procedures. Many of the computed eye locations were visibly off the eye, and the average difference between the computed and FaceIt eyes was 6 pixels.

## 7.4 In the End: Measuring Quality

As mentioned in the previous section, an obvious guess is that weather and atmospherics reduce the raw “image quality,” which is why they reduce performance. In order to study this potential impact, we formally defined quality and tried to study its relation to performance. We experimented with multiple measures of “quality,” including blur and contrast in various ways. We eventually defined a blind signal-to-noise ratio estimator for facial image quality, based on concepts from [28]. The



**Fig. 7.12** Window around eyes for various images qualities

concept is that statistical properties of edge images change with quality and have been shown to be correlated with underlying signal-to-noise ratios. In our experiments, our derived measure is, under general conditions, better correlated with recognition rates than the other quality measures examined.

To derive this measure, suppose the probability density function of an edge intensity image,  $\|\nabla I\|$ , is given by  $f \|\nabla I\|(\cdot)$  which is assumed to have mean  $\mu$ . The histogram of edge intensity image  $I$  can be modeled as a mixture of Rayleigh probability density functions, and that can be used to show that an estimate of the signal-to-noise ratio (SNR) is given by

$$QS = \int_{2\mu}^{\infty} f \|\nabla I\|(r) dr$$

It has been proven that the value of QS for a given noisy image is always smaller than the value of QS for that image with less noise. Zhang and Blum also show that it can estimate blur and is overall correlated with signal-to-noise ratio.

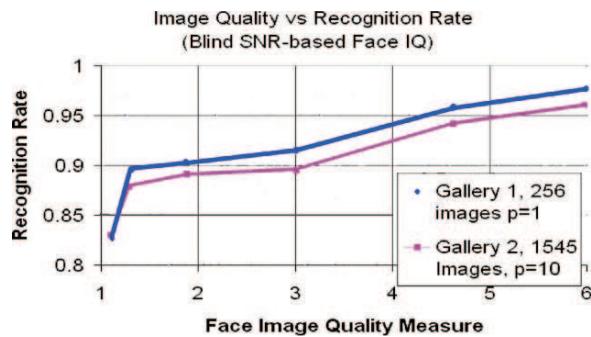
Choosing a fixed-sized window around the eyes (examples are shown in Fig. 7.12), we can define the Face SNR image quality as

$$Q' = \frac{\sum \text{edge above } 2\mu \text{'s pixels}}{\sum \text{edge pixels}} \simeq \int_{2\mu}^{\infty} f \|\nabla I\|(r) dr \quad (7.3)$$

This Face SNR IQ estimate, the ratio of number of pixels above twice the mean strength to the total number of edge pixels, is easily calculated and can be shown to be a good approximation to  $QS$ .

The results from this estimator are well correlated with recognition rate. We took the images and classified them into five bins using  $Q'$ , and then examined the recognition rate for each subset. Figure 7.13 shows the correlation between quality

**Fig. 7.13** In this plot, larger is better for quality. Correlations for blind SNR-based face image quality to recognition rate are 0.922 and 0.930. Experiments were also performed with multiple levels of blur, contrast, and multi-metric fusion. None were better than the blind SNR estimate



and recognition rate, with overall correlations of 0.922 and 0.930 for two different galleries of photo-head data. Beyond the Face SNR IQ estimate, we also performed experiments with multiple levels of blur, contrast, and multi-metric fusion – none were better than the blind SNR estimate. While at first this might seem significant, you should recall we were looking to understand/mitigate the impacts of atmospherics and wanted to use the quality predict, on a per image basis, if an image was going to be successful for quality. Unfortunately, a strong correlation was not sufficient for a good predictor.

We concluded that “quality” is indeed found in the recognition performance, not on what we “like” to imagine in some preconceived concept of quality or even our blind SNR estimates. Interestingly, recent NIST studies [2, 7] on quality assessment come to this same conclusion. For the iris work in [7], three different quality assessment algorithms lacked correlation in resulting recognition performance, indicating a lack of consensus on what image quality actually is. In the face recognition work [2], out-of-focus imagery was shown to produce better match scores.

We had already shown that on an individual image level both perceived and measured quality could be inversely related to rank, but also showed that quality was positively correlated with overall recognition scores. We are not alone in this observation. We note that more recently [2] showed that, on a per instance basis, what is visually of poor quality produced good recognition results; good, it was not sufficient for per image predictor. Reflecting upon this issue of quality a bit deeper, we began to wonder how to predict if an image would be successful and also how to compare different measures of “quality” for face recognition.

The concept is using some measure of the system to predict if a particular input image will be (or is) successfully classified by the system. That is, we could threshold on quality and say any quality less than 2 will fail. With such a model we can compare the usefulness of different image quality measures.

The question then is to measure the effectiveness of each predictor. Since this is measuring system performance, this then suggests that for a comparison of measures what is needed is some form of a receiver operator characteristic (ROC) analysis on the prediction/classification performance. In [12] and [21] we define four cases that can be used as the basis of such an analysis. Let us define the following:

1. “False Accept,” when the prediction is that the recognition system will succeed but the ground truth shows it will not. Type I error of the failure prediction and Type I or Type II error of the recognition system.
2. “False Reject,” when the prediction is that the recognition system will fail but the ground truth shows that it will be successful. Type II error of failure prediction.
3. “True Accept,” wherein the underlying recognition system and the prediction indicate that the match will be successful.
4. “True Reject,” when the prediction system predicts correctly that the system will fail. Type I or Type II error of the recognition system.

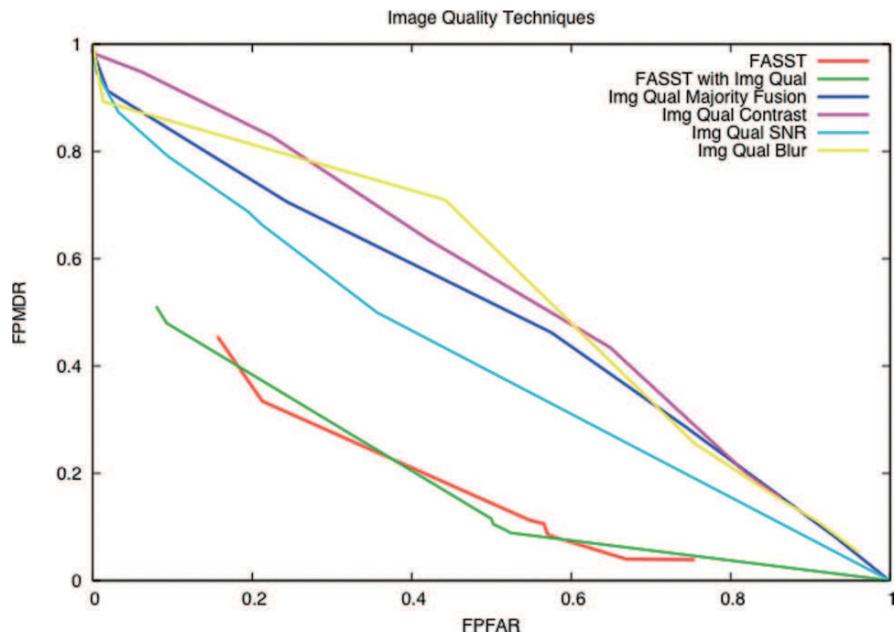
The two cases of most interest are Case 2 (system predicts they will not be recognized, but they are) and Case 1 (system predicts that they will be recognized but they are not). From these two cases we can define the failure prediction false accept

rate (FPFAR), and failure prediction miss detection rate (FPMDR) ( $= 1 - \text{FPFRR}$  (failure prediction false reject rate)) as.

$$\text{FPFAR} = \frac{|\text{Case2}|}{|\text{Case2}| + |\text{Case3}|} \quad (7.4)$$

$$\text{FPMDR} = \frac{|\text{Case1}|}{|\text{Case1}| + |\text{Case4}|} \quad (7.5)$$

With these definitions, the performance of the different reliability measures, and their induced classifier, can then be represented in a failure prediction receiver operating characteristic (FPROC) curve, of which an example is shown in Fig. 7.14. Implicitly, various thresholds are points along the curve and as the quality/performance threshold is varied, predictions of failure change the FPFAR and FPMDR just as changing the threshold in a biometric verification system varies the false accept rate and the miss detect rate (or false reject rate). High-quality data, which usually match better, will generally be toward the upper right, with low failure prediction false alarms (and lower failures overall), but when good-quality data do fail it is harder to predict it, so more are missed. Lowest quality data are usually toward the bottom right, with few missed failure predictions, but more false predictions, as poor quality more often results in marginal but correct matches.



**Fig. 7.14** FPROC for four different image quality techniques on 12,000 images, compared with the post-recognition failure analysis from similarity surface theory (FASST) technique, with and without image quality as a feature dimension

The advantage of using the FPROC curve as opposed to simple CMC or ROC curves with the data segmented by quality (or any other predictor variable) is twofold: First it allows for a more direct comparison of different measures on the same population, or the same quality measure on different sensors/groups. Second, segmentation of data to generated CMC/ROC curves inflates the measure since it means the quality  $i$  data are not interacting with quality  $j$  data. Furthermore, it is not practical to compare measures or sensors when each one generates multiple ROC curves, especially if trying to compare multiple different “quality” measures. The FPROC evaluation approach allows us to vary the quality threshold over the gallery and see how it impacts prediction, while still maintaining a mixed gallery of qualities. The FPROC curve requires an “evaluation” gallery and depends on the underlying recognition system’s tuning, sensors, and decision-making process.

The impact of switching approaches from a standard multiple CMC/ROC evaluation of image quality the FPROC representation is noted in Fig. 7.14, where three different image quality techniques and a simple image quality fusion scheme are plotted. The underlying data are 12,000 images obtained in varied weather conditions outdoors. As can be seen, while our Face SNR estimate outperforms the other quality measures in prediction, none of the image quality techniques are very powerful at predicting failure. Thus, while image quality is well correlated with recognition overall, it can fare poorly on a per image basis where significant pose, lighting, contrast, and compression are allowed, in essence, any unconstrained setting where data collection is taking place.

Early on in our “quality” analysis, we introduced a compelling alternative approach [12], which was to learn to predict when a system fails and when it succeeds and classify individual recognition instances using the learning as a basis. Based on the decisions made by a machine-learning classification system, a failure prediction receiver operator characteristic curve can be plotted, allowing the system operator to vary a quality threshold in a meaningful way. Failure prediction analysis of this sort has been shown to be quite effective for single modalities [12], fusion across sensors for a single modality [26], and across different machine-learning techniques [19, 21]. The FPROC quality prediction results of [12] are compared with basic image quality predictions in Fig. 7.14 and are clearly significantly better.

Since the early observation on image quality, we have continued to build the alternative approach in the form of post-recognition analysis of the recognition score distributions. We call this analysis *Failure Analysis from Similarity Surface Theory*. Let  $S$  be an  $n$ -dimensional similarity surface composed of  $k$ -dimensional feature data computed from similarity scores. The surface  $S$  can be parameterized by  $n$  different characteristics and the features may be from matching data, non-matching data, or a mixed set of both.

**Similarity Surface Theorem 7.4.1** *For a recognition system, there exists a similarity surface  $S$ , such that surface analysis around a hypothesized “match” can be used to predict failure of that hypothesis with high accuracy.*

While the (empirical) Similarity Surface Theorem 7.4.1 suggests that shape analysis should predict failure, the details of the shapes and their potential for prediction are unknown functions of the data space. Because of the nature of biometric spaces, the similarity surface often contains features at multiple scales caused by matching with sub-clusters of related data (for example, multiple samples from the same individual over time, from family members, or from people in similar demographic populations). What might be “peaked” in a low-noise system, where the inter-subject variations are small compared to intra-subject variations, might be flat in a system with significant inter-subject variations and a large population. These variations are functions of the underlying population, the biometric algorithms, and the collection system. Thus, with Theorem 7.4.1 as a basis, the system “learns” the appropriate similarity shape information for a particular system installation. We have applied the FASST technique to a variety of different data sets, with implementations utilizing different learning techniques and underlying features generated from the recognition scores [12, 19, 21, 26]. Even if we cannot get good predictors from just image face quality data, the “quality” of face data for recognition *can* be learned from the distribution of scores after matching. Further, we have demonstrated a multi-modal fusion approach [20] for this sort of failure prediction, which is able to enhance recognition performance beyond the best-performing multi-modal fusion algorithms.

## 7.5 Conclusions

In this chapter, we looked at the issues in image acquisition that must be considered for effective long-range facial recognition. As we have seen, both obvious and very non-obvious issues arise in all aspects of the image acquisition process. We discussed working volume and resolution issues that designer must consider. Lighting is always a challenge for outdoor acquisition, and problems multiply in low-light conditions. We have had good success measuring “face luminance” as opposed to scene lux or illuminance. Further, a megapixel EMCCD sensor with high resolution has provided us with images of sufficient quality and spatial resolution for standard face recognition, which overcomes many of the problems faced by LWIR systems. In general, with today’s technology, cheaper components (low resolution, interlaced sensors, rolling shutters, cheap lenses) will often hurt performance, in spite of their bargain price tag.

Designing a long-range facial recognition system requires extensive testing for validation. Our photo-head setup provided much insight into the effects of weather and atmospherics on long-range data acquisition. Not only did we learn of the impact on raw recognition scores, but also the limitations of image quality, which has been a traditional indicator of performance. Our observations have led us to define a new paradigm for image assessment based on post-recognition score analysis. We believe that this post-recognition analysis is a critical component to enhance performance, along with proper equipment selection and system design.

**Acknowledgments** This work was supported in part by the DARPA HID program, ONR contract #N00014-00-1-0388, by NSF PFI Award # 0650251, DHS SBIR NBCHC080054, ONR STTR N00014-07-M-0421, and ONR MURI N00014-08-1-0638.

## Proposed Questions and Exercises

- Using the standard formulas for illumination and luminance (or irradiance and radiance, your choice), sketch out the steps needed to determine the amount of light reaching the sensor for a face that is 100-m away from the sensor. Using this determine if a 2-Megapixel camera, with  $11\text{-}\mu\text{m}$  pixels, fitted with a Cannon EOS 400-mm f2.8 lens with a 2x adapter could operate at different scene light levels.
- Derive a formula for the necessary focal length for long-range face recognition, at distance  $d$ , using a sensor with  $1280 \times 1024$  pixels each of  $p$  microns across. State any assumptions you need to make along the way.
- Derive a formula for the operational volume, both width of the FOV and depth of field, for long-range face recognition, at distance  $d$ , using a sensor with  $1280 \times 1024$  pixels each of  $p$  microns across with a Cannon EOS 400-mm lens with a 2x adapter set at maximum aperture.
- For a camera at 200 m from a subject of interest, who is exiting a door, what is the necessary sensor size to have at least 3 s of video with sufficient resolution for face recognition (and temporal fusion). It is useful to ask for 3 s to ensure some frames at the top of the gate where there is minimal motion blur. State any assumptions you need to make along the way.
- For subjects walking at normal speed, determine the shutter speed to ensure their walking does not produce more than a 0.5 pixel motion blur.
- Consider the design of the photo-head experiment. List four limitations of the experimental design and suggest alternative designs that overcome these limitations (while on a university/student budget :-)).
- In the definition of Face SNR IQ, we constrained it to a narrow region around the eyes and nose. Discuss the advantages and disadvantages of this windowing.

## References

1. Adini, Y., Moses, Y. and Ullman, S.: Face Recognition: The Problem of Compensating for Changes in Illumination Direction. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **19** (1997), no. 7, 721–732.
2. Beveridge, R.: Face Recognition Vendor Test 2006 Experiment 4 Covariate Study. Presentation at the NIST MBGC Kick-off Workshop (2008).
3. Canon: Optical Terminology. The EF Lens Work III, Canon Inc., Lens Products Group, 2006, 192–216.
4. Chen, X., Flynn, P.J. and Bowyer, K.W.: IR and Visible Light Face Recognition. *Computer Image and Vision Understanding* **99** (2005), no. 3, 332–358.
5. Chen, T., Yin, W., Zhou, X., Comaniciu, D. and Huang T.: Total Variation Models for Variable Lighting Face Recognition. *IEEE Trans. Pattern Analysis Machine Intelligence* **28** (2006), no. 9, 1519–1524.

6. Dou, M.S., Zhang, C., Hao, P.W. and Li, J.: Converting Thermal Infrared Face Images into Normal Gray-Level Images. The 2007 Asian Conference on Computer Vision, 2007, II: 722–732.
7. Flynn, P.: ICE Mining: Quality and Demographic Investigations of ICE 2006 Performance Results. Presentation at the NIST MBGC Kick-off Workshop (2008).
8. Georgiades, A.S., Kriegman, D.J. and Belhumeur, P.N.: Illumination Cones for Recognition under Variable Lighting: Faces. Proc. of 1998 IEEE Conf. on Computer Vision and Pattern Recognition, 1998, 52–58.
9. Hoist, G.C.: CCD Array, Cameras, and Displays. SPIE Optical Engineering, Bellingham 1996.
10. Jacobs, D.W., Belhumeur, P.N. and Basri, R.: Comparing Images Under Variable Illumination. Proc. of 1998 IEEE Conf. on Computer Vision and Pattern Recognition, 1998, 610–617.
11. Kong, S.G., Heo, J., Abidi, B.R., Paik, J.K. and Abidi, M.A.: Recent Advances in Visual and Infrared Face Recognition: A Review. Computer Vision and Image Understanding **97** (2005), no. 1, 103–135.
12. Li, W., Gao, X. and Boult, T.: Predicting Biometric System Failure. Proc. of the IEEE Conference on Computational Intelligence for Homeland Security and Personal Safety (CIHSPS 2005), 2005.
13. Micheals, R. and Boult, T.: Efficient evaluation of classification and recognition systems. Proc. of 2001 IEEE Conf. on Computer Vision and Pattern Recognition, 2001, I: 50:57.
14. Marasco, P. and Task, H.: The Impact of Target Luminance and Radiance on Night Vision Device Visual Performance Testing. Helmet- and Head-Mounted Displays VIII: Technologies and Applications. Edited by Rash, Clarence E.; Reese, Colin E. Proceedings of the SPIE, **5079** (2003), 174–183 .
15. Narasimhan, S. and Nayar, S.: Contrast Restoration of Weather Degraded Images. IEEE Trans. on Pattern Analysis and Machine Intelligence **25** (2003), no. 6, 713–724.
16. Phillips, P.J., Grother, P., Micheals, R., Blackburn, D., Tabassi, E. and Bone, M.: Face Recognition Vendor Test 2002 (FRVT 2002). National Institute of Standards and Technology, NISTIR 6965, 2003.
17. Phillips, P.J. and Vardi, Y.: Efficient Illumination Normalization of Facial Images. Pattern Recognition Letters **17** (1996), no. 8, 921–927.
18. Riopka, T. and Boult, T.: The Eyes Have It. ACM Biometrics Methods and Applications Workshop, 2003, 33–40.
19. Riopka, T. and Boult, T.: Classification Enhancement via Biometric Pattern Perturbation. IAPR Conference on Audio- and Video-based Biometric Person Authentication (Springer Lecture Notes in Computer Science) **3546** (2005), 850–859.
20. Scheirer, W. and Boult, T.: A Fusion Based Approach to Enhancing Multi-Modal Biometric Recognition System Failure and Overall Performance. In Proc. of the Second IEEE Conference on Biometrics: Theory, Applications, and Systems, 2008.
21. Scheirer, W., Bendale, A. and Boult, T.: Predicting Biometric Facial Recognition Failure With Similarity Surfaces and Support Vector Machines. In Proc. of the IEEE Computer Society Workshop on Biometrics, 2008.
22. Socolinsky, D., Wolff, L. and Lundberg, A.: Image Intensification for Low-Light Face Recognition. In Proc. of the IEEE Computer Society Workshop on Biometrics, 2006.
23. Socolinsky, D., Wolff, L., Neuheisel, J. and Eveland, C.: Illumination Invariant Face Recognition Using Thermal Infrared Imagery. Proc. of 2001 IEEE Conf. on Computer Vision and Pattern Recognition, 2001, I: 527:534.
24. Vogelsong, T., Boult, T., Gardner, D., Woodworth, R., Johnson, R. C. and Hefflin, B.: 24/7 Security System: 60-FPS Color EMCCD Camera With Integral Human Recognition. Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense VI. Edited by Carapezza, Edward M. Proceedings of the SPIE **6538** (2007), 65381S.

25. Wilder, J., Phillips, P.J., Jiang, C. and Wiener, S.: Comparison of Visible and Infra-red Imagery for Face Recognition. Proc. of the IEEE Conf. on Automated Face and Gesture Recognition, 1996, 182–187.
26. Xie, B., Boult, T., Ramesh, V. and Zhu, Y.: Multi-Camera Face Recognition by Reliability-Based Selection. Proc. of the IEEE Conference on Computational Intelligence for Homeland Security and Personal Safety (CIHSPS 2006), 2006.
27. Yitzhaky, Y., Dror, I. and Kopeika, N.: Restoration of Atmospherically Blurred Images According to Weather Predicted Atmospheric Modulation Transfer Function (MTF). Optical Engineering **36** (1997), no. 11.
28. Zhang, Z. and Blum, R.: On Estimating the Quality of Noisy Images. Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 1998, 2897–2900.
29. Zhao, W. and Chellappa, R.: Illumination-Insensitive Face Recognition using Symmetric Shape-from-Shading. Proc. of 2000 IEEE Conf. on Computer Vision and Pattern Recognition, 2000, I: 286–293.

# **Chapter 8**

## **A Review of Video-Based Face Recognition Algorithms**

**Rama Chellappa, Manuele Bicego, and Pavan Turaga**

**Abstract** Traditional face recognition systems have relied on a gallery of still images for learning and a probe of still images for recognition. While the advantage of using motion information in face videos has been widely recognized, computational models for video-based face recognition have only recently gained attention. This chapter reviews some recent advances in this novel framework. In particular, the utility of videos in enhancing performance of image-based tasks (such as recognition or localization) will be summarized. Subsequently, spatiotemporal video-based face recognition systems based on particle filters, hidden Markov models, and system theoretic approaches will be presented. Further, some useful face databases employable by researchers interested in this field will be described. Finally, some open research issues will be proposed and discussed.

### **8.1 Introduction**

Faces are articulating three-dimensional (3D) objects. Recognition of objects usually relies on estimates of the 3D structure that is inferred from observing the object. Extraction of 3D structure can be performed in various ways including the use of stereo, shading, and motion. Algorithms for estimating the structure of objects from physical constraints on the measurements induced by stereo, shading, and motion are well studied in computer vision and are broadly classified as SfX algorithms (structure/shape from X, where X stands for stereo, shading, motion, etc.). In the case of faces, motion has special significance since it encodes far richer information than simply the 3D structure of the face. This extra information is in the form of behavioral cues such as idiosyncratic head movements and gestures which can potentially aid in recognition tasks. Faces are thus characterized by both 3D structure and dynamic information.

---

R. Chellappa (✉)

Department of Electrical and Computer Engineering and Center for Automation Research,  
UMIACS University of Maryland, College Park, MD 20742, USA  
e-mail: Rama@umiacs.umd.edu

Psychophysical studies in human recognition of faces [32] have found evidence that when both structure information and dynamics information are available, humans tend to rely more on dynamics under non-optimal viewing conditions (such as low spatial resolution and harsh illumination conditions). Dynamics also aids in recognition of familiar faces [36]. In the psychology literature, there are two hypotheses that attempt to explain the role of motion in face recognition – supplemental information hypothesis and the representation enhancement hypothesis [32]. According to the supplemental information hypothesis, humans represent characteristic facial motion and gestures in addition to the 3D structure, i.e., humans use two distinct sources of information to recognize faces – structure and dynamics. On the other hand, the representation enhancement hypothesis suggests that the motion of faces is used primarily to refine the estimates of the 3D structure of a face.

Traditional face recognition systems have relied on a gallery of still images for learning and a probe of still images for recognition. While the advantage of using motion information in face videos has been widely recognized, computational models for video-based face recognition have only recently gained attention.

The rest of the chapter is organized as follows. In Section 8.2, we will first describe the utility of videos in enhancing performance of image-based tasks (such as recognition or localization). In Section 8.3, we discuss spatiotemporal video-based face recognition systems based on particle filters, hidden Markov models, and system theoretic approaches. In Section 8.4, we suggest some useful face databases employable by researchers interested in this field. Finally, in Section 8.5, we discuss open research issues.

## 8.2 Utility of Video in Enhancing Performance of Image-Based Task

### 8.2.1 Enhancing Performance of 2D Matchers

An immediate possible utilization of temporal information for video-based face recognition is to fuse the results obtained by a two-dimensional (2D) face classifier on each frame of the sequence. The video sequence can be seen as an unordered set of images to be used for both training and testing phases. During testing one can use the sequence as a set of probes, each of them providing a clue on the identity of the person. Appropriate fusion techniques can then be applied to provide the final identity. Following the extensive work in the field of multi-classifier systems (MCS – see, e.g., [21]) – with its direct application to biometrics known as multimodal biometrics [37] – we observe that fusion can be performed at three different levels: feature level, score level, and decision level. In the first case, the features extracted from each frame are combined together to form a ‘super feature vector.’ This is typically achieved by simple feature concatenation. It should be noted that this approach is not suited to the current problem for the following reasons: first, concatenation collapses each video into a single vector, thus losing the appealing redundancy and abundance of video data; second, concatenation produces feature vectors of different lengths when the number of frames is not the same in different video sequences.

For this reason, fusion at the score or at the decision level is more appropriate. In the score-level case, the matching scores obtained at each frame are combined in order to obtain a single final score. Simple rules such as sum, product, maximum, and minimum have been shown to be quite useful, as well as more sophisticated supervised techniques (which require more data in order to learn how to combine the scores). For video-based face recognition, all scores are typically generated by a single face recognizer; hence the problem of score normalization (making all scores in a comparable range) is avoided. Finally, the decision-level fusion brings frame-level decisions (e.g., authenticated/not authenticated) into a final decision. Examples in this context are the voting approach and the more complex Borda count and Highest Rank approach [16]. This scenario could be of course enhanced by using multiple 2D classifiers or different classifiers for different parts of the video.

### 8.2.2 Enhancing Facial Appearance Models

Most face recognition approaches rely on a model of appearance for each individual subject. The simplest appearance model is simply a static image of the person. Such appearance models are rather limited in utility in video-based face recognition where subjects may be imaged under varying viewpoints, illuminations, expressions, etc. Thus, instead of using a static image as an appearance model, a sufficiently long video which encompasses several variations in facial appearance can lend itself to building more robust appearance models.

Several methods have been proposed for extracting more descriptive appearance models from videos. In their work [25], Lee et al. consider a facial video as a sequence of images sampled from an ‘appearance manifold.’ In principle, the appearance manifold of a subject contains all possible appearances of the subject. In practice, the appearance manifold for each person is estimated from training data of videos. For ease of estimation, the appearance manifold is considered to be a collection of affine subspaces, where each subspace encodes a set of similar appearances of the subject. Temporal variation of appearance in a given video sequence is then modeled as transitions between the appearance subspaces. This method is robust to large appearance changes if sufficient pose and illumination variations are present in the training set. Further, the tracking problem can be integrated into this framework by searching for a bounding box on the test image that minimizes the distance of the cropped region to the learnt appearance manifold.

In a related work, Arandjelovic and Cipolla [4] propose to explain the appearance variations due to shape and illumination on human faces. They make the assumption that the ‘shape–illumination manifold’ of all possible illuminations and head poses is generic for human faces. This means that the shape–illumination manifold can be estimated using a set of subjects exclusive of the test set. They show that the effects of face shape and illumination can be learnt using probabilistic principal component analysis (PCA) from a small, unlabeled set of video sequences of faces in randomly varying lighting conditions. Given a novel sequence, the learnt model is used to decompose the face appearance manifold into albedo and shape–illumination manifolds, producing the classification decision using robust likelihood estimation.

### 8.2.3 Enhancing Face Localization Performance

In a generic face recognition setting, given a test video of a moving face the first step is to track the facial features across all the frames of the video. From the tracked features, one can extract a few key frames which can be used for matching with the exemplars in the gallery.

There has been significant work on facial tracking using 2D appearance-based models (cf. [15, 23, 45]). The 2D approaches do not provide the 3D configuration of the head, hence are not robust to large changes in pose or viewpoint. Recently, several methods have been developed for 3D face tracking. A closed-loop approach to estimate 3D structure using a structure from motion algorithm was proposed in [18]. A cylindrical face model for face tracking has been used in [7]. In their formulation, it is assumed that the inter-frame warping function is locally linear and that the inter-frame pose change occurs only in one of the six degrees of freedom of the rigid cylindrical model. As an extension to this approach, a method was presented in [2] which does not use information about the camera calibration parameters but uses particle filters for state estimation. We shall first provide a brief review of the particle filtering methodology and then show how it is adapted to this task.

#### 8.2.3.1 Particle Filtering for Dynamical Inference

We will assume that a certain feature representation for spatiotemporal patterns of moving faces has been made. We will also assume that there exists a set of hidden parameters, constituting the state vector, which govern how the spatiotemporal patterns evolve in time. The state vector encodes information such as motion parameters that can be used for tracking and identity parameters that can be used for recognition. Given a set of features, we need inference algorithms for estimating these hidden parameters. When the state observation description of the system is linear and Gaussian, the parameters can be estimated using the Kalman filter. But the design of the Kalman filter becomes complicated for intrinsically non-linear problems and is not suited for estimating posterior densities that are non-Gaussian. Particle filtering [12, 13, 27] is a method for estimating arbitrary posterior densities by representing them by a set of weighted particles. We will first describe the state estimation problem and then show how particle filtering can be used to solve video-based face recognition.

#### 8.2.3.2 Problem Statement

Consider a system with parameters  $\theta$ . We assume that the system parameters evolve in time according to dynamics given by  $F_t(\theta, D, N)$ .

*System Dynamics:*

$$\theta_t = F_t(\theta_{t-1}, D_t, N_t) \quad (8.1)$$

where  $N$  is the process noise in the system. The auxiliary variable  $D$  indexes the set of motion models or behavior exhibited by the object and is usually omitted in typical tracking applications. This auxiliary variable assumes importance

in problems such as activity recognition or behavioral analysis. Each frame of the video contains pixel intensities which act as observations  $Z$ .

*Observation Equation:*

$$Z_t = G(\theta_t, I, W_t) \quad (8.2)$$

where  $W$  represents the observation noise. The auxiliary variable  $I$  indexes the various object classes being modeled.

The problem now is to obtain estimates of the hidden state parameters recursively as and when the observations become available. Quantitatively, we are interested in estimating the posterior density of the state parameters given the observations, i.e.,  $P(\theta_t | Z_{0:t})$ .

The particle filter approximates the desired posterior probability density function (pdf)  $P(\theta_t | Z_{0:t})$  by a set of weighted particles  $\{\theta_t^{(j)}, w_t^{(j)}\}_{j=1}^M$ , where  $M$  denotes the number of particles. Here, we shall briefly discuss the method of sequential importance sampling (SIS) for estimating the state posterior density.

### 8.2.3.3 Inference by Sequential Importance Sampling

Consider a general time series state-space model fully determined by (i) the overall state transition probability  $p(\theta_t | \theta_{t-1})$ , (ii) the observation likelihood  $p(Z_t | \theta_t)$ , and (iii) prior probability  $p(\theta_0 | Z_0)$ . We also assume statistical independence among all noise variables. We wish to compute the posterior probability  $p(\theta_t | Z_{0:t})$ . If the model is linear with Gaussian noise, it is analytically solvable by a Kalman filter which propagates the mean and covariance of a Gaussian distribution over time. For non-linear and non-Gaussian cases, an extended Kalman filter (EKF) and its variants have been used to arrive at an approximate analytic solution [3]. Recently, the SIS technique, a special case of Monte Carlo method [11, 17, 20, 27], has been used to provide a numerical solution and propagate an arbitrary distribution over time.

The essence of the Monte Carlo method is to approximately represent an arbitrary probability distribution  $\pi(\theta)$  closely by a set of discrete samples. Ideally, one would like to draw i.i.d. samples  $\{\theta^{(m)}\}_{m=1}^M$  from  $\pi(\theta)$ . However, this is often difficult to implement, especially for nontrivial distributions. Instead, a set of samples  $\{\theta^{(m)}\}_{m=1}^M$  is drawn from an importance function  $g(\theta)$  which is easy to sample from, and a weight  $w^{(m)} = \pi(\theta^{(m)})/g(\theta^{(m)})$  is assigned to each sample. This technique is called importance sampling (IS). It can be shown [27] that the importance sample set  $S = (\theta^{(m)}, w^{(m)})$  is properly weighted to the target distribution  $\pi(\theta)$ . To accommodate a video, importance sampling is used in a sequential fashion, which leads to SIS. SIS propagates according to the sequential importance function, say  $g(\theta_t | \theta_{t-1})$  and calculates the weight using

$$w_t = w_{t-1} p(Z_t | \theta_t) p(\theta_t | \theta_{t-1}) / g(\theta_t | \theta_{t-1}) \quad (8.3)$$

In the CONDENSATION algorithm [17],  $g(\theta_t | \theta_{t-1})$  is taken to be  $p(\theta_t | \theta_{t-1})$  and (8.3) becomes

$$w_t = w_{t-1} p(Z_t | \theta_t) \quad (8.4)$$

In practice, (8.4) is implemented by first re-sampling the sample set  $S_{t-1}$  according to  $w_{t-1}$  and then updating the weight  $w_t$  using  $p(Z_t|\theta_t)$ . The interested reader is referred to [12, 13, 27] for a complete treatment of particle filtering. The state estimate  $\hat{\theta}_t$  can be recovered from the pdf as the maximum likelihood (ML) estimate or the minimum mean squared error (MMSE) estimate or any other suitable estimate based on the probability density function.

#### 8.2.3.4 Tracking Faces in Uncalibrated Videos

A method to recover the 3D configuration of faces in uncalibrated videos using the particle filtering framework is now described [2]. The 3D configuration consists of the three translational parameters and the three orientation parameters which correspond to the yaw, pitch, and roll of the face. The approach combines the structural advantages of geometric modeling with the statistical advantages of a particle filter-based inference. The face is modeled as the curved surface of a cylinder which is free to translate and rotate arbitrarily. The geometric modeling takes care of pose and self-occlusion while statistical modeling handles moderate occlusion and illumination variations.

#### 8.2.3.5 Tracking Framework

Once the structural model and feature vector have been chosen, the goal is to estimate the configuration (or pose) of the moving face in each frame of a given video. This can be viewed as a recursive state estimation problem. Particle filtering can now be used for estimating the unknown state  $\theta$  of the system from a sequence of noisy observations  $Z_{1:t}$ .

First, specific forms for the hidden dynamics and the observation equations (8.1) and (8.2) need to be chosen. Without making any assumptions about the nature of the dynamics such as constant velocity or constant acceleration, a random-walk motion model can be used:

$$\theta_t = \theta_{t-1} + N_t \quad (8.5)$$

where  $N_t$  is normally distributed with zero mean. Based on domain knowledge, one can impose more meaningful motion models that will require fewer particles for estimating the posterior density.

The observation model involves the feature vector and can be written as

$$Z_t = \Gamma(y_t, \theta_t) = F_t + W_t \quad (8.6)$$

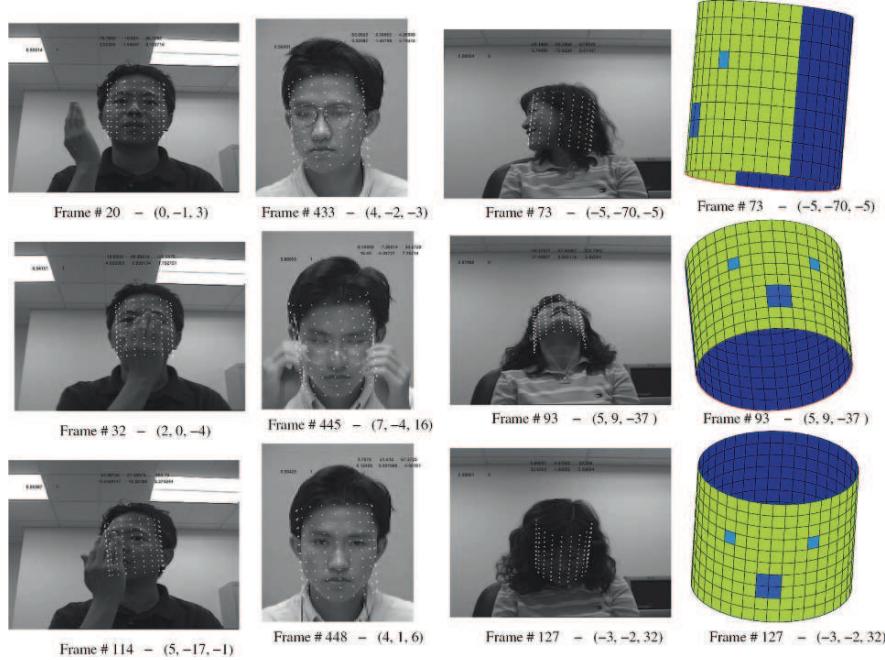
where  $y_t$  is the current frame (the gray scale image) and  $\Gamma$  is the mapping that computes the feature vector given an image  $y_t$  and a configuration  $\theta_t$ .  $Z_t$  is the computed feature vector and  $F_t$  is the feature model. The feature model is used to compute the likelihood of the particles (which correspond to the various proposed configurations of the face). For each particle, the likelihood is computed using the

average sum of squared difference (SSD) between the feature model and the mean vector  $Z_t$  corresponding to the particle. The feature model can be a fixed template, or one can use a dynamic template such as  $F_t = \hat{Z}_{t-1}$ . The fixed template  $F_t = F_0$  is referred to as the static model, while the dynamic template  $F_t = \hat{Z}_{t-1}$  is referred to as the wander model. The wander model is capable of handling appearance changes due to illumination, expression, etc., as the face translates/rotates in the real world, while the static model is resistant against drifts. One can use both the models simultaneously and compute the final likelihood of a particle as the maximum of the likelihoods using the static and the wander models. This provides the capability to handle appearance changes and to correct the estimation if the wander model drifts.

### 8.2.3.6 Experiments

#### Tracking Under Extreme Poses

Here, results of tracking experiments on three data sets (Honda/UCSD data set [24], BU data set [7], and Li data set [26]) are shown. These data sets have numerous sequences in which there are significant illumination changes, expression variations, and people are free to move their heads arbitrarily. Figure 8.1 shows a few frames

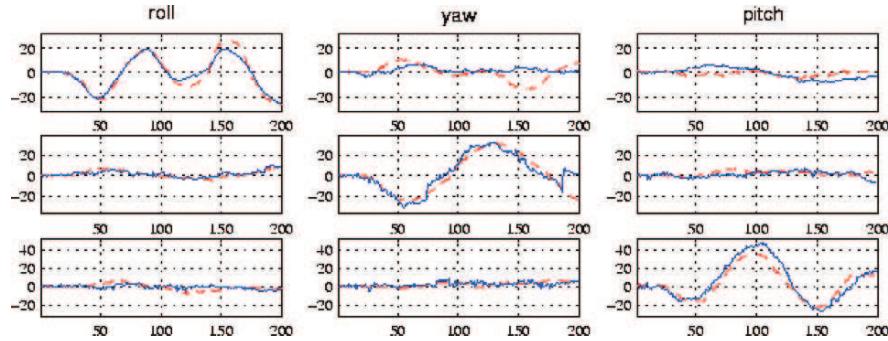


**Fig. 8.1** Tracking results under severe occlusion, extreme poses, and different illumination conditions. The cylindrical grid is overlaid on the image plane to display the results. The 3-tuple shows the estimated orientation (roll, yaw, pitch). The last column shows the cylindrical model in the pose estimated for the sequence in the third column (courtesy [2])

from three videos with grid points on the estimated cylinders overlaid on the image frame. The first and second columns show the robustness of the tracker against considerable occlusion. The tracker does well even when confronted with extreme poses as shown in the third column. Moderate expressions do not affect the feature since it is the mean intensity within a small surface patch on the face. The tracker is able to maintain the track all along the sequences.

#### Ground Truth Comparison

The BU data set [7] provides ground truth for the pose of the face in each frame. Results of a tracking experiment on the BU data set and comparison of the estimates of yaw, pitch, and roll with the ground truth are shown in Fig. 8.2. It can be seen that the tracker accurately estimates the pose of the face in most frames.



**Fig. 8.2** Each row shows the three orientation parameters. The *red/dashed* curve depicts the ground truth while the *blue/solid* curve depicts the estimated values (courtesy [2])

## 8.3 Spatiotemporal Face Recognition Systems

### 8.3.1 Joint Tracking and Recognition Using Particle Filters

Temporal information in videos can be exploited for simultaneous tracking and recognition of faces without the need to perform these tasks in a sequential manner. There are several advantages in performing these tasks simultaneously. In the *tracking-then-recognition* framework, estimation of parameters for registration between a test face and a template face, compensating for appearance variations due to changes in viewpoint, illumination, etc., and voting on the individual recognition results might be an ad hoc solution. By effectively exploiting temporal information, the *tracking-and-recognition* framework performs all these steps in an integrated manner.

Temporal information was exploited in [26] to design a facial feature tracking algorithm which could be used in two modes – pure tracking mode and tracking for verification mode. In the pure tracking mode, the tracker is initialized by computing a set of features based on Gabor jets from the first frame. However, the proposed

parameterization also allows features from a given template face to be tracked in a given sequence (even though the test and template may belong to different subjects). Moreover, the authors note that when the template and the test sequence belong to the same person, the tracked features exhibit coherent motion patterns. On the other hand, if the template and the test sequence belong to different persons, no coherent motion patterns are seen thus providing a cue for recognizing the person.

In this section, a framework for joint recognition and tracking of human faces using a gallery of still or video images and a probe set of videos is presented using the particle filter framework [44]. In still-to-video recognition where the gallery consists of still images, a time series state-space model is proposed to fuse temporal information in a probe video, which simultaneously characterizes the kinematics and identity using a motion vector and an identity variable, respectively. The joint posterior distribution of the motion vector and the identity variable is estimated at each time instant and then propagated to the next time instant. Marginalization over the motion vector yields a robust estimate of the posterior distribution of the identity variable. In this technique motion is modeled using a rigid deformation of a face template. Hence, this method does not explicitly model the behavioral aspects of faces.

### 8.3.1.1 Joint Motion and Identity Model

For the joint motion and identity model, the hidden state vector consists of two components – the motion component  $m$  and the identity component  $n$ . Thus, the overall state vector becomes  $\theta = (m, n)$ . Each component is governed by a different model as follows.

#### *Motion Equation*

$$m_t = m_{t-1} + w_t \quad (8.7)$$

The choice of  $m_t$  is application dependent. Affine motion parameters are often used when there is no significant pose variation available in the video sequence. However, if a 3D face model is chosen, 3D motion parameters should be used accordingly as discussed in Section 2.3.

#### *Identity Equation*

$$n_t = n_{t-1} \quad (8.8)$$

assuming that the identity does not change as time proceeds.

#### *Observation Equation*

By assuming that the transformed observation is a noise-corrupted version of a still template in the gallery, the observation equation can be written as

$$\mathbf{T}_{m_t} \{Z_t\} = I_{n_t} + v_t \quad (8.9)$$

where  $v_t$  is observation noise at time  $t$ , whose distribution determines the observation likelihood  $p(Z_t|m_t, n_t)$ , and  $\mathbf{T}_{m_t} \{Z_t\}$  is a transformed version of the observation  $Z_t$ .

We assume statistical independence between all noise variables and prior knowledge on the distributions  $p(m_0|Z_0)$  and  $p(n_0|Z_0)$ . Recalling that the overall state vector is  $\theta_t = (m_t, n_t)$ , (8.7) and (8.8) can be combined into a single state equation which is completely described by the overall state transition probability

$$p(\theta_t|\theta_{t-1}) = p(m_t|m_{t-1})p(n_t|n_{t-1})$$

For recognition, the goal is to compute the posterior probability  $p(n_t|Z_{0:t})$ , which is in fact a marginal of the overall state posterior probability  $p(\theta_t|Z_{0:t}) = p(m_t, n_t|Z_{0:t})$ , where the marginalization is performed over the motion parameters. These required posteriors can be computed using the particle filtering framework based on the SIS algorithm. Further, by exploiting the fact that the identity variable can only take values in a finite index set (corresponding to the number of distinct humans in the gallery), more efficient inference algorithms can be devised than direct application of SIS. We refer the reader to [44] for more detail on how this is done.

### 8.3.1.2 Experiments

In this section, we describe a still-to-video recognition scenario where the gallery consists of still images and the probe is a video. A few frames from tracking results for a probe video are shown in Fig. 8.3.

In Fig. 8.4, we show how the posterior probability of identity of a person changes over time. Suppose the correct identity for Fig. 8.3 is  $l$ . From Fig. 8.4, we observe that the posterior probability  $p(n_t = l|Z_{0:t})$  increases as time proceeds and eventually approaches 1, and all others  $p(n_t = j|Z_{0:t})$ ,  $j \neq l$  go to 0.

## 8.3.2 Hidden Markov Models: Basic Models, Advanced Models

The temporal and motion information is an important cue for video-based recognition. The hidden Markov model (HMM) has been successfully applied to model temporal information for applications such as speech recognition, gesture recognition, expression recognition, and many others. Due to its importance in the field of video-based face recognition, we will briefly review its fundamentals.

A discrete-time first-order HMM [35] is a stochastic finite state machine defined over a set of  $K$  states  $S = \{S_1, S_2, \dots, S_K\}$ . The states are hidden, i.e., not directly observable. Each state has an associated probability density function encoding the probability of observing a certain symbol being output from that state.

Let  $Q = (Q_1, Q_2, \dots, Q_T)$  be a fixed state sequence of length  $T$  with the corresponding observations  $O = (O_1, O_2, \dots, O_T)$ . An HMM is described by a model  $\lambda$ , determined by a triple  $(A, B, \pi)$  such that

$A = (a_{ij})$  is a matrix of transition probabilities, in which  $a_{ij} = P(Q_t = S_j | Q_{t-1} = S_i)$  denotes the probability of state  $S_j$  following state  $S_i$ .

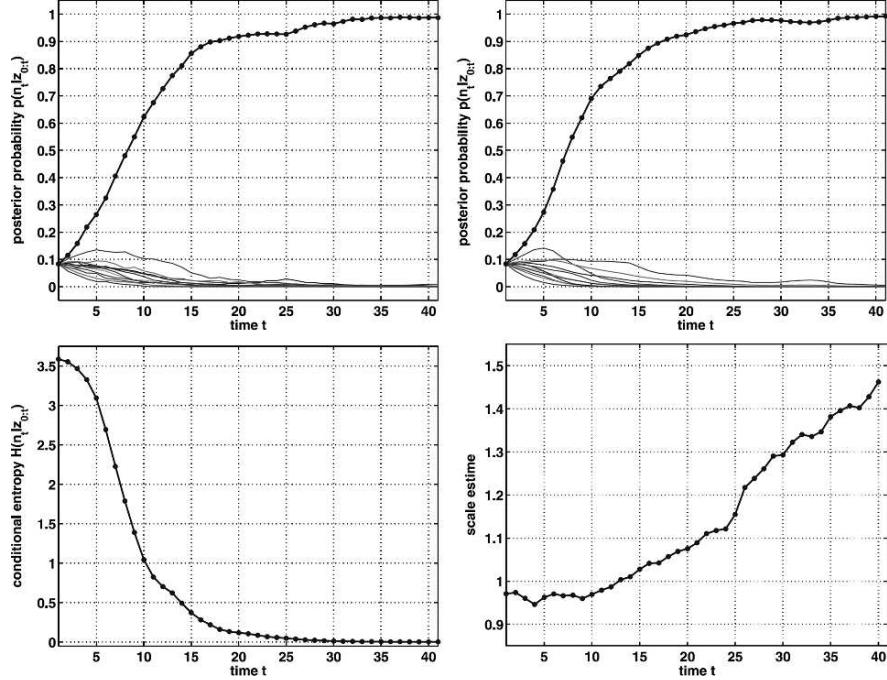


**Fig. 8.3** *Top:* 2D appearance models for the individuals in the gallery. *Bottom:* A few images from a video sequence in which a person is walking. The target's face is being tracked and the image within the bounding box of the tracked face is matched with the 2D appearance models in the gallery in order to perform recognition

$\mathbf{B} = (b_j(o))$  consists of emission probabilities, in which  $b_j(o) = P(O_t = o | Q_t = S_j)$  is the probability of emitting the symbol  $o$  when being in state  $S_j$ . In a discrete HMM, the emitted symbol comes from a finite alphabet, and in a continuous HMM, the emission probability density is modeled by a continuous function such as a mixture of Gaussians.

$\pi = (\pi_i)$  is the initial state probability distribution, i.e.,  $\pi_i = P(Q_1 = S_i)$ .

There are two additional assumptions that an HMM is based on – first-order Markov property and output independence. The former property states that information in the current state is conditionally independent from the previous states, i.e.,  $P(Q_t | Q_{t-1}, \dots, Q_1) = P(Q_t | Q_{t-1})$ . The second assumption,  $P(O_t | Q_t, Q_{t-1}, \dots, Q_1, O_{t-1}, \dots, O_1) = P(O_t | Q_t)$  states that the output at time  $t$  depends only on the current state.



**Fig. 8.4** Posterior probability of identity against time  $t$ , obtained by the CONDENSATION algorithm (top left) and the proposed algorithm (top right). Conditional entropy (bottom left) and MMSE estimate of scale parameter  $sc$  (bottom right) against time  $t$ . These plots are for the NIST video data used in [44]

Given a set of sequences  $\{\mathbf{O}_k\}$ , the training of the model is usually performed using the standard *Baum–Welch reestimation* method [35]. During the training phase, the parameters  $(\mathbf{A}, \mathbf{B}, \pi)$  that maximize the probability  $P(\{\mathbf{O}_k\}|\lambda)$  are computed. The evaluation step (i.e., the computation of the probability  $P(\mathbf{O}|\lambda)$ , given a model  $\lambda$  and a sequence  $\mathbf{O}$ ) is performed using the *forward–backward* procedure [35].

In a typical video-based face recognition system, during the training process the statistics of training sequences and their temporal dynamics are learned by an HMM. Different features can be extracted from each face and used as observation vectors (e.g., pixels values and DCT coefficients). During the recognition process, the temporal characteristics of the test video sequence are analyzed over time by the HMM corresponding to each subject. This approach can learn the dynamic information and improve the recognition performance compared to conventional methods that simply utilize the majority voting of image-based recognition results [14].

Some extensions to this basic scheme have been proposed in the past. For example, in [28], an adaptive step was added to the HMM. In that work, each frame in the video sequence was considered as one observation (whose dimensions were reduced with PCA). The adaptation was performed during the recognition process:

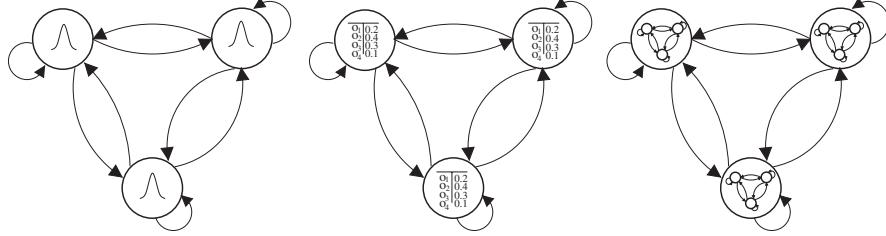
after recognizing one test sequence as one subject, it was used to update the HMM of that subject, learning the new appearance in that sequence and providing an enhanced model of that subject. A standard MAP adaptation technique was employed.

Another interesting extension has been recently proposed in [41], where pseudo-hierarchical HMMs have been introduced: they represent HMMs in which the pdf inside each state is an HMM itself: the inside-state HMMs model the appearance of the face, while the outer chain models the temporal evolution. More information on this technique is given in the following section.

### 8.3.2.1 Appearance and Behavioral Modeling Using HMMs

As explained in the previous parts of the chapter, the use of dynamic information is extremely important for humans in visual perception of biological forms and motion. Apart from aiding in the computation of the structure of the viewed objects, motion itself conveys far more information, which helps understand the scene. For video-based face recognition, not only physical features but also behavioral features should be accounted for in the face representation. While physical features are obtained from the subject's face appearance, behavioral features are obtained from the individual motion and articulation of the face. The recently introduced approach by [41] combines both physical and behavioral features. While physical features are obtained from the subject's face appearance, behavioral features are obtained by asking the subject to vocalize a given sentence, such as counting from 1 to 10 or vocalizing his/her name (feasible when the subject is cooperative, such as in authentication). Each individual has his/her own characteristic way of vocalizing a given sentence, which could change both the appearance of the face and the temporal evolution of the visual patterns. These differences are mainly due to typical accents, pronunciation, and speed of conversation. By including these behavioral features the characteristic dynamic features in the video stream are enhanced. The system presented here is based on pseudo-hierarchical hidden Markov models (PH-HMM). HMMs are quite appropriate for the representation of dynamic data; nonetheless, the emission probability function of a standard continuous HMM (Gaussians or mixture of Gaussians) is not sufficient to fully represent the variability in the appearance of the face. In this case, it is more appropriate to apply a more complex model, such as another HMM, which has been proven to be very accurate to represent variations in the face appearance [6, 22]. In summary, the proposed method is based on the modeling of the entire video sequence using an HMM in which the emission probability function of each state consists of another HMM (see Fig. 8.5), resulting in a pseudo-hierarchical HMM.

Now we discuss the PH-HMM framework in more detail. Given a set of video sequences corresponding to the subject's face, the enrolment phase aims at determining the best PH-HMM that explains the subject's facial appearance. This model encompasses both the invariant aspects of the face and its changeable features. Identity verification is performed by projecting a captured face video sequence on the PH-HMM model belonging to the claimed identity.



**Fig. 8.5** Differences between standard HMMs and PH-HMM, where emission probabilities are shown within each state: (*left*) standard Gaussian emission; (*center*) standard discrete emission; (*right*) pseudo-hierarchical HMM: in the PH-HMM the emissions are represented by HMMs

The enrolment process consists of a series of sequential steps:

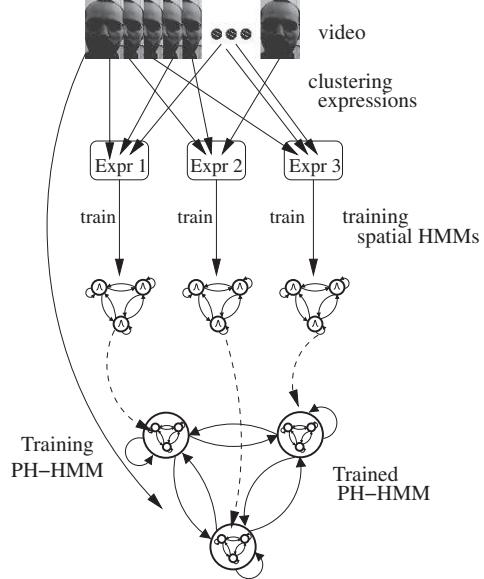
1. The video sequences are ‘unrolled’ (i.e., considered as an unordered set of images) and analyzed to detect all faces sharing similar expression, i.e., find clusters of expressions. Then, each face image of the video sequence is modeled using a standard spatial HMM [6, 22]. The resulting face HMM models are clustered in different groups based on their similarities, using a model-based approach to clustering [40]. Faces in the sequence with similar expressions are grouped together. The number of different expressions is automatically determined from the data using Bayesian information criterion [19, 38].
2. For each expression cluster, a spatial face HMM is trained which models the appearance of a single face (as in [22] or in [6]). In this phase all the sequences of the cluster are used to train the HMM. At the end of the process, K-HMMs are trained. Each spatial HMM models a particular expression of the face in the video sequences. These models represent the emission probability functions of the PH-HMM.
3. The transition matrix and the initial state probability of the PH-HMM are estimated from the sequences using the Baum–Welch procedure [35] and the emission probabilities found in the previous step. This process aims at determining the temporal evolution of facial expressions over time. The number of states is fixed to the number of discovered clusters.

In summary, the main objective of the PH-HMM representation is to determine the dominant facial expressions in the video sequence and modeling each of them with a spatial HMM. The temporal variation in expressions is then modeled by the transition matrix of the PH-HMM (as sketched in Fig. 8.6).

### 8.3.3 System Theoretic Approach for Appearance and Behavioral Modeling

As discussed in the introduction the dynamic signature in the form of idiosyncratic gestures or expressions of the face also plays an important role in identifying faces [32]. In this section, a system theoretic approach for modeling facial appearance and

**Fig. 8.6** Sketch of the enrolment phase



facial dynamics jointly [1, 42] will be presented. Since this is a joint appearance and dynamics framework, it directly matches a probe video to a gallery video in terms of both facial appearance and facial dynamic signatures.

Given a video of a moving face, we first assume that a set of features are extracted for each frame of the video. The set of features could be one of several choices. The most common ones include 2D locations of facial features such as eyes, nose, and mouth or a set of responses derived by convolving the image with a filter bank. We then model the evolution of features in time using a parametric linear dynamic system (LDS) as given below:

$$\begin{aligned} x(t+1) &= Ax(t) + w(t) \\ f(t) &= Cx(t) + v(t) \end{aligned} \quad (8.10)$$

where  $x$  is the hidden state vector,  $A$  is the transition matrix, and  $C$  is the measurement matrix.  $w$  and  $v$  are noise components modeled as normal with 0 mean and covariance matrices  $R$  and  $Q$ , respectively. In this model, the matrix  $C$  encodes the overall facial appearance and the matrix  $A$  encodes the person-specific facial dynamics such as idiosyncratic gestures and expressions.

Tools from system identification theory can be used to estimate the model parameters for each video. The most popular model estimation algorithms are N4SID [34], EM [8], and PCA-ID [39]. N4SID is a subspace identification algorithm and is an asymptotically optimal solution. For large dimensions the computational requirements make this method prohibitive. The learning problem can also be posed as a maximum likelihood estimation of the model parameters that maximize the

likelihood of the observations which can be solved by expectation-maximization (EM). PCA-ID is a suboptimal solution to the learning problem. It makes the assumption that filtering in space and time is separable, which makes it possible to estimate the parameters of the model efficiently via principal component analysis. First, PCA is used for dimensionality reduction. The principal vectors form the columns of the C matrix. The lower dimensional PCA subspace is considered as the hidden state-space. The transition matrix A can now be estimated from the projections of the feature sequence onto the lower dimensional subspace. We refer the interested reader to [8, 34, 39] for details regarding the model estimation algorithms.

### 8.3.3.1 Measuring Distance Between Models

Facial video sequences are modeled using the LDS framework and the model parameters ( $\hat{A}$ ,  $\hat{C}$ ) are estimated for both the gallery and probe videos. Now, for each probe video we need to find the closest exemplar in the gallery. This can be achieved by comparing the probe's model parameters with those of the gallery. This requires appropriate distance metrics on the space of LDSs. Several distance metrics exist to measure the distance between LDSs. The simplest method to measure distance is the  $L_2$  norm between model parameters. Martin [29] proposed a more principled method to measure the distance between LDSs based on cepstral coefficients. A unifying framework based on subspace angles of observability matrices was presented in [9] to measure the distance between LDSs. Specific metrics such as the Frobenius norm and the Martin metric can be derived as special cases based on the subspace angles. Recently, Vishwanathan et al. [43] presented a framework to extend the Cauchy–Binet kernels to the space of dynamical systems and incorporated the dependence on initial conditions of the dynamical system as well. The subspace angles based kernel has proved popular due to its ease of computation. Subspace angles ( $\theta_i, i = 1, 2, \dots, n$ ) between two LDSs are defined in [9] as the principal angles between the column spaces generated by the observability matrices of the two models extended with the observability matrices of the inverse models. The subspace angles between the range spaces of two matrices A and B are recursively defined as follows [9]:

$$\begin{aligned} \cos \theta_1 &= \max_{x,y} \frac{x^T A^T B y}{\|Ax\|_2 \|By\|_2} = \frac{x_1^T A^T B y_1}{\|Ax_1\|_2 \|By_1\|_2} \\ \cos \theta_k &= \max_{x,y} \frac{x^T A^T B y}{\|Ax\|_2 \|By\|_2} = \frac{x_k^T A^T B y_k}{\|Ax_k\|_2 \|By_k\|_2} k = 2, 3, \dots \end{aligned} \quad (8.11)$$

subject to the constraints  $x_i^T A^T A x_k = 0$  and  $y_i^T B^T B y_k = 0$ . For the case of LDS, the subspace angles can be solved efficiently by solving a discrete Lyapunov equation involving the model parameters [9]. Using these subspace angles ( $\theta_i, i = 1, 2, \dots, n$ ), three distances, namely, Martin distance ( $d_M$ ), gap distance

$(d_g)$ , and Frobenius distance  $(d_F)$ , between the LDS are defined as follows:

$$d_M^2 = \ln \prod_{i=1}^n \frac{1}{\cos^2 \theta_i}, \quad d_g = \sin \theta_{\max}, \quad d_F^2 = \sum_{i=1}^n \sin^2 \theta_i \quad (8.12)$$

Now using any of these distance metrics, we can compute the nearest exemplar in the gallery to the given probe.

### 8.3.3.2 Exploiting Multiple Exemplars Using Statistics on the Grassmann Manifold

In the previous section, we defined distance metrics on the LDS parameter space. These distance metrics can be used directly to implement nearest neighbor classifiers. In order to develop accurate inference algorithms on the LDS parameter space, we need to understand the geometric structure of this space and derive appropriate distance measures and probability distribution functions (pdf) that are consistent with this geometric structure. In the following, we show how the space of LDS parameters can be considered as a Grassmann manifold.

A key observation is that a given LDS with parameters  $M = (A, C)$  can be alternately identified by the column space generated by its extended observability matrix given by

$$O_M(\infty) = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} \quad (8.13)$$

If the infinite observability matrix is approximated by the finite observability matrix of large enough dimension (large  $n$ )

$$O_M(n) = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \quad (8.14)$$

then the space spanned by the columns of this matrix is a linear subspace of a high-dimensional Euclidean space. Linear subspaces of Euclidean spaces are well understood as points on the *Grassmann* manifold.

**The Grassmann Manifold**  $G_{k,m-k}$  [10]: The Grassmann manifold  $G_{k,m-k}$  is the space whose points are  $k$ -planes or  $k$ -dimensional hyperplanes (containing the origin) in  $R^m$ .

### 8.3.3.3 Numerical Representation of Points on the Grassmann Manifold

#### Projection Matrix Representation

An equivalent definition of the Grassmann manifold is as follows [10]. To each  $k$ -plane  $v$  in  $G_{k,m-k}$  corresponds a unique  $m \times m$  orthogonal projection matrix  $P$  idempotent of rank  $k$  onto  $v$ . If the columns of an  $m \times k$  orthonormal matrix  $Y$  spans  $v$  (i.e., if columns of  $Y$  form an orthonormal basis for the  $k$ -plane  $v$ ), then  $YY^T = P$ .

While the projection matrix representation is a unique representation of points on the manifold, the projection matrix is a large  $m \times m$  matrix which requires large computational and memory requirements to store and manipulate.

#### Procrustes Representation

To each  $k$ -plane  $v$  in  $G_{k,m-k}$ , we associate an orthonormal matrix  $Y$  such that columns of  $Y$  span  $v$  (i.e., columns of  $Y$  form an orthonormal basis for the  $k$ -plane  $v$ ) [10], together with the equivalence class of matrices  $YR$  where  $R$  is a full rank  $k \times k$  real matrix.

The Procrustes representation leads to efficient distance computations and non-parametric methods for learning pdfs.

### 8.3.3.4 Distances and Statistical Modeling

The squared Procrustes distance for two points  $P_1, P_2$  on the Grassmann manifold can be computed using orthonormal basis  $X_1$  and  $X_2$  of the respective subspaces as the smallest squared Euclidean distance between any pair of matrices in the corresponding equivalence classes. Hence

$$d^2(P_1, P_2) = \min_R \text{tr}(X_1 - X_2 R)^T (X_1 - X_2 R) \quad (8.15)$$

This minimization problem can be solved in closed form [10]. For the case when  $R$  varies over all  $k \times k$  real matrices, the Procrustes distance is given by

$$d^2(P_1, P_2) = \text{tr}(I_k - A^T A)$$

where  $A = X_1^T X_2$  and  $I_k$  is a  $k \times k$  identity matrix.

### 8.3.3.5 Density Estimation Using Kernels

Kernel methods for estimating probability densities have proved extremely popular in several pattern recognition problems in recent years driven by improvements in computational power. Kernel methods provide a better fit to the available data than simpler parametric forms. Given several examples from a class on the Grassmann manifold (numerically represented using orthonormal basis  $(X_1, X_2, \dots, X_n)$ ), the class conditional density can be estimated using an appropriate kernel function.

Using the Procrustes representation of points, the density estimate is given as [10]

$$\hat{f}(X; M) = \frac{1}{n} C(M) \sum_{i=1}^n K \left[ M^{-1/2} (I_k - X_i^T X X^T X_i) M^{-1/2} \right] \quad (8.16)$$

where  $K(T)$  is the kernel function,  $M$  is a  $k \times k$  positive definite matrix which plays the role of the kernel width or a smoothing parameter.  $C(M)$  is a normalizing factor chosen so that the estimated density integrates to unity. The matrix-valued kernel function  $K(T)$  can be chosen in several ways. The exponential kernel  $K(T) = \exp(-\text{tr}(T))$  is widely used due to closed-form expression for its integration constant  $C(M)$  [10].

### 8.3.3.6 Experiments

Results of a recognition experiment using the Li data set [26] are shown here. The data set consists of face videos for 16 subjects with 2 sequences per subject. Subjects arbitrarily change head orientation and expressions. The illumination conditions differed widely for the two sequences of each subject. For each subject, one sequence was used as the gallery while the other formed the probe. The experiment was repeated by swapping the gallery and the probe data. The recognition results are reported in Table 8.1. For kernel density estimation, the available gallery sequence for each actor was split into three distinct sequences. As seen in the last column, the kernel-based method outperforms the other approaches.

**Table 8.1** Comparison of video-based face recognition approaches: (a) LDS distance using subspace angles, (b) Procrustes distance, and (c) manifold kernel density

	Test condition	Subspace angles	Procrustes distance	Kernel density
1	Gallery1, Probe2	81.25%	93.75%	93.75%
2	Gallery2, Probe1	68.75%	81.25%	93.75%
3	Average	75%	87.5%	93.75%

## 8.4 Databases

### 8.4.1 The Honda/UCSD Video Database (University of California San Diego – US)

<http://vision.ucsd.edu/~leekc/HondaUCSDVideoDatabase/HondaUCSD.html>

This database [25] is intended for evaluating face tracking/recognition algorithms. Each video sequence is recorded in an indoor environment at 15 frames per second and lasts for at least 15 s. The resolution of each video sequence is  $640 \times 480$ . Every individual is recorded in at least two video sequences. All the video sequences contain significant 2D (in-plane) and 3D (out-of-plane) head rotations. The Honda/

UCSD Video Database contains two data sets. The first includes three different subsets – training, testing, and occlusion testing. Each subset contains 20, 42, 13 videos, respectively, from 20 human subjects. The second data set includes two subsets – training and testing of 30 videos from 15 different human subjects.

#### **8.4.2 The BANCA Database (University of Surrey – UK)**

<http://www.ee.surrey.ac.uk/CVSSP/banca/>

The BANCA database represents a multimodal database intended for training and testing multimodal verification systems using face and voice. Video and speech data were collected for 52 subjects (26 male and 26 female) in 4 different languages (English, French, Italian, and Spanish), i.e., a total of 208 combinations. The English part is now available to the research community. Each subject recorded 12 sessions, each of these sessions containing 2 recordings: 1 true *client access* and 1 informed (the actual subject knew the text that the claimed identity subject was supposed to utter) *impostor attack*. The 12 sessions were separated into 3 different scenarios – *controlled*, *degraded*, and *adverse* for sessions 9–12. A webcam was used in the degraded scenario, while an expensive camera was used in the controlled and adverse scenarios. During each recording, the subject was prompted to say a random 12 digit number, his/her name, their address, and date of birth. Each recording took an average of 20 s. An evaluation protocol is defined in [5] which permits comparison of results with other researchers – this is also done in official competitions (ICPR 2004 [30] and ICBA 2004 [31]).

#### **8.4.3 The Database of Moving Faces and People (University of Texas at Dallas)**

<http://bbs.utdallas.edu/facelab/otoole/database.htm>

The Database of Moving Faces and People [33] was constructed with the goal of collecting high-quality images and videos of faces and people. The database consists of a series of close and moderate range images and videos of people. The close-range videos capture dynamic information about the face across a range of emotional expressions and poses and include lip movements due to speech. The moderate-range videos include dynamic information about individuals who are either actively or passively involved in a conversation. Videos of subjects walking include information about the face, posture, and gait of the individual. A second duplicate session is available for most subjects, allowing for testing recognition algorithms that make use of images and videos in which the subject may have a different hairstyle and different clothing, and in general look different in appearance. This database is useful for testing the performance of humans and machines for the tasks of face/person recognition, tracking, and computer graphics modeling of natural human motions.

#### **8.4.4 The Bogazici University (BU) Database**

<http://www.cs.bu.edu/groups/ivc/HeadTracking/>

This database [7] consists of two classes of sequences. One set of sequences was collected under uniform illumination conditions. The other set was collected under time-varying illumination. The time-varying illumination had a uniform component and a sinusoidal directional component. All the sequences are 200 frames long (approximately 7 s) and contain free head motion of several subjects. The first set consists of 45 sequences (9 sequences each of 5 subjects) taken under uniform illumination where the subjects perform free head motion including translations and both in-plane and out-of-plane rotations. The second set consists of 27 sequences (9 sequences each of 3 subjects) taken under time-varying illumination and where the subjects perform free head motion.

#### **8.4.5 Other Minor Databases**

The VidTIMIT Audio–Video Dataset (Sanderson), Face Video Database (Max Planck Institute for Biological Cybernetics).

### **8.5 Conclusions and Open Issues**

The advantages in using motion information for face recognition have been acknowledged in both neurophysiologic and computational studies. As confirmed by recent neurophysiologic studies, the use of dynamic information is extremely important for humans in visual perception of biological forms and motion. Apart from the mere computation of the structure of the scene, motion conveys more information about the behavior of objects in the scene. The analysis of video streams of face images has received increasing attention in biometrics. An immediate advantage in using video information is the possibility of employing redundancy present in the video sequence to improve still image systems. Moreover, as described in this chapter, great advantages could be obtained in localization of faces (by means of tracking) and by simultaneously addressing recognition and tracking.

Since this is a relatively new research area, there are many unaddressed issues. The progressively wider deployment of surveillance cameras set in public places provides us with large amounts of data. Smart methods for acquiring, pre-processing, and analyzing video streams are needed in order to extract useful information from the otherwise voluminous data.

Face recognition is well understood when the subjects are in a canonical pose. Dealing with pose and illumination becomes crucial in real-life application scenarios, where there are multiple pan–tilt–zoom cameras monitoring the scene. Moreover, cameras are usually ceiling-mounted which prevents capturing of faces in canonical pose. These factors may drastically alter the appearance of a face. Thus,

exploiting dynamic information could help in mitigating the decreased reliability of static features.

Another crucial issue that needs to be addressed is robustness to low-resolution images. Standard face recognition methods work well if enough high-quality images are available, but would fail when enough resolution is not present. In this case too, effective utilization of dynamic information may hold the key to recognize a person. Such methods will be useful for designing face recognition systems that can be effective at large ranges.

Another direction of future research is the possibility of registering multiple flows, i.e., to consider subjects captured in different cameras at the same time. Synchronized cameras, as well as synchronized recognition systems, may be effectively exploited to recover from partial occlusions or low-resolution images.

Finally, there is much work to be done in order to realize methods that reflect how humans recognize faces and optimally make use of the temporal evolution of the appearance of the face for recognition.

**Acknowledgments** The work of Rama Chellappa and Pavan Turaga was supported by the IARPA VACE program at University of Maryland.

## References

1. Aggarwal, G., Chowdhury, A.R., and Chellappa, R.: A system identification approach for video-based face recognition, In: Proceedings of the 17th International Conference on Pattern Recognition (ICPR), Cambridge, UK, 175–178, 2004.
2. Aggarwal, G., Veeraraghavan, A., and Chellappa, R.: 3D facial pose tracking in uncalibrated videos, In: International Conference on Pattern Recognition and Machine Intelligence (PReMI), 2005.
3. Anderson, B. and Moore, J.: Optimal Filtering, Prentice Hall, Englewood Cliffs, New Jersey, 1979.
4. Arandjelovic, O., and Cipolla, R.: Face recognition from video using the generic shape-illumination manifold, In: Proc. 9th European Conference on Computer Vision, Graz (Austria) (May) Edited by A. Leonardis, H. Bischof and A. Pinz, volume LNCS 3954, 27–40, Springer, 2006.
5. Bailly-Bailliére, E., Bengio, S., Bimbot, F., Hamouz, M., Kittler, J., Mariéthoz, J., Matas, J., Messer, K., Popovici, V., Porée, F., Ruiz, B., and Thiran, J.P.: The BANCA database and evaluation protocol. Proc. of Audio Video-based Person Authentication, 625–638, 2003.
6. Bicego, M., Castellani, U., and Murino, V.: Using hidden Markov models and wavelets for face recognition. Proc. of IEEE Int. Conf. on Image Analysis and Processing 52–56, 2003.
7. Cascia, M.L., Sclaroff, S., and Athitsos, V.: Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models, In: IEEE Trans. on Pattern Analysis and Machine Intelligence, **22** 322–336, 2000.
8. Chan, A.B. and Vasconcelos, N.: Probabilistic kernels for the classification of auto-regressive visual processes, In: IEEE Conference on Computer Vision and Pattern Recognition, **1** 846–851, 2005.
9. Cock, K.D. and Moor, B.D.: Subspace angles between ARMA models, Systems and Control Letters, **46** 265–270, 2002.
10. Chikuse, Y.: Statistics on special manifolds, Lecture Notes in Statistics. Springer, New York, 2003.

11. Doucet, A., Godsill, S.J., and Andrieu, C.: On sequential Monte Carlo sampling methods for Bayesian filtering, In: *Statistical Computing*, **10**(3) 197–209, 2000.
12. Doucet, A., Freitas, N.D., and Gordon, N.: *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, 2001.
13. Gordon, N.J., Salmond, D.J., and Smith, A.F.M.: Novel approach to nonlinear/non-gaussian Bayesian state estimation, In: *IEE Proceedings on Radar and Signal Processing*, **140** 107–113, 1993.
14. Hadid, A., and Pietikainen, M.: An experimental investigation about the integration of facial dynamics in video-based face recognition, *Electronic Letters on Computer Vision and Image Analysis*, **5**(1):1–13, 2005
15. Hager, G.D. and Belhumeur, P.N.: Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20** 1025–1039, 1998.
16. Ho, T.K., Hull, J.J., and Srihari, S.N.: Decision combination in multiple classifier systems, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **16**(1), 66–75, 1994.
17. Isard, M. and Blake, A.: Contour tracking by stochastic propagation of conditional density, In: *Proceedings of European Conference on Computer Vision* 343–356, 1996.
18. Jebara, T.S. and Pentland, A.: Parameterized structure from motion for 3D adaptive feedback tracking of faces, In: *IEEE Conference on Computer Vision and Pattern Recognition*, 1997.
19. Kashyap, R.L.: A Bayesian comparison of different classes of dynamic models using empirical data, *IEEE Transactions on Automatic Control*, **AC-22**(5) 715–727, 1977.
20. Kitagawa, G.: Monte carlo filter and smoother for non-gaussian nonlinear state space models, In: *Journal of Computational and Graphical Statistics*, **5** 1–25, 1996.
21. Kittler, J., Hatef, M., Duin, R.P.W. and Matas J.: On combining classifiers, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(3) 226–239, 1998.
22. Kohir, V.V. and Desai, U.B.: Face recognition using DCT-HMM approach, Proc. Workshop on Advances in Facial Image Analysis and Recognition Technology, 1998.
23. Lanitis, A., Taylor, C., and Cootes, T.: Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19** 743–756, 1997.
24. Lee, K.C., Ho, J., Yang, M.H., and Kriegman, D.: Video-based face recognition using probabilistic appearance manifolds, In: *IEEE Conference on Computer Vision and Pattern Recognition*, 2003.
25. Lee, K.C., Ho, J., Yang, M.H., and Kriegman, D.: Visual tracking and recognition using probabilistic appearance manifolds. *Computer Vision and Image Understanding*, **99**(3): 303–331, 2005.
26. Li, B. and Chellappa, R.: Face verification through tracking facial features, In: *Journal of the Optical Society of America A*, **18** 2969–2981, 2001.
27. Liu, J.S. and Chen, R.: Sequential Monte Carlo for dynamic systems, In: *Journal of the American Statistical Association*, **93** 1031–1041, 1998.
28. Liu, X. and Chen, T.: Video-based face recognition using adaptive hidden Markov models, *IEEE Conference on Computer Vision and Pattern Recognition*, 2003.
29. Martin, R.J.: A metric for ARMA processes, In: *IEEE Transactions on Signal Processing*, **48**(4), 1164–1170, 2000.
30. Messer, K., Kittler, J., Sadeghi, M., Hamouz, M., Kostin, A., Cardinaux, F., Marcel, S., Bengio, S., Sanderson, C., Poh, N., Rodriguez, Y., Czyz, J., Vandendorpe, L., McCool, C., Lowther, S., Sridharan, S., Chandran, V., Palacios, R.P., Vidal, E., Li Bai, LinLin Shen, Yan Wang, Chiang Yueh-Hsuan, Liu Hsien-Chang, Hung Yi-Ping, Heinrichs, A., Muller, M., Tewes, A., von der Malsburg, C., Wurtz, R., Zhenger Wang, Feng Xue, Yong Ma, Qiong Yang, Chi Fang, Xiaoqing Ding, Lucey, S., Goss, R., and Schneiderman, H.: Face Authentication Test on the BANCA Database. *ICPR*, **4** 523–532, 2004.
31. Messer, K., Kittler, J., Sadeghi, M., Hamouz, M., Kostin, A., Marcel, S., Bengio, S., Cardinaux, F., Sanderson, C., Poh, N., Rodriguez, Y., Kryszczuk, K., Czyz, J., Vandendorpe, L.,

- Ng, J., Cheung, H., and Tang, B.: Face authentication competition on the BANCA database. ICBA 8–15, 2004.
- 32. O'Toole, A.J., Roark, D., and Abdi, H.: Recognizing moving faces: A Psychological and Neural Synthesis, In: Trends in Cognitive Sciences, **6**, 261–266, 2002.
  - 33. O'Toole, A.J., Harms, J., Snow, S.L., Hurst, D.R., Pappas, M.R., and Abdi, H.: A video database of moving faces and people. IEEE Transactions on Pattern Analysis and Machine Intelligence, **27**(5) 812–816, 2005.
  - 34. Overschee, P.V. and Moor, B.D.: N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems, In: Automatica, **30** 75–93, 1994.
  - 35. Rabiner, L.: A tutorial on Hidden Markov Models and selected applications in speech recognition, Proc. of IEEE, **77**(2) 257–286, 1989.
  - 36. Roark, D.A., Barrett, S.E., O'Toole, A.J., and Abdi, H.: Learning the moves: The effect of familiarity and facial motion on person recognition across large changes in viewing format, In: Perception, **35** 761–773, 2006.
  - 37. Ross, A., Nandakumar, K., and Jain, A.K.: Handbook of Multibiometrics, Springer, New York, 2006.
  - 38. Schwarz, G.: Estimating the dimension of a model. Annals of Statistics, **6**(2):461–464, 1978.
  - 39. Soatto, S., Doretto, G., and Wu, Y.N.: Dynamic textures, In: International Conference on Computer Vision, **2** 439–446, 2001.
  - 40. Smyth, P.: Clustering sequences with hidden Markov models, In: M. Mozer, M. Jordan, T. Petsche (Eds.), Advances in Neural Information Processing Systems, **9**, MIT Press, Cambridge, MA, p. 648, 1997.
  - 41. Tistarelli, M., Bicego, M., and Grossi, E.: Dynamic face recognition: From Human to Machine Vision, Image and Vision Computing, **27**(3) 222–232, 2009.
  - 42. Turaga, P., Veeraraghavan, A., and Chellappa, R.: Statistical analysis on Stiefel and Grassmann manifolds with applications in computer vision, In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.
  - 43. Vishwanathan, S.V.N., Smola, A.J., and Vidal, R.: Binet-Cauchy kernels on dynamical systems and its application to the analysis of dynamic scenes, In: International Journal of Computer Vision, **73**(1), 95–119, 2007.
  - 44. Zhou, S., Krueger, V., and Chellappa, R.: Probabilistic recognition of human faces from video, In: Computer Vision and Image Understanding (CVIU) (special issue on Face Recognition) **91** 214–245, 2003.
  - 45. Zhou, S.K., Chellappa, R., and Moghaddam, B.: Visual tracking and recognition using appearance-adaptive models in particle filters, In: IEEE Transactions on Image Processing **13**(11) 1491–1506, 2004.

# **Chapter 9**

## **3D Face Recognition: Technology and Applications**

**Berk Gökberk, Albert Ali Salah, Neşe Alyüz, and Lale Akarun**

**Abstract** 3D face recognition has received a lot of attention in the last decade, leading to improved sensors and algorithms that promise to enable large-scale deployment of biometric systems that rely on this modality. This chapter discusses advances in 3D face recognition with respect to current research and technology trends, together with its open challenges. Five real-world scenarios are described for application of 3D face biometrics. Then we provide a comparative overview of the currently available commercial sensors, and point out to research databases acquired with each technology. The algorithmic aspects of 3D face recognition are broadly covered; we inspect automatic landmarking and automatic registration as sine qua non parts of a complete 3D facial biometric system. We summarize major coordinated actions in evaluating 3D face recognition algorithms, and conclude with a case study on a recent and challenging database.

### **9.1 Introduction**

Face is the natural assertion of identity: We show our face as proof of who we are. Due to this widely accepted cultural convention, face is the most widely accepted biometric modality.

Face recognition has been a specialty of human vision: Something humans are so good at that even a days-old baby can track and recognize faces. Computer vision has long strived to imitate the success of human vision and in most cases has come nowhere near its performance. However, the recent Face Recognition Vendor Test (FRVT06) has shown that automatic algorithms have caught up with the performance of humans in face recognition [76].

How has this increase in performance come about? This can partly be attributed to advances in 3D face recognition in the last decade. Three-dimensional face recognition has important advantages over 2D: It makes use of shape and texture channels simultaneously, where the texture channel carries 2D image information. However,

---

B. Gökberk (✉)

Department of Electrical Engineering, Mathematics and Computer Science, University of Twente,  
Enschede, The Netherlands  
e-mail: b.gokberk@ewi.utwente.nl

texture is registered with the shape channel, and intensity can now be associated with shape attributes such as the surface normal. The shape channel does not suffer from certain problems that the texture suffers from, such as poor illumination or pose changes. Recent research in 3D face recognition has shown that shape carries significant information about identity. At the same time, the shape information makes it easier to eliminate the effects of illumination and pose from the texture. Processed together, the shape and the texture make it possible to achieve high performances under different illumination and pose conditions.

Although 3D offers additional information that can be exploited to infer the identity of the subject, this is still not a trivial task: *External factors* such as illumination and camera pose have been cited as complicating factors. However, there are *internal factors* as well: Faces are highly deformable objects, changing shape and appearance with speech and expressions. Humans use the mouth and the vocal tract to produce speech; and the whole set of facial muscles to produce facial expressions. Human vision can deal with face recognition under these conditions. Automatic systems are still trying to devise strategies to tackle expressions. A third dimension complicating face recognition is the *time dimension*. Human faces change primarily due to two factors. The first factor is ageing: All humans naturally age. This happens very fast at childhood, somewhat slower once adulthood is reached. The other factor is intentional: Humans try to change the appearance of their faces through hair style, makeup, and accessories. Although the intention is usually to enhance the beauty of the individual, the detrimental effects for automatic face recognition are obvious.

This chapter will discuss advances in 3D face recognition together with open challenges and ongoing research to overcome these. In Section 9.2, we discuss real-world scenarios and acquisition technologies. In Section 9.3, we overview and compare 3D face recognition algorithms. In Section 9.4, we outline outstanding challenges and present a case study from our own work; and in Section 9.5, we present conclusions and suggestions for research directions. A number of questions touching on the important points of the chapter can be found at the end.

## 9.2 Technology and Applications

### 9.2.1 Acquisition Technology

Among biometric alternatives, facial images offer a good trade-off between acceptability and reliability. Even though iris and fingerprint biometrics provide accurate authentication and are more established as biometric technologies, the acceptability of face as a biometric makes it more convenient. Three-dimensional face recognition aims at bolstering the accuracy of the face modality, thereby creating a reliable and non-intrusive biometric.

There exist a wide range of 3D acquisition technologies, with different cost and operation characteristics. The most cost-effective solution is to use several calibrated

2D cameras to acquire images simultaneously and to reconstruct a 3D surface. This method is called *stereo acquisition*, even though the number of cameras can be more than two. An advantage of these type of systems is that the acquisition is fast, and the distance to the cameras can be adjusted via calibration settings, but these systems require good and constant illumination conditions.

The reconstruction process for stereo acquisition can be made easier by projecting a structured light pattern on the facial surface during acquisition. The structured light methods can work with a single camera, but require a projection apparatus. This usually entails a larger cost when compared to stereo systems, but a higher scan accuracy. The potential drawbacks of structured light systems are their sensitivity to external lighting conditions and the requirement of a specific acquisition distance for which the system is calibrated. Another problem associated with structured light is that the projected light interferes with the color image and needs to be turned off to generate it. Some sensors avoid this problem by using near-infrared light.

Yet a third category of scanners relies on active sensing: A laser beam reflected from the surface indicates the distance, producing a range image. These types of laser sensors, used in combination with a high-resolution color camera, give high accuracies, but sensing takes time.

The typical acquisition distance for 3D scanners varies between 50 and 150 cm, and laser scanners are usually able to work with longer distances (up to 250 cm) when compared to stereo and structured light systems. Structured light and laser scanners require the subject to be motionless for a short duration (0.8–2.5 s in the currently available systems), and the effect of motion artifacts can be much more detrimental for 3D in comparison to 2D. The Cyberware scanner takes a longer scan time (8–34 s), but it acquires an 360° scan within this time.

Depending on the surface normal, laser scanners are able to provide 20–100  $\mu\text{m}$  accuracy in the acquired points. The actual point accuracy is smaller for structural light systems, but the feature accuracy can be similar. For instance, Breuckmann FaceScan III uses a fringe projection technique including the phase shift method and thus can detect object details with a feature accuracy of  $\pm 200 \mu\text{m}$ . The presence of strong motion artifacts would make a strong smoothing necessary, which will dispel the benefits of having such a great accuracy. Simultaneous acquisition of a 2D image is an asset, as it enables fusion of 2D and 3D methods to potentially greater accuracy. The amount of collected data affects scan times, but also the time of transfer to the host computer, which can be significant. For instance, a Minolta 910 scanner requires 0.3 s to scan the target in the fast mode (about 76 K points) and about 1 s to transfer it to the computer. Longer scan times also result in motion-related problems, including poor 2D–3D correspondence. Some scanners (e.g., 3dMDface) work with two viewpoints and can simultaneously capture 180° (ear-to-ear coverage / sides of the nose) of face data, as opposed to the more frequently seen single-view scanners. The two-view systems can create more data points, but also require true 3D representations (i.e., range maps will lose the additional information). Table 9.1 lists properties of some commercial sensors.

**Table 9.1** Three-dimensional scanners used for face recognition

Scanner	Scanning technology	Scan time	Range (m)	Field of view	Accuracy	Databases	Web site
3dMDface static	Stereo photogrammetry, unstructured light	1.5 ms	1	180° coverage	0.2 mm	BU-3DFE, ASU PRISM	<a href="http://www.3dmd.com">www.3dmd.com</a>
3dMDface dynamic (4D)	Stereo photogrammetry, unstructured light	60 frames per second	1	180° coverage	0.5 mm	Univ. Surrey, Univ. Houston, Cardiff Univ., UCLAN, UNC Chapel Hill	<a href="http://www.3dmd.com">www.3dmd.com</a>
Konica Minolta 300/700/9i/910	Laser scanner	0.3-2.5 s	0.6-1.2 m	463 × 347 × 500	0.16 mm	FRGC, $I^2V^2$ , GavabDB, FRAV3D	<a href="http://www.konicaminolta-3d.com">www.konicaminolta-3d.com</a>
Inspeck MegaCapturor II	Structured light	0.7 s	1.5	435 × 350 × 450	0.3 mm	Bosphorus	<a href="http://www.inspectk.com">www.inspectk.com</a>
Geometrix Facevision (ALIVE Tech)	Stereo camera	1 s	1	Not specified	Not specified	IDENT	<a href="http://www.geometrix.com">www.geometrix.com</a>
Bioscrypt VisionAccess (formerly A4)	Near-infrared light	< 1 s	0.9-1.8	Not specified	Not specified	N/A	<a href="http://www.bioscrypt.com">www.bioscrypt.com</a>
Cyberware PX	Laser scanner	17 s	0.5-1	440 × 360 × 249 (360°)	0.05-0.015 mm	N/A	<a href="http://www.cyberware.com">www.cyberware.com</a>
Cyberware 3030	Laser scanner	17 s	<0.5 m	300 × 340 × 300 (360°)	0.075-0.3 mm	BJUT-3D	<a href="http://www.cyberware.com">www.cyberware.com</a>
Genex 3D FaceCam	Stereo-structured light	0.5 s	1	510 × 400 × 300	0.6 mm	N/A	<a href="http://www.genextech.com">www.genextech.com</a>
Breuckmann FaceScan III	Structured light	0.8 s	1	600 × 460 × 400	0.43 mm	N/A	<a href="http://www.breuckmann.com">www.breuckmann.com</a>

### ***9.2.2 Application Scenarios***

**Scenario 1 – Border Control:** Since 3D sensing technology is relatively costly, its primary application is the high-security, high-accuracy authentication setting, for instance, the control point of an airport. In this scenario, the individual briefly stops in front of the scanner for acquisition. The full face scan can contain between 5.000 and 100.000 3D points, depending on the scanner technology. This data are processed to produce a biometric template of the desired size for the given application. Template security considerations and the storage of the biometric templates are important issues. Biometric databases tend to grow as they are used; the FBI fingerprint database contains about 55 million templates.

In verification applications, the storage problem is not so vital since templates are stored in the cards such as e-passports. In verification, the biometric is used to verify that the scanned person is the person who supplied the biometric in the e-passport, but extra measures are necessary to ensure that the e-passport is not tampered with. With powerful hardware, it is possible to include a screening application to this setting, where the acquired image is compared to a small set of individuals. However, for a recognition setting where the individual is searched among a large set of templates, biometric templates should be compact.

Another challenge for civil ID applications that assume enrollment of the whole population is the deployment of biometric acquisition facilities, which can be very costly if the sensors are expensive. This cost is even greater if multiple biometrics are to be collected and used in conjunction.

**Scenario 2 – Access Control:** Another application scenario is the control of a building, or an office, with a manageable size of registered (and authorized) users. Depending on the technology, a few thousand users can be managed, and many commercial systems are scalable in terms of users with appropriate increase in hardware cost. In this scenario, the 3D face technology can be combined with RFID to have the template stored on a card together with the unique RFID tag. Here, the biometric is used to authenticate the card holder given his/her unique tag.

**Scenario 3 – Criminal ID:** In this scenario, face scans are acquired from registered criminals by a government-sanctioned entity, and suspects are searched in a database or in videos coming from surveillance cameras. This scenario would benefit most from advances in 2D–3D conversion methods. If 2D images can be reliably used to generate 3D models, the gallery can be enhanced with 3D models created from 2D images of criminals, and acquired 2D images from potential criminals can be used to initiate search in the gallery.

**Scenario 4 – Identification at a Distance:** For Scenarios 1–3, available commercial systems can be employed. A more challenging scenario is identification at a distance, when the subject is sensed in an arbitrary situation. In this scenario, people can be far away from the camera, unaware of the sensors. In such cases, challenge stems from uncooperative users. Assuming that the template of the subject is acquired with a neutral expression, it is straightforward for a person who tries to avoid being detected to change parts of his or her facial surface by a smiling or open-mouthed expression. Similarly, growing a moustache or a beard, or wearing

glasses can make the job of identifying the person difficult. A potential solution to this problem is to use only the rigid parts of the face, most notably the nose area, for recognition. However, restricting the input data to such a small area means that a lot of useful information will be lost and the overall accuracy will decrease. Furthermore, certain facial expressions affect the nose and subsequently cause a drop in the recognition accuracy.

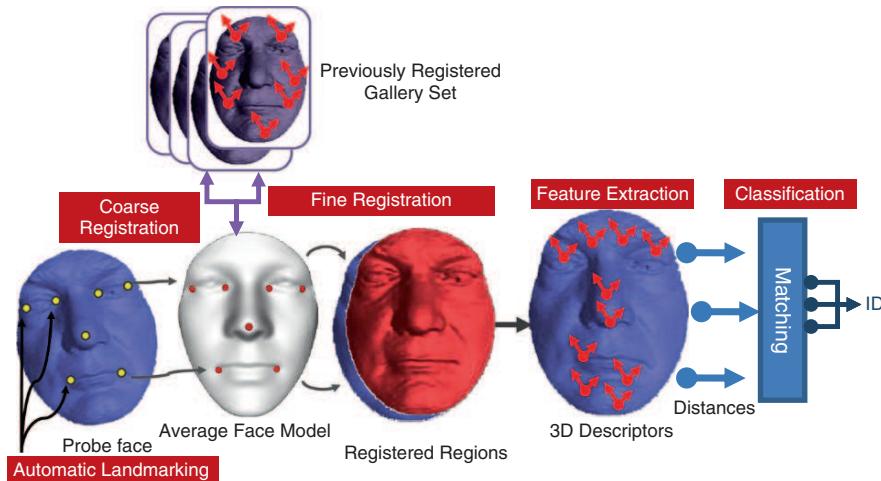
This scenario also includes consumer identification, where a commercial entity identifies a customer for personalized services. Since convenience is of utmost importance in this case, face biometrics are preferable to most alternatives.

**Scenario 5 – Access to Consumer Applications:** Finally, a host of potential applications are related with appliances and technological tools that can be proofed against theft with the help of biometrics. For this type of scenario, the overall system cost and user convenience are more important than recognition accuracy. Therefore, stereo camera-based systems are more suited for these types of applications. Computers or even cell phones with stereo cameras can be protected with this technology. Automatic identification has additional benefits that can increase the usefulness of such systems. For instance, a driver authentication system using 3D facial characteristics may provide customization for multiple users in addition to ensuring security. Once the face acquisition and analysis tools are in place, this system can also be employed for opportunistic purposes, for instance, to determine drowsiness of the driver by facial expression analysis.

### 9.3 Three-Dimensional Face Recognition Technology

A 3D face recognition system usually consists of the following stages: (1) preprocessing of raw 3D facial data, (2) registration of faces, (3) feature extraction, and (4) matching [4, 7, 23, 34, 73, 86]. Figure 9.1 illustrates the main components of a typical 3D face recognition system. Prior to these steps, the 3D face should be localized in a given 3D image. However, currently available 3D face acquisition devices have a very limited sensing range and the acquired image usually contains only the facial area. Under such circumstances, recognition systems do not need face detection modules. With the availability of more advanced 3D sensors that have large range of view, we foresee the development of highly accurate face detection systems that use 3D facial shape data together with the 2D texture information. For instance, in [29], a 3D face detector that can localize the upper facial part under occlusions is proposed.

The preprocessing stage usually involves simple but critical operations such as surface smoothing, noise removal, and hole filling. Depending on the type of the 3D sensor, the acquired facial data may contain significant amount of local surface perturbations and/or spikes. If the sensor relies on reflected light for 3D reconstruction, dark facial regions such as eyebrows and eye pupils do not produce 3D data, whereas specular surfaces scatter the light; as a result, these areas may contain holes. In addition, noise and spike removal algorithms also produce holes. These holes should be filled at the preprocessing phase.



**Fig. 9.1** Overall pipeline of a typical 3D face recognition system

After obtaining noise-free facial regions, the most important phases in the 3D face recognition pipeline are the registration and feature extraction phases. Since human faces are similar to each other, accurate registration is vital for extracting discriminative features. Face registration usually starts with acceptable initial conditions. For this purpose, facial landmarks are usually used to pre-align faces. However, facial feature localization is not an easy task under realistic conditions. Here, we survey methods that are proposed for 3D landmarking, registration, feature extraction, and matching.

### 9.3.1 Automatic Landmarking

Robust localization of facial landmarks is an important step in 2D and 3D face recognition. When guided by accurately located landmarks, it is possible to coarsely register facial images, increasing the success of subsequent fine registration.

The most frequently used approach to facial landmark detection is to devise a number of heuristics that seem to work for the experimental conditions at hand [13, 20, 51, 58, 99, 101]. These can be simple rules, such as taking the point closest to the camera as the tip of the nose [28, 102], or using contrast differences to detect eye regions [59, 102]. For a particular data set, these methods can produce very accurate results. However, for a new setting, these methods are not always applicable. Another typical approach in landmark localization is to detect the easiest landmark first and to use it in constraining the location of the next landmark [13, 20, 28]. The problem with these methods is that one erroneously located landmark makes the localization of the next landmark more difficult, if not impossible.

The second popular approach avoids error accumulation by jointly optimizing structural relationships between landmark locations and local feature constraints [87, 98]. In [98], local features are modeled with Gabor jets, and a template

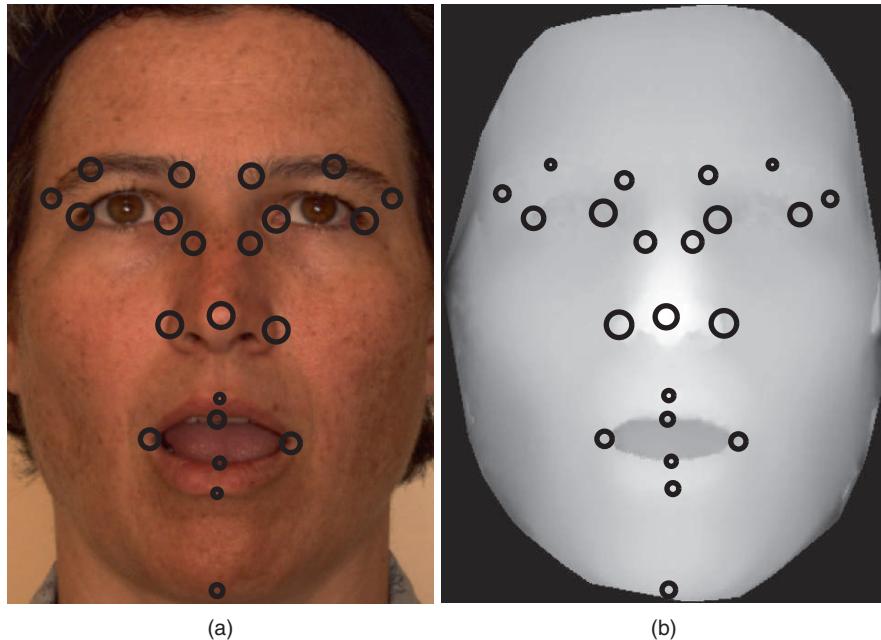
library (called *the bunch*) is exhaustively searched for the best match at each feature location. A canonic graph serves as a template for the inter-feature distances, and deviations from this template are penalized by increases in internal energy. In [46], an attentive scheme is employed to constrain the detailed search to smaller areas. A feature graph is generated from the feature point candidates, and a simulated annealing scheme is used to find the distortion relative to the canonic graph that results in the best match. A large number of facial landmarks (typically 30–40) are used for these methods and the optimization is difficult as the matching function exhibits many local minima. Most of the landmarks used in this scenario do not have sufficiently discriminating local features associated with them. For instance, landmarks along the face boundary produce very similar features.

The third approach is the adaptation of feature-based face detection algorithms to the problem of landmarking [16, 33]. Originally, these methods are aimed at finding a bounding box around the face. Their application to the problem of exact facial landmarking calls for fine-tuning steps.

The problem is no less formidable in 3D, although the prominence of the nose makes it a relatively easy candidate for fast, heuristic-based approaches. If the symmetry axis can be found, it is relatively easy to find the eye and mouth corners [51]. However, the search for the symmetry axis can be costly without the guiding landmarks. Curvature-based features seem to be promising in 3D due to their invariance to several transformations [28, 35, 58]. Especially, Gaussian and mean curvatures are frequently used to locate and segment facial parts. For instance, in [5], multi-scale curvature features are used to localize several salient points such as eye pits and nose. However, curvature-based descriptors suffer from a number of problems. Reliable estimation of curvature requires a strong preprocessing that eliminates surface irregularities, especially near eye and mouth corners. Two problems are associated with this preprocessing: the computational cost is high; and the smoothing destroys local feature information to a great extent, producing many points with similar curvature values in each local neighborhood. One issue that makes consistent landmarking difficult is that the anatomical landmarks are defined in structural relations to each other, and the local feature information is sometimes not sufficient to determine them correctly. For flat-nosed persons, the “tip of the nose” is not a point, but a whole area of points with similar curvature. More elaborate 3D methods, like spin images, are very costly in practice [30].

It is possible to use 3D information in conjunction with 2D for landmark localization [6, 20, 28]. Although the illumination sensitivity of the 2D features will have a detrimental effect on the joint model, one can use features that are relatively robust to changes in illumination. In [28], 2D Harris corners are used together with 3D shape indices. In some cases 3D is just used to constrain the 2D search [6, 20]. Under consistent illumination conditions, 2D is richer in discriminative information, but 3D methods are found to be more robust under changing illumination conditions [81].

In [81] and [82] statistical feature models are used to detect each facial feature independently on 3D range images. The advantage of this method is that no heuristics are used to tailor the detection to each landmark separately. A structural



**Fig. 9.2** The amount of statistical information available for independent detection of different landmarks in (a) 2D face images and (b) 3D face images. Marker sizes are proportional to localization accuracy (varies between 32 and 97%). The 2D images are assumed to be acquired under controlled illumination. The method given in [81] is used on Bosphorus dataset [84]

analysis subsystem is used between coarse and fine detections. Separating structural analysis and local feature analysis avoids high computational load and local minima issues faced by joint optimization approaches [83]. Figure 9.2 shows the amount of statistical information available in 2D and 3D face images for independent detection of different landmarks.

### 9.3.2 Automatic Registration

Registration of facial scans is guided by automatically detected landmarks and greatly influences the subsequent recognition. Three-dimensional face recognition research is dominated by dense registration-based methods, which establish point-to-point correspondences between two given faces. For recognition methods based on point cloud representation, this type of registration is the standard procedure, but even range image-based methods benefit from this type of registration.

Registration is potentially the most expensive phase of a 3D face recognition process. We distinguish between *rigid* and *non-rigid* registration, where the former aligns facial scans by an affine transformation, and the latter applies deformations to align facial structures more closely. For any test scan, registration needs to be performed only once for an authentication setting. For the recognition setting, the two

extreme approaches are registering the query face to all the faces in the gallery or to a single average face model (AFM), which automatically establishes correspondence with all the gallery faces which have been registered with the AFM during enrollment. In between these extremes, a few category-specific AFMs (for instance, one AFM for males and one for females) can be beneficial to accuracy and still be computationally feasible [82].

For *rigid registration*, the standard technique is the *iterative closest point* (ICP) algorithm [15]. For registering shape  $S_1$  to a coarsely aligned shape  $S_2$ , the ICP procedure first finds the closest points in  $S_2$  for all the points on  $S_1$  and computes the rotation and translation vectors that will minimize the total distance between these corresponding points. The procedure is applied iteratively, until a convergence criterion is met. Practical implementations follow a coarse-to-fine approach, where a subset of points in  $S_1$  are used initially. Once two shapes are put into dense correspondence with ICP, it is straightforward to obtain the total distance of the shapes, as this is the value minimized by ICP. This value can be employed for both authentication and recognition purposes.

Previous work on ICP show that a good initialization is necessary for fast convergence and an accurate end result. In [82], four approaches for coarse alignment to an AFM are contrasted:

1. Assume that the point with the greatest depth value is the tip of the nose, and find the translation to align it to the nose tip of the AFM. This heuristic is used in [28].
2. Use the manually annotated nose tip.
3. Use seven automatically located landmarks on the face (eye corners, nose tip, mouth corners), and use Procrustes analysis to align them to the AFM. Procrustes analysis finds a least squares alignment between two sets of landmark points and can also be used to generate a mean shape from multiple sets of landmarks [45].
4. Use seven manually annotated landmarks with Procrustes analysis.

On the FRGC benchmark dataset, it was shown that the nose tip heuristic performed the worst (resulting in 82.85% rank 1 recognition rate), followed by automatically located landmarks with Procrustes alignment (87.86%), manually annotated nose tip (90.60%) and manually annotated landmarks with Procrustes alignment (92.11%) [82]. These results also confirmed that the nose tip is the most important landmark for 3D face registration.

*Non-rigid registration techniques* have been used for registration as well as for synthesis applications. Blanz and Vetter [19] have used deformable models to register faces to a 3D model, which can then be used to synthesize faces with a specific expression and pose. The latter are subsequently used for recognition. Deformable approaches employ a common model to register faces. This face model can be conveniently annotated and thus allows automatic annotation of any face after establishing dense correspondence. It has been shown that the construction of the common model is critical for the success of the registration [82]. Many of the techniques for deformable registration employ the thin plate spline algorithm [22] to deform the surface so that a set of landmark points are brought in

correspondence [50]. Most non-rigid registration techniques in the literature (such as [43, 50, 51, 62, 90]) are derived from the work of Bookstein on thin plate spline (TPS) models [21]. This method simulates the bending of a thin metal plate that is fixed by several anchor points. For a set of such points  $P_i = (x_i, y_i), i = 1 \dots n$ , the TPS interpolation is a vector-valued function  $f(x, y) = [f_x(x, y), f_y(x, y)]$  that maps the anchor points to their specified homologues  $P'_i = (x'_i, y'_i), i = 1 \dots n$  and specifies a surface which has the least possible bending, as measured by an integral bending norm. The mapping for the anchor points (i.e., specified landmarks on the facial surface) is exact, whereas the rest of the points are smoothly interpolated. This type of registration strongly depends on the number and accuracy of landmarks.

If the landmark set is large, all surfaces will eventually resemble the AFM and lose their individuality. To avoid this problem, Mao et al. [62] deform the AFM rather than the individual facial surfaces. Tena et al. [90] optimize this algorithm with the help of facial symmetry and multiresolution analysis. A compromise between individuality and surface conformance is achieved through the minimization of an energy function combining internal and external forces [54]. The success of the fitting is highly dependent upon the initial pose alignment, the construction of the AFM, and the mesh optimization. Kakadiaris et al. [53] use an anthropomorphically correct AFM and obtain a very good accuracy after careful optimization.

### 9.3.3 Feature Extraction and Matching

Feature extraction techniques usually depend on the particular registration method employed. As explained in the previous section, most registration approaches register facial surfaces onto a common model (the AFM), which serves to facilitate dense correspondence between facial points. The one-to-all ICP technique fails to do that: Surface pairs are registered to each other rather than to a common model. Therefore, the ICP error, which serves as a measure of how well the surfaces match, is directly usable for matching. Many early systems for 3D face recognition use this convention [12, 23, 65]. Some representations such as the point signatures [27, 95], spin images [96], or histogram-based approaches [100] are special in that they do not require prior registration. In the rest of this section, we will assume that the surfaces are densely registered and a dense one-to-one mapping exists between facial surfaces.

The point cloud feature, which is simply the set of 3D coordinates of surface points of densely registered faces, is the simplest feature one can use, and the point cloud and point set difference, used directly as a feature, are analogous to the ICP error. Principal component analysis (PCA) has been applied to the point cloud feature by [72, 79].

Geometrical features rely on the differences between facial landmark points located on the facial surfaces [57, 78]. The number of landmark points used in such systems is quite variable: ranging from 19 [72] to as many as 73 [37]. Riccio and Dugelay [77] use 3D geometrical invariants derived from MPEG4 feature points.

Facial surfaces are often called 2.5D data since there exists only one  $z$ -value for a given  $(x, y)$  pair. Therefore, a unique projection along the  $z$ -axis provides a unique depth image, sometimes called a range image, which can then be used to extract features. Common feature extraction techniques are subspace projection techniques such as PCA, linear discriminant analysis (LDA), independent component analysis (ICA), discrete Fourier transform (DFT), discrete cosine transform (DCT), or nonnegative matrix factorization (NNMF) [36]. Many researchers have applied these standard techniques [26, 42, 48], as well as proposed other techniques such as discriminant common vectors [106]. Other 2D face feature extraction techniques are also applicable [32, 80]. Most of the statistical feature extraction-based methods treat faces globally. However, it is sometimes beneficial to perform local analysis, especially under adverse situations. For instance, in [36], authors perform local region-based DCT analysis on the depth images and construct final biometric templates by concatenating local DCT features. Similarly, DCT features derived from overlapping local windows placed over the upper facial region are employed in [64].

While depth images rely on a projection to obtain an image from a surface, one can intersect the surface with planes to generate a set of curves [17, 104]. One-dimensional curve representation techniques can then be applied to represent the curves.

Curvature-based surface descriptors are among the most successful 3D surface representations. They have been used for facial surface segmentation [69] as well as representation [89]. Commonly used descriptors are maximum and minimum principal directions [89], and normal maps [2, 3]. Kakadiaris et al. [52] have fused Haar and pyramid features of normal maps. Gökberk et al. [42] have used shape indices, principal directions, mean and Gaussian curvatures and have concluded that principal directions perform the best.

The combination of different representations has also attracted widespread interest. Some approaches use feature level fusion: A typical example is given in [71], where shape and texture information are merged at the point cloud level, thus producing 4D point features. Osaimi et al. [8] fuse local and global fields in a histogram.

Score fusion is more commonly used to combine shape and texture information [41, 44]. Tsalakanidou et al. [91, 92] propose a classic approach, where shape and texture images are coded using PCA and their scores are fused at the decision level. Malassiotis and Strintzis [61] use an embedded hidden Markov model-based (EHMM) classifier that produces similarity scores. Chang et al. [26] use PCA-based matchers for shape (depth image) and texture modalities. In both cases, the outputs of the matchers are fused by a weighted sum rule. BenAbdelkader and Griffin [14] concatenate depth image pixels with texture image pixels for data level fusion. Linear discriminant analysis (LDA) is then applied to the concatenated feature vectors to extract features. Wang and Chua [94] select 2D Gabor wavelet features as local descriptors for the texture modality, use point signatures as local 3D shape descriptors, and use score-level fusion using weighted sum rule.

A two-level sequential combination idea was used in [60] for 2D texture images, where the ICP-based surface matcher eliminates the unlikely classes at the first round; and at the second round, LDA analysis is performed on the texture information to finalize the identification.

To deal with degradation in the recognition performance due to facial expressions, part-based classification approaches are considered, where the similarity scores from individual classifiers are fused for the final classification [9, 10].

In [56], the sign of mean and Gaussian curvatures are calculated at each point for a range image, and these values are used to segment a face into convex regions. Extended Gaussian images (EGI) corresponding to each region are created and correlation between EGIs is used for regional classification. In [69], Moreno et al. segment the 3D facial surface using mean and Gaussian curvatures and extract various descriptors for each segment. Cook et al. [31] use Log-Gabor templates (LGT) on range images and divide a range image into 147 regions. Classification is handled by fusing the scores of each individual classifier.

In [25], Chang et al. use multiple overlapping regions around the nose area. Individual regional surfaces are registered with ICP and the regional similarity measures are fused with sum, min, or product rules. In [38], Faltemier et al. extend the use of multiple regions of [25] and utilize seven overlapping nose regions. ICP is used for individual alignment of facial segments. Threshold values determined for regions are utilized in committee-voting fusion approach. In [39], the work of Faltemier et al. is expanded to utilize 38 regions segmented from the whole facial surface. The regional classifiers based on ICP alignment are fused with the modified Borda count method.

In [52], a deformable facial model is used to describe a facial surface. The face model is segmented into regions and after the alignment, 3D geometry images and normal maps are constructed for regions of test images. The regional representations are analyzed with a wavelet transform and individual classifiers are fused with a weighted sum rule. Deformation information can also be used as a face descriptor. Instead of allowing deformations for better registration, that deformation field may uniquely represent a person. Zou et al. [107] follow this approach by selecting several prototype faces from the gallery set and then learn the *warping space* from the training set. A given probe face is then warped to a generic face template, where the warping parameters found at this stage are linear combinations of the previously learned warpings.

Deformation invariance can be accomplished with the use of *geodesic distances* [24, 70]. It has been shown that the geodesic distance between two points over the facial surface does not change significantly when facial surface deforms slightly [24]. In [70], facial surface is represented using geodesic polar parametrization to cope with facial deformations. When a face is represented by geodesic polar coordinates, intrinsic properties are preserved and a deformation invariant representation is obtained. Using this representation, the face is assumed to contain 2D information embedded in 3D space. For recognition, 2D PCA classifiers in color and shape space are fused.

Mian et al. [67] extract inflection points around the nose tip and utilize these points for segmenting the face into forehead–eye and nose regions. The individual regions are less affected under expression variations and separately matched with ICP. The similarity scores are fused at the metric level. In order to handle the time complexity problem of matching a probe face to every gallery face, authors propose a rejection classifier that eliminates unlikely classes prior to region-based ICP-matching algorithm. The rejection classifier consists of two matchers: the first one uses spherical histogram of point cloud data for the 3D modality (spherical face representation, SFR); and the second one employs scale-invariant feature transform-based (SIFT) 2D texture features. By fusing each matcher’s similarity scores, the rejection classifier is able to eliminate 97% of the gallery faces, which speeds up the ICP-based matching at the second phase.

### 9.3.3.1 Evaluation Campaigns for 3D Face Recognition

We have seen that there are many alternatives at each stage of a 3D face recognition system: The resulting combinations present abundant possibilities. Many 3D face recognition systems have been proposed over the years and performance has gradually increased to rival the performance of 2D face recognition techniques. Table 9.2 lists commonly used 3D face databases together with some statistics (such as the number of subjects and the total number of 3D scans present). In the presence of literally hundreds of alternative systems, independent benchmarks are needed to evaluate alternative algorithms and to assess the viability of 3D face against other biometric modalities such as high-resolution 2D faces, fingerprints, and iris scans. Face Recognition Grand Challenge (FRGC) [75] and Face Recognition Vendor Test 2006 (FRVT’06) [76] are the two important evaluations, where the 3D face modality is present.

#### Face Recognition Grand Challenge

FRGC is the first evaluation campaign that focuses expressly on face: 2D face at different resolutions and illumination conditions and 3D face, alone or in combination with 2D [74, 75]. The FRGC data corpus contains 50,000 images, where the 3D part is divided into two sets: *development set* (943 images) and *evaluation set* (4007 images collected from 466 subjects). The evaluation set is composed of *target* and *query images*. Face images in the target set are to be used for enrollment, whereas face images in the query set represent the test images. Faces were acquired under controlled illumination conditions using a Minolta Vivid 900/910 sensor, which is a structured light sensor with a range resolution of  $640 \times 480$  and produces a registered color image at the same time.

FRGC has three sets of 3D verification experiments: shape and texture together (Experiment 3), shape only (Experiment 3s), and texture only (Experiment 3t). The baseline algorithm for the 3D shape+texture experiment uses PCA applied to the shape and texture channels separately, the scores of which are fused to

**Table 9.2** List of popular 3D face databases. The UND database is a subset of the FRGC v.2. Pose labels: L: left, R: right, U: up, and D: down

Database	Subject count	Sample count	Total scans	Expressions	Pose
ND2006 [40]	888	1–63	13450	Neutral, happiness, sadness, surprise, disgust, other	–
York [47]	350	15	5250	Happy, angry, eyes closed, eyebrows raised	U,D
FRGC v.2 [75]	466	1–22	4007	Angry, happy, sad, surprised, disguised, and puffy	–
BU-3DFE [103]	100	25	2500	Happiness, disgust, fear, angry, surprise, sadness (four levels)	–
CASIA [105]	123	15	1845	Smile, laugh, anger, surprise and closed eyes	–
UND [93]	275	1–8	943	Smile	–
3DRMMA [18]	120	6	720	–	L,R,U,D
GavabDB [68]	61	9	549	Smile, frontal laugh, frontal random gesture	L,R,U,D
Bosphorus [84, 85]	81	31–53	3396	34 expressions (28 AUs and 6 emotional expressions including happiness, surprise, fear, sadness, anger, disgust)	13 poses
BUT-3D [1]	500	1	500	–	–
Extended M2VTS database [66]	295	4	1180	–	–
MIT-CBCL [97]	10	324	3240	–	Some pose variations
ASU [88]	117	5–10	421	Smile, anger, surprise	–
FRAV3D [30]	106	16	1696	Smile, open mouth	8 poses

obtain the final scores. At the FAR rate of 0.1%, verification rate of the baseline system is found to be 54%. The best reported performance is 97% at FAR rate of 0.1% [74, 75]. Table 9.3 summarizes the results of several published papers in the literature for an FAR rate of 0.001% using the FRGC v.2 database. The FRGC 3D experiments have shown that the individual performance of the texture channel is better than the shape channel. However, fusing shape and texture channels together always results in better performance. Comparing 2D and 3D, high-resolution 2D images obtain slightly better verification rates than the 3D modality. However, at low resolution and extreme illumination conditions, 3D has a definite advantage.

#### Face Recognition Vendor Test 2006

The FRVT 2006 is an independent large-scale evaluation campaign that aims to look at performance of high-resolution 2D and 3D modalities [76] together with other modalities. The competition was open to academia and companies. The objectives of the FRVT 2006 tests were to compare face recognition performance to that of top-performing modalities. Another objective was to compare the performance to that of face recognition by humans.

Submitted algorithms were tested on sequestered data collected from 330 subjects (3,589 3D scans). The participants of the 3D part were *Cognitec*, *Viisage*, *Tsinghua University*, *Geometrics*, and *University of Houston*. The best performers for the 3D modality have an FRR interquartile range of 0.005–0.015 at an FAR of 0.001 for the *Viisage* normalization algorithm and an FRR interquartile range of 0.016–0.031 at an FAR of 0.001 for the *Viisage* 3D one-to-one algorithm. In FRVT 2006, it has been concluded that (1) 2D, 3D, and iris biometrics are all comparable in terms of verification rates, (2) there is a decrease in the error rate by at least an order of magnitude over what was observed in the FRVT 2002. This decrease in error rate was achieved by still and 3D face recognition algorithms, and (3) at low false alarm rates for humans, automatic face recognition algorithms were comparable to or better than humans in recognizing faces under different illumination conditions.

**Table 9.3** Verification rates in % of various algorithms at FAR rate of 0.001% on the FRGC v.2 dataset

System	Neutral vs all		neutral vs neutral		neutral vs non-neutral	
	3D	3D+2D	3D	3D+2D	3D	3D+2D
Mian et al. [67]	98.5	99.3	99.4	99.7	97.0	98.3
Kakadiaris et al. [52]	95.2	97.3	NA	99.0	NA	95.6
Husken et al. [49]	89.5	97.3	NA	NA	NA	NA
Maurer et al. [63]	86.5	95.8	97.8	99.2	NA	NA
FRGC baseline	45.0	54.0	NA	82.0	40.0	43.0

## 9.4 Challenges and a Case Study

### 9.4.1 Challenges

The scientific work of the last 20 years on 3D face recognition, the large evaluation campaigns organized, and the abundance of products available in the market all suggest that 3D face recognition is becoming available as a viable biometric identification technology. However, there are many technical challenges to be solved for 3D face recognition to be used widely in all application scenarios mentioned in the beginning of the chapter. The limitations can be grouped as follows:

**Restrictions due to scanners:** The first important restriction is cost: Reliable 3D face recognition still requires a high-cost, high-precision scanner; and that restricts its use to only very limited applications. A second limitation is the acquisition environment: current scanners require the object to stand at a fixed distance away, with controlled pose. Furthermore, most scanners require that the subject be motionless for a short time, since acquisition usually takes some time. As scanners get faster, not only will this requirement be relaxed, but other modes, such as 3D video will become available.

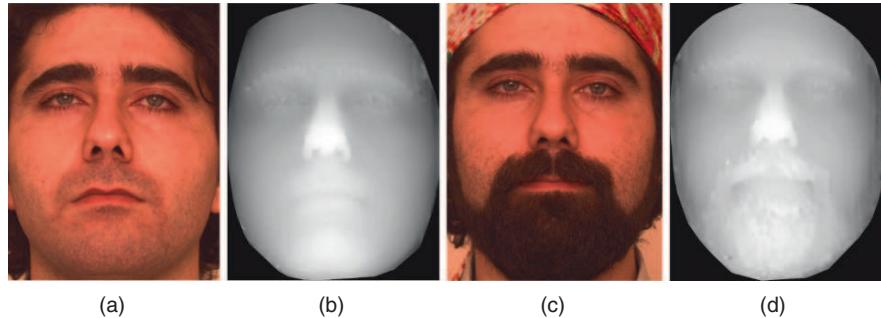
**Restrictions due to algorithms:** Most studies have been conducted on datasets acquired in controlled environments, with controlled poses and expressions. Some datasets have incorporated illumination variances and some have incorporated varying expressions. However, until recently, there was no database with joint pose, expression, and illumination differences and no studies on robust algorithms to withstand all these variations. There is almost no work on occlusions caused by glasses and hair and on surface irregularities caused by facial hair. In order for ubiquitous 3D face recognition scenarios to work, the recognition system should incorporate:

- 3D face detection in a cluttered environment
- 3D landmark detection and pose correction
- 3D face recognition under varying facial deformations
- 3D face recognition under occlusion

FRVT2006 has shown that significant progress has been made in dealing with external factors such as illumination and pose. The internal factors are now being addressed as well: In recent years, significant research effort has focused on expression-invariant 3D face recognition, and new databases that incorporate expression variations have become available. The time factor and deception attacks are yet to be addressed. Here, we point to the outstanding challenges and go over a case study for expression-invariant face recognition.

#### 9.4.1.1 Challenge 1: How to Deal with Changes in Appearance in Time

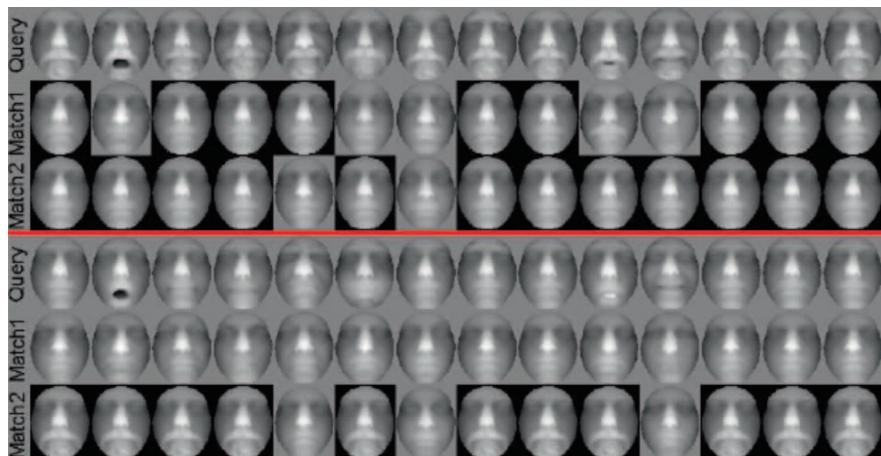
The first factor that comes to mind when time is mentioned is naturally occurring ageing. A few attempts have been made to model ageing [55]. However, intentional



**Fig. 9.3** 2D and 3D images of a subject with and without beard and corresponding depth images

or cosmetic attempts to change the appearance pose serious challenges as well: Beards, glasses, hairstyle, and makeup all hamper the operation of face recognition algorithms. Assume the following case where we have a subject that grows a beard from time to time. Figure 9.3 shows sample 2D and 3D images of a person with or without beard. The gallery image of the subject can be bearded or not, but these cases are not symmetrical.

Figure 9.4 shows the matching results for an experiment conducted with a 48-image gallery from the Bosphorus database enhanced with such a subject. In the first case (the first three lines), the query is bearded and the gallery image is not. Depending on the expression of the query (the first line), the correct image is mostly located in the gallery (the second line). The third line gives rank-1 match for an eye region-based registration and matching, and it is more successful. The second batch



**Fig. 9.4** The effect of beard for 3D face recognition. For each experiment, three lines of images are given: the query, rank-1 matched face with a complete facial matching, and rank-1 matched face with an eye region-based matching. The correct matches are shown with black borders. See text for more details.

of experiments tells a different story. Now the gallery image is bearded, whereas the query (the fourth line) is not. This time, the full-scan registration fails to retrieve the correct gallery image for the whole range of queries (the fifth line). The reason for this failure is the change in query image–gallery image distances. The total distance between the bearded and non-bearded images of the subject does not change, but it is large when compared to the distance between a non-bearded query image and a non-bearded gallery image belonging to a different subject. Thus, in the first experiment, the query produces large distances to all gallery images, from which the correct one can be retrieved; but in the second experiment, non-bearded false positives dominate because of their generally smaller distances. Subsequently, it is better to have the non-bearded face in the gallery. Alternatively, bearded subjects can be pre-identified and matched with region-based methods. The sixth line of Fig. 9.4 shows that the eye region-based matcher correctly identifies the query in most of the cases.

This experiment demonstrates the necessity of carefully controlled and analyzed experimental conditions for testing appearance changes.

#### 9.4.1.2 Challenge 2: How to Deal with Internal Factors

The facial surface is highly deformable. It has a single joint, the jaw, which is used to open the mouth; and sets of facial muscles that are used to open and close the eyes and the mouth, and to move the facial surface. The principal objective of mouth movements is speech and the secondary objective of all facial deformations is to express emotions. Face recognition has largely ignored movement and assumed that the face is still. In recent years, many researchers have focused on expression-invariant face recognition.

Here, we present a case study showing how expressions change the facial surface on a database collected for this purpose, and an example system designed to deal with these variations.

### 9.4.2 A Case Study

We have outlined two challenges above: Dealing with internal variations such as facial expressions and dealing with deception attacks, especially occlusions. To develop robust algorithms that can operate under these challenges, one needs to work with a special database that includes both a vast range of expression changes and occlusions. In this section, we will first introduce a database collected for these purposes and then describe an example part-based system that is designed to deal with variations on the facial surface.

#### 9.4.2.1 Bosphorus DB

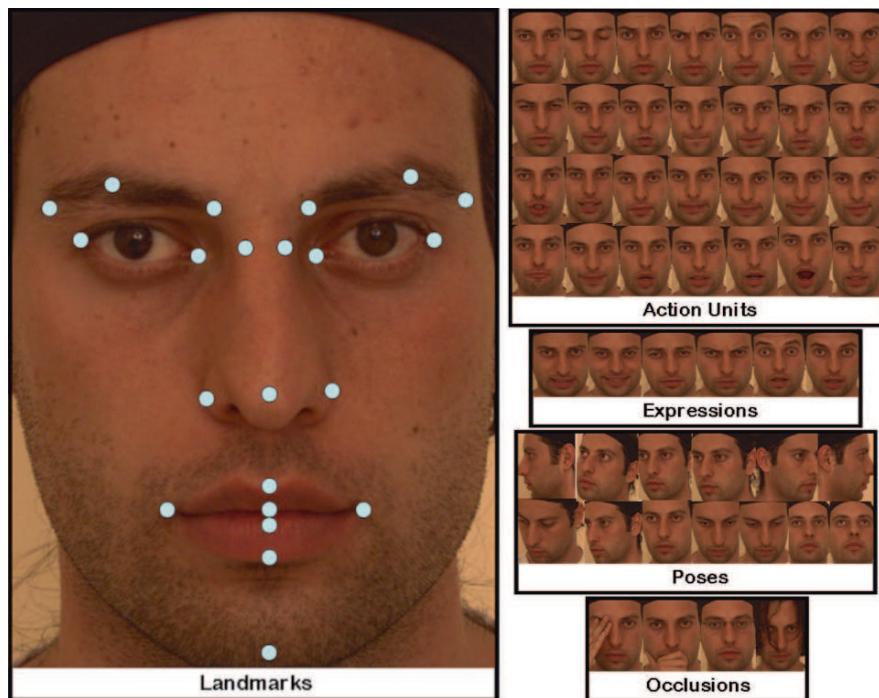
The Bosphorus database is a 2D–3D face database including extreme and realistic expression, pose, and occlusion variations that may occur in real life [84, 85].

For facial data acquisition, a structured light-based 3D digitizer device, Inspec Megacapturor II 3D, is utilized. During acquisition, vertical straight lines are projected on the facial surface, and the reflections are used for information extraction. For 3D model reconstruction, a region of interest including the central facial region is manually selected, thus the background clutter is removed.

The 3D sensor has 0.3, 0.3, and 0.4 mm sensitivity in  $x$ ,  $y$ , and  $z$ -axes respectively, and a typical preprocessed scan consists of approximately 35 K points. The texture images are of high resolution ( $1600 \times 1200$ ) with perfect illumination conditions.

After the reconstruction and preprocessing phases, 22 fiducial points have been manually labeled on both 2D and 3D images, as shown in Fig. 9.5.

The Bosphorus database contains a total of 3396 facial scans acquired from 81 subjects, 51 men and 30 women. Majority of the subjects are Caucasian and aged between 25 and 35. The Bosphorus database has two parts: the first part, Bosphorus v.1, contains 34 subjects and each of these subjects has 31 scans: 10 types of expressions, 13 different poses, 4 occlusions, and 4 neutral/frontal scans. The second part, Bosphorus v.2, has more expression variations. In the Bosphorus v.2, there are 47 subjects, each subject having 34 scans for different expressions including 6 emotional expressions and 28 facial action units, 13 scans for pose variations,



**Fig. 9.5** Manually located landmark points and typical variations for the Bosphorus database

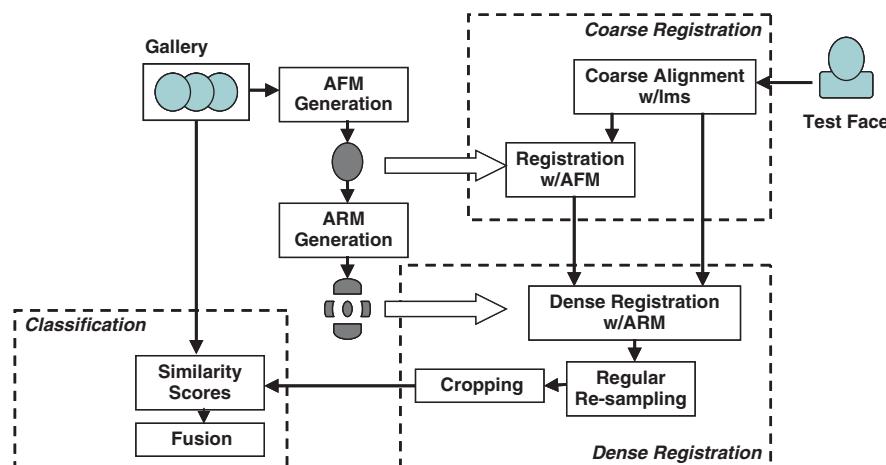
4 occlusions, and 1 or 2 frontal/neutral faces. Thirty of these 47 subjects are professional actors/actresses. Figure 9.5 shows the total scan variations included in the Bosphorus v.2.

#### 9.4.2.2 Example System

The rigid registration approaches are highly affected by facial expression diversities [9–11]. To deal with deformations caused by expressions, we apply rigid registration in a regional manner. Registering all gallery faces to a common AFM off-line decreases run time cost. Motivated from this approach, we proposed to use regional models for component-based dense registration. The average regional models (ARMs) are constructed by manually segmenting an AFM. These ARMs are used for indexing the regions on gallery and probe faces. After regions are extracted, the facial segments can be used for recognition.

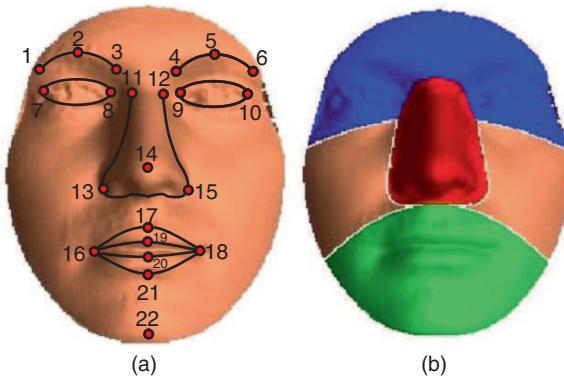
**ARM-based registration:** In regional registration, each region is considered separately when aligning two faces. For fast registration, we have adapted the AFM-based registration for regions, where regional models act as index files. ARMs are obtained by manually segmenting a whole facial model. The average model is constructed using the gallery set. In this study, we have divided the face into four basic logical regions: forehead–eyes, nose, cheeks, mouth–chin. In Fig. 9.7, the AFM for the Bosphorus v.1 database and the constructed ARMs are given.

In ARM-based registration, a test face is registered individually to each regional model and the related part is labeled and cropped. Registering a test face to the whole gallery consists of four individual alignments, one specific for each region.



**Fig. 9.6** The outline of the proposed system which consists of several steps: facial model construction, dense registration, coarse registration, and classification

**Fig. 9.7** (a) AFM for the Bosphorus v.1 gallery set and (b) four ARMs for forehead–eyes, nose, cheeks, and mouth–chin regions



**Part-based recognition:** After registering test faces to ARMs, the cropped 3D point clouds are regularly re-sampled, hence the point set difference calculation is reduced to a computation between only the depth vectors.

As a result of point set difference calculations, four dissimilarity measure sets are obtained to represent distance between gallery and test faces. Each regional registration is considered as an individual classifier and fusion techniques are applied to combine regional classification results. In this study, we have utilized various fusion approaches from different levels: standard and modified plurality voting at the abstract level; sum rule and product rule at the score level [42].

In plurality voting (PLUR), each classifier votes for the nearest gallery identity and the identity with the highest vote is assigned as the final label. When there are ties among closest classes, the final label is assigned randomly among these class labels. In modified plurality voting, each classifier votes for the nearest gallery identity and the identity with the highest vote is assigned as the final label. A value to define the confidence of a classifier is also present and when there are ties, the label of the class with the highest confidence is chosen as the final decision. More details on confidence-aided fusion methods can be found in [41, 43, 44].

At the score level, SUM and PRODUCT rules are tested, where similarity scores of individual classifiers are fused using simple arithmetic operations. For these approaches, the scores are normalized with the min–max normalization method prior to fusion.

**Experiments:** In our experiments, we have utilized both v.1 and v.2 of the Bosphorus database. For each version, we have grouped the first neutral scan of each subject into the gallery. The scans containing expression and AU variations are grouped into the probe set. The number of scans in each gallery and probe set is given in Table 9.4. It is observed that when expressions are present, the baseline AFM-based classifier's performance drops by about 30%.

The ICP registration is greatly affected by the accuracy of the coarse alignment of surfaces. To analyze the effect, we have proposed two different coarse alignment approaches for the ARM-based registration. The first method is referred to as *the one-pass registration*, where coarse alignment of facial surfaces is handled

**Table 9.4** Gallery and probe sets for Bosphorus DB v.1 and v.2. AFM-based ICP accuracies are also shown for each version

Bosphorus		Gallery	Probe	AFM results (%)
v.1	Neutral scans	34	102	100.00
	Expression scans	—	339	71.39
	Neutral scans	47	—	—
v.2	Expression scans	—	1508	67.67

by Procrustes analysis of 22 manual landmarks. In the second approach, namely *the two-pass registration*, dense alignment with the AFM is obtained before registering with the regional models. In Table 9.5, recognition accuracies obtained for ARM-based registration using these approaches are given. As observed in the results, when expression diversity is large, as in v.2, better results are obtained by utilizing the two-pass registration.

As the results in Table 9.5 exhibit, the nose and forehead–eyes regions are less affected by deformations caused by facial expressions and, therefore, these regional classifiers yield better results. However, different expressions affect different facial regions, and fusing the results of all regions always yields better results than relying on a single region. Table 9.6 shows the results of fusion using different fusion rules on scores obtained by the two-pass registration. It is observed that the best performance is achieved by the product rule, which is greater than 95% for both datasets. Compared to the accuracy of the best regional classifier (the nose region), fusion improves the system performance by about 10%.

The accuracy of the MOD-PLUR method, which utilizes the classifier confidences, follows the performance of the product rule. The second score-level fusion method we have used, the sum rule, does not perform as good as the product rule or the confidence-aided fusion schemes. The accuracy of the sum rule can be improved by weighting the effect of regional classifiers. For the weighted sum rule, the weights are calculated from an independent set: We have used Bosphorus v.1 set to calculate the weights and tested the fusion algorithm on the v.2 set. The optimal weights calculated from the v.1 database are:  $w_{\text{nose}} = 0.40$ ,  $w_{\text{eye}} = 0.30$ ,  $w_{\text{cheek}} = 0.10$ , and  $w_{\text{chin}} = 0.20$ . Due to the weights chosen, nose and forehead–eyes regions have greater contribution to total recognition performance. The contribution of weighting scheme is reflected in the SUM rules' performance increase from 88.79 to 93.51%, and from 91.78 to 93.50% for v.1 and v.2 sets, respectively.

**Table 9.5** Comparison of coarse alignment approaches

ARM	One pass		Two pass	
	v.1	v.2	v.1	v.2
Forehead–eyes	82.89	82.16	82.89	83.09
Nose	85.55	82.23	85.84	83.95
Cheeks	53.39	52.12	54.57	51.72
Mouth–chin	42.48	34.55	45.72	34.95

**Table 9.6** Recognition rates (%) for fusion techniques

Fusion method	v.1	v.2
MOD-PLUR	94.40	94.03
SUM	88.79	91.78
Weighted SUM	93.51	93.50
PROD	<b>95.87</b>	<b>95.29</b>

## 9.5 Conclusions

Three-dimensional face recognition has matured to match the performance of 2D face recognition. When used together with 2D, it makes the face modality a very strong biometric: Face as a biometric modality is widely acceptable for the general public, and face recognition technology is able to meet the accuracy demands of a wide range of applications.

While the accuracy of algorithms have met requirements in controlled tests, 3D face recognition systems have yet to be tested in real application scenarios. For certain application scenarios, such as airport screening and access control, systems are being tested in the field. The algorithms in these application scenarios will need to be improved to perform robustly under time changes and uncooperative users. For other application scenarios, such as convenience and consumer applications, the technology is not yet appropriate: The sensors should get faster, cheaper, and less intrusive; and algorithms should adapt to the new sensor technologies to yield good performance with coarser and noisier data.

One property of 3D face recognition sets it apart from other biometric modalities: It is inherently a multimodal biometric, comprising texture and shape. Therefore, a lot of research effort has gone into the fusion of 2D and 3D information. There are yet areas to be explored in the interplay of 2D and 3D: How to obtain one from the other; how to match one to the other, and how to use one to constrain the other. In the future, with the widespread use of 3D video, the time dimension will open new possibilities for research, and it will be possible to combine 3D face with behavioral biometrics expressed in the time dimension.

**Acknowledgments** This research is supported by EU COST 2101 and the Dutch BSIK/BRICKS projects.

## Proposed Questions and Exercises

- What are the advantages of 3D over 2D for face recognition, and vice versa? Would a 2D+3D system overcome the drawbacks of each of these systems, or suffer under all these drawbacks?
- Consider the five scenarios presented in the first section. What are the security vulnerabilities for each of these scenarios? How would you overcome these vulnerabilities?

- Propose a method for a 3D face-based biometric authentication system for banking applications. Which sensor technology is appropriate? How would the biometric templates be defined? Where would they be stored? What would be the processing requirements?
- Discuss the complexity of an airport security system, in terms of memory size and processing load, under different system alternatives.
- If the 3D data acquired from a sensor is noisy, what can be done?
- How many landmark points are needed for a 3D face recognition system? Where would they be chosen?
- What are the pros and cons of deformable registration vs rigid registration?
- Propose an analysis-by-synthesis approach for 3D face recognition.
- Is feature extraction possible before registration? If yes, propose a method.
- Suppose a 3D face recognition system represents shape features by mean and Gaussian curvatures, and texture features by Gabor features. Which fusion approach is appropriate? Data level or decision level fusion? Discuss and propose a fusion method.
- What would you do to deceive a 3D face recognizer? What would you add to the face recognition system to overcome your deception attacks?

## References

1. *The BJUT-3D Large-Scale Chinese Face Database, MISKL-TR-05-FMFR-001*, 2005.
2. A.F. Abate, M. Nappi, S. Ricciardi, and G. Sabatino. Fast 3D face recognition based on normal map. In *IEEE Int. Conf. on Image Processing*, pages 946–949, 2005.
3. A.F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 3D face recognition using normal sphere and general Fourier descriptor. In *Proc. Int. Conf. on Pattern Recognition*, 2006.
4. A.F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2D and 3D face recognition: A survey. *Pattern Recognition Letters*, 28:1885–1906, 2007.
5. E. Akagündüz and I. Ulusoy. 3D object representation using transform and scale invariant 3D features. In *Int. Conf. on Computer Vision*, pages 1–8, 2007.
6. H.Ç. Akakin, A.A. Salah, L. Akarun, and B. Sankur. 2D/3D facial feature extraction. In *Proc. SPIE*, volume 6064, pages 441–452. SPIE, 2006.
7. L. Akarun, B. Gökberk, and A.A. Salah. 3D face recognition for biometric applications. In *Proc. European Signal Processing Conference*, Antalya, Turkey, 2005.
8. F.R. Al-Osaimi, M. Bennamoun, and A. Mian. Integration of local and global geometrical cues for 3D face recognition. *Pattern Recognition*, 41(2):1030–1040, 2008.
9. N. Alyüz, B. Gökberk, and L. Akarun. A 3D face recognition system for expression and occlusion invariance. *IEEE Second Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS'08)*, 2008.
10. N. Alyüz, B. Gökberk, H. Dibeklioğlu, and L. Akarun. Component-based registration with curvature descriptors for expression insensitive 3D face recognition. *8th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2008.
11. N. Alyüz, B. Gökberk, H. Dibeklioğlu, A. Savran, A.A. Salah, L. Akarun, and B. Sankur. 3D face recognition benchmarks on the Bosphorus database with focus on facial expressions. In *Proc. First COST 2101 Workshop on Biometrics and Identity Management (BIOID)*, Denmark, May 2008.
12. B.B. Amor, M. Ardabilian, and L. Chen. New experiments on ICP-based 3D face recognition and authentication. In *Int. Conf. on Pattern Recognition*, 2006.

13. S. Arca, P. Campadelli, and R. Lanzarotti. A face recognition system based on automatically determined facial fiducial points. *Pattern Recognition*, 39:432–443, 2006.
14. C. BenAbdelkader and P.A. Griffin. Comparing and combining depth and texture cues for face recognition. *Image and Vision Computing*, 23(3):339–352, 2005.
15. P. Besl and N. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.
16. G.M. Beumer, Q. Tao, A.M. Bazen, and R.J.N. Veldhuis. A landmark paper in face recognition. *Proc. 7th Int. Conf. on Automatic Face and Gesture Recognition*, pages 73–78, 2006.
17. C. Beumier and M. Achery. Automatic 3D face authentication. *Image and Vision Computing*, 18(4):315–321, 2000.
18. C. Beumier and M. Achery. Face verification from 3D and grey level cues. *Pattern Recognition Letters*, 22:1321–1329, 2001.
19. V. Blanz and T. Vetter. Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.
20. C. Boehnen and T. Russ. A fast multi-modal approach to facial feature detection. In *Proc. 7th IEEE Workshop on Applications of Computer Vision*, pages 135–142, 2005.
21. F.L. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:567–585, 1989.
22. F.L. Bookstein. Shape and the information in medical images: A decade of the morphometric synthesis. *Computer Vision and Image Understanding*, 66(2):97–118, 1997.
23. K. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition. *Computer Vision and Image Understanding*, 101:1–15, 2006.
24. A.M. Bronstein, M.M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, 64(1):5–30, 2005.
25. K. I. Chang, K.W. Bowyer, and P.J. Flynn. Adaptive rigid multi-region selection for handling expression variation in 3D face recognition. In *2005 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'05)*, pages 157–164, 2005.
26. K.I. Chang, K.W. Bowyer, and P.J. Flynn. An evaluation of multi-modal 2D+3D face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):619–624, 2005.
27. C.S. Chua, F. Han, and Y.K. Ho. 3D human face recognition using point signature. In *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pages 233–238, 2000.
28. D. Colbry, G. Stockman, and A.K. Jain. Detection of anchor points for 3D face verification. In *Proc. IEEE Workshop on Advanced 3D Imaging for Safety and Security*, 2005.
29. A. Colombo, C. Cusano, and R. Schettini. 3D face detection using curvature analysis. *Pattern Recognition*, 39(3):444–455, 2006.
30. C. Conde, A. Serrano, L.J. Rodríguez-Aragón, and E. Cabello. 3D facial normalization with spin images and influence of range data calculation over face verification. In *IEEE Conf. Computer Vision and Pattern Recognition*, 2005.
31. J. Cook, V. Chandran, and C. Fookes. 3D face recognition using log-Gabor templates. In *Biritch Machine Vision Conference*, pages 83–92, 2006.
32. J. Cook, V. Chandran, S. Sridharan, and C. Fookes. Gabor filter bank representation for 3D face recognition. In *Proc. Digital Imaging Computing: Techniques and Applications*, 2005.
33. D. Cristinacce and T.F. Cootes. Facial feature detection and tracking with automatic template selection. In *Proc. 7th Int. Conf. on Automatic Face and Gesture Recognition*, pages 429–434, 2006.
34. K. Delac and M. Grgic. *Face Recognition*. I-Tech Education and Publishing, Vienna, Austria, 2007.
35. H. Dibeklioğlu, A.A. Salah, and L. Akarun. 3D facial landmarking under expression, pose, and occlusion variations. *IEEE Second Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS'08)*, 2008.

36. H.K. Ekenel, H. Gao, and R. Stiefelhagen. 3-D face recognition using local appearance-based models. *IEEE Transactions on Information Forensics and Security*, 2(3):630–636, 2007.
37. A.H. Eraslan. 3D universal face-identification technology: Knowledge-based composite-photogrammetry. In *Biometrics Consortium*, 2004.
38. T. Faltemier, K.W. Bowyer, and P.J. Flynn. 3D face recognition with region committee voting. In *Proc. 3DPVT*, pages 318–325, 2006.
39. T. Faltemier, K.W. Bowyer, and P.J. Flynn. A region ensemble for 3D face recognition. *IEEE Transactions on Information Forensics and Security*, 3(1):62–73, 2007.
40. T. Faltemier, K.W. Bowyer, and P.J. Flynn. Using a multi-instance enrollment representation to improve 3D face recognition. In *Proc. of Biometrics: Theory, Applications, and Systems, (BTAS)*, pages 1–6, 2007.
41. B. Gökberk and L. Akarun. Comparative analysis of decision-level fusion algorithms for 3D face recognition. In *Proc. Int. Conf. on Pattern Recognition*, pages 1018–1021, 2006.
42. B. Gökberk, H. Dutagaci, A. Ulaş, L. Akarun, and B. Sankur. Representation plurality and fusion for 3D face recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 38(1):155–173, 2008.
43. B. Gökberk, M.O. İrfanoğlu, and L. Akarun. 3D shape-based face representation and feature extraction for face recognition. *Image and Vision Computing*, 24(8):857–869, 2006.
44. B. Gökberk, A.A. Salah, and L. Akarun. Rank-based decision fusion for 3D shape-based face recognition. In T. Kanade, A. Jain, and N.K. Ratha, editors, *Proc. Int. Conf. on Audio- and Video-based Biometric Person Authentication, LNCS*, volume 3546, pages 1019–1028, 2005.
45. C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society B*, 53(2):285–339, 1991.
46. R. Herpers and G. Sommer. An attentive processing strategy for the analysis of facial features. *Face Recognition: From Theory to Applications, NATO ASI Series F, Springer Verlag*, 163:457–468, 1998.
47. T. Heseltine, N. Pears, and J. Austin. Three-dimensional face recognition using combinations of surface feature map subspace components. *Image and Vision Computing*, 26:382–396, March 2008.
48. C. Hesher, A. Srivastava, and G. Erlebacher. A novel technique for face recognition using range imaging. In *Seventh Int. Symposium on Signal Processing and Its Applications*, pages 201–204, 2003.
49. M. Hüskens, M. Brauckmann, S. Gehlen, and C. von der Malsburg. Strategies and benefits of fusion of 2D and 3D face recognition. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2005.
50. T. Hutton, B. Buxton, and P. Hammond. Dense surface point distribution models of the human face. In *IEEE Workshop on Mathematical Methods in Biomedic Image Analysis*, pages 153–160, 2001.
51. M. O. İrfanoğlu, B. Gökberk, and L. Akarun. 3D shape-based face recognition using automatically registered facial surfaces. In *Proc. Int. Conf. on Pattern Recognition*, volume 4, pages 183–186, 2004.
52. I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatiakis, and T. Theoharis. Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):640–649, 2007.
53. I.A. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis, and N. Murtuza. Multimodal face recognition: combination of geometry with physiological information. In *Proc. Computer Vision and Pattern Recognition Conference*, pages 1022–1029, 2005.
54. M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. Journal of Computer Vision*, 1(4):321–331, 1988.
55. A. Lanitis, C.J. Taylor, and T.F. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):442–455, 2002.

56. J.C. Lee and E. Milios. Matching range images of human faces. In *Int. Conf. on Computer Vision*, pages 722–726, 1990.
57. Y. Lee, H. Song, U. Yang, and H. Shin K. Sohn. Local feature based 3D face recognition. In *Int. Conf. on Audio- and Video-based Biometric Person Authentication (AVBPA 2005)*, pages 909–918, 2005.
58. P. Li, B.D. Corner, and S. Paquette. Automatic landmark extraction from three-dimensional head scan data. In *Proc. SPIE*, volume 4661, page 169. SPIE, 2002.
59. C.T. Liao, Y.K. Wu, and S.H. Lai. Locating facial feature points using support vector machines. In *Proc. 9th Int. Workshop on Cellular Neural Networks and Their Applications*, pages 296–299, 2005.
60. X. Lu, A. Jain, and D. Colbry. Matching 2.5D face scans to 3D models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):31–43, 2006.
61. S. Malassiotis and M.G. Strintzis. Pose and illumination compensation for 3D face recognition. In *Proc. Int. Conf. on Image Processing*, 2004.
62. Z. Mao, P. Siebert, P. Cockshott, and A. Ayoub. Constructing dense correspondences to analyze 3D facial change. In *Int. Conf. on Pattern Recognition*, pages 144–148, 2004.
63. T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni. Performance of Geometrix ActiveID 3D face recognition engine on the FRGC data. In *Proc. IEEE Workshop Face Recognition Grand Challenge Experiments*, 2005.
64. C. McCool, V. Chandran, S. Sridharan, and C. Fookes. 3D face verification using a free-parts approach. *Pattern Recogn. Lett.* 29(9):1190–1196, 2008.
65. G. Medioni and R. Waupotitsch. Face recognition and modeling in 3D. In *IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures*, pages 232–233, 2003.
66. K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In *Proc. 2nd Int. Conf. on Audio and Video-based Biometric Person Authentication*, 1999.
67. A.S. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2D-3D hybrid approach to automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11):1927–1943, 2007.
68. A.B. Moreno and A. Sanchez. GavabDB: A 3D face database. In *Proc. 2nd COST275 Workshop on Biometrics on the Internet*, 2004.
69. A.B. Moreno, A. Sanchez, J.F. Velez, and F.J. Diaz. Face recognition using 3D surface-extracted descriptors. In *Irish Machine Vision and Image Processing Conf. (IMVIP 2003)*, 2003.
70. I. Mpiperis, S. Malassiotis, and M.G. Strintzis. 3-D face recognition with the geodesic polar representation. *IEEE Transactions on Information Forensics and Security*, 2(3 Part 2): 537–547, 2007.
71. T. Papathodorou and D. Rueckert. Evaluation of automatic 4D face recognition using surface and texture registration. In *Proc. AFGR*, pages 321–326, 2004.
72. T. Papathodorou and D. Rueckert. Evaluation of 3D face recognition using registration and PCA. In *AVBPA05*, page 997, 2005.
73. D. Petrovska-Delacrétaz, G. Chollet, and B. Dorizzi. *Guide to Biometric Reference Systems and Performance Evaluation (in publication)*. Springer-Verlag, London, 2008.
74. P.J. Phillips, P.J. Flynn, T. Scruggs, K.W. Bowyer, and W. Worek. Preliminary Face Recognition Grand Challenge results. In *Proc. 7th Int. Conf. on Automatic Face and Gesture Recognition*, pages 15–24, 2006.
75. P.J. Phillips, P.J. Flynn, W.T. Scruggs, K.W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W.J. Worek. Overview of the face recognition grand challenge. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, volume 1, pages 947–954, 2005.
76. P.J. Phillips, W.T. Scruggs, A.J. O'Toole, P.J. Flynn, K.W. Bowyer, C.L. Schott, and M. Sharpe. *FRVT 2006 and ICE 2006 Large-Scale Results (NISTIR 7408)*, March 2007.
77. D. Riccio and J.-L. Dugelay. Geometric invariants for 2D/3D face recognition. *Pattern Recognition Letters*, 28(14):1907–1914, 2007.

78. D. Riccio and J.L. Dugelay. Asymmetric 3D/2D processing: a novel approach for face recognition. In *13th Int. Conf. on Image Analysis and Processing LNCS*, volume 3617, pages 986–993, 2005.
79. T. Russ, C. Boehnen, and T. Peters. 3D face recognition using 3D alignment for PCA. In *Proc. of the IEEE Computer Vision and Pattern Recognition (CVPR06)*, 2006.
80. T. Russ, M. Koch, and C. Little. A 2D range Hausdorff approach for 3D face recognition. In *IEEE Workshop on Face Recognition Grand Challenge Experiments*, 2005.
81. A.A. Salah and L. Akarun. 3D facial feature localization for registration. In *Proc. Int. Workshop on Multimedia Content Representation, Classification and Security LNCS*, volume 4105/2006, pages 338–345, 2006.
82. A.A. Salah, N. Alyüz, and L. Akarun. Registration of three-dimensional face scans with average face models. *Journal of Electronic Imaging*, 17(1), 2008.
83. A.A. Salah, H. Cinar, L. Akarun, and B. Sankur. Robust facial landmarking for registration. *Annals of Telecommunications*, 62(1-2):1608–1633, 2007.
84. A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, L. Akarun, and B. Sankur. Bosphorus database for 3D face analysis. In *First European Workshop on Biometrics and Identity Management Workshop (BioID 2008)*, 2008.
85. A. Savran, O. Çeliktutan, A. Akyol, J. Trojanova, H. Dibeklioğlu, S. Esenlik, N. Bozkurt, C. Demirkir, E. Akagündüz, K. Çalışkan, N. Alyüz, B. Sankur, İ. Ulusoy, L. Akarun, and T.M. Sezgin. 3D face recognition performance under adversarial conditions. In *Proc. eINTERFACE'07 Workshop on Multimodal Interfaces*, 2007.
86. A. Scheenstra, A. Ruifrok, and R.C. Veltkamp. A survey of 3D face recognition methods. In *Proc. Int. Conf. on Audio and Video-Based Biometric Person Authentication (AVBPA)*. Springer, 2005.
87. R. Senaratne and S. Halgamuge. Optimised landmark model matching for face recognition. In *Proc. 7th Int. Conf. on Automatic Face and Gesture Recognition*, pages 120–125, 2006.
88. M. Soo-Bae, A. Razdan, and G. Farin. Automated 3D face authentication and recognition. In *IEEE Int. Conf. on Advanced Video and Signal based Surveillance*, 2007.
89. H.T. Tanaka, M. Ikeda, and H. Chiaki. Curvature-based face surface recognition using spherical correlation principal directions for curved object recognition. In *Third Int. Conf. on Automated Face and Gesture Recognition*, pages 372–377, 1998.
90. J.R. Tena, M. Hamouz, A. Hilton, and J. Illingworth. A validated method for dense non-rigid 3D face registration. In *Int. Conf. on Video and Signal Based Surveillance*, pages 81–81, 2006.
91. F. Tsalakanidou, S. Malassiotis, and M. Strintzis. Integration of 2D and 3D images for enhanced face authentication. In *Proc. AFGR*, pages 266–271, 2004.
92. F. Tsalakanidou, D. Tzovaras, and M. Strintzis. Use of depth and colour eigenfaces for face recognition. *Pattern Recognition Letters*, 24:1427–1435, 2003.
93. University of Notre Dame (UND) Face Database. <http://www.nd.edu/cvrl/>.
94. Y. Wang and C.-S. Chua. Face recognition from 2D and 3D images using 3D Gabor filters. *Image and Vision Computing*, 23(11):1018–1028, 2005.
95. Y. Wang and C.S. Chua. Robust face recognition from 2D and 3D images using structural Hausdorff distance. *Image and Vision Computing*, 24(2):176–185, 2006.
96. Y. Wang, G. Pan, Z. Wu, and S. Han. Sphere-spin-image: A viewpoint-invariant surface representation for 3D face recognition. In *ICCS, LNCS 3037*, pages 427–434, 2004.
97. B. Weyrauch, J. Huang, B. Heisele, and V. Blanz. Component-based face recognition with 3D morphable models. In *Proc. First IEEE Workshop on Face Processing in Video*, 2004.
98. L. Wiskott, J.-M Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.
99. K. Wong, K. Lam, and W. Siu. An efficient algorithm for human face detection and facial feature extraction under different conditions. *Pattern Recognition*, 34:1993–2004, 2001.
100. Z. Wu, Y. Wang, and G. Pan. 3D face recognition using local shape map. In *Proceedings of the Int. Conf. on Image Processing*, pages 2003–2006, 2004.

101. C. Xu, T. Tan, Y. Wang, and L. Quan. Combining local features for robust nose location in 3D facial data. *Pattern Recognition Letters*, 27(13):1487–1494, 2006.
102. Y. Yan and K. Challapali. A system for the automatic extraction of 3-d facial feature points for face model calibration. In *Proc. Int. Conf. on Image Processing*, volume 2, pages 223–226, 2000.
103. L. Yin, X. Wei, Y. Sun, J. Wang, and M.J. Rosato. A 3D facial expression database for facial behavior research. In *Proc. 7th Int. Conf. on Automatic Face and Gesture Recognition*, pages 211–216, 2006.
104. L. Zhang, A. Razdan, G. Farin, J. Femiani, M. Bae, , and C. Lockwood. 3D face authentication and recognition based on bilateral symmetry analysis. *The Visual Computer*, 22(1):43–55, 2006.
105. C. Zhong, Z. Sun, and T. Tan. Robust 3D face recognition using learned visual codebook. In *Proc. Computer Vision and Pattern Recognition*, pages 1–6, 2007.
106. C. Zhong, T. Tan, C. Xu, and J. Li. Automatic 3D face recognition using discriminant common vectors. In *Proc. Int. Conf. on Biometrics*, volume 3832, pages 85–91, 2006.
107. S.C. Le Zou, Z. Xiong, M. Lu, and K.R. Castleman. 3-D face recognition based on warped example faces. *IEEE Transactions on Information Forensics and Security*, 2(3):513–528, Sept. 2007.

# Chapter 10

## Machine Learning Techniques for Biometrics

Francesca Odone, Massimiliano Pontil, and Alessandro Verri

**Abstract** This chapter reports recent advances in the statistical learning literature that may be of interest for biometrics. In particular we discuss two different algorithmic settings, binary classification and multi-task learning, and analyze the two closely related problems of feature selection and feature learning. In the binary case the theoretical and algorithmic advances to feature selection are applied to solve face detection and face authentication problems. In the multi-task case we show how the data structure described by a group of features common to the various tasks can be effectively learned, and then we discuss how this approach could be used to address face recognition.

### 10.1 Introduction

Biometrics is intrinsically a classification problem. In its general formulation (i.e., if no prior on the person identity is available) it can be seen as a multi-class problem. Instead, in the authentication/validation setting, where a prior on the person identity is available, we may consider a binary classifier that discriminates between the genuine user and impostors. In this context a quite general pipeline consists of a data representation phase and a classification phase. If the biometrics considered presents a high intra-class variability one can resort to the learning from examples paradigm. Statistical learning is an ideal framework, providing tools for data representation (for instance, by means of data-driven dimensionality reduction techniques) and for decision making.

In this chapter we review some recent advances on learning features from data: we first analyze an iterative approach to regularized feature selection based on a 1-norm penalty term. We then report an approach to feature learning for multi-task applications that allows us to automatically extract feature groups relevant

---

F. Odone (✉)  
DISI Università degli Studi di Genova, Via Dodecaneso 35, 16146 Genova, Italy  
e-mail: odone@disi.unige.it

for different tasks. These techniques are suitable to deal with various biometric applications, even though they are more appropriate to address biometries that lack performance and distinctiveness robustness, like face, gait, or signature biometrics [29], where the learning from examples paradigm has already been shown effective. Here we focus on 2D face biometry. In this case the initial representation can be derived from the computer vision literature and will not represent the focus of this chapter. Instead, we will analyze how feature selection may be adopted as an automatic tool for obtaining compact data representations, useful for both improving classification performance and time complexity. Indeed, many image descriptions carrying the intra-class descriptiveness are often high dimensional, thus dimensionality reduction techniques are popular in this application domain (see for instance [8, 13, 19, 26, 53, 56]).

Feature selection recently emerged as an effective way of extracting meaningful information from examples. In the single-task (binary) case the feature selection problem has been extensively studied, and a number of methods are emerging for their effectiveness [24, 51, 57, 61]. Among them, it is certainly the case to mention the impressive amount of work on the use of Adaboost methods [21] for feature selection [22, 33, 56]. An alternative interesting way to cope with feature selection in the learning by examples framework is to resort to regularization techniques based on penalty terms of 1-norm type [51]. A theoretical support to this strategy can be derived from [11]. More details on the approach discussed in this chapter may be found in [13, 14].

While the problem of learning the meaningful information contained in a given data representation has been extensively studied either for single-task supervised learning or for unsupervised learning, there has been only limited work in the multi-task supervised learning setting [3, 7, 30]. In this chapter we analyze the problem of learning a common structure shared by a set of supervised tasks. This is particularly useful to improve generalization performance, in particular when the tasks are many but the available data are relatively few. Also, a knowledge on the common structure may facilitate learning of new tasks. More details on the reported approach may be found in [4, 6]. In the case of multi-task learning we address the more general problem of feature learning, i.e., the estimation of a set of new features obtained as a combination of input variables (similar to PCA in the unsupervised setting). In this formulation feature selection can be seen as a special case.

It may be the case to remark that in both cases under analysis the complexity of the solution is controlled by the regularization parameter: in single-task feature selection it controls the sparsity of the solution, while in multi-task feature learning it controls the number of features common among the various tasks learned.

The structure of the chapter is as follows. Section 13.2 reports the state of the art on face biometrics, with particular reference to machine learning approaches. Section 13.3 describes two regularized methods that may be applied with success in the biometrics domain, namely example-based feature selection in the binary and feature learning in the multi-task case. Section 13.4 reports two applications of feature selection relevant to face biometrics and an exhaustive experimental analysis of multi-task feature learning. Section 13.5 is left to a final discussion.

## 10.2 State of the Art on Face Biometrics

In this chapter we mainly focus on face biometry as it has benefited from machine learning advances more than other biometries. In face biometry we mainly focus on the *preprocessing stage* – where face detection and the detection of face features have been studied in depth – and on the actual *face recognition stage*.

The learning from examples paradigm has been often adopted in the literature to solve recognition problems or detection problems (of faces or face parts) as an effective way to model implicitly class variations.

In the case of face detection (that may be considered as the most important pre-processing phase of a face identification system) we deal with a binary classification problem, and we rely on a data set of positive (face) and negative (non-face) examples – for an introduction to the problem see, for instance, the survey in [59] and references therein, or the Face Detection homepage.<sup>1</sup> Early approaches to face detection were based on applying machine learning algorithms, such as neural networks or support vector machines, to representations derived by the whole image [40, 44, 45, 50]. The data sets were selected so as to express the degree of variability required by the system. On this respect, view-based approaches were often used to capture the variability with respect to viewpoint [43, 46]. Later on, component-based approaches highlighted the fact that local areas are often more appropriate to deal with occlusions and deformations [28, 37, 54]. A recent trend, made popular by the influential work of Viola and Jones [56], is based on the adoption of overcomplete dictionaries of local features, able to capture local structures and medium range symmetries, and on the use of feature selection strategies that reduce the redundancy of the initial description. Such methods are popular for their effectiveness and because they may be applied with success in real-world applications.

The literature on face recognition is also vast – see the comprehensive survey [60] or the material available on the Face Recognition Homepage.<sup>2</sup> There are two different ways to address the actual recognition phase: (1) *identification or recognition* where the test or probe datum is available, but no prior on the related identity is known; therefore the probe is compared with all identities belonging to the gallery of enrolled people. (2) *Authentication or verification*: a claimed identity is available with the probe datum; in this case the probe is compared with the description of the individual carrying the claimed identity. Face authentication is again a binary classification problem, and the objective is to build a classifier per each identity of the gallery that discriminates between the genuine user and impostors. Face recognition is instead a multi-class classification problem, where the number of classes grows proportionally to the size of the gallery. As a multi-class problem characterized by a relatively small inter-class distance, difficult to address in high-dimensional spaces, it has often been dealt with dimensionality reduction approaches.

---

<sup>1</sup> <http://www.facedetection.com/>

<sup>2</sup> <http://www.face-rec.org>

Face recognition methods based on intensity images may be roughly divided into global or holistic approaches and local or feature-based approaches. Eigenfaces [53] are without doubt one of the most popular holistic approaches to face detection and recognition. In a training phase a relatively small set of data are used to estimate a subspace that permits to reduce the dimensionality of data with a quite small loss of information. The subspace is obtained with principal components analysis (PCA). The motivation of the approach is that we can reasonably assume different face images lie in a rather compact manifold in the space of all the possible images of a certain size. Finding such a manifold is useful both to reduce the dimensionality of the problem, working more efficiently with a more compact set of data, and to take into account the peculiar structure of the object of interest, in our case human faces. After the subspace is found all images of the data set may be represented as weight vectors, obtained by projecting the image into the subspace, often referred to as eigenface space (universal eigenspace). In the test phase, a new image is projected in the space and it is associated with the identity of the image whose projection is closer to it. In [36] the eigenface method is extended to use a probabilistic similarity that models the *intra-personal* variations versus the *extra-personal* variations. The former are related to variations within the same individual, due to illumination, viewpoint, expression changes; the latter refers to the difference between an individual and another. The authors use a statistical approach learning which types of variations are observed in different images of the same individual and compare them to the distribution of extra-personal variation. The intra-personal and extra-personal distributions estimated are assumed to be Gaussian and are approximated with principal eigenvectors which are global. This approach, which was adopted by many other authors, essentially aims at capturing descriptions that do not vary between images of the same individual, for all individuals represented in the universal eigenspace. An alternative approach of modeling *individual eigenspaces* each one describing a given individual has been proposed and shown effective for face authentication (see, for instance, [35]). The use of Fisher discriminant analysis as classifiers for face recognition is widely spread [8, 19, 60]. In [60] it is reported that according to [47] high-frequency components, while they are not crucial for instance for the sex judgment task, may be important for face recognition, where finer details may help in discriminating different identities. At the same time it has been shown that, especially when dealing with higher resolution images, global approaches are outperformed by component-based methods that are more stable to local changes [27]. Feature-based approaches include methods for precise location of specific points (usually called *fiducial points*) [34, 42, 58] and methods based on templates, where each area of interest is described by approximately located patches. Brunelli and Poggio [10] compare the two approaches showing that the template-based approach is simpler and better performing, while the approach based on precise features may lead to more compressed descriptions. The latter approach is explored in this chapter, mainly to the purpose of automatically extracting areas of interest for a given recognition task. We notice that this approach is more appropriate for medium-low-quality images. Possibly the first local approach to face recognition is again due to Pentland and his co-workers

[41]: here the features are obtained by performing PCA to local face regions independently. Automatic feature extraction may be obtained extracting meaningful keypoints from the face image and describing them so as to maximize the extra-class variance. We mention one of the first attempts to our knowledge to export the very popular SIFT descriptors to face recognition [9] and the widely used local binary patterns (LBP) [2] of which a number of variants and combinations are currently available.

We conclude this section recalling that, as pointed out in [60], face recognition and identification from a video sequence is an important problem, mainly for the impact it has on video-surveillance applications. Besides market requirements, another reason for exploring video-based face recognition lies in the fact that dynamics (expression) information may be an important cue for recognition. Only in recent years a few attempts of exploiting the redundant information coming from video-sequences have been proposed. The first approaches proposed mainly exploited the abundance of information in videos, selecting the most appropriate frames to the purpose of recognition and then adopting an image-based approach. In this respect we mention [23]. More recently the face dynamics has been used: in [18] active appearance models are exploited to gather evidence from image sequences, in [32] hidden Markov models are the starting point for a dynamic face descriptor, similarly in [1, 48] autoregressive moving average (ARMA) models are proposed.

Finally we report an extension of the LBP features to videos [25]: image neighborhoods used in the conventional LBP are substituted by spatio-temporal prisms. A very high number of features may be generated from a video sequence, therefore the authors resort to an Adaboost scheme to extract the most representative features (capturing both appearance and dynamics) of a given individual.

### 10.3 Recent Advances to Machine Learning

This section reports recent advances of the statistical learning theory to learn descriptions from data. These methods are effective ways to deal *automatically* with the redundancy of image data and the complexity of multi-class problems.

#### Notation

We begin by introducing the notation used in the remainder of the section. We let  $\mathbb{R}$  be the set of real numbers and  $\mathbb{R}_+$  the subset of nonnegative ones. For every  $n \in \mathbb{N}$ , we let  $\mathbb{N}_n := \{1, 2, \dots, n\}$ . If  $w, u \in \mathbb{R}^d$ , we define  $\langle w, u \rangle := \sum_{i=1}^d w_i u_i$  and  $\|w\|_2 = \sqrt{\langle w, w \rangle}$ . If  $A$  is a  $d \times T$  matrix we denote by  $a^i \in \mathbb{R}^T$  and  $a_t \in \mathbb{R}^d$  the  $i$ th row and the  $t$ th column of  $A$ , respectively. We denote by  $\mathbf{S}_{++}^d$  the set of symmetric and positive definite matrices. If  $D$  is a  $d \times d$  matrix, we define  $\text{trace}(D) := \sum_{i=1}^d D_{ii}$ . If  $w \in \mathbb{R}^d$ , we denote by  $\text{Diag}(w)$  or  $\text{Diag}(w_i)_{i=1}^d$  the diagonal matrix having the components of vector  $w$  on the diagonal. We let  $\mathbf{O}^d$  be the set of  $d \times d$  orthogonal matrices.

### 10.3.1 Learning Features for Binary Classification

We consider the problem of selecting a set of features descriptive of a given binary classification task. In this context we review a recently proposed regularization method for learning features from a set of examples of two different classes. Such a method is based on the so-called Lasso scheme[51].

#### 10.3.1.1 Problem Formulation

We consider a binary classification problem and assume that we are given a training set of  $m$  (positive and negative) elements. Each example is represented by a dictionary of  $d$  features. We look for a compact representation of input data for the problem of interest by means of feature selection. Our solution associates a weight to each feature: features with nonzero weights are relevant to model the diversity between the two classes under consideration.

We consider the case of a *linear* dependence between input and output data. The problem can be thus formulated as the solution of the following linear system of equations:

$$y = Xa \quad (10.1)$$

where  $X = \{x_{ij}\}$ ,  $i = 1, \dots, m$ ;  $j = 1, \dots, d$  is the  $m \times d$  features matrix obtained representing the training set of the  $m$  inputs  $x_i$  with a dictionary of  $d$  features. The  $m$  vector  $y = (y_1, \dots, y_m)^\top$  contains the output labels: since we focus on a binary classification setting all data are associated to a label  $y_i \in \{-1, 1\}$ .  $a = (a_1, \dots, a_d)^\top$  is the vector of the unknown weights, each entry is associated to one feature and intuitively describes the importance of the feature in determining the membership of a given feature vector to one of the two classes.

In the application domains we consider the dimensions of  $X$  to be large, thus the usual approaches for solving the algebraic system (10.1) turn out to be unfeasible. Moreover the typical number of features  $d$  is much larger than the dimension  $m$  of the training set, so that the system is hugely under-determined. Also, because of the redundancy of the feature set, we may have to deal with the collinearities between feature vectors that are responsible for severe ill-conditioning. Both difficulties call for some form of regularization and can be obviated by turning problem (10.1) into a penalized least-squares problem.

Classical regularization such as the so-called ridge regression (also referred to as Tikhonov regularization) uses a quadratic penalty, typically the 2-norm of the vector  $a$ :  $\|a\|_2^2 = \sum_j a_j^2$ . Such quadratic penalties, however, do not provide feature selection in the sense that the solution of the penalized least-squares problem will typically yield a vector  $a$  with all weights  $a_j$  different from zero. This is the reason why the replacement of quadratic penalties by sparsity-enforcing penalties has been advocated in recent literature. This means that the penalty will automatically enforce the presence of (many) zero weights in the vector  $a$ . Among such zero-enforcing penalties, the 1-norm of  $a$  is the only convex one, hence providing feasible

algorithms for high-dimensional data. Thus we consider the following problem, usually referred to as *Lasso regression* [51]:

$$a_L = \arg \min_a \{\|y - Xa\|^2 + 2\tau \|a\|_1\}, \quad (10.2)$$

where  $\|a\|_1 = \sum_j |a_j|$  is the 1-norm of  $a$  and  $\tau > 0$  is a regularization parameter regulating the balance between the data misfit and the penalty. In feature selection problems, this parameter also allows to vary the degree of sparsity (number of true zero weights) of the vector  $a$ .

### 10.3.1.2 Learning Algorithm

The presence of the 1-norm penalty in problem (10.2) makes the dependence of Lasso solutions on  $y$  non-linear. Hence the computation of 1-norm penalized solutions is more difficult than with 2-norm penalties. To solve (10.2) we adopt a simple iterative strategy proposed in [12]:

$$a_L^{(n+1)} = S_\tau[a_L^{(n)} + X^\top(y - Xa_L^{(n)})] \quad n = 0, 1, \dots \quad (10.3)$$

with arbitrary initial vector  $a_L^{(0)}$ , where  $S_\tau$  is the following “soft-thresholding”

$$(S_\tau h)_j = \begin{cases} h_j - \tau \text{sign}(h_j) & \text{if } |h_j| \geq \tau \\ 0 & \text{otherwise} \end{cases}.$$

In the absence of soft-thresholding ( $\tau = 0$ ) this scheme is known as the Landweber iteration, which converges to the generalized solution (minimum-norm least-squares solution) of (10.1). The soft-thresholded Landweber scheme (10.3) has been proven in [12] to converge to a minimizer of (10.2), provided the norm of the matrix  $X$  is renormalized to a value strictly smaller than 1.

Experimental evidence showed that the choice of the initialization vector is not crucial, therefore we always initialize the weight vector  $a$  with zeros:  $a^{(0)} = \mathbf{0}^\top$ . The stopping rule of the iterative process is related to the stability of the solution reached and it is based on comparing the solution obtained at the  $n$ th iteration  $a^{(n)}$  with the previous one. Algorithm in Table 10.1 reports the pseudo-code of the iterative scheme.

### 10.3.1.3 Feature Selection for Large Problems

As we will see in Section 13.4, the problems that one could build in a biometric setting will be rather large. For this reason applying the iterative algorithm described in Equation (10.3) directly to the whole matrix may not be feasible on all PCs: the matrix multiplication needs to be implemented carefully so that we do not keep in primary memory the entire matrix  $X$ .

**Table 10.1** Algorithm 1: Binary feature selection

---

**Input:** training set (in the appropriate representation)  
 $\{(x_i, y_i)\}_{i=1}^m \quad y_i = \{-1, 1\}, x_i = (X_{i1}, \dots, X_{id})^\top \quad X = (x_1, \dots, x_m)^\top$

**Parameters:** regularization parameter  $\tau$

**Output:** the sparse weights vector  $a$

**Initialization:** **for all**  $j$   $a_j = 0$ ; **end for**  
normalize matrix  $X$  (see text)

**while** exit conditions are met **do**  
    update  $a = a + X^\top(y - Xa)$   
    **for**  $i = 1, \dots, m$  **do**  
        **if**  $|a_i| \geq \tau$  **then**  
             $a_i = a_i = \tau \text{sign}(a_i)$   
        **else**  
             $a_i = 0;$   
        **end if**  
    **end for**  
**end while**

---

To address this computational problem we devise a strategy based on resampling the feature set and obtaining many smaller problems: we build  $S$  feature subsets *each time* extracting with replacement  $p$  features from the *original* set of size  $d$  ( $p << d$ ), we then obtain  $S$  smaller linear sub-problems of the type:  $X_s a_s = y$  for  $s = 1, \dots, S$ , where  $X_s$  is a submatrix of  $X$  containing the columns relative to the features in  $s$ ;  $a_s$  is computed accordingly. As for the choice of the number  $S$  of sub-problems and their size we observe that the subset size should be big enough to be descriptive, small enough to handle the matrix easily; thus, we consider subsets with about 10% of the original feature set size. To choose the number of sub-problems  $S$ , we rely on the binomial distribution and estimate how many extractions are needed so that each feature is extracted at least 10 times with high probability [13].

After we build the  $S$  sub-problems we look for the  $S$  solutions running  $S$  iterative methods as in Equation (10.3). At the end of the process we are left with  $S$  overlapping sets of features. The final set is obtained choosing the features that have been selected *each time they appear in the subsets*.

The computational complexity of the iterative algorithm (10.3) is caused by the two matrix-vector multiplications and thus it is  $O(md)$  for each iteration and a fixed  $\tau$ . Since the number of iterations  $I$  is not negligible we consider  $O(mdI)$ . In the model selection phase this should be repeated as many times as the number of  $\tau$  that we evaluate.

It is worth mentioning the fact that the resampled version of the method is suitable for parallel computation and would allow for a saving of computation time in inverse relation to the number of processors used. In the case a single processor only is available one should design carefully the model selection phase in order to limit the computational cost of the training phase. In our experiments we consider two methods: (*i*) to choose  $\tau$  that includes a fixed number of 0s in the solution (or, equivalently, that selects a given number features) in about  $I$  iterations and (*ii*) to choose  $\tau$  on the basis of the generalization performance, using cross-validation: we select  $\tau$ s leading to classification rates below a certain threshold and then choose among them, the value providing the smallest number of features. Here Equation (10.1) is used as a classifier.

#### 10.3.1.4 Evaluation Methods

We conclude this section by briefly discussing how we evaluate the quality of the obtained subset of features – for an introduction to this topic we refer the reader to [24]. We first recall that we are interested in selecting features useful to build a good predictor (classifier). Therefore, we evaluate the generalization performance of a classifier trained on a data set represented with the selected features.

As a classification algorithm we adopt a *linear support vector machine (SVM)* [55], which will also be used to implement the final detection and authentication modules. The performance evaluations will be based on receiver operating characteristic (ROC) curves, built by varying the SVM offset value  $b$ . On the ROC curve we consider, in particular, the hit rate corresponding to 0.5% false positives (this is particularly important for detection problems where the number of negatives is very high) and the equal error rate (EER), i.e., the value producing an equal number of false positives and false negatives.

### 10.3.2 Learning Shared Representations for Multiple Tasks

In this section, we study the problem of learning multiple regression or classification tasks simultaneously. In particular, we review a method for learning a set of features which are shared across the tasks [4, 5]. This is a problem of interest in many research areas, whose potential was discussed in [52] for the object detection case. Images of different objects may share a number of features that are different from the pixel representation of images. In the biometry domain this approach can be advantageous when dealing with the multi-class problem of face recognition: different individuals may share a number of features; thus, learning these common features may facilitate the different recognition tasks.

#### 10.3.2.1 Problem Formulation

We are given  $T$  supervised learning tasks. For every  $t \in \mathbb{N}_T$ , the corresponding task is identified by a function  $f_t : \mathbb{R}^d \rightarrow \mathbb{R}$  (e.g., a regressor or margin classifier). For each task  $t$ , we are given a set of  $m$  input/output examples which, similar to

the previous section, we organize in a  $m \times d$  matrix  $X_t$  and in a  $m$  vector  $y_t$  for  $t = 1, \dots, T$ .

We wish to use the available examples in order to uncover *particular* relationships across the tasks. Our working assumption is that the tasks *all share a small set of features*, namely the functions  $f_t$  can be represented as a linear combination of a few feature functions. For simplicity, we consider linear homogeneous features, each of which is represented by a vector  $u_i \in \mathbb{R}^d$  – extensions to non-linear features are dealt with in [4]. Furthermore, we assume that the vectors  $u_i$  are orthogonal and, so, we consider only up to  $d$  of such vectors.

If we denote by  $U \in \mathbf{O}_d$  the matrix whose columns are the vectors  $u_i$ , the task functions can be written as  $f_t(x) = \langle a_t, U^\top x \rangle$ ,  $x \in \mathbb{R}^d$ , where  $a_t = (a_{t1}, \dots, a_{td})^\top$  is the vector of regression coefficients for the  $t$ th task.

Our assumption that the tasks share a “small” set of features means that the matrix  $A$  has “many” rows which are identically equal to zero and, so, the corresponding features (columns of matrix  $U$ ) will not be used by any task.

Our method for multi-task feature learning is to solve the optimization problem

$$\min \{\mathcal{E}(A, U) : U \in \mathbf{O}_d, A \in \mathbb{R}^{d \times T}\}, \quad (10.4)$$

$$\mathcal{E}(A, U) = \sum_{t=1}^T \|y_t - X_t U a_t\|^2 + \gamma \|A\|_{2,1}^2, \quad (10.5)$$

where  $\gamma > 0$  is a regularization parameter.

In the first term in (10.5) the square loss could be replaced with a loss function  $L : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_+$  convex in the second argument. The second term is a regularization term which penalizes the  $(2, 1)$ -norm of matrix  $A$ . It is obtained by first computing the 2-norms of the (across the tasks) rows  $a^i$  (corresponding to feature  $i$ ) and then the 1-norm of the vector  $(\|a^1\|_2, \dots, \|a^d\|_2)$ . The magnitudes of the components of this vector indicate how important each feature is.

If the matrix  $U$  is not learned and we set  $U = I_{d \times d}$ , problem (10.4) selects a “small” set of variables, common across the tasks. Furthermore, if  $T = 1$ , function (10.5) reduces to the 1-norm regularization problem discussed in the first part of this section.

The  $(2, 1)$ -norm above favors a small number of nonzero rows in the matrix  $A$ , thereby ensuring that few common features will be learned across the tasks. Of course the number of features learned depends on the value of the parameter  $\gamma$  and it will typically not be increasing with  $\gamma$ .

### 10.3.2.2 Equivalent Convex Problem

Solving problem (10.4) is challenging for two main reasons. First, it is a non-convex problem, although it is separately convex in each of the variables  $A$  and  $U$ . Second, the regularizer  $\|A\|_{2,1}^2$  is not smooth, which makes the optimization problem more difficult to solve.

Fortunately, problem (10.4) can be transformed into an equivalent convex problem. To describe this result, for every  $W \in \mathbb{R}^{d \times T}$  with columns  $w_t$  and  $D \in \mathbf{S}_{++}^d$ , we define the function

$$\mathcal{R}(W, D) = \sum_{t=1}^T \sum_{i=1}^m L(y_{ti}, \langle w_t, x_{ti} \rangle) + \gamma \text{trace}(D^{-1} W W^\top) \quad (10.6)$$

in which we replace the square loss with a more general loss  $L$ .

It is then possible to show that problem (10.4) is equivalent to the convex optimization problem

$$\inf \{ \mathcal{R}(W, D) : W \in \mathbb{R}^{d \times T}, D \in \mathbf{S}_{++}^d, \text{trace}(D) \leq 1 \}. \quad (10.7)$$

In particular, any minimizing sequence of problem (10.7) converges to a minimizer of problem (10.4) and (10.5). Moreover, the solutions  $(\hat{A}, \hat{U})$  and  $(\hat{W}, \hat{D})$  of problems (10.4) and (10.7), respectively, are related by the formula

$$(\hat{W}, \hat{D}) = \left( \hat{U} \hat{A}, \hat{U} \text{Diag} \left( \frac{\|\hat{a}^i\|_2}{\|\hat{A}\|_{2,1}} \right)_{i=1}^d \hat{U}^\top \right).$$

We refer the reader to [4, Section 3] for more information on this observation.

Note that in problem (10.7) we have bounded the trace of matrix  $D$  from above, because otherwise the optimal solution would be to simply set  $D = \infty$  and only minimize the empirical error term in the right-hand side of Equation (10.6).

Returning to the discussion of Section 13.2 on the  $(2, 1)$ -norm, the rank of the optimal matrix  $D$  indicates how many common relevant features the tasks share. Indeed, it is clear from the above discussion that the rank of matrix  $\hat{D}$  equals the number of nonzero rows of matrix  $\hat{A}$ .

### 10.3.2.3 Learning Algorithm

We now discuss an algorithm for solving the convex optimization problem (10.7). The algorithm minimizes a perturbation of the objective function (10.6), in which a perturbation  $\epsilon I$  is added to the matrix  $WW^\top$ , appearing in the second term in the right-hand side of (10.7), where  $\epsilon > 0$  and  $I$  is the identity matrix. This perturbation keeps  $D$  nonsingular and ensures that the infimum over  $D$  is always attained.

We now describe the two steps of Algorithm in Table 10.2 for solving this perturbed problem. In the first step, we keep  $D$  fixed and minimize over  $W$ . This step can be carried out independently across the tasks since the regularizer decouples when  $D$  is fixed. More specifically, introducing new variables for  $D^{-\frac{1}{2}}w_t$  yields a standard 2-norm regularization problem for each task with the same kernel  $K(x, x') = x^\top D x'$ . In the second step, we keep matrix  $W$  fixed and minimize

**Table 10.2** Algorithm 2: Multi-Task Feature Learning

---

**Input:** training sets  $\{(x_{ti}, y_{ti})\}_{i=1}^m, t \in \mathbb{N}_T$

**Parameters:** regularization parameter  $\gamma$ , tolerances  $\varepsilon, tol$

**Output:**  $d \times d$  matrix  $D$ ,  $d \times T$  regression matrix  $W = [w_1, \dots, w_T]$

**Initialization:** set  $D = \frac{I_d}{d}$

**while**  $\|W - W_{prev}\| > tol$  **do**

**for**  $t = 1, \dots, T$  **do**

compute  $w_t = \operatorname{argmin} \left\{ \sum_{i=1}^m L(y_{ti}, \langle w, x_{ti} \rangle) + \gamma \langle w, D^{-1}w \rangle : w \in \mathbb{R}^d \right\}$

**end for**

set  $D = \frac{(WW^\top + \varepsilon I_d)^{\frac{1}{2}}}{\operatorname{trace}(WW^\top + \varepsilon I_d)^{\frac{1}{2}}}$

**end while**

---

with respect to  $D$ . One can show that partial minimization with respect to  $D$  has a closed-form solution given by

$$D_\varepsilon(W) = \frac{(WW^\top + \varepsilon I_d)^{\frac{1}{2}}}{\operatorname{trace}(WW^\top + \varepsilon I_d)^{\frac{1}{2}}}. \quad (10.8)$$

Algorithm can be interpreted as alternately performing a supervised and an unsupervised step. In the former step we learn task-specific functions (namely the vectors  $w_t$ ) using a common representation across the tasks. This is because  $D$  encapsulates the features  $u_i$  and thus the feature representation is kept fixed. In the unsupervised step, the regression functions are fixed and we learn the common representation.

In [4] an analysis of the above algorithm is provided. In particular, it is shown that for every  $\varepsilon > 0$  the algorithm converges to a solution of the corresponding perturbed problem. Moreover, as  $\varepsilon$  goes, any limiting point of the sequence of such solutions solves problem (10.7)).

We proceed with a few remarks on an alternative formulation for the problem. By substituting Equation (10.8) with  $\varepsilon = 0$  in Equation (10.7), we obtain a regularization problem in  $W$  only, in which the regularization term is the square of the trace norm of matrix  $W$ , namely the sum of singular values of  $W$ . As shown in [20], the trace norm is the convex envelope of  $\operatorname{rank}(W)$  in the unit ball. We also note here that a similar problem has been studied in [49] in the context of collaborative filtering.

## 10.4 Applications: Learning Face Features

In this section we describe how we apply the statistical learning methods described in Section 13.3 to face biometry. After a brief discussion on the peculiarity of the data under consideration, we show how feature selection may be adopted for both

face detection and face authentication, then we present numerical simulations and discuss how multi-task feature learning may be applied to recognition problems.

### **10.4.1 The Peculiarity of Face Features**

Image features are characterized by a high degree of redundancy. Such redundancy does not necessarily compromise generalization performance of a classifier, but the curse of dimensionality [17] may degrade the obtained results, in case only a relatively small training set is available. Also, and not less importantly, it affects the computational efficiency of the classifier.

Dealing with information redundancy via feature selection means selecting one or few delegates for each group of correlated features to represent the other elements of the group. As for the choice of the delegate, unlike in other application domains (such as micro-array data analysis) in most image-related problems one could choose a random delegate for the correlated feature sets. Image features are not important *per se* but for the appearance information that they carry, which is resemblant to all other members of their group.

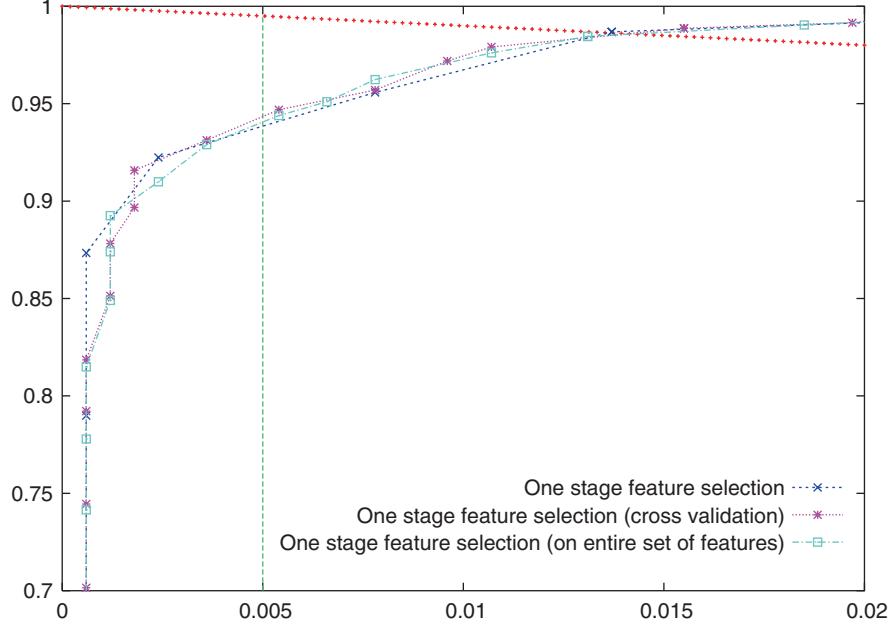
On this respect a remark is in order. Face features may be correlated not only because of the intrinsic short-range correlation of all natural images, or because the chosen description is redundant, but also because of dependencies related to the class of faces (which contain multiple occurrences of similar patterns at different locations, for example the eyes). Correlation due to the representation chosen produces redundant descriptions, while correlation due to the class of interest may carry important information on its peculiarities. A feature selection scheme should take into consideration both these needs, although often, for computational reasons, a small set of features is preferred.

### **10.4.2 Face Detection**

In this section we focus on face feature selection and then briefly discuss the design of an efficient face detector, starting off from an image representation based on the rectangle features [56] which widely retained a good starting point for many object detection problems.

The raw data we consider are image patches of size  $N \times N$  ( $N = 19$ ). We compute rectangle features over all locations, sizes, and aspect ratios for each image or image patch under consideration. We therefore compute about 64000 features per image/patch. Let us now describe how we apply the iterative algorithm described in Section 13.3 to the selection of face features.

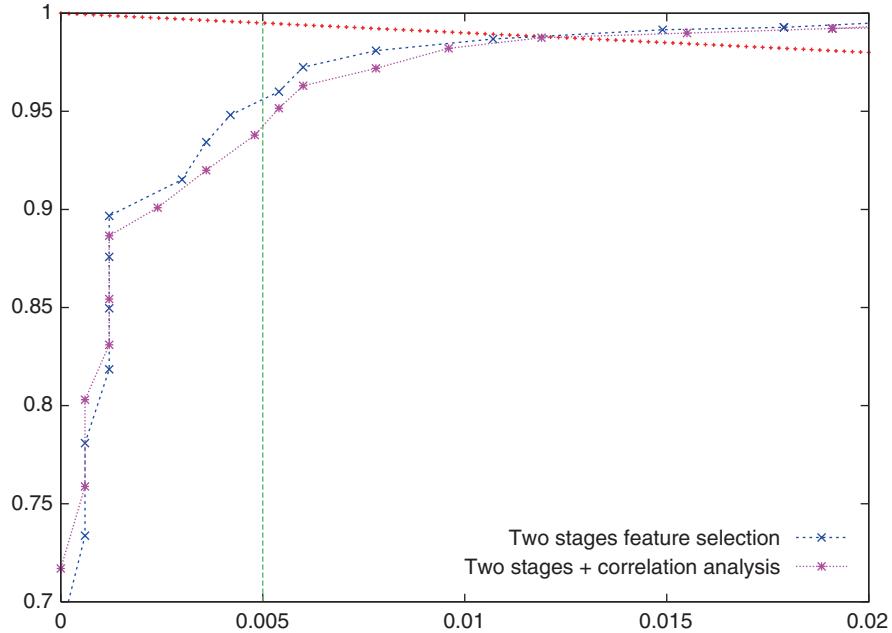
We consider a data set of 4000 training data, evenly distributed between positive (approximately frontal faces) and negative (non-faces), 2000 validation, and 3400 test data. The linear system that we build for feature selection is rather big, the data matrix  $A$  is  $4000 \times 64000$  (since each entry is stored in single precision, the total



**Fig. 10.1** Comparison between (1) a direct solution of the linear problem (—o—) and (2) the resampling strategy (with two different parameter choices, setting sparsity level (—x—) or through cross-validation (—\*—))

space required for matrix  $A$  is of about 1 GB). In [13] we report an exhaustive set of experiments confirming the appropriateness of the resampling strategy when the data matrix size grows. Figure 10.1 summarizes the results obtained with and without the resampling strategy, and it shows that not only there is no loss when applying the resampling strategy but there is actually a small gain. For what concerns model selection, having to deal with a high number of different problems, we choose  $\tau$  so as to keep a fixed number of features (i.e., find solutions with a given percentage of 0 entries). Figure 10.1 also shows the performance obtained with a model selection based on cross-validation (the results are comparable in this first stage).

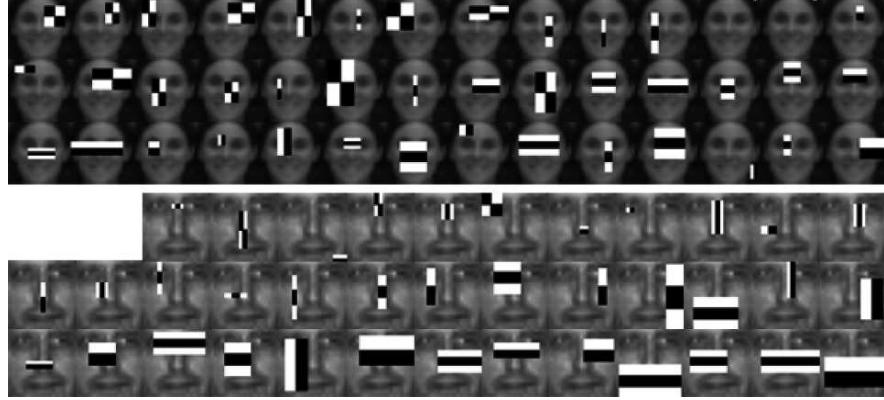
At the end of this feature selection stage the remaining features  $S_1$  are a good synthesis of the original description as confirmed by the classification results on our test set and they maintain all the descriptiveness, representing all meaningful areas of a face. Nonetheless, the number of selected features is still high (about 4500 in our experiments). To reduce the number of features, instead than choosing a value for  $\tau$  that force a higher number of zeros (this choice lead to a considerable decrease in performance, see [13]), we further apply the feature selection procedure to set  $S_1$  looking for a new, sparser, solution. The new data matrix is obtained selecting from  $A$  the columns corresponding to  $S_1$ , and  $f_{S_1}^{(0)}$  is initialized again to a vector of



**Fig. 10.2** Results obtained with a two-stage selection procedure, with (—\*) and without (—x—) the further correlation analysis

zeros. As for the parameter tuning strategy we choose cross-validation to be used in this second stage. Cross-validation takes explicitly into account generalization and therefore is more appropriate whenever it is computationally feasible. The two-stage selection procedure leaves us with a set  $S_2$  of 247 features with no decrease in performance with respect to the first stage – see Figure 10.2. Figure 10.2 also shows the classification results of the feature set  $S_3$  obtained by applying a simple correlation analysis to  $S_2$  that keeps only one delegate from sets of multiple features of the same type, strong correlation, and spatially overlapping.

The results we report here have been obtained from a data set gathered by means of a monitoring system installed in our department [16], and manually labeled in positive and negative examples. In this case the detected faces were approximately frontal, but no manual registration of facial features was applied. In a different run of experiments we used the CBCL-CMU frontal faces data set, in this case positive examples were accurately registered. The results obtained in the two cases are very different, as they reflect the different nature of the data sets. Notice how in the case of CBCL-CMU there is a predominance of features with a vertical symmetry, due to the fact that the data set is nicely registered and all faces are exactly frontal. When using our data set, the faces are only approximately registered and data contain moderate viewpoint variations. Therefore, horizontal symmetries are preserved, while vertical symmetries are not distinctive of the data set (see Fig. 10.3).



**Fig. 10.3** *Top:* the selected features from the DISI data set. *Bottom:* selected features from the CBCL-CMU data set (see text)

We conclude with an account on the final face detector and the performance obtained (for details see [13, 14, 16]). The features in  $S_3$  are used to build a cascade of small SVMs that is able to process video frames in real time. The cascade of classifiers analyzes an image according to standard coarse-to-fine approach. Each image patch at a certain location and scale is the input of the classifiers cascade: if the patch fails one test it is immediately discarded; if it passes all tests a face is detected. Each classifier of the cascade is built by starting with three mutually distant features, training a linear SVM on such features, and adding further features until a target performance is obtained on a validation set. Target performance is chosen so that each classifier will not be likely to miss faces: we set the minimum hit rate to 99.5% and the maximum false-positive rate to 50%. Assuming a cascade of 10 classifiers, we would get as overall performance:  $HR = 0.995^{10} \sim 0.9$  and  $FPR = 0.5^{10} \sim 3 \times 10^{-5}$  [56].

Table 10.3 shows its detection performance as a face detector of our real-time system in two different situations. The first row of the table refers to the results obtained on images acquired in controlled situations (people were asked to walk towards the camera one at a time), while the second row refers to uncontrolled detection: we manually labeled the events occurring in a 5-hour recording of a busy week day; the recording was entirely out of our control and it included changes of the scene, people stopping for unpredictable time, lateral faces. Notice that at run time, the classifiers have been tuned so as to minimize the number of false-positives.

**Table 10.3** The performance of the face detection system on a controlled set of images (top row) and on a 5-hour unconstrained live video (bottom row)

Test data	False-positive rate (%)	hit rate (%)
Test images	0.1	94
5 hours live	$4 \times 10^{-7}$	76

We observe that the amount of data analyzed is huge since, on average, the detector analyzes 20000 patches per image or frame.

### 10.4.3 Face Authentication

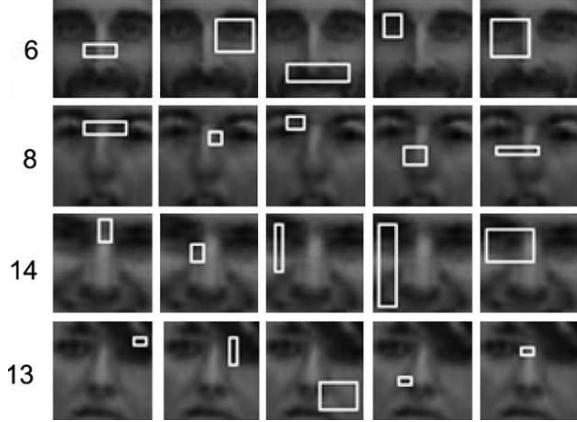
We now consider a face authentication problem that aims at discriminating examples of the person whose identity was declared versus examples of other people. In particular we consider a problem setting that models, for each individual, his/her *intra-class versus extra-class* variability. We consider an image description based on LBP features [2] at a fixed scale.

We assume we have a data set of  $N$  individuals  $\mathcal{I}_1, \dots, \mathcal{I}_N$ , let us briefly describe how we represent each individual  $\mathcal{I}$ :

- We start off from a set of  $40 \times 40$  positive images  $I_p$ ,  $p = 1, \dots, P$  of individual  $\mathcal{I}$  and a set of negative images  $I_n$ ,  $n = 1, \dots, N$  randomly sampled from other individuals.
- For each image  $I_i$ , each neighborhood is represented as a LBP. The neighborhood size that we consider is obtained choosing eight sampling points on a circle of radius 2 [2]. Then, for each image, we compute  $L$  LBP histograms on rectangular regions of at least  $3 \times 3$  pixels, on all locations and aspect ratios; thus the description of image  $I_i$  is a list of histograms  $H_i^1, \dots, H_i^L$ . Notice that a feature selection procedure is important, since the size of feature vectors is rather big: we obtained  $L = 23409$  LBP histograms.
- We resort again to the regularized feature selection of Section 13.2. The linear system is built so as to express the intra-personal and extra-personal variation of each histogram. Similar to [25] we compute the feature vectors (in our case the rows of the matrix  $A$ ) as follows: for each pair of images  $I_A$  and  $I_B$  we compare corresponding LBP histograms (after normalization) using the  $\chi^2$  distance. That is, for each pair of input images we obtain a feature vector  $x_i$  whose elements are  $\chi^2$  distances:  $x_i = (\chi^2(I_A^1, I_B^1), \dots, \chi^2(I_A^L, I_B^L))$ . We associate a label  $g_i \in \{+1, -1\}$ , where  $g_i = 1$  if both  $I_A$  and  $I_B$  are positive images,  $g_i = -1$  if either  $I_A$  or  $I_B$  is negative.

For a given set of examples, the number of feature vectors that we may compute is very high (for a set of positive images of size  $P$  we will have  $\binom{P}{2}$  positive feature vectors, and the negative in general will be much higher). In order to obtain balanced systems of reasonable size we randomly sample at most 2000 positive and 2000 negative feature vectors and build matrix  $A$ . The vector  $g$  is composed of the corresponding labels  $g_i$ , and the vector  $f$  of unknowns will weight the importance of each LBP histogram for intra-person and extra-person comparison. Once we have built a system for a given individual, according to the procedure previously described, we select features following the same protocol illustrated for face detection. Notice that for each individual we obtain a different set of distinctive features, the ones that best represent his/her peculiarity.

**Fig. 10.4** Top five features for some individuals: they often capture distinctive face features (beard, hair, or a clown nose; (bottom row)

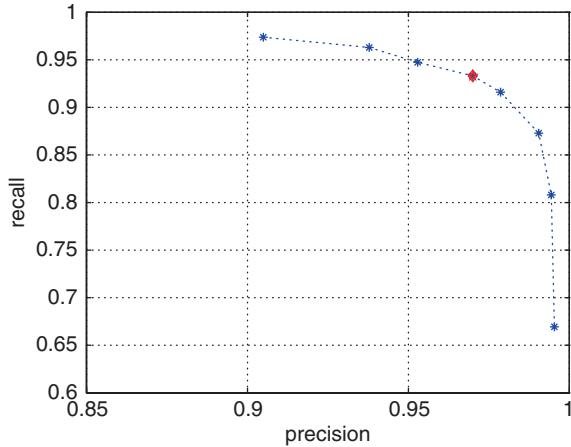


Again, we evaluate the goodness of a feature set in terms of its generalization ability. The data set we consider for face authentication is gathered and registered automatically by the face detection module. In particular, the experiments reported in this chapter are obtained with our monitoring system [16] from several weeks of acquisition: at the end of 2 weeks we manually labeled the stored videos (a video is stored if faces are detected in it) and built models for all the individuals that have a rich data set. From the third week onward data were gathered and labeled for testing: individuals that did not previously appear were stored to be used as negative examples (impostors) for all models. In total, 16 individuals were included in the training phase, while a total of 64 individuals were gathered for testing. Figure 10.4 shows the top five features extracted for some individuals. Notice how the features localize in the most distinctive areas of the face for each individual. For person 6, for instance, the beard is spotted. The top two features of person 13 are localized where there is a fringe. Most of the features extracted in this case are in the nose area.

Let us now describe briefly the actual testing procedure. After we have selected the appropriate set of features for each individual, we are ready to train and tune a classifier to discriminate between pairs of images of the individual  $\mathcal{I}$  under consideration, and pairs of him/her with an impostor. The rationale of this classifier is that when a probe claims to be  $\mathcal{I}$  it will be coupled with all the elements in the gallery of  $\mathcal{I}$ , and as many test pairs as the size of the gallery will be built. These test pairs will be classified to see whether they represent the same person  $\mathcal{I}$ ; thanks to feature selection the classification will be entirely based on the features that are important for  $\mathcal{I}$ . As a classifier, we consider again linear SVMs, this time without the need for a coarse-to-fine strategy, as we perform only one evaluation per frame.

To test the effectiveness of our authentication we run each classifier on the test set that we gathered, containing about 1700 probes, assuming an equal a priori probability of genuine people and impostors (similar to the evaluation procedure proposed in [31]): we use all positive data available for each classifier, and an equal number of negatives randomly selected. Each test image  $T$  fed on a given classifier  $\mathcal{I}_i$  produces  $M$  test pairs. In order to associate a unique output to each probe

**Fig. 10.5** The trend of precision-recall obtained by varying from 20 to 90 the percentage of positive test data obtained from a single probe image. The *darker spot* indicates the precision-recall at 50% currently used by our system



we compute the percentage of the test pairs generated by it that were classified as positives.

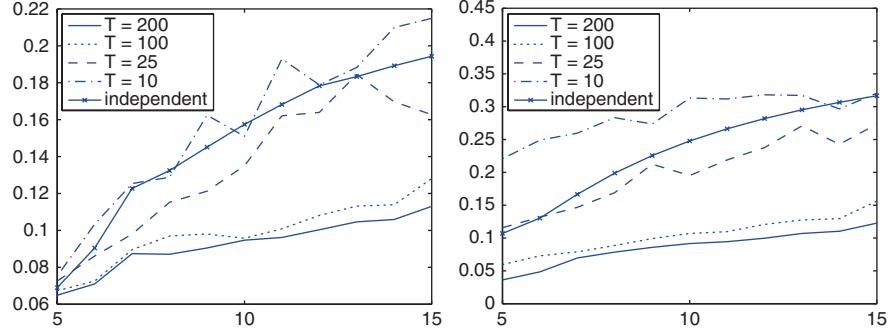
Figure 10.5 shows the trend of average precision-recall by varying the percentage of positive scores  $q$  in the range 20–90%.

#### 10.4.4 Numerical Evaluation of Multi-task Learning

As we pointed out earlier, multi-task learning may be used to learn groups of shared features among different tasks that could represent, for instance, different individuals. In this section we present numerical simulations showing the performance and the quality of the features learned over a growing number of tasks.

We use the square loss function and automatically tuned the regularization parameter  $\gamma$  with cross-validation. We consider up to  $T = 200$  regression tasks. Each of the  $w_t$  parameters of these tasks was sampled from a 5D Gaussian distribution with zero mean and covariance  $Cov = \text{Diag}(1, 0.64, 0.49, 0.36, 0.25)$ . To these 5D  $w_t$ 's we kept adding up to 20 irrelevant dimensions which are exactly zero. The training and test data were generated uniformly from  $[0, 1]^d$  where  $d$  ranged from 5 to 25. The outputs  $y_{ti}$  were computed from the  $w_t$  and  $x_{ti}$  as  $y_{ti} = \langle w_t, x_{ti} \rangle + \vartheta$ , where  $\vartheta$  is zero-mean Gaussian noise with standard deviation equal to 0.1. Thus, the true features  $\langle u_i, x \rangle$  we wish to learn were in this case just the input variables. The desired result is a feature matrix  $U$  which is close to the identity matrix (on five columns) and a matrix  $D$  approximately proportional to the covariance  $Cov$  used to generate the task parameters (on a  $5 \times 5$  principal submatrix).

We generated 5 and 20 examples per task for training and testing, respectively. To test the effect of the number of jointly learned tasks on the test performance and (more importantly) on the quality of the features learned, we used our methods with  $T = 10, 25, 100, 200$  tasks. For  $T = 10, 25$  and 100, we averaged the performance metrics over randomly selected subsets of the 200 tasks, so that our estimates



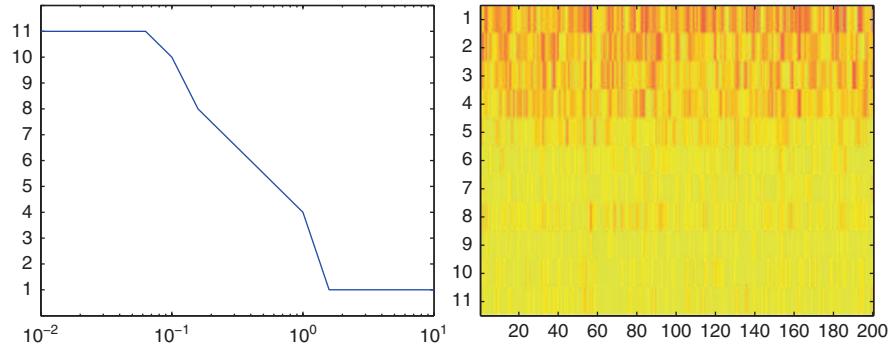
**Fig. 10.6** *Left:* test error versus the number of irrelevant variables, as the number of tasks simultaneously learned changes. *Right:* Frobenius norm of the difference of the learned and actual matrices  $D$  versus the number of irrelevant variables, as the number of tasks simultaneously learned changes. This is a measure of the quality of the learned features

have comparable variance. We also estimated each of the 200 tasks independently using standard ridge regressions.

We present, in Figure 10.6, the impact of the number of tasks simultaneously learned on the test performance as well as the quality of the features learned, as the number of irrelevant variables increases. First, as the left plot shows, in agreement with past empirical and theoretical evidence – see e.g., [7] – learning multiple tasks together significantly improves on learning the tasks independently, as the tasks are indeed related in this case. Moreover, performance improves as the number of tasks increases. More important, this improvement increases with the number of irrelevant variables.

The plot on the right of Fig. 10.6 is the most relevant one for our purposes. It shows the distance of the learned features from the actual ones used to generate the data. More specifically, we depict the Frobenius norm of the difference of the learned  $5 \times 5$  principal submatrix of  $D$  and the actual  $Cov$  matrix (normalized to have trace 1). We observe that adding more tasks leads to better estimates of the underlying features, a key contribution of this chapter. Moreover, like for the test performance, the relative (as the number of tasks increases) quality of the features learned increases with the number of irrelevant variables. Similar results were obtained by plotting the residual of the learned  $U$  from the actual one, which is the identity matrix in this case.

We also tested the effect of the regularization parameter  $\gamma$  on the number of features learned (as measured by  $\text{rank}(D)$ ) for six irrelevant variables. We show the results on the left plot of Fig. 10.7. As expected, the number of features learned decreases with  $\gamma$ . Finally, the right plot in Fig. 10.7 shows the absolute values of the elements of matrix  $A$  learned using the parameter  $\gamma$  selected by leave-one-out cross-validation. This is the resulting matrix for 6 irrelevant variables and all 200 simultaneously learned tasks. This plot indicates that our algorithm learns a matrix  $A$  with the expected structure: there are only five important features. The (normalized) 2-norms of the corresponding rows are 0.31, 0.21, 0.12, 0.10, and 0.09,



**Fig. 10.7** Linear synthetic data. *Left:* number of features learned versus the regularization parameter  $\gamma$  (see text for description). *Right:* matrix  $A$  learned, indicating the importance of the learned features – the first five rows correspond to the true features (see text). The color scale ranges from yellow (low values) to purple (high values)

respectively, while the true values (diagonal elements of  $Cov$  scaled to have trace 1) are 0.36, 0.23, 0.18, 0.13, and 0.09, respectively.

## 10.5 Discussion

Over the past decades machine learning proposed effective solutions to the challenging problem of biometrics. These solutions were particularly important in the face biometrics case, whose major drawback is the lack of robustness to appearance variations. The last few years witnessed many interesting advances on the statistical learning theory, in particular for what concerns feature learning. Today the field is mature to adopt these theoretically grounded solutions in many application fields.

This chapter analyses two recent advances concerning features learning from data: in the first case we describe a feature selection process, designed to deal with binary problems, and in the second case we evaluate a method that learns groups of shared features in a multi-task setting.

The binary feature selection approach is based on the hypothesis of a linear relationship between input and output. It assumes that we are looking for a sparse solution with no explicit account on the presence of correlated features. In the applications domain that we have explored in this chapter such assumption is appropriate, but we mention the fact that, in order to obtain a more principled way to treat correlated features, a mixed (1–2)-norm penalty may be applied, to the price of a more redundant description (see, for instance, [15, 38]).

For what concerns multi-task learning the theoretical advances are ready to be applied to real-world domains. The face recognition problem is a challenging domain where it would be interesting to understand possible face structures common to different individuals.

**Acknowledgments** Most of the work presented in this chapter is a joint effort with A. Argyriou, A. Destrero, and C. De Mol. The authors thank L. Rosasco and E. De Vito for many useful discussions. This work has been partially supported by the FIRB project LEAP RBIN04PRL.

## Proposed Questions and Exercises

- Implement a Matlab version of the soft thresholding iterative scheme to solve the 1-norm penalty (Lasso) feature selection or adapt the scheme to solve the  $(1 - 2)$ -norm penalty available online<sup>3</sup> to the pure 1-norm case
- Implement a Matlab version of the rectangle features [56]
- Apply the feature selection method to the *CBCL – CMU* data set choosing among one of the following descriptions: gray-levels, Haar wavelets, rectangle features. Are gray-levels appropriate for this scheme? Motivate.
- Write a pseudo-code that adopts the extended LBP [39] to represent faces in the authentication setting; compute the number of features obtained with this extension considering a reasonable number of different scales (what is “reasonable” with a  $40 \times 40$  image?)
- Show how the 1-norm penalty feature selection algorithm is a special case of multi-task feature learning.
- Implement a Matlab version of the multi-task feature learning algorithm reported in Table 10.2. Reproduce the experimental evaluation proposed in [6].

## References

1. G. Aggarwal, A. Chowdhury, and R. Chellappa. A system identification approach for video-based face recognition. In *Proc. International Conference on Pattern Recognition*, pages 175–178, 2004.
2. T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: application to face recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(12):2037–2041, 2006.
3. R. K. Ando and T. Zhang. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6:1817–1853, 2005.
4. A. Argyriou, T. Evgeniou, and M. Pontil. Convex multi-task feature learning. *Machine Learning*, 73(3): 243–272, 2008.
5. A. Argyriou, T. Evgeniou, and M. Pontil. Multi-task feature learning. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*. MIT Press, 2006.
6. A. Argyriou, T. Evgeniou, and M. Pontil. Multi-task feature learning. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19*. MIT Press, 2007. In press.
7. J. Baxter. A model for inductive bias learning. *Journal of Artificial Intelligence Research*, 12:149–198, 2000.
8. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces versus fisherfaces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19:711–720, 1997.

---

<sup>3</sup> <http://slipguru.disi.unige.it/>

9. M. Bicego, A. Lagorio, E. Grossi, and M. Tistarelli. On the use of sift features for face authentication. In *Proc. of IEEE Int Workshop on Biometrics, in association with CVPR06*, page 35ff, 2006.
10. R. Brunelli and T. Poggio. Face recognition: features versus templates. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15:1042–1052, 1993.
11. S. S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal of Scientific Computing*, 20(1), 1998.
12. I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure Applied Mathematics*, 57, 2004.
13. A. Destrero, C. De Mol, F. Odone, and A. Verri. A regularized approach to feature selection for face detection. Technical Report DISI-TR-07-01, Dipartimento di informatica e scienze dell'informazione, Universita' di Genova, 2007.
14. A. Destrero, C. De Mol, F. Odone, and A. Verri. A regularized approach to feature selection for face detection. In Y. Yagi et al., editor, *Proc. of the Asian Conference on Computer Vision, ACCV*, LNCS 4844, pages 881–890, 2007.
15. A. Destrero, S. Mosci, C. De Mol, A. Verri, and F. Odone. Feature selection for high dimensional data. *Computational Management Science* 6(1): 25–40 (2009).
16. A. Destrero, F. Odone, and A. Verri. A system for face detection and tracking in unconstrained environments. In *IEEE International Conference on Advanced Video and Signal-based Surveillance*, In Proceedings IEEE AVSS 2007, pages 499–504, 2007.
17. D. Donoho. High-dimensional data analysis: The curses and blessings of dimensionality. Aide-Memoire of a Lecture at AMS conference on Math Challenges of 21st Century. Available at <http://www-stat.stanford.edu/~donoho/Lectures/AMS2000/AMS2000.html>
18. G. J. Edwards, C. J. Taylor, and T. F. Cootes. Improving identification performance by integrating evidence from sequences. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, vol. 1, pp. 1486, 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'99) – Volume 1, 1999.
19. K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images. *Journal of the Optical Society of America A*, 14:1724–1733, 1997.
20. M. Fazel, H. Hindi, and S. P. Boyd. A rank minimization heuristic with application to minimum order system approximation. In *Proceedings, American Control Conference*, 4734–4739, 2001.
21. Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *European Conference on Computational Learning Theory*, pages 23–37, 1995.
22. J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting, 1998.
23. D. O. Gorodnichy. On importance of nose for face tracking. In *IEEE International conference on automatic face and gesture recognition*, pages 181–186, 2002.
24. I. Guyon and E. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003.
25. A. Hadid, M. Pietikäinen, and S. Z. Li. Learning personal specific facial dynamics for face recognition from videos. In *Analysis and Modeling of Faces and Gestures*, pages 1–15, Springer LNCS 4778, 2007.
26. X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using laplacianfaces. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 27:328–340, 2005.
27. B. Heisele, P. Ho, J. Wu, and T. Poggio. Face recognition: component-based versus global approaches. *Computer Vision and Image Understanding*, 91:6–21, 2003.
28. B. Heisele, T. Serre, M. Pontil, and T. Poggio. Component-based face detection. In *CVPR*, 2001.
29. A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Trans. on Circuits and Systems for Video Technology*, 14(1), 2004.

30. T. Jebara. Multi-task feature and kernel selection for SVMs. In *Proceedings of the 21st International Conference on Machine Learning*, 2004.
31. K. Messer, J. Kittler, M. Sadeghi, M. Hamouz, A. Kostyn, S. Marcel, S. Bengio, F. Cardinaux, C. Sanderson, N. Poh, Y. Rodriguez, K. Kryszczuk, J. Czyz, L. Vandendorpe, J. Ng, H. Cheung, and B. Tang. Face authentication competition on the banca database. In *Biometric Authentication*, LNCS 3072, 2004.
32. B. Li and R. Chellappa. Face verification through tracking facial features. *Journal of the Optical Society of America*, JOSA-A, 18(12): 2969–2981, 2001.
33. S. Z. Li and Z. Q. Zhang. FloatBoost learning and statistical face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9), 2004.
34. R. Lanzarotti S. Arca, P. Campadelli. A face recognition system based on automatically determined facial fiducial points. *Pattern Recognition*, 39(3):432–443, 2006.
35. X. Liu, T. Chen, and B. V. K. Vijaya Kumar. On modeling variations for face authentication. *Pattern Recognition*, 36(2):313–328, 2003.
36. B. Moghaddam and A. Pentland. Probabilistic visual learning for object representation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19:696–710, 1997.
37. A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. *IEEE Trans. on PAMI*, 23(4):349–361, 2001.
38. C. de Mol, S. Mosci, M. S. Traskine, and A. Verri. Sparsity enforcing and correlation preserving algorithm for microarray data analysis. Technical Report DISI-TR-07-04, Dipartimento di informatica e scienze dell'informazione, Universita' di Genova, 2007.
39. T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(7), 2002.
40. E. Osuna, R. Freund, and F. Girosi. Training support vector machines: an application to face detection. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on (CVPR'97), pp. 130, 1997.
41. A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, page 84-91, 1994.
42. A. Pentland, B. Moghaddam, and T. Starner. Estimation of eye, eyebrow and nose features in videophone sequences. In *International Workshop on Very Low Bitrate Video Coding (VLBV 98)*, page 101-104, 1998.
43. M. Pontil and A. Verri. Support vector machines for 3-d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:637–646, 1998.
44. D. Roth, M. Yang, and N. Ahuja. A snowbased face detector. *Neural Information Processing*, 12, 2000.
45. H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20:22–38, 1998.
46. H. Schneiderman and T. Kanade. A statistical method for 3D object detection applied to faces and cars. In *International Conference on Computer Vision*, 2000.
47. J. Sergent. Microgenesis of face perception. In H. D. Ellis, M. A. Jeeves, F. Newcombe, and A. M. Young, editors, *Aspects of face processing* (pp. 17–33). Dordrecht: Martinus Nijhoff, 1986.
48. S. Soatto, G. Doretto, and Y. Wu. Dynamic textures. In *Proc of the International Conference on Computer Vision*, pages 439–446, 2001.
49. N. Srebro, J. D. M. Rennie, and T. S. Jaakkola. Maximum-margin matrix factorization. In *Advances in Neural Information Processing Systems*, 17, pages 1329–1336. MIT Press, 2005.
50. K. Sung and T. Poggio. Example-based learning for view-based face detection. *IEEE Transactions on PAMI*, 20, 1998.
51. R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58(1):267–288, 1996.

52. A. Torralba, K. P. Murphy, and W. T. Freeman. Sharing features: efficient boosting procedures for multiclass object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2: 762–769, 2004.
53. M. A. Turk and A. P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.
54. S. Ullman, M. Vidal-Naquet, and E. Sali. Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, 5(7), 2002.
55. V. N. Vapnik. *Statistical Learning Theory*. Wiley, 1998.
56. P. Viola and M. J. Jones. Robust real-time face detection. *International Journal on Computer Vision*, 57(2):137–154, 2004.
57. J. Weston, A. Elisseeff, B. Scholkopf, and M. Tipping. The use of zero-norm with linear models and kernel methods. *Journal of Machine Learning Research*, 3, 2003.
58. L. Wiskott, J. Fellous, N. Kuiger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:775–779, 1997.
59. M.-H. Yang, D. J. Kriegman, and N. Ahuja. Detecting faces in images: a survey. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.
60. W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips. Face recognition: a literature survey. *ACM Computing Surveys*, 35(4): 399–458, 2003.
61. J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani. 1-norm support vector machines. In *Advances in Neural Information Processing Systems*, 16. MIT Press, 2004.

# **Chapter 11**

## **Multibiometric Systems: Overview, Case Studies, and Open Issues**

**Arun Ross and Norman Poh**

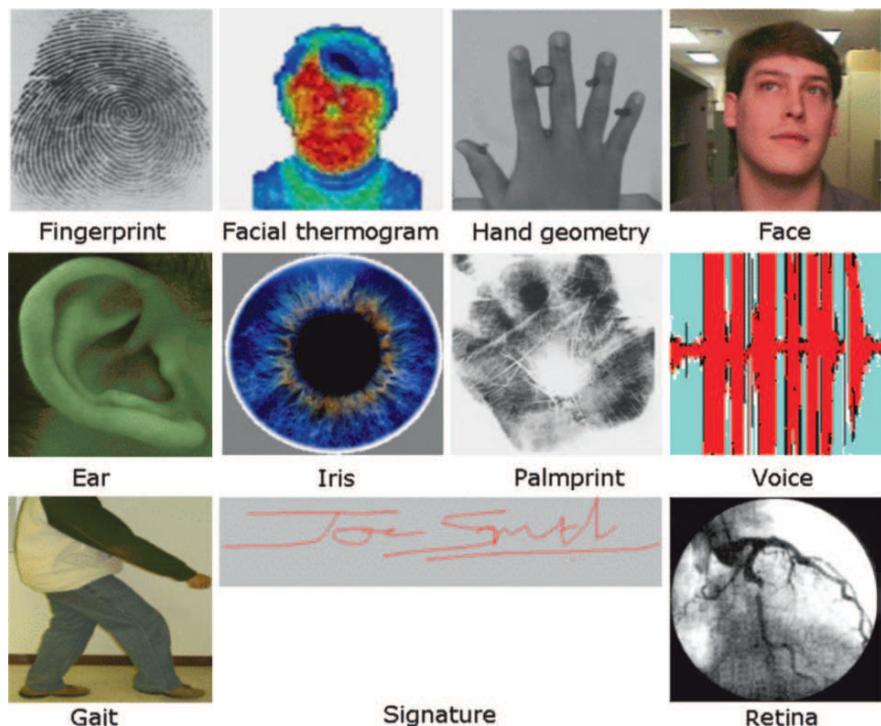
**Abstract** Information fusion refers to the reconciliation of evidence presented by multiple sources of information in order to generate a decision. In the context of biometrics, evidence reconciliation plays a pivotal role in enhancing the recognition accuracy of human authentication systems and is referred to as multibiometrics. Multibiometric systems combine the information presented by multiple biometric sensors, algorithms, samples, units, or traits in order to establish the identity of an individual. Besides enhancing matching performance, these systems are expected to improve population coverage, deter spoofing, facilitate continuous monitoring, and impart fault tolerance to biometric applications. This chapter introduces the topic of multibiometrics and enumerates the various sources of biometric information that can be consolidated as well as the different levels of fusion that are possible in a biometric system. The role of using ancillary information such as biometric data quality and soft biometric traits (e.g., height) to enhance the performance of these systems is discussed. Three case studies demonstrating the benefits of a multibiometric system and the factors impacting its architecture are also presented. Finally, some of the open challenges in multibiometric system design and implementation are enumerated.

### **11.1 Introduction**

A reliable identity management system is a critical component in several applications that render services to only legitimately enrolled users. Examples of such applications include sharing networked computer resources, granting access to nuclear facilities, performing remote financial transactions, or boarding a commercial flight. The proliferation of web-based services (e.g., online banking) and the deployment of decentralized customer service centers (e.g., credit cards) have further enhanced the need for reliable identity management systems. Traditional

---

A. Ross (✉)  
West Virginia University, Morgantown, West Virginia, USA  
e-mail: arun.ross@mail.wvu.edu



**Fig. 11.1** Examples of biometric traits that can be used for authenticating an individual

methods of establishing a person's identity include knowledge-based (e.g., passwords) and token-based (e.g., ID cards) mechanisms, but these surrogate representations of identity can be easily lost, shared, manipulated, or stolen thereby undermining the intended security. Biometrics offers a natural and reliable solution to certain aspects of identity management by utilizing fully automated or semi-automated schemes to recognize individuals based on their inherent physical and/or behavioral characteristics [28]. By using biometrics (see Fig. 11.1) it is possible to establish an identity based on *who you are*, rather than by *what you possess*, such as an ID card, or *what you remember*, such as a password.

Most biometric systems that are presently in use, typically use a single biometric trait to establish identity (i.e., they are unibiometric systems). Some of the challenges commonly encountered by biometric systems are listed below:

1. Noise in sensed data: The biometric data being presented to the system may be contaminated by noise due to imperfect acquisition conditions or subtle variations in the biometric itself. For example, a scar can change a subject's fingerprint while the common cold can alter the voice characteristics of a speaker. Similarly, unfavorable illumination conditions may significantly affect the face and

- iris images acquired from an individual. Noisy data can result in an individual being incorrectly labeled as an impostor thereby increasing the false reject rate (FRR) of the system.
2. Non-universality: The biometric system may not be able to acquire meaningful biometric data from a subset of individuals resulting in a failure-to-enroll (FTE) error. For example, a fingerprint system may fail to image the friction ridge structure of some individuals due to the poor quality of their fingerprints. Similarly, an iris recognition system may be unable to obtain the iris information of a subject with long eyelashes, drooping eyelids, or certain pathological conditions of the eye. Exception processing will be necessary in order to include such users into the authentication system.
  3. Upper bound on identification accuracy: The matching performance of a unibiometric system cannot be continuously improved by tuning the feature extraction and matching modules. There is an implicit upper bound on the number of distinguishable patterns (i.e., the number of distinct biometric feature sets) that can be represented using a template. The capacity of a template is constrained by the variations observed in the feature set of each subject (i.e., *intra-class* variations) and the variations between feature sets of different subjects (i.e., *inter-class* variations).
  4. Spoof attacks: Behavioral traits such as voice [15] and signature [21] are vulnerable to spoof attacks by an impostor attempting to mimic the traits corresponding to legitimately enrolled subjects. Physical traits such as fingerprints can also be spoofed by inscribing ridge-like structures on synthetic material such as gelatine and play-doh [41, 55]. Targeted spoof attacks can undermine the security afforded by the biometric system and, consequently, mitigate its benefits [56].

Some of the limitations of a unibiometric system can be addressed by designing a system that consolidates *multiple* sources of biometric information. This can be accomplished by fusing, for example, multiple traits of an individual or multiple feature extraction and matching algorithms operating on the same biometric. Such systems, known as multibiometric systems [27, 60], can improve the matching accuracy of a biometric system while increasing population coverage and deterring spoof attacks.

The rest of the chapter is structured as follows. Section 11.2 lists some of the advantages of using multibiometric systems; Section 11.3 presents the taxonomy used to characterize these systems; Section 11.4 provides an overview of the various levels of fusion that are possible; Section 11.5 discusses the possibility of incorporating ancillary features such as soft biometrics and quality in order to enhance the matching performance of fusion systems; Section 11.6 presents three case studies highlighting the benefits of fusion and the factors impacting its architecture; Section 11.7 lists some of the open challenges in biometric fusion; Section 11.8 summarizes the contributions of this chapter.

## 11.2 Advantages of Multibiometric Systems

Besides enhancing matching accuracy, the other advantages of multibiometric systems over traditional unibiometric systems are enumerated below [60]:

1. Multibiometric systems address the issue of non-universality (i.e., limited population coverage) encountered by unibiometric systems. If a subject's dry finger prevents her from successfully enrolling into a fingerprint system, then the availability of another biometric trait, for instance her iris, can aid in the inclusion of the individual in the biometric system. A certain degree of flexibility is achieved when a user enrolls into the system using several different traits (e.g., face, voice, fingerprint, iris, hand) while only a subset of these traits (e.g., face and voice) is requested during authentication based on the nature of the application under consideration and the convenience of the user.
2. Multibiometric systems can facilitate the filtering or indexing of large-scale biometric data bases. For example, in a bimodal system consisting of face and fingerprint, the face feature set may be used to compute an index value for extracting a candidate list of potential identities from a large data base of subjects. The fingerprint modality can then determine the final identity from this limited candidate list.
3. It becomes increasingly difficult (if not impossible) for an impostor to spoof multiple biometric traits of a legitimately enrolled individual. If each subsystem indicates the probability that a particular trait is a "spoof," then appropriate fusion schemes can be employed to determine if the user is, in fact, an impostor or not. Furthermore, by asking the user to present a random subset of traits at the point of acquisition, a multibiometric system facilitates a challenge-response type of mechanism, thereby ensuring that the system is interacting with a *live* user. Note that a challenge-response mechanism can be initiated in unibiometric systems also (e.g., system prompts "Please say 1-2-5-7," "Blink twice and move your eyes to the right," and "Change your facial expression by smiling").
4. Multibiometric systems also effectively address the problem of noisy data. When the biometric signal acquired from a single trait is corrupted with noise, the availability of other (less noisy) traits may aid in the reliable determination of identity. Some systems take into account the *quality* of the individual biometric signals during the fusion process. This is especially important when recognition has to take place in adverse conditions where certain biometric traits cannot be reliably extracted. For example, in the presence of ambient acoustic noise, when an individual's voice characteristics cannot be accurately measured, the facial characteristics may be used by the multibiometric system to perform authentication. Estimating the quality of the acquired data is in itself a challenging problem but, when appropriately done, can reap significant benefits in a multibiometric system.
5. These systems also help in the *continuous* monitoring or tracking of an individual in situations when a single trait is not sufficient. Consider a biometric system that uses a 2D camera to procure the face and gait information of a person walking

down a crowded aisle. Depending upon the distance and pose of the subject with respect to the camera, both these characteristics may or may not be simultaneously available. Therefore, either (or both) of these traits can be used depending upon the location of the individual with respect to the acquisition system thereby permitting the continuous monitoring of the individual.

6. A multibiometric system may also be viewed as a fault-tolerant system which continues to operate even when certain biometric sources become unreliable due to sensor or software malfunction or deliberate user manipulation. The notion of fault tolerance is especially useful in large-scale authentication systems involving a large number of subjects (such as a border control application).

### 11.3 Taxonomy of Multibiometric Systems

A multibiometric system relies on the evidence presented by multiple sources of biometric information. Based on the nature of these sources, a multibiometric system can be classified into one of the following six categories [60]: multi-sensor, multi-algorithm, multi-instance, multi-sample, multimodal, or hybrid (see Fig. 11.2).

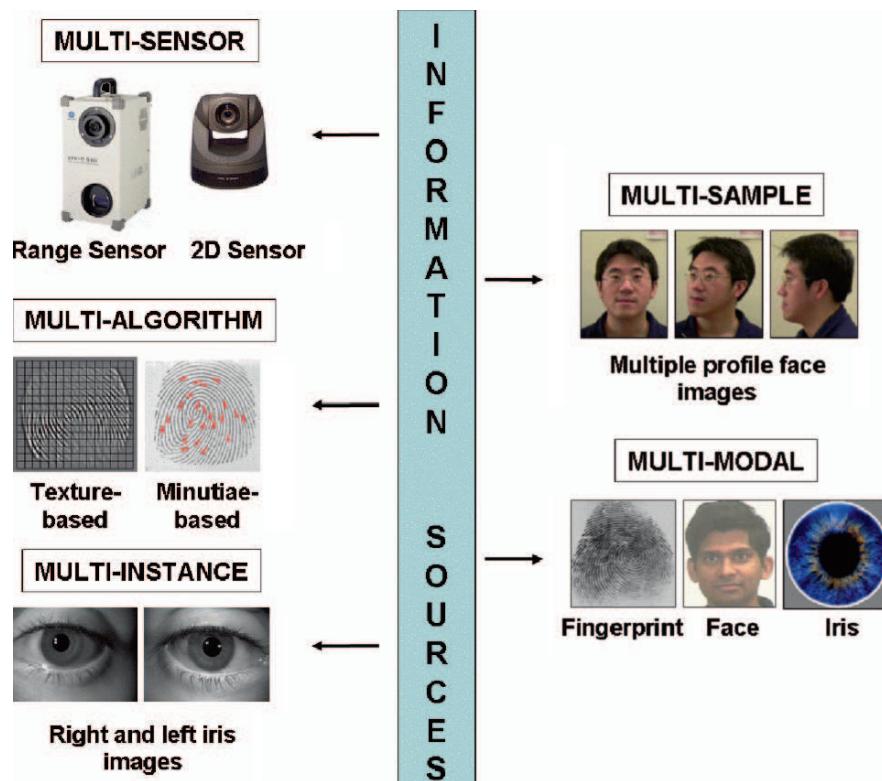


Fig. 11.2 Sources of information for biometric fusion

1. Multi-sensor systems: Multi-sensor systems employ multiple sensors to capture a single biometric trait of an individual. For example, a face recognition system may deploy multiple 2D cameras to acquire the face image of a subject [37]; an infrared sensor may be used in conjunction with a visible-light sensor to acquire the subsurface information of a person's face [8, 31, 67]; a multispectral camera may be used to acquire images of the iris, face, or finger [48, 62]; or optical and capacitive sensors may be used to image the fingerprint of a subject [40]. The use of multiple sensors, in some instances, can result in the acquisition of complementary information that can enhance the recognition ability of the system. For example, based on the nature of illumination due to ambient lighting, the infrared and visible-light images of a person's face can present different levels of information resulting in enhanced matching accuracy. Similarly, the performance of a 2D face matching system can be improved by utilizing the shape information presented by 3D range images.
2. Multi-algorithm systems: In some cases, invoking multiple feature extraction and/or matching algorithms on the same biometric data can result in improved matching performance. Multi-algorithm systems consolidate the output of multiple feature extraction algorithms or that of multiple matchers operating on the same feature set. These systems do not necessitate the deployment of new sensors and, hence, are cost-effective compared to other types of multibiometric systems. But on the other hand, the introduction of new feature extraction and matching modules can increase the computational complexity of these systems. Ross et al. [59] describe a fingerprint recognition system that utilizes minutiae as well as texture information to represent and match fingerprint images. Lu et al. [39] discuss a face recognition system that combines three different feature extraction schemes (principal component analysis (PCA), independent component analysis (ICA), and linear discriminant analysis (LDA)).
3. Multi-instance systems: These systems use multiple instances of the same body trait and have also been referred to as multi-unit systems in the literature. For example, the left and right index fingers, or the left and right irises of an individual, may be used to verify an individual's identity [29, 54]. FBI's IAFIS combines the evidence of all ten fingers to determine a matching identity in the database. These systems can be cost-effective if a single sensor is used to acquire the multi-unit data in a sequential fashion, but this can increase data acquisition time thereby causing inconvenience to the user. Thus, in some instances, it may be desirable to obtain the multi-unit data simultaneously thereby demanding the design of an effective (and possibly more expensive) acquisition device.
4. Multi-sample systems: A single sensor may be used to acquire multiple samples of the same biometric trait in order to account for the variations that can occur in the trait or to obtain a more complete representation of the underlying trait. A face system, for example, may capture (and store) the frontal profile of a person's face along with the left and right profiles in order to account for variations in the facial pose. Similarly, a fingerprint system equipped with a small size sensor may acquire multiple dab prints of an individual's finger in order to obtain images of various regions of the fingerprint. A mosaicing scheme may then be used to stitch the multiple impressions and create a composite image. One of the key issues in

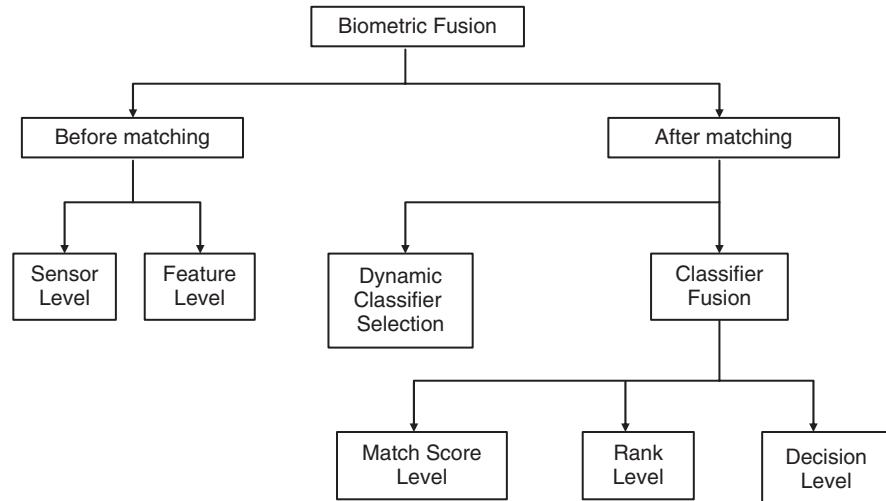
a multi-sample system is determining the *number* of samples that have to be acquired from an individual. It is important that the procured samples represent the *variability* as well as the *typicality* of the individual's biometric data. To this end, the desired relationship between the samples has to be established beforehand in order to optimize the benefits of the integration strategy. For example, a face recognition system utilizing both the frontal- and side-profile images of an individual may stipulate that the side-profile image should be a three-quarter view of the face [22, 47]. Alternately, given a set of biometric samples, the system should be able to automatically select the “optimal” subset that would best represent the individual's variability. Uludag et al. [71] discuss two such schemes in the context of fingerprint recognition.

5. Multimodal systems: Multimodal systems establish identity based on the evidence of multiple biometric traits. For example, some of the earliest multimodal biometric systems utilized face and voice features to establish the identity of an individual [4, 6, 11]. Physically uncorrelated traits (e.g., fingerprint and iris) are expected to result in better *improvement* in performance than correlated traits (e.g., voice and lip movement). The cost of deploying these systems is substantially more due to the requirement of new sensors and, consequently, the development of appropriate user interfaces. The identification accuracy can be significantly improved by utilizing an increasing number of traits although the *curse-of-dimensionality* phenomenon would impose a bound on this number. The number of traits used in a specific application will also be restricted by practical considerations such as the cost of deployment, enrollment time, throughput time, expected error rate, and user habituation issues.
6. Hybrid systems: Chang et al. [7] use the term *hybrid* to describe systems that integrate a subset of the five scenarios discussed above. For example, Brunelli et al. [6] discuss an arrangement in which two speaker recognition algorithms are combined with three face recognition algorithms at the match score and rank levels via a HyperBF network. Thus, the system is multi-algorithmic as well as multimodal in its design.

## 11.4 Levels of Fusion

Based on the type of information available in a certain module, different levels of fusion may be defined. Sanderson and Paliwal [64] categorize the various levels of fusion into two broad categories: pre-classification or fusion *before* matching and post-classification or fusion *after* matching (see Fig. 11.3). Such a categorization is necessary since the amount of information available for fusion reduces drastically once the matcher has been invoked. Pre-classification fusion schemes typically require the development of new matching techniques (since the matchers used by the individual sources may no longer be relevant) thereby introducing additional challenges. Pre-classification schemes include fusion at the sensor (or raw data) and the feature levels while post-classification schemes include fusion at the match score, rank, and decision levels.

1. Sensor-level fusion: The raw biometric data (e.g., a face image) acquired from an individual represents the richest source of information although it is expected



**Fig. 11.3** Fusion can be accomplished at various levels in a biometric system

to be contaminated by noise (e.g., non-uniform illumination, and background clutter). Sensor-level fusion refers to the consolidation of (a) raw data obtained using multiple sensors or (b) multiple snapshots of a biometric using a single sensor [61, 65].

2. Feature-level fusion: In feature-level fusion, the feature sets originating from multiple biometric algorithms are consolidated into a single feature set by the application of appropriate feature normalization, transformation, and reduction schemes. The primary benefit of feature-level fusion is the detection of correlated feature values generated by different biometric algorithms and, in the process, identifying a salient set of features that can improve recognition accuracy. Eliciting this feature set typically requires the use of dimensionality reduction methods and, therefore, feature-level fusion assumes the availability of a large number of training data. Also, the feature sets being fused are typically expected to reside in commensurate vector space in order to permit the application of a suitable matching technique upon consolidating the feature sets [58, 68].
3. Score-level fusion: In score-level fusion the match scores output by multiple biometric matchers are combined to generate a new match score (a scalar) that can be subsequently used by the verification or identification modules for rendering an identity decision. Fusion at this level is the most commonly discussed approach in the biometric literature primarily due to the ease of accessing and processing match scores (compared to the raw biometric data or the feature set extracted from the data). Fusion methods at this level can be broadly classified into three categories [60]: density-based schemes [12, 70], transformation-based schemes [26], and classifier-based schemes [72].
4. Rank-level fusion: When a biometric system operates in the identification mode, the output of the system can be viewed as a ranking of the enrolled identities.

In this case, the output indicates the set of possible matching identities sorted in decreasing order of confidence. The goal of rank-level fusion schemes is to consolidate the ranks output by the individual biometric subsystems in order to derive a consensus rank for each identity. Ranks provide more insight into the decision-making process of the matcher compared to just the identity of the best match, but they reveal less information than match scores. However, unlike match scores, the rankings output by multiple biometric systems are comparable. As a result, no normalization is needed and this makes rank-level fusion schemes simpler to implement compared to the score-level fusion techniques [23].

5. Decision-level fusion: Many commercial off-the-shelf (COTS) biometric matchers provide access only to the final recognition decision. When such COTS matchers are used to build a multibiometric system, only decision-level fusion is feasible. Methods proposed in the literature for decision-level fusion include “AND” and “OR” rules [13], majority voting [36], weighted majority voting [34], Bayesian decision fusion [74], the Dempster–Shafer theory of evidence [74], and behavior knowledge space [24].

## 11.5 Incorporating Ancillary Information

Another category of multibiometric systems combine primary biometric identifiers (such as face and fingerprint) with soft biometric attributes (such as gender, height, weight, and eye color). Soft biometric traits cannot be used to distinguish individuals reliably since the same attribute is likely to be shared by several different people in the target population. However, when used in conjunction with primary biometric traits, the performance of the authentication system can be significantly enhanced [25]. Soft biometric attributes also help in filtering (or indexing) large biometric databases by limiting the number of entries to be searched in the database. For example, if it is determined (automatically or manually) that the subject is an “Asian Male,” then the system can constrain its search to only those identities in the database labeled with these attributes. Alternately, soft biometric traits can be used in surveillance applications to decide if all primary biometric information has to be acquired from a certain individual. Automated techniques to estimate soft biometric characteristics is an ongoing area of research and is likely to benefit law enforcement and border control biometric applications.

Some biometric systems incorporate data quality into the fusion process. This strategy is referred to as *quality-based fusion*. The purpose is to (a) automatically assign weights to the participating modalities thereby mitigating the errors introduced by poor-quality input data [45] or (b) appropriately invoke the modalities in a cascade fashion thereby maximizing recognition accuracy [16]. Soft biometric data and quality indices are referred to as ancillary information in the context of biometric fusion.

In quality-based fusion (e.g., [5, 18, 30, 32, 45]), the quality associated with the template (i.e., gallery) as well as the query (i.e., probe) biometric sample are taken into account. For assessing the quality of a biometric sample, a number of measures

have been proposed in the literature (e.g., fingerprint [10, 19], iris [9], face [20], speech [46], signature [44], and classifier-dependent confidence measures [3, 50]). These quality measures, in general, aim to quantify the degree of excellence or conformance of biometric samples to some predefined criteria known to influence the performance of the system.

Depending on their role, there are at least two ways in which quality measures can be incorporated into a fusion classifier – either as a control parameter (primary role) or as an evidence (secondary role). In their primary role, quality measures are used to modify the way a fusion classifier is trained or tested. For example, quality measures have been incorporated into the following classifiers: Bayesian-based classifier [5], reduced polynomial classifier [69], support vector machine [18], and fixed-rule fusion [17]. In their secondary role, quality measures are often concatenated with the outputs of individual matchers (such as match scores) and the ensuing “feature” is input to a fusion classifier as discussed in [30] (logistic regression classifier) and [45] (Bayesian classifier). The use of Bayesian networks to gauge the complex relationship between expert outputs and quality measures (e.g., [42]) has also been explored. The work in [52] takes into account an array of quality measures rather than summarizing the quality as a scalar value. By means of grouping the multi-faceted quality measures into multiple clusters, a unique fusion strategy can be devised for each cluster.

Quality measures have also been used to improve biometric device interoperability [1, 51]. Such an approach is commonly used in speaker verification where different strategies are invoked based on the microphone type [2].

The notion of quality is closely related to that of *reliability*. In [33], the estimated reliability for each biometric modality was used for combining symbolic-level decisions and in [38, 50, 57, 73], score-level fusion was considered. In [38, 57, 73], the term “failure prediction” is used instead of “reliability.” Such information derived solely from the output of individual matchers (rather than explicit quality measures) has been demonstrated to be effective for (a) single biometric modalities, (b) fusion across multiple sensors for a single biometric modality, and (c) fusion across different machine learning techniques. In [50], the notion of reliability was represented by the *margin*, a concept used in large-margin classifiers [66]. However, a precise definition for reliability and the procedure used for estimating it are open research issues and further work is essential in this regard.

## 11.6 Benefits of Combining Multiple Biometric Experts: Three Case Studies

The literature on biometrics is replete with examples demonstrating the benefits of multibiometric fusion.<sup>1</sup> In this section, three examples illustrating the benefits of multibiometric systems as well as the design issues involved are presented. All the

---

<sup>1</sup> The term “expert” is sometimes used to refer to individual biometric matchers or modalities used in a multibiometric system.

examples are based on results reported in the literature on public data sets. The first case illustrates the potential of multibiometric systems to improve matching accuracy. The second case illustrates the benefit of using quality measures in the fusion scheme. The third case demonstrates the possibility of optimizing the cost of authentication for a given target accuracy by managing the choice of biometric traits and matchers. Such a cost-based analysis enables one to decide, for instance, if combining multiple biometric traits is better than combining multiple samples of the same biometric (by reusing the same device), as the latter is less expensive.

### **11.6.1 On the Complementarity of Multimodal Experts**

The first case study which illustrates the merits of both multimodal and multi-algorithm fusion [63] involves fusing three different modalities: face, voice, and lip movement. Two face recognition algorithms, two speaker recognition algorithms, and one lip dynamic recognition algorithm were used. This multibiometric system was evaluated on the XM2VTS database [43] producing match scores according to the Lausanne Experimental Protocol in Configuration I [43] (see Table 11.1). Although the performances of the individual experts were unremarkable (with the exception of one speaker recognition algorithm), the fusion of these biometric experts by simple weighted averaging resulted in improved performance, as summarized in Table 11.2. The results show that multimodal fusion has the potential to ameliorate the performance of the single best expert even if some of the individual recognition algorithms have error rates that are an order of magnitude worse than the best expert. Interestingly, the combination of the best experts from each of the three modalities is only marginally better than the best performing speaker recognition algorithm. In the second row of Table 11.2, it is observed that the weights assigned to the weaker algorithms are greater than those associated with the best algorithm of

**Table 11.1** Performance of multiple biometric algorithms (i.e., experts) on the test set (Configuration I). Results are from the work reported in [63]

Algorithm	Threshold	FRR (%)	FAR (%)
Lips	0.50	14.00	12.67
Face 1	0.21	5.00	4.45
Face 2	0.50	6.00	8.12
Voice 1	0.50	7.00	1.42
Voice 2	0.50	0.00	1.48

**Table 11.2** Fusion using the simple sum rule (Configuration I). Results are from the work reported in [63]

Experts	Weights	Threshold	FRR (%)	FAR (%)
Lips, Face 1, Voice 2	0.27, 0.23, 0.50	0.51	0.00	1.31
Face 1, Face 2, Voice 1, Voice 2	0.02, 0.06, 0.87, 0.05	0.50	0.00	0.52
Lips, Face 1, Face 2, Voice 1, Voice 2	0.03, 0.01, 0.04, 0.89, 0.03	0.50	0.00	0.29

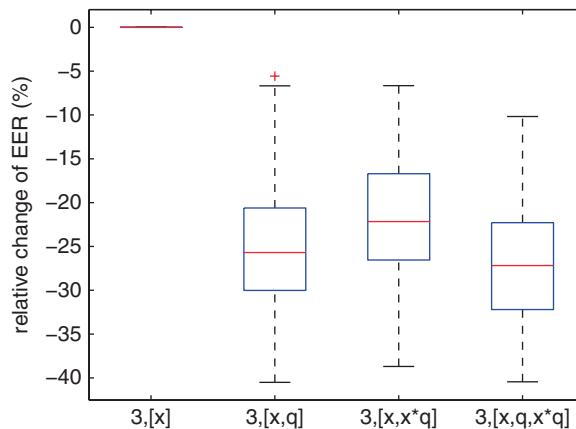
each modality. It appears that the diversity offered by these weaker algorithms has led to an improved matching accuracy after fusion.

### 11.6.2 Benefits of Quality-Based Fusion

The second example demonstrates the benefits of using quality measures in fusion [30]. In the referenced study, logistic regression was used as a fusion classifier, producing an output that approximates the posterior probability of observing a vector of match scores (denoted as  $x$ ); the elements of the vector correspond to the match scores generated by the individual experts.

Quality measures (denoted as  $q$ ) are then considered as additional inputs to the fusion module. The interaction of quality measures and match scores is explicitly modeled by feeding three variants of input to the fusion classifier:  $[x, q]$  (i.e., augmenting  $x$  and  $q$  via concatenation),  $[x, x \otimes q]$  (where  $\otimes$  denotes a tensor product), and  $[x, q, x \otimes q]$ . If there are  $N_q$  terms in  $q$  and  $N_x$  terms in  $x$ , the tensor product between  $q$  and  $x$  produces  $N_q \times N_x$  elements, hence, providing the fusion classifier with an additional degree of freedom to model the pair-wise product elements generated by the two vectors. Thus, the input to the fusion module has four possible arrangements:  $x$ ,  $[x, q]$ ,  $[x, x \otimes q]$ , and  $[x, q, x \otimes q]$ .

The results of using these input arrangements on the XM2VTS database with both standard and degraded data are shown in Fig. 11.4. Six face recognition algorithms and one speaker recognition algorithm were used in the experiment.



**Fig. 11.4** Relative change in a posteriori EER (%) when combining the outputs of multiple biometric algorithms on the “good” and “degraded” subsets of the XM2VTS face database according to the modified Lausanne Protocol Configuration I [30]. Each bar shows the distribution of 63 values corresponding to the 63 possible face and speaker fusion tasks. The first and third quantiles are depicted by a bounding box and the median value by a horizontal red bar. In arrangement  $[x]$ , the quality information is not used. The mean absolute performance of the first bar is about 3% (EER) whereas that of the remaining three quality-based fusion systems is about 2%

Since the voice modality was included in all fusion tasks, the number of ways in which the experts can be combined is  $2^6 - 1 = 63$ . Each bar in this figure represents a statistic measuring the relative difference in error between a fusion system that does *not* use quality and one of the arrangements mentioned above. As can be observed, the use of quality measures can reduce the verification error (measured using the equal error rate (EER)) by as much as 40%. Such an improvement is possible especially when both the face and voice biometric modalities contain significantly different quality types. In the experimental setting, the speech was artificially corrupted by uniformly distributed additive noise of varying magnitude (from 0 to 20 dB) whereas the face data contained either well-illuminated or side-illuminated face images. More research is needed to account for cases when the noise characteristics are ill-defined or unknown.

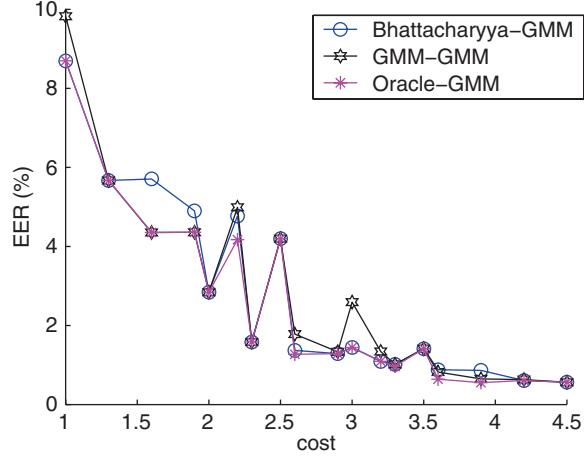
### 11.6.3 Multibiometric Expert Selection

The third example casts the fusion problem as an optimization problem. Since using more than one biometric modality often implies the need for additional hardware and software resources as well as authentication time (thus inconveniencing the user), it is reasonable to attribute an abstract cost to each additional biometric device. The goal of optimization in this context can be formulated as follows: determine the subset of candidate biometric experts that minimizes the overall cost of operation while exhibiting reasonable matching accuracy. Ideally, such a cost-sensitive optimization criterion should be robust to population mismatch, where one attempts to design a fusion classifier based on a development set of subjects and apply the resultant classifier to a target set of subjects not present in the development set. One possible objective function (criterion) for this optimization problem could be the error rate of the fusion classifier.

An example of a cost-sensitive performance curve for a verification experiment is shown in Fig. 11.5. This experiment was conducted on the Biosecure DS2 database.<sup>2</sup> Match scores were generated using eight different recognition algorithms pertaining to the face, fingerprint, and iris modalities. In order to carry out the experiment, the match score database was partitioned into two: the development set, for training, and the evaluation set, for testing. The genuine and impostor user populations were different for these two sets. Two theoretical error measures (the Chernoff and Bhattacharyya bounds) and two empirical EER measures (corresponding to two Bayesian classifiers) were used as objective functions for the optimization problem. The first Bayesian classifier was designed by estimating the genuine and impostor densities using a Gaussian mixture model and using the likelihood ratio test statistic to classify an arbitrary score (GMM classifier). The second classifier was designed by estimating a decision boundary to separate the genuine and impostor scores; the

---

<sup>2</sup> The data set is available for download at <http://face.ee.surrey.ac.uk/qfusion>.



**Fig. 11.5** The rank-one cost-sensitive performance curve using the GMM Bayesian classifier on all 255 combinations in the Biosecure DS2 fusion database. The legend “Bhattacharyya–GMM” refers to optimization using the Bhattacharyya criterion on the development set and measuring the performance of the GMM Bayesian classifier on the evaluation set. The curve “GMM–GMM” should be interpreted similarly. The Oracle–GMM is the performance of the GMM Bayesian classifier on the evaluation set. The experimental results using the QDA Bayesian classifier are similar to this figure (not shown here)

boundary itself was estimated using quadratic discriminant analysis (QDA classifier). See [53] for details. For estimating the empirical error rate on the development set, a twofold cross-validation procedure was employed. The average EER of the twofolds was used as an indicator of the error on the development set. The classifier trained using the development set was then used to assess the empirical error rate on the evaluation set in a similar way. In order to compute the theoretical error bounds, the genuine and impostor match scores were modeled using normal distributions whose parameters were estimated using the development/evaluation data set.

These error measures were computed for both the development and evaluation sets for various combinations of experts. The cost of each combination was also computed. In the fusion problem considered here (see Fig. 11.5), the cost ranges from 1 (using a single expert) to 4.5 (using all eight experts). Costs were assigned as follows: the use of one system is charged a unit cost; subsequent reuse of the system (e.g., multiple fingers in a multi-unit system) is charged 0.3 units. Thus, using a face, an iris, and two fingers will incur a cost of  $1 + 1 + 1 + 0.3 = 3.3$  units.

When eight experts (i.e., face, one iris, and six fingers) are used, the search space has  $2^8 - 1 = 255$  elements. We plot here a “rank-one” *cost-sensitive* performance curve (performance versus cost) where the performance has been computed on the evaluation set. Since the goal is to achieve minimum error with minimum cost, a curve toward the lower left corner is the ideal one. This curve is called a

rank-one curve because it uses the development set to determine the best performing combination-of-experts at each cost value. Similarly, a rank-two cost-sensitive curve would be computed by using the development set to determine the top two performing combination-of-experts at each cost value, and then reporting the best of these two combinations on the evaluation set. With enough rank order, the performance curve will tend to the oracle (the ideal curve with error estimated on the evaluation set). While the rank-one curve using the Bhattacharyya bound as the objective function is satisfactory, the rank-three curve was observed to exhibit exactly the same characteristics as the oracle for the QDA classifier, and the rank-five curve was observed to exhibit exactly the same characteristics as the oracle for the GMM classifier. When the empirical error rate was used as the objective function, a rank-six curve was needed to achieve the performance of the oracle for the QDA classifier while more than rank-ten was needed for achieving the oracle for the GMM classifier.

## 11.7 Open Issues and Challenges

In this section, we will highlight some of the issues in multibiometrics which call for further research.

- **Fusion Architecture:** The range of possible fusion configurations encompassing serial (cascade), parallel, and hybrid modes of operation is very large. While the parallel fusion strategy is most commonly used, there are additional advantages in exploring serial fusion where the experts are considered one at a time. It offers the possibility of making reliable decisions with a few experts, leaving only difficult problems to be handled by the remaining (possibly more expensive) experts. However, automatically deducing these configurations is an open problem and requires more research.
- **Correlated Experts:** An important consideration when adopting a fusion strategy is to model the statistical dependency among the expert outputs. For instance, in a multi-algorithm setting, several experts may rely on the same biometric sample and so higher dependency is expected among the expert outputs. On the other hand, in a multimodal setting, the pool of experts is likely to be statistically independent. Assessing the impact of correlated experts on overall matching accuracy is an interesting problem that has received very little attention in the biometric literature (see [30]).
- **Expert selection:** Expert selection can be cast as a feature selection problem, as illustrated in [49]. However, directly applying such a technique to biometric authentication is difficult for several reasons. In Section 11.6, for instance, we have seen that the optimal set of experts determined using a development population of users may not be the best for the target set of users. Further, the “Doddington Zoo” effect would result in an asymmetrical distribution of errors across the user population [14]. Another issue is raised by cost considerations. Conciliating both the operational cost and matching accuracy into a single criterion is a difficult task.

A related problem, known as dynamic expert selection, arises in the context of serial fusion. In dynamic expert selection, a fusion classifier may decide which expert would be the *most informative* even before the biometric data are acquired from the user. In the recent multimodal biometric benchmark evaluation organized by the Biosecure (EU-funded) project, the use of a dynamic fusion strategy proved to be very promising in achieving good performance while minimizing costs.

Other topics of research in multibiometrics include (a) protecting multibiometric templates; (b) indexing multimodal databases; (c) consolidating biometric sources in highly unconstrained environments; (d) designing dynamic fusion algorithms to address the problem of incomplete input data; (e) predicting the matching performance of a multibiometric system; and (f) continuous monitoring of an individual using multiple traits.

## 11.8 Summary

Multibiometric systems are expected to enhance the recognition accuracy of a personal authentication system by reconciling the evidence presented by multiple sources of information. In this chapter, the different sources of biometric information as well as the type of information that can be consolidated was presented. Typically, early integration strategies (e.g., feature-level) are expected to result in better performance than late integration (e.g., score-level) strategies. However, it is difficult to predict the performance gain due to each of these strategies prior to invoking the fusion methodology. The use of ancillary information, such as soft biometrics and data quality, can further improve the performance of a multibiometric system if an appropriate fusion strategy is used. The three case studies discussed in this chapter highlight the benefits of fusion.

While the *availability* of multiple sources of biometric information (pertaining either to a single trait or to multiple traits) may present a compelling case for fusion, the *correlation* between the sources has to be examined before determining their suitability for fusion. Combining uncorrelated or negatively correlated sources is expected to result in a better improvement in matching performance than combining positively correlated sources [35]. However, defining an appropriate diversity measure to predict fusion performance has been elusive. Thus, there are several open challenges in the multibiometric field that require further research. Nevertheless, it is becoming increasingly apparent that multibiometric systems will have a profound impact on how identity is established in the 21st century.

**Acknowledgments** Norman Poh was supported by the advanced research fellowship PA0022 121477 of the Swiss National Science Foundation and by the EU-funded Mobio project grant IST-214324. Arun Ross was supported by US NSF CAREER grant number IIS 0642554.

## References

1. F. Alonso-Fernandez, J. Fierrez, D. Ramos, and J. Ortega-Garcia. Dealing with sensor interoperability in multi-biometrics: The upm experience at the biosecure multimodal evaluation 2007. In *Proc. of SPIE Defense and Security Symposium, Workshop on Biometric Technology for Human Identification*, 2008.
2. R. Auckenthaler, M. Carey, and H. Lloyd-Thomas. Score normalization for text-independent speaker verification systems. *Digital Signal Processing (DSP) Journal*, 10:42–54, 2000.
3. S. Bengio, C. Marcel, S. Marcel, and J. Marithoz. Confidence measures for multimodal identity verification. *Information Fusion*, 3(4):267–276, 2002.
4. E. S. Bigun, J. Bigun, B. Duc, and S. Fischer. Expert conciliation for multimodal person authentication systems using bayesian statistics. In *First International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 291–300, Crans-Montana, Switzerland, March 1997.
5. J. Bigun, J. Fierrez-Aguilar, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Multimodal biometric authentication using quality signals in mobile communications. In *12th Int'l Conf. on Image Analysis and Processing*, pages 2–13, Mantova, 2003.
6. R. Brunelli and D. Falavigna. Person identification using multiple cues. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10):955–966, October 1995.
7. K. I. Chang, K. W. Bowyer, and P. J. Flynn. An evaluation of multimodal 2D+3D face biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4):619–624, April 2005.
8. X. Chen, P. J. Flynn, and K. W. Bowyer. IR and Visible light face recognition. *Computer Vision and Image Understanding*, 99(3):332–358, September 2005.
9. Y. Chen, S. Dass, and A. Jain. Localized iris image quality using 2-d wavelets. In *Proc. Int'l Conf. on Biometrics (ICB)*, pages 373–381, Hong Kong, 2006.
10. Y. Chen, S.C. Dass, and A.K. Jain. Fingerprint quality indices for predicting authentication performance. In *LNCS 3546, 5th Int'l. Conf. Audio- and Video-Based Biometric Person Authentication (AVBPA 2005)*, pages 160–170, New York, 2005.
11. C. C. Chibelushi, J. S. D. Mason, and F. Deravi. Feature-level data fusion for bimodal person recognition. In *Proceedings of the Sixth International Conference on Image Processing and Its Applications*, 1: 399–403, Dublin, Ireland, July 1997.
12. S. C. Dass, K. Nandakumar, and A. K. Jain. A Principled approach to score level fusion in multimodal biometric systems. In *Proceedings of Fifth International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, pages 1049–1058, Rye Brook, USA, July 2005.
13. J. Daugman. Combining Multiple Biometrics. Available at <http://www.cl.cam.ac.uk/users/jgd1000>, 2000.
14. G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds. Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. In *Int'l Conf. Spoken Language Processing (ICSLP)*, Sydney, 1998.
15. A. Eriksson and P. Wretling. How flexible is the human voice? A case study of mimicry. In *Proceedings of the European Conference on Speech Technology*, pages 1043–1046, Rhodes, 1997.
16. E. Erzin, Y. Yemez, and A. M. Tekalp. Multimodal speaker identification using an adaptive classifier cascade based on modality reliability. *IEEE Transactions on Multimedia*, 7(5):840–852, October 2005.
17. O. Fatukasi, J. Kittler, and N. Poh. Quality controlled multimodal fusion of biometric experts. In *12th Iberoamerican Congress on Pattern Recognition CIARP*, pages 881–890, Via del Mar-Valparaiso, Chile, 2007.
18. J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, and J. Bigun. Kernel-based multimodal biometric verification using quality signals. In *Proc. of SPIE Defense and Security*

- Symposium, Workshop on Biometric Technology for Human Identification*, 5404: 544–554, 2004.
19. H. Fronthaler, K. Kollreider, J. Bigun, J. Fierrez, F. Alonso-Fernandez, J. Ortega-Garcia, and J. Gonzalez-Rodriguez. Fingerprint image-quality estimation and its application to multialgorithm verification. *IEEE Trans. on Information Forensics and Security*, 3:331–338, 2008.
  20. X. Gao, R. Liu, S. Z. Li, and P. Zhang. Standardization of face image sample quality. In *LNCS 4642, Proc. Int'l Conf. Biometrics (ICB'07)*, pages 242–251, Seoul, 2007.
  21. W. R. Harrison. *Suspect Documents, Their Scientific Examination*. Nelson-Hall Publishers, 1981.
  22. H. Hill, P. G. Schyns, and S. Akamatsu. Information and viewpoint dependence in face recognition. *Cognition*, 62(2):201–222, February 1997.
  23. T. K. Ho, J. J. Hull, and S. N. Srihari. Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(1):66–75, January 1994.
  24. Y. S. Huang and C. Y. Suen. Method of combining multiple experts for the recognition of unconstrained handwritten numerals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(1):90–94, January 1995.
  25. A. K. Jain, K. Nandakumar, X. Lu, and U. Park. Integrating faces, fingerprints and soft biometric traits for user recognition. In *Proceedings of ECCV International Workshop on Biometric Authentication (BioAW)*, LNCS 3087: 259–269, Prague, Czech Republic, May 2004. Springer.
  26. A. K. Jain, K. Nandakumar, and A. Ross. Score normalization in multimodal biometric systems. *Pattern Recognition*, 38(12):2270–2285, December 2005.
  27. A. K. Jain and A. Ross. Multibiometric systems. *Communications of the ACM, Special Issue on Multimodal Interfaces*, 47(1):34–40, January 2004.
  28. A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics*, 14(1):4–20, January 2004.
  29. J. Jang, K. R. Park, J. Son, and Y. Lee. Multi-unit Iris Recognition System by Image Check Algorithm. In *Proceedings of International Conference on Biometric Authentication (ICBA)*, pages 450–457, Hong Kong, July 2004.
  30. J. Kittler, N. Poh, O. Fatukasi, K. Messer, K. Kryszczuk, J. Richiardi, and A. Drygajlo. Quality dependent fusion of intramodal and multimodal biometric experts. In *Proc. of SPIE Defense and Security Symposium, Workshop on Biometric Technology for Human Identification*, volume 6539, 2007.
  31. A. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi. Recent advances in visual and infrared face recognition – a review. *Computer Vision and Image Understanding*, 97(1):103–135, January 2005.
  32. K. Kryszczuk and A. Drygajlo. Credence estimation and error prediction in biometric identity verification. *Signal Processing*, 88:916–925, 2008.
  33. K. Kryszczuk, J. Richiardi, P. Prodanov, and A. Drygajlo. Reliability-based decision fusion in multimodal biometric verification systems. *EURASIP Journal of Advances in Signal Processing*, 2007.
  34. L. I. Kuncheva. *Combining Pattern Classifiers – Methods and Algorithms*. Wiley, New York 2004.
  35. L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin. Is Independence Good for Combining Classifiers? In *Proceedings of International Conference on Pattern Recognition (ICPR)*, 2: 168–171, Barcelona, Spain, 2000.
  36. L. Lam and C. Y. Suen. Application of majority voting to pattern recognition: An analysis of its behavior and performance. *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, 27(5):553–568, 1997.
  37. J. Lee, B. Moghaddam, H. Pfister, and R. Machiraju. Finding Optimal Views for 3D Face Shape Modeling. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, pages 31–36, Seoul, Korea, May 2004.

38. W. Li, X. Gao, and T.E. Boult. Predicting biometric system failure. *Proceedings of the IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety*, pages 57–64, 31 2005-April 1 2005.
39. X. Lu, Y. Wang, and A. K. Jain. Combining classifiers for face recognition. In *IEEE International Conference on Multimedia and Expo (ICME)*, 3: 13–16, Baltimore, USA, July 2003.
40. G. L. Marcialis and F. Roli. Fingerprint verification by fusion of optical and capacitive sensors. *Pattern Recognition Letters*, 25(11):1315–1322, August 2004.
41. T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. Impact of artificial gummy fingers on fingerprint systems. In *Optical Security and Counterfeit Deterrence Techniques IV, Proceedings of SPIE*, 4677: 275–289, San Jose, USA, January 2002.
42. D. E. Mauren and J. P. Baker. Fusing multimodal biometrics with quality estimates via a Bayesian belief network. *Pattern Recognition*, 41(3):821–832, 2007.
43. K Messer, J Matas, J Kittler, J Luettin, and G Maitre. Xm2vtsdb: The extended m2vts database. In *Second International Conference on Audio and Video-based Biometric Person Authentication*, 1999.
44. S. Muller and O. Henniger. Evaluating the biometric sample quality of handwritten signatures. In *LNCS 3832, Proc. Int'l Conf. Biometrics (ICB'07)*, pages 407–414, 2007.
45. K. Nandakumar, Y. Chen, S. C. Dass, and A. K. Jain. Likelihood ratio based biometric score fusion. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30:342–347, 2008.
46. National Institute of Standards and Technology. Nist Speech Quality Assurance Package 2.3 Documentation.
47. A. O'Toole, H. Bulthoff, N. Troje, and T. Vetter. Face recognition across large viewpoint changes. In *Proceedings of the International Workshop on Automatic Face- and Gesture-Recognition (IWAFFGR)*, pages 326–331, Zurich, Switzerland, June 1995.
48. Z. Pan, G. Healey, M. Prasad, and B. Tromberg. Face recognition in hyperspectral images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1552–1560, December 2003.
49. N. Poh and S. Bengio. Towards predicting optimal subsets of base-experts in biometric authentication task. In *LNCS 3361, 1st Joint AMI/PASCAL/IM2/M4 Workshop on Multimodal Interaction and Related Machine Learning Algorithms MLMI*, pages 159–172, Martigny, 2004.
50. N. Poh and S. Bengio. Improving fusion with margin-derived confidence in biometric authentication tasks. In *LNCS 3546, 5th Int'l. Conf. Audio- and Video-Based Biometric Person Authentication (AVBPA 2005)*, pages 474–483, New York, 2005.
51. N. Poh, T. Bourlai, and J. Kittler. Improving biometric device interoperability by likelihood ratio-based quality dependent score normalization. In *IEEE Conf. on Biometrics: Theory, Applications and Systems*, pages 1–5, Washington, D.C., 2007.
52. N. Poh, G. Heusch, and J. Kittler. On combination of face authentication experts by a mixture of quality dependent fusion classifiers. In *LNCS 4472, Multiple Classifiers System (MCS)*, pages 344–356, Prague, 2007.
53. N. Poh and J. Kittler. On using error bounds to optimize cost-sensitive multimodal biometric authentication. In *Proc. 19th Int'l Conf. Pattern Recognition (ICPR)*, 2008.
54. S. Prabhakar and A. K. Jain. Decision-level fusion in fingerprint verification. Technical Report MSU-CSE-00-24, Michigan State University, October 2000.
55. T. Putte and J. Keuning. Biometrical fingerprint recognition: Don't get your fingers burned. In *Proceedings of IFIP TC8/WG8.8 Fourth Working Conference on Smart Card Research and Advanced Applications*, pages 289–303, 2000.
56. N. K. Ratha, J. H. Connell, and R. M. Bolle. An analysis of minutiae matching strength. In *Proceedings of Third International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 223–228, Halmstad, Sweden, June 2001.
57. T. P. Riopka and T. E. Boult. Classification enhancement via biometric pattern perturbation. In *Proc. of Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 850–859, 2005.

58. A. Ross and R. Govindarajan. Feature level fusion using hand and face biometrics. In *Proceedings of SPIE Conference on Biometric Technology for Human Identification II*, 5779: 196–204, Orlando, USA, March 2005.
59. A. Ross, A. K. Jain, and J. Reisman. A hybrid fingerprint matcher. *Pattern Recognition*, 36(7):1661–1673, July 2003.
60. A. Ross, K. Nandakumar, and A. K. Jain. *Handbook of Multibiometrics*. Springer, New York, USA, 1st edition, 2006.
61. A. Ross, S. Shah, and J. Shah. Image versus feature mosaicing: A case study in fingerprints. In *Proceedings of SPIE Conference on Biometric Technology for Human Identification III*, pages 620208–1 – 620208–12, Orlando, USA, April 2006.
62. R. K. Rowe and K. A. Nixon. Fingerprint enhancement using a multispectral sensor. In *Proceedings of SPIE Conference on Biometric Technology for Human Identification II*, 5779: 81–93, March 2005.
63. U.R. Sanchez and J. Kittler. Fusion of talking face biometric modalities for personal identity verification. In *IEEE Int'l Conf. Acoustics, Speech, and Signal Processing*, 5: V–V, 2006.
64. C. Sanderson and K. K. Paliwal. Information fusion and person verification using speech and face information. Research Paper IDIAP-RR 02-33, IDIAP, September 2002.
65. R. Singh, M. Vatsa, A. Ross, and A. Noore. Performance enhancement of 2D face recognition via mosaicing. In *Proceedings of the 4th IEEE Workshop on Automatic Identification Advanced Technologies (AutID)*, pages 63–68, Buffalo, USA, October 2005.
66. A. J. Smola and P. J. Bartlett, editors. *Advances in Large Margin Classifiers*. MIT Press, Cambridge, MA, 2000.
67. D. A. Socolinsky, A. Selinger, and J. D. Neuheisel. Face recognition with visible and thermal infrared imagery. *Computer Vision and Image Understanding*, 91(1-2):72–114, July-August 2003.
68. B. Son and Y. Lee. Biometric authentication system using reduced joint feature vector of Iris and Face. In *Proceedings of Fifth International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 513–522, Rye Brook, USA, July 2005.
69. K-A. Toh, W-Y. Yau, E. Lim, L. Chen, and C-H. Ng. Fusion of auxiliary information for multimodal biometric authentication. In *LNCS 3072, Int'l Conf. on Biometric Authentication (ICBA)*, pages 678–685, Hong Kong, 2004.
70. B. Ulery, A. Hicklin, C. Watson, W. Fellner, and P. Hallinan. Studies of biometric fusion. Technical Report NISTIR 7346, NIST, September 2006.
71. U. Uludag, A. Ross, and A. K. Jain. Biometric template selection and update: A case study in fingerprints. *Pattern Recognition*, 37(7):1533–1542, July 2004.
72. P. Verlinde and G. Cholet. Comparing decision fusion paradigms using k-NN based classifiers, decision trees and logistic regression in a multi-modal identity verification application. In *Proceedings of Second International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 188–193, Washington D.C., USA, March 1999.
73. B. Xie, T. Boult, V. Ramesh, and Y. Zhu. Multi-camera face recognition by reliability-based selection. *Proceedings of the IEEE International Conference on Computational Intelligence for Homeland Security and Personal Safety*, pages 18–23, Oct. 2006.
74. L. Xu, A. Krzyzak, and C. Y. Suen. Methods for combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 22(3):418–435, 1992.

# **Chapter 12**

## **Ethics and Policy of Biometrics**

**Emilio Mordini**

**Abstract** This chapter describes the ethical and privacy implications concerning biometric technology. Two emerging issues related to biometrics, function creep and informatization of the body, are discussed. Because function creep results from the over-generation of data, the argument is made that, by design, biometric applications are unlikely to cease the collection and processing of surplus amounts of personal data. Concerning informatization of the body, biometrics can be seen in both a positive and a negative light. When biometric technology is used to give a personal identity to a previously unidentified person, an increased sense of personal empowerment through the attainment of identity is witnessed. However, when biometrics is instead used to offer an identity to individuals solely for the purpose of categorization, we can then consider this to be an unwelcome risk of this technology. Thus, care must be taken in the application of biometric technology.

### **12.1 Introduction**

Scientific literature on ethical and privacy implications of biometrics is becoming increasingly important. A sharp debate is emerging about whether biometric technology offers society any significant advantages over conventional forms of identification and whether it constitutes a threat to privacy and a potential weapon in the hands of authoritarian governments. Main issues at stakes concern large-scale applications, biometric databases, remote and covert biometrics, respect for fair information principles, medical applications, enrolment of vulnerable and disabled groups, information sharing and system interoperability, technology convergence and behavioural biometrics surveillance. It is, however, arguable as to whether it makes sense to discuss all these issues together without differentiating between different biometrics and applications. As a matter of fact, biometrics encompasses so many different technologies and applications that it is hardly thinkable to develop arguments which are valid in all circumstances. Yet there are two interrelated issues

---

E. Mordini (✉)

Centre for Science, Society and Citizenship, Piazza Capo di Ferro 23 00186 Rome, Italy  
e-mail: emilio.mordini@cssc.eu

about biometrics that are worth discussing in general terms. They are “function creep” and the so-called “informatization of the body”.

## 12.2 Function Creep

“Function creep” is the term used to describe the expansion of a process or system, where data collected for one specific purpose are subsequently used for another unintended or unauthorized purpose. Since 2001, function creep has been in the limelight of the ethical and privacy debate [29]. Function creep represents a serious breach of the ethical tenet that requires a moral agent to be honest and responsible for her actions; from a political perspective function creep is a serious lesion of public trust and menaces to destroy confidence in biometric systems. Most problems connected with respect to privacy and data protection in biometric applications are actually rooted in the issue of function creep.

Generally speaking, function creep refers to information that is collected for one purpose but being used for another. Consequently, in the context of biometric identification it should refer to any use of biometric data different from mere identification. Function creep in the field of automated personal recognition may be motivated by several reasons (from state intelligence to commercial purposes) and it is not limited to biometric identification. Most examples of function creep are indeed rather innocuous. The “Social Security Number” in the United States is an often cited example of function creep. Although the original social security cards bore the warning that the SSN could not be used for identification, in the 1960s the Internal Revenue Service started using the SSN for tax payer identification and today the SSN has been the main identity document used by most US citizens. Function creep usually involves three elements: (1) a policy vacuum; (2) an unsatisfied demand for a given function; and (3) a slippery slope effect, or a covert application. A policy vacuum is probably the most important element to determine the risk of function creep. When organizations (be it small companies, large industries, or governmental agencies) adopt new information technologies, or new information schemes, failing to create specific policies, these technologies end up being driven only by the different interests of various stakeholders. As a result, the new scheme may develop in a quite different sense – sometimes even opposite – from that primarily intended. Civil liberties advocates have the tendency to implicitly blame people because they adopt new technologies without considering the downside. Yet it is evident that the main responsibility rests with policy makers who should create the opportune policy frame for innovations.

An unsatisfied demand for a given function is the second important element that has the potential for creating a function creep. Information collected for one purpose is used for another when there is a need that is not properly met. In the example of the SSN, it is evident that the lack of a suitable tax payer identifier was the main driver for the SSN’s mission change.

Finally, function creep must develop almost unnoticed. Usually, function creep happens for one of two reasons: either the new functions develop little by little,

quite innocently, because of the incremental effect of several minor changes of mission/technology misuse, or the new functions are the result of a hidden agenda or of various undisclosed purposes. Warrantless cell phone tracking by law enforcement is a good example of this latter kind of function creep, which is obviously the most ethically and politically worrisome.

Analysing function creep more in depth, and in the context of biometric applications, one should distinguish between two different situations: the first being when biometric identification is used beyond the limits for which the system was officially adopted, the second being when biometric identification is used to generate extra information which is not per se relevant to identification.

As per the first point, it is evident that any identification scheme can be carried out with a hidden agenda (e.g. sorting out some social groups, collecting extra information about people and eliciting the feeling of being under observation) and biometrics does not make an exception. For instance a security agency could decide to carry out a covert screening programme parasitic to a standard identification/verification programme. People enrolled for identification or verification purposes could also be covertly screened against, say, a database of terrorists. In such a case, biometrics would still be used for identification purposes, but these purposes would be partly different from those officially claimed. More correctly, in these cases one should refer to a “subversive use” of a biometric system, say, to an attempt to subvert the correct and intended system policy (ISO SC37 Harmonized Biometric Vocabulary – Standing Document 2 Version 8 – dated 2007-08-22<sup>1</sup>).

There is, however, also the possibility that biometric systems are used to generate details which are not relevant to personal recognition, which is a proper case of function creep. To date, there is no evidence that any biometric application has ever been systematically used to reveal a subject’s personal details beyond those necessary for personal recognition (of course, one could argue that some ID schemes adopted in countries such as Pakistan or Malaysia have been inherently designed in order to produce extra information and to track citizens, but this is a rather different issue because it involves the whole policy of a state rather than specific technology misuse). No one doubts that biometrics can be used to generate extra information, but until now details that could be elicited by biometrics could also be obtained in easier ways. Consequently, there is no point in obtaining such information through the use of biometric applications. In practice, biometrics has the potential for being misused, but the cost/benefit ratio of such a misuse is still discouraging to any Big Brother candidate. Yet, this is likely to change in the near future with second-generation biometrics and large-scale adoption of multimodal systems. One of the possible variants of Murphy’s law states that if any technology can be misused, it will be, and there is no reason to assume that biometrics might escape this rule.

Function creep is facilitated by a surplus of information. The principle of data minimization, or limitation, should then be the cornerstone of any biometric policy, which is respectful of privacy tenets. Unfortunately, this is not the case with most

---

<sup>1</sup> <http://isotc.iso.org/livelink/livelink?func=ll&objId=2299739&objAction=browse>

biometric systems, which are redundant and end up generating more data than necessary. What is worse – and this is my argument – is that it is unlikely that biometric applications may really succeed in minimizing data capture and processing. In the following paragraphs I shall try to explain why.

### 12.3 Biometrics, Body and Identity

Measurable physical and behavioural attributes can be used for studying patterns in individuals and populations, for looking for change or consistency over time, and in different contexts. Decisions about attributes to be measured and the algorithm to be adopted are never neutral, in the sense that they are crucial decisions which are influencing the way in which the whole biometric system works and its degree of “informational intrusiveness”.

Not all biometrics are good for all purposes, yet no single biometrics is suited for one purpose alone. In other words, there are no biometrics that are completely “clean”, as is exemplified by the fact that the same biometrics can be used in a variety of ways, say, for population statistics and genetics research, for medical monitoring, for clustering human communities and ethnic groups and for individual identification. Some attributes are more appropriate for individual recognition (e.g. fingerprint, iris), others are more suited for group categorization (e.g. face geometry, body shape) and still others are more apt to medical monitoring (e.g. thermal images, voice), yet all biometrics may generate data that are not strictly relevant only to personal identification.<sup>2</sup> This data surplus runs the risk of becoming available for “unintended or unauthorized purposes”. If a biometric system generated only those details which are indispensable for making the system work, function creep – at least in the sense of a malicious use of biometric applications – would be hardly feasible. Unfortunately, this is likely to be an impossible mission.

A modern biometric system consists of some basic modules. Each biometric module may generate extra data because of its operational structure. Sensors unavoidably generate data about time and location. They may also collect shadow information, for instance face recognition sensors end up inescapably collecting extra information on a person’s age, gender and ethnicity [16, 17]. Given that facial expressions are topological configurations that can be measured, facial recognition can also become an invaluable source of information on emotional states. Sensors can also elicit details on the medical history of the identifying person. For instance injuries or changes in health can prevent recognition and then be recorded. Although most current technologies have no capability of determining the causes of the recognition failure, no one can exclude that future applications may be designed to identify these causes. It is also indisputable that most biometric techniques may potentially reveal medical information [30]. There can be medical systems that

---

<sup>2</sup> Unless otherwise stated, we use here the term “personal identification” as a generic term, a synonym of recognition, including either positive and negative identifications or authentication.

capture similar images to biometric systems, but they use the information for diagnosis of disease and not identification. Yet it is hard to predict what could arrive if someone starts combining biometric medical data with biometric identifying data in a data fusion centre. Measuring body features to search for consistency over time – as we do for identification/verification purposes – is not radically different from measuring to look for patterns of change as medical doctors do. Biometric images (e.g. face, fingerprint, eye images or voice signals) acquired by the system may show features that can reveal health information. Knowing that certain medical disorders are associated with specific biometric patterns, researchers might actively investigate such questions as whether biometric patterns can be linked to behavioural characteristics, or predispositions to medical conditions. For instance, certain chromosomal disorders – such as Down's syndrome, Turner's syndrome and Klinefelter's syndrome – are known to be associated with characteristic fingerprint patterns in a person. Moreover, by comparing selected biometric data captured during initial enrolment and subsequent entries with the current data, biometric technologies may detect several medical conditions, in particular, when the operator keeps the original images, or in other cases, when some information remains in the template (e.g. if a template stores a compressed version of the image). Finally, most second-generation biometrics (e.g. DNA biometrics, behavioural biometrics and body odour) have the potential for disclosure of medical data. For instance, many are using thermography, which is the use of cameras that are sensitive within the infrared spectrum. Thermography provides highly secure, rapid and non-contact identification that can be achieved with no cooperation from the identifying person. Infrared cameras can easily, and covertly, detect dental reconstruction and plastic surgery (e.g. adding or subtracting skin tissue, adding inter materials, body implants, scar removing, resurfacing the skin via laser and removing tattoos) because the temperature distribution across reconstructed and artificial tissues is different from normal. Also, people's awareness of the presence of biometric sensors is important. Some of the fears surrounding biometric information include that it will be gathered without permission and public knowledge. Warning people that they are entering into a biometrically controlled area does not seem to be a sufficient measure in order to ensure a fair procedure. We all know to what extent it is easy to become acquainted with these kinds of warning labels, with the result being that they are no longer perceived.

Biometrics is measuring attributes of a living body.<sup>3</sup> The aliveness detection module measures a person's physiological signs of life in order to avoid being cheated by artificial attributes. The aliveness detection module is still more critical in being potentially able to generate extra data. The most obvious way to check aliveness is to elicit a physiological response. This, however, generates unintended information on a subject's physiology and on her medical conditions. The spectrum of medical details that can be elicited is particularly vast, as they range from trivial information on body size to more sophisticated data on quite a number of diseases

---

<sup>3</sup> A few biometrics can also be used for identifying corpses, yet this is not a standard practice.

[19]. It is, however, possible with certain biometrics to detect aliveness by using features that hardly disclose medical and physiological information [26]. This point should always be carefully considered in application design.

The quality checker module performs a quality check on raw measurements and indicates whether the characteristic should be sensed again. Also, the quality check module may become responsible for producing extra data. The most important element of a quality metric is its utility, say, the best features collected are those that result in a better identification, which does not always correspond to raw measures with the highest resolution. In other words, the sample resolution should not be higher than necessary to produce the most effective biometrics in order to avoid the risk of including too many details in the sample.

The feature-generator module extracts the set of discriminatory features from the raw measurements and generates a mathematical representation of the biometric features, which is called “live template”. This is also an important passage, though less important than the public tends to believe. Very often, privacy advocates look at templates as though they were the most personal information generated by biometric systems. This is hardly tenable. A good template contains only information necessary to identify or authenticate an individual. Theoretically, this information could also be redundant, that is, revealing more details than necessary, but this is almost never the case because this would defeat the chief purpose of the template, which is the efficient storage of only identifying data (templates designed to include information irrelevant to identification purposes would be false biometric templates, which is quite different from function creep). This is convincingly demonstrated by the impossibility to re-create original characteristics of an individual from her biometric templates [23]. Many “lay” people fear that it is possible to reconstruct the original biological and behavioural characteristic of an individual from a template. This is an urban myth. To understand this, it is important to have clearly in mind the sequence of data generated by a biometric system. The system – or more precisely the sensor(s) – targets any physical property of body parts, physiological and behavioural processes and any combinations of these. Then the sensor produces an analogical or digital representation of the characteristic, which is called “biometric sample”. At this point, the feature extraction process isolates from the sample repeatable and distinctive numbers or labels, which are usually called “biometric features”. A whole set of these features then constitute the template. In the last analysis, both biometric features and templates are just measures of original characteristics. You cannot infer further personal details from them, at least no more than knowing one’s measurement sheet (no doubt that Sherlock Holmes could unravel the whole of one’s life from a tailor-made costume, but this could be applied to any piece of information and cannot be used to argue that templates are highly sensitive data). Indeed template reverse engineering is not possible because most data, which would be necessary to re-create the original attribute, have been discarded and are not present anymore in the template. Templates could be used to re-create artefacts that might be exploited for spoofing the system but this concerns security and only indirectly privacy and ethics (one could argue that security is an ethical issue, but this is evidently a different point). Of course, it is important that compressed

biometric samples are not stored in the system or – what would be even worse – included in the template because, if they were, they could reveal a lot of personal details.

The matcher module compares the live template against one or more templates previously stored. The decision module takes the final decision about identity according to the system threshold for acceptable matching. Extra data can hardly be generated by these modules; their ethical relevance chiefly concerns the ratio between false rejections and false acceptance rates, which is a theme that is out of the scope of this chapter.

In conclusion, biometric systems do generate extra information and consequently are at risk of function creep. What is worse is that the more biometric applications become sophisticated, and exploit the rich set of soft biometrics to increase precision and to avoid being spoofed (i.e. the more biometrics tends to imitate human processes for personal recognition), the more biometric systems become redundant. This is not due to imperfect technologies, or to procedures which still need to be refined, but it depends on the very nature of the human body. Biometrics is based on measuring physical attributes, trying to isolate those measurements that could be used only for recognition purposes. Such metrics simply do not exist because human bodies are full of stories and are stories themselves; they are biographies. In other words, the search for a technology that may collect only identifying details is probably destined to fail.

### ***12.3.1 The Communicational Structure of the Human Body***

Human beings continuously emit signals about themselves. Human bodies are never only bodies, they are also biographies, narratives. Biological and behavioural characteristic of an individual, such as height, weight, hair, skin colour, gender, odours, gestures, posture, prosody and so, send non-verbal messages during interaction. Details communicated include the following:

1. Aliveness details: One is alive and present now and here, this is the first piece of information that one gives to the other by non-verbal communication. The communicational structure of this basic information is demonstrated by the fact that one can pretend to be dead, as frequently occurs with soldiers on the battlefield to escape the enemy.
2. Human details: One also communicates that she is a human being. Also this basic piece of information often passes unnoticed. Yet this is evident when one tries to communicate with other species and is sometimes obliged to mitigate signals about her belonging to the human species (e.g. body odour, posture and skin colour). If ever we happened to live in a “Blade Runner” like world, where biometrics is used to distinguish between humans and androids, the function of non-verbal language to communicate species details could become still more evident.

3. Gender details: Another group of details communicated by using non-verbal languages concerns gender identity. Human beings are a species with a low dimorphism and consequently are able to manipulate and govern, more than other species, this kind of information about themselves. As a matter of fact, our "social" gender is more a result of what we communicate to others than the effect of our genes.
4. Category details: Non-verbal languages also transmit most information on culture, ethnicity, age or social groups to which the individual belongs. Non-verbal languages are probably the most powerful instrument one has to inform others about the various real and virtual communities (networks and category of people) in which she has grown up and lives. The communicational nature of these pieces of information is well illustrated by the fact that we continuously try to control this communication and use it for our purposes.
5. Individual details: Finally, non-verbal languages may also inform others about an individual's personal identity. Scars, wrinkles, body posture, voice prosody, idiosyncratic behaviours, memories, all of these elements reveal something about a particular person, about her biography, her oneness and her specific identity.

Standard personal recognition processes are generated by the interplay between all these communicational levels and an ongoing negotiation between what one wants to communicate to others, what one actually communicates beyond her voluntary control, and what the other(s) – the receiver(s) – are able to understand and interpret at both conscious and unconscious levels. Non-verbal communication is always mixed with identification processes as it is vividly illustrated by infant research that shows how, even moments after birth, the newborn seeks out the mother's eyes and face not only for recognizing her but also as his initial source of information [8] about the world. We are all used to thinking of recognition as a process in which an (active) subject (or devise) recognizes a (passive) individual by searching for some identifiers. This model is hardly tenable. In the real world, inter-human recognition is closer to a conversation rather than an investigation. As each body speaks, we are words made of flesh. She who recognizes someone else listens to signals emitted by the other rather than just detecting some identifiers (identifiers do exist, but they are not passive elements; on the contrary they are often complex symbolic elements, such as tattoos or seals).

Systems for automated recognition of individuals cannot adopt such a sophisticated scheme and they need to decrease variables in order to be able to process data. Most automated systems reach this objective by using electronic labels (e.g. smart tags and RFIDs) which include only those pieces of information required by the system to work. Electronic labels have various downsides: they can get lost, broken, or even sold, and they can be counterfeit. Biometric identification is based on the simple assumption that human beings can be automatically recognized by using a scheme which is rather close to human interaction. This makes biometrics scientifically challenging and practically highly effective, but this is also the reason why biometrics can become troublesome. Biometric systems need to digitalize behaviours and physical appearances in order to process them. In essence, biometric

technologies try to crystallize the human body and to remove from it any biographical dimension which is not relevant to recognition. Ideally biometrics aims to turn persons into mere living beings, biographical life into pure biological existence, which can be measured and matched with other biological objects. This leads to the dramatic distinction between zoe and bios, natural life and political life, which is – according to some thinkers<sup>4</sup> – one of the decisive events of modernity.

At this point, one can also better understand the strong opposition against biometric applications raised by political philosophers such as Giorgio Agamben.<sup>5</sup> Agamben draws our attention to “bare life”, a state of existence which might be defined as life no longer cohering, no longer invested in any form but the very basic component of being alive [1, 2]. This zombie-like condition should not be thought of – as Agamben’s readers usually think and as Agamben himself seems to suggest by referring to Primo Levi’s *The Drowned and the Saved* – only in terms of extreme conditions, as Guantanamo Bay prisoners, but it should be understood as the potential condition of any electronic citizen, more and more unaware of her biographical identity, of her humanity, and increasingly entangled in a virtual web made up by modern information technologies that paradoxically prevent, rather than facilitate, communication.

This leads us to consider the second point, the so-called “informatization of the body”.

### 12.3.2 *Informatization of the Body*

Together with function creep, informatization of the body is the other general issue which concerns biometric policy and ethics. Scholars speak of “informatization of the body” to point out the digitalization of physical and behavioural attributes of a person and their distribution across the global information network [27]. According to a popular aphorism, biometrics is turning the human body into a passport or a password. As usual, aphorisms say more than they intend. Taking the dictum seriously, we would be two: we and our body. Who are we, if we are not our body? And what is our body without us? Briefly at the core of the notion of “informatization of the body” there is a concern for the ways in which digitalization of physical features – through medical imaging, genetics, biometrics and so – may affect the representation of ourselves and may produce processes of “disembodiment”. As privacy advocates and civil liberty organizations are concerned with the risk of function creep, philosophers are often concerned with informatization of the body, because it would touch our inner nature, the “human essence”.

Some scholars [14] see this as a promise of transmigration from our “biological body” to the “cyborg” when individuals may become free to create, or simply to

<sup>4</sup> For example, Foucault [11]; Agamben [1]; Muller [20]; Nelkin and Andrews [21].

<sup>5</sup> See, for instance, “No to Bio-Political Tattooing.” From: La Monde, 10 January 2004. Infoshop-News. <http://www.infoshop.org/inews/stories.php?story=04/01/17/2017978>

remake, themselves and to change their identities as though they were clothing [15]. Other, more critical perspectives [22] have questioned the ways in which information technologies articulate themselves as technologies of immateriality. Baudrillard [4] describes a process of dematerialization, which starts from thing, to commodity, to sign, to mere information. Baudrillard's analysis derives from the famous Marx's notion of "commodity fetishism" and indeed the concept of informatization of the body owes much to early theorization on the fetish. The fetish is an object endowed with a special force, a magical power, inhabited by a spirit [28]. Processes of disembodiment – as those carried out by biometric technologies – would end up turning the body into a fetish inhabited by "us". This has also led to (overly emphatic) questions about whether biometrics risks dehumanizing the body and offends human dignity.<sup>6</sup>

Finally, a number of more pragmatic theories have addressed the ways in which information paradigms pervade biological descriptions. For instance Irma van der Ploeg [28] argues that "the human body is co-defined by, and in co-evolution with, the technologies applied to it. [...] the dominant view of what the body is, what it is made of and how it functions, is determined and defined by the practices, technologies and knowledge production methods applied to it [...] Seen in this light, biometrics appear as a key technology in a contemporary redefinition of the body in terms of information".

## 12.4 Could Biometrics Become a "Liberating" Technology?

The co-evolution between technologies and the body has various reasons but one deserves to be emphasized: all technologies regard the body because their ultimate scope is to enhance our "imperfect" nature, to alleviate the tyranny of human material constitution, its physical limitation, its space–temporal constraints and its limited capacity to perform actions. Technology is in its essence power or, as Mesthene puts it, "technology is nothing if not liberating" [18]. This is also what biometrics does, or, at least, promises to do.

Biometrics promises to free human beings from at least two kinds of limitations.

First, biometrics promises to free humans from centuries of identification systems based on physical marks, tokens, passwords, passes, visas and so. Most probably the need for recognition schemes started at the very beginning of human civilization, with the first urban societies in the Middle East and China, when societies became as complex as to require frequent interactions between people who did know each other.<sup>7</sup> Obviously, most people used to live within the borders of their

---

<sup>6</sup> For example, the French National Consultative Ethics Committee for Health and Life Sciences [10]. "Do the various biometric data that we have just considered constitute authentic human identification? Or do they contribute, on the contrary, to instrumentalizing the body and in a dehumanizing way by reducing a person to an assortment of biometric measurements?"

<sup>7</sup> Yet, it is worth noting that most primitive recognition schemes chiefly had a religious sense (e.g., tattoos, circumcisions, ritual scarifications), as, at that time, the primary point of recognition was

village, or town, and did not need any identifier. Yet, persons that travelled outside of the confines of their home (e.g. military, sailors, traders) needed to be recognized and recognize others in return. A recorded description of physical appearances (e.g. body size and shape, skin and hair colour, face shape, any physical deformity or particularity, wrinkles and scars) was probably the first way of recognizing someone else, and to be recognized. However, as the body gets older, faces change, voices can be altered, scares fade; brief descriptions of physical appearances alone probably became inadequate as human interactions became more and more frequent and complex. The first recognition schemes [6, 7] were probably based on artificial and more permanent body modifications (e.g. branding, tattooing, scarifications)<sup>8</sup> and analogical identifiers. An analogical identifier is a token, a symbol, which could be either a physical object (e.g. a pass, a seal, a ring) or a mental content (e.g. a password, a memory, a poem) which may be linked with only an individual or a category of individuals.

The term “symbol” means “to bring together” and originally the Greek word for “symbol” meant a plank, which was broken, in order for friends to recognize each other by mail. For example, if a messenger comes from a friend to ask for help, he was to bring the second part of the broken plank, and if it was matching the first part, then indeed it was a meeting with a friend. The Roman Empire was the first cosmopolitan society in the West and was also the first example of a universal system for people recognition, which was mainly based on badges and written documents. During the Middle Ages, in Europe – where the majority of the population never went outside the immediate area of their home or villages – individuals were chiefly identified through passes and safe-conducts issued by religious and civil authorities. The genuineness of these documents was witnessed chiefly by seals and handwriting. The birth of large-scale societies and the increased mobility associated with urbanization imposed new recognition schemes. The first passports were issued in France by Luis XIV in 1669 [24] and by the end of the 17th century passports and ID documents had become standard. Yet, only by the end of the 19th century was a true passport system for controlling people’s movement between states universally established. Various ID documents, passes, safe-conducts, seals and other tokens remained the main instrument to ascertain people’s identity in everyday life until World War I [24]. In the 20th century passports and ID cards – incorporating face photography, and, in some cases, also fingerprinting – became

---

first to be recognized by the God(s). Interestingly enough in *Genesis 3:8–10*, after eating from the tree of the knowledge, humans try to escape God’s gaze, to avoid being recognized by him, and the alliance between God and Abraham is sealed by a sign of recognition, the circumcision.

<sup>8</sup> The word “tattoo” is a borrowing of the Samoan word tatau, meaning to “strike something”, but also to “mark someone” [12]. In Pacific cultures, tattooing had great historic significance: the full face tattoo called “moko” was a mark of distinction, which communicated status, lines of descent, and tribal affiliation. In the period of early contact between Maori and Europeans, Maori chiefs sometimes drew their “moko” on documents in place of a signature. In the West, Romans considered tattooing to be barbaric. The Latin word for tattoo was “stigma”, which this tells a lot about the meaning of tattoos in Roman civilization. Romans used to mark criminals and slaves; during the early Roman Empire all slaves exported to Asia were tattooed with the word “tax paid”.

the primary tool for people recognition also within states, at least in those countries that made ID documents mandatory. Finally, in late 1960s, Automatic Identification and Data Capture Technologies (AIDC)<sup>9</sup> emerged as the first true innovation from the birth of photographic passports. It took some time, however, before people understood that biometrics had a very special status among other AIDCs. Biometrics could overcome – or at least have the potential for overcoming – all previous human recognition schemes. Biometrics does not imply any artificial modification of the body as tattoos do. Nor are biometric systems based on analogical representations (biometrics is not icons). A biometric system measures body parts, physiological and behavioural processes. Biometric systems generate digitalized representations of personal characteristics, say, digitalized tokens which link the individual observed here and now with reference data stored in a document, such as a travel document, or in a database. This is the real novelty of biometrics and what makes this technology revolutionary. For the first time in the history of the human species, human beings have really enhanced their capacity for recognizing other people by amplifying – through technical devices – their natural, physiological and recognition scheme, which is based on the appreciation of a complex web of physical and behavioural appearances. Biometric technology aims to solidify this scheme, which would naturally be fluctuating, liquid, unpredictable and even arbitrary.<sup>10</sup>

Second, biometric technologies promise to liberate citizens from the “tyranny” of nation states and create new global, decentralized, rhizomatic schemes for personal recognition. States keep in their hands the power to establish national identities, to fix genders, names, surnames, parental relationships and to assign rights and obligations to individual subjects according to the names written on their identity documents. In his fascinating book on the history of passports [7], John Torpey argues that “modern states, and the international state system of which they are a part, have expropriated from individuals and private entities the legitimate means of movement” (p. 4). Beginning with the French Revolution, there has been both conceptually and historically an indivisible unity of national citizenship and personal identification. On 4 August 1794, 5 years after the revolution, France enacted the first law in the West that fixed identity and citizenship to a birth certificate [3]. The birth certificate is an official document that proves the fact of birth, parentage and family relationship and establishes the place and the date of birth. The original birth certificate is usually stored at a government record office, and one of the main tasks of modern states is to register birth certificates and to secure their authenticity. According to Torpey, nation states have generated “the worldwide development of techniques for uniquely and unambiguously identifying each and every person on

<sup>9</sup> AIDC encompasses a diverse group of technologies (e.g., RFID, matrix bar code, biometrics, smart cards, OCR and magnetic strips) and systems that automate the capture and communication of data. AIDC technologies can be used both to identify items (like bar codes in retail) and to recognize, track and monitor individuals.

<sup>10</sup> J. Lacan, the French psychoanalyst, speaks of the *instant du regard* (the instant of the gaze) as the moment in which recognition and understanding merge.

the face of the globe, from birth to death; the construction of bureaucracies designed to implement this regime of identification and to scrutinize persons and documents in order to verify identities, and the creation of a body of legal norms designed to adjudicate claims by individuals to entry into particular spaces and territories" (p. 7). This state of affairs could now be radically challenged. After small-scale societies and large-scale, industrial societies, globalization is generating the third period of personal identification schemes. Globalization is about the stretching of connections, relations and networks between human communities, their increase in intensity and a general speeding up of all these phenomena. This has important implications for personal recognition as well.

Globalization is characterized by the development of technologies (fibre-optic cables, jet planes, audiovisual transmissions, digital TV, computer networks, the Internet, satellites, credit cards, faxes, electronic point-of-sale terminals, mobile phones, electronic stock exchanges, high-speed trains and virtual reality) which dramatically transcend national control and regulation, and thus also the traditional identification scheme. Moreover, the globalized world is confronted with a huge mass of people with weak or absent identities. Most developing countries have weak and unreliable documents and the poorest in these countries do not have even those unreliable documents. In 2000, UNICEF calculated that 50 million babies (41% of births worldwide) were not registered at birth and thus were without any identification document. Pakistan, Bangladesh and Nepal have not yet made child registration at birth mandatory.<sup>11</sup> In this scenario a personal identity scheme based on citizenship and birth certificates is less and less tenable. The tourist who wants to use the same credit card in any part of the world, the asylum seeker who wants to access social benefits in the host country, the banker who moves in real time huge amounts of money from one stock market to another, they all have the same need. They must prove their identities, and they must be certain of others' identities. But, they can no longer rely on traditional means for proving identities such as birth certificates, passports or ID cards, because these schemes are not dependable enough in most part of the world and unfit for global networks.

By providing global networks with the means to establish trusted identities, the deployment of biometric applications is both a consequence and a building block of the new global wave. Moreover, biometric systems are the only large-scale identification systems that could be run also by small private actors and independent agencies instead of heavy governmental structures. This presents the possibility to create a global system for personal recognition, which would be closer to the Internet than to the Leviathan. Certainly, any process of personal identification implies that individuals are recognized subjects of rights and obligations, and this could be seen as a limitation of individual liberty. Yet, there would be no right, no liberty, without personal identities. One can claim her rights, including the right to be left alone, and the right to refuse to be identified, only if she is an identifiable subject, if she has a public identity. Even if one is identified only for being unjustly arrested,

---

<sup>11</sup> UNICEF, available at [25]

this still means that there are some rules. Personal recognition always implies a certain respect for the law (of course a law can be horrible, but this is a different issue) because it implicitly affirms the principle of personal responsibility. Erasing individual identity and substituting it with a group identity has always been an important instrument for dehumanizing people. People recognition implies two separate concepts, namely that an individual belongs to categories and that she is distinguished by other persons and understood as one. In other words, there are two different aspects involved in people recognition: (1) distinguishing between individuals and (2) distinguishing between sets of people. The latter is likely to be the real issue. Dictatorships of any kind and totalitarian regimes have always ruled by categorizing people and by creating different classes of subjects. There are at least two causes for this: (1) from a psychological point of view, it is easier to induce cruelty against groups that are somehow abstract entities, rather than against single, identified, individuals, and (2) from a social and political point of view this allows for a process known as “pseudospeciation”.

Pseudospeciation is a process which turns social and cultural differences into biological diversities. It promotes cooperation within social groups, overpowering the selfish interests of individuals in favour of collective interests, yet inhibits cooperation between groups, and it fosters conflict and mistrust. Erik Erikson, the great child psychoanalyst known for his studies on the child's identity, was the first to use this term. He lamented that pseudospeciation produces atrocities and brutality. “What is at stake here“, wrote Erikson, “is nothing less than the realization of the fact and the obligation of man's specieshood. Great religious leaders have attempted to break through the resistance against this awareness, but their churches have tended to join rather than shun man's deep-seated conviction that some providence has made his tribe and race or class, caste, or religion ‘naturally’ superior to others. This seems to be part of a psychosocial evolution by which he has developed into pseudo-species . . . for man is not only apt to lose all sense of species, but also to turn on another subgroup with a ferocity generally alien to the ‘social’ animal world.”<sup>12</sup> Raids, concentration camps, mass deportations and executions, which have caused the most horrible manslaughters, are all acts based on pseudospeciation, which requires that people are sorted out according to some shared attributes (e.g. skin colour, cultural or religious belonging, nationality, physical disabilities, social class, location) and that their individual identities are cancelled.

## 12.5 Final Considerations

This leads us to a final consideration, rather than a true conclusion.

General public and privacy advocates are often worried that large-scale biometric systems can turn democratic states into states of police. It is certainly true

---

<sup>12</sup> Erikson [9]

that biometric applications can be misused and I have illustrated why; biometric systems tend unavoidably to collect extra details that are at permanent risk to facilitate or even encourage function creep. Yet, until biometric applications are used for specific individual recognition they are rather safe. Personal recognition per se does not threaten basic liberties or infringe upon the private sphere. Individual recognition has very rarely been used in mass surveillance because – apart from any other consideration – it would be too expensive and ineffective. Of course, one can be legitimately worried by giving too much power to governments, but this is not the most troublesome issue, particularly since biometric technologies are global technologies which promise, in the mid-term, to weaken rather than to strengthen nation state authority. On the contrary, personal recognition may empower people, not only because it offers a ground for entitlements, but also because identity is what establishes and protects fundamental human rights.<sup>13</sup>

Yet biometrics, in particular second-generation biometrics and soft biometrics [5, 13, 16], has also the potentiality to be powerful instrument for people categorization and this is its “dark side”, what could make biometrics ethically and politically worrisome. There is then a certain irony in the fact that enthusiasms and concerns surrounding biometrics are both justified, but they are probably both misplaced.

**Acknowledgments** This work has been funded by a grant from the European Commission – DG Research – Contract 2008- 217762 HIDE (HOMELAND SECURITY, BIOMETRICS AND PERSONAL IDENTIFICATION ETHICS).

## Proposed Questions

- What are the main reasons for ethical concerns on the development and deployment of biometric applications?
- What does “function creep” mean?
- In your opinion, why is function creep ethically meaningful?
- What are the main elements of function creep?
- In your opinion, which biometric is most likely to facilitate function creep?
- Read the paper Mordini E [19] listed in the references and report the main bioethical ethical issues of biometrics in biomedicine.
- Are non-verbal languages important in personal recognition?
- What does “informatization of the body” mean?
- What are the main arguments about the “liberating” function of biometrics?
- What does “categorization” mean?
- How would biometrics identify sets of people rather than individuals?
- Propose a solution to mitigate the risk that biometrics are used for social sorting and profiling.

---

<sup>13</sup> For instance, think of the *habeas corpus*, which would not be possible without certain identities.

## References

1. Agamben G, (1998) *Homo Sacer: Sovereign Power and Bare Life*. trans. Daniel Heller-Roazen, Stanford UP, Palo Alto
2. Agamben G, No to Bio-Political Tattooing. *La Monde*, 10 January 2004. Infoshop News. <http://www.truthout.org/article/le-monde-no-bio-political-tattooing>
3. Arendt H, (1976) 9th ed *The Origin of Totalitarianism*. Harvest Book-Harcourt Inc., San Diego
4. Baudrillard J, (1990) *Fatal Strategies: Revenge of the Crystal*. Power Institute Pub., Sydney
5. Brunelli R, Falavigna D, (1995) Person identification using multiple cues. *IEEE Trans. Pattern Analysis and Machine Intelligence* 17, pp. 955–966
6. Burguière A, Revel J, (eds.) (1989) *L'Espace français*. Editions du Seuil, Paris
7. Caplan J, Torpy J, (eds.) (2001) *Documenting Individual Identity*. Princeton UP, Princeton
8. Cohen L, Gelber ER, (1975) Infant visual memory. In: Cohen L. and Salapatek P. (eds.) *Infant Perception: From Sensation to Cognition*. Vol. 1, pp. 347–403. Academic Press, New York
9. Erikson ER, (1964) *Insight and responsibility*. Norton, New York, p. 66
10. French National Consultative Ethics Committee for Health and Life Sciences (2007) OPINION N° 98 Biometrics, identifying data and human rights. <http://www.ccne-ethique.fr/docs/Avis104Final.%20ang.pdf>
11. Foucault M, (1977) *Discipline and Punish: The Birth of the Prison*. Vintage, Essex
12. Gilbert SG, (2001) *Tattoo History. A Source Book*. Juno Books, Rockville
13. Givens G, Beveridge JR, Draper BA, Bolme D, (2003) A Statistical Assessment of Subject Factors in the PCA Recognition of Human Subjects. In: Proceedings of CVPR Workshop: Statistical Analysis in Computer Vision.
14. Haraway DJ, (ed) (1991) *Simians, Cyborgs and Women: The Reinvention of Nature*. Routledge, New York
15. Hardt M, Negri A, (2000) *Empire*. Harvard UP, Boston
16. Jain AK, Bolle R, Pankanti S, (eds.) (1999) *Biometrics: Personal Identification in Networked Society*. Kluwer, Dordrecht
17. Jain AK, Lu X, (2004) Ethnicity Identification from Face Images. In: Proceedings of SPIE International Symposium on Defense and Security: Biometric Technology for Human Identification (To appear) <http://citeseer.ist.psu.edu/jain04ethnicity.html>
18. Mesthene EG, (1970) *Technological Change: its Impact on Man and Society*. Harvard Univ Press, Boston p.20
19. Mordini E, (2008) Biometrics, Human Body and Medicine: A Controversial History. In: Duquenoy P, George C, Kimppa K, (eds.) *Ethical, Legal and Social Issues in Medical Informatics*. Idea Group Inc, Toronto
20. Muller B J, (2004,) (Dis)Qualified Bodies: Securitization, Citizenship and “Identity Management”. In: *Citizenship Studies* 8(3): 279–294
21. Nelkin D, Andrews L, (2003) Surveillance Creep in the Genetic Age. In: Lyon D (ed.) *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination*. Routledge, New York, pp. 94–110
22. Pietz W, (1985) The Problem of the Fetish. In: *Res: Anthropology and Aesthetics* 9 (Spring, 1985): 5–17
23. Statham P, (2006) Issues and Concerns in Biometrics IT Security. In: Bigdoli H, (ed.) *Handbook of Information Security*. John Wiley and Sons, New York, pp. 471–501
24. Torpey J, (2000) The invention of the Passport- Surveillance, Citizenship and the State. Cambridge UP, Cambridge
25. UNICEF, [http://www.unicef.org/protection/files/Birth\\_Registration.pdf](http://www.unicef.org/protection/files/Birth_Registration.pdf)
26. Valencia V, (2002) Biometric Liveness Testing. In: Woodward JD Jr., Orlans NM, Higgins PT, (eds.) *Biometrics*. Osborne McGraw Hill, New York
27. van der Ploeg I, (2005) *The Machine-Readable Body. Essays on Biometrics and the Informationization of the Body*. Shaker, Germany

28. van der Ploeg I, (2008) Machine-Readable Bodies, Biometrics, Informatization and, Surveillance. In: Mordini E, Green M, (eds.) Identity, Security and Democracy. In: NATO Science for Peace and Security Series: Human and Societal Dynamics. IOS Press, Amsterdam
29. Woodward JD Jr, Watkins Webb K, Newton EM, et al. (2001) Army Biometric Applications, Identifying and Addressing Sociocultural Concerns. Document Number: MR-1237-A. [http://www.rand.org/pubs/monograph\\_reports/MR1237/](http://www.rand.org/pubs/monograph_reports/MR1237/)
30. Zhang D (ed.), (2008) Medical Biometrics. Springer, New York

**Part II**

**Selected Contributions from Students**

**of the International Summer**

**School on Biometrics**

# **Chapter 13**

## **Assessment of a Footstep Biometric Verification System**

**Rubén Vera Rodríguez, John S.D. Mason, and Nicholas W.D. Evans**

**Abstract** This chapter reports some novel experiments which assess the potential of footsteps as a biometric. We present a semi-automatic capture system and report results on a large database of footprint signals with independent development and evaluation data sets comprised of more than 3000 footsteps collected from 41 persons. An optimisation of geometric and holistic feature extraction approaches is reported. Following best practice we report some of the most statistically meaningful and best verification scores ever reported on footprint recognition. An equal error rate of 10% is obtained with holistic features classified with a support vector machine. As an added benefit of the work, the footprint database is freely available to the research community. Currently, the research focus is on features extraction on a new high spatial density footsteps database.

### **13.1 Introduction**

Different biometrics have been used for many years to verify the identity of persons. Some of the most researched such as fingerprints or faces have been included in passports and ID cards. Iris recognition has been introduced in airports and palm vein recognition in cash machines. These methods belong to the group of the physiological biometrics as they do not exhibit a large variance over time. On the other hand, behavioural biometrics are more likely to change throughout different recording sessions. Voice recognition is the most popular of these biometrics because of its application in telephony used worldwide.

Gait and footsteps are also considered to be behavioural biometrics. Gait recognition has been investigated over the past decades for medical applications as well as for the sport shoe industry. Gait recognition is based on the study of the way persons walk by camera recordings, whereas footsteps recognition is based on the study of signals captured from persons walking over a sensing area. Both techniques are very related and could be easily fused in the same environment.

---

R.V. Rodríguez (✉)  
Swansea University, Singleton Park, Swansea, SA2 8PP, UK,  
e-mail: r.vera-rodriguez.405831@swansea.ac.uk

Footstep recognition was proposed as a new biometric 10 years ago, but it has been studied only by a small number of researchers. The main benefit of footsteps over the more well-known biometrics is the fact that footstep signals can be collected covertly, and therefore, the sensing system is less likely to induce behavioural changes as well as presenting less of an inconvenience to the user. Also, footsteps are not as susceptible to environmental noise as in the case of speaker recognition or lighting variability in the case of face recognition. As we review in Section 13.2, different techniques have been developed using different sensors, features and classifiers. Results achieved are promising and give an idea of the potential of footsteps as a biometric; however, these results are related to small databases in a number of persons and footsteps and this is a limitation of the work to date.

In this chapter we present results achieved using a database comprised of more than 3000 footsteps from 41 persons. As described in Section 13.3, this database has been further divided into independent development and evaluation data sets adopting a standard, best practice evaluation strategy allowing us to present more statistically meaningful results and potentially more reliable predictions of performance. In addition, we describe the development of a semi-automatic footstep capture system used to gather the database, which is publicly available to the research community [1].

Preliminary work with geometric and holistic feature extraction methods was presented in [2]. Extending this previously published work, this chapter presents an optimisation of the two feature approaches. A discriminative-based classifier in the form of a support vector machine (SVM) is used to obtain an equal error rate (EER) of 9.5% for development set and 13.5% for evaluation set for the holistic feature approach as described in Section 13.4. Section 13.5 describes the focus of our current work and presents a new footstep capture system with a high density of sensors. Finally our conclusions are presented in Section 13.6.

## 13.2 Review of Footsteps as a Biometric

Footstep recognition is a relatively new biometric certainly judged in terms of published work. Table 13.1 summarises the material in the open literature.

One of the first investigations into footstep recognition was reported by UK researchers in 1997 [3] (first row in Table 13.1). They reported experiments on a database of 300 footstep signals that were captured from 15 walkers from load cells measuring the ground reaction force (GRF). An identification accuracy of 91% was achieved with an HMM classifier and samples from the GRF as features.

In 2000, using a similar sensor approach a group in the United States [4] reported results on a database of 1680 footstep signals collected from 15 persons. Signals were collected from both left and right feet and different footwear. Ten features were extracted from the GRF signal: the mean value, the standard deviation, maxima and minima, etc. An identification accuracy of 93% was reported using a nearest neighbour classifier.

Whilst focused towards the study of gait, in 2002 a group from Switzerland [5] developed a system fusing data acquired from three tiles of four piezo force

**Table 13.1** A comparison of different approaches to footstep recognition 1997-2007

Group / year	Database (steps / persons)	Technology	Features	Classifier	Results
The ORL Active Floor/1997 [3]	300/15	Load cells	Subsampled GRF	HMM	ID rate: 91%
The Smart Floor (USA)/2000 [4]	1680/15	Load cells	Geometric feature from GRF	NN	ID rate: 93%
ETH Zurich/2002 [5]	480/16	Piezo force sensors	Power spectral density	Euclidean distance	Verif. EER: 9.4%
Ubifloor (Korea)/2003 [6]	500/10	Switch sensors	Position of several steps	MLP neural network	ID rate: 92%
EMFi Floor (Finland)/2004 [7]	440/11	Electro mechanical film	Geometric feature from GRF and FFT	MLP neural network	ID rate: 79%; and 92% combining three consecutive steps
Southampton University (UK)/2005 [8]	180/15	Resistive (switch) sensors	Stride length, stride cadence and heel-to-toe ratio	Euclidean distance	ID rate: 80%
Southampton University (UK)/2006 [9]	400/11	Load cells	Geometric feature from GRF	NN	ID rate: 94%
Swansea University (UK)/2007 [2]	3174/41	Piezoelectric sensors	Geometric and holistic features	SVM	Verif. EER: 9.5% for Development 11.5% for Evaluation

sensors each and video cameras. A database of 480 footsteps was collected from 16 persons. They extracted different geometric features from GRF as reported in [4] and the phase plane. The best verification performance was achieved using the power spectral density of the footstep signals, with an Euclidean distance classifier obtaining an EER of 9.4%.

A Korean group reported a system in 2003 [6] that used 144 simple ON/OFF switch sensors. Stride data (connected footsteps) were collected from 10 persons, each contributing 50 footsteps resulting in a database of 500 signals. An accuracy of 92% was reported with a multilayer-perceptron neural network used as an experimental identification method.

In 2005 a group from Finland investigated footstep recognition using electro mechanical film (EMFi) [10]. Long strips of the sensor material were laid over an area covering 100 m<sup>2</sup>. A database of 440 footstep signals was collected from 11 persons. They presented experiments [7] combining different feature sets using a two-level classifier. On the first level three different feature sets were extracted from a single footstep as geometric features from the GRF, similar to [10], FFT of GRF

with PCA and FFT of the derivate GRF with PCA. Then, a product rule was used to combine the three results obtained. On the second level different footsteps from the same person were combined using an average strategy. These experiments were done for two classifiers: LVQ and an MLP neural network. Results were better for the MLP classifier in all cases, having a recognition rate of 79% for the case of a single footprint and a 92% for three consecutive footsteps.

In 2005 a group from Southampton (UK) [8] reported trials with a system comprising 1536 sensors each covering an area of  $3\text{ cm}^2$ . A database of 180 signals was collected from 15 people without wearing footwear. Three features were extracted: stride length, stride cadence and heel-to-toe ratio. An identification accuracy of 80% was reported using a Euclidean distance classifier.

In 2006 another group from Southampton [9] investigated a system similar to the work in [3, 4]. A database of 400 signals was collected from 11 people. Using geometric features extracted from GRF profiles as in [4], an identification accuracy of 94% was achieved using a nearest neighbour classifier.

More recently, in 2007, our research group presented [2] experiments obtained with a database comprised of 3174 footsteps of 41 persons and divided into development and evaluation sets. Geometric and holistic features were extracted from the footprint signals, and NN and SVM classifiers were compared. Equal error rates of 9.5% and 11.5% were obtained respectively for the development and evaluation sets using holistic features and an SVM classifier.

Table 13.1 summarises the material in the open literature. The second column shows that relatively small database sizes are a common characteristic of the earlier work certainly judged in relation to other biometric evaluations where persons are normally counted in hundreds or thousands and the number of tests perhaps in many thousands. A maximum number of 16 persons and 1680 footprint examples were gathered in all cases except in [2] which reports results on 3147 footsteps and 41 persons. In each case the databases are divided into training and testing sets; however, with the exception of [2], none use independent development and evaluation sets, a limitation which makes performance predictions both difficult and unreliable. Identification, rather than verification, was the task considered in all but three of the cases, the exceptions being [2, 5]. Identification has the benefit of utilising the available data to a maximum but suffers from well-known scalability problems in terms of the number of classes in the set.

### 13.3 Data Capture System and Database

The footprint data capture system has been designed to facilitate the capture of many thousands of footprint signals over a relatively short time period. Two piezoelectric transducers inserted into the underside of a rubber floor tile are used to capture footprint signals. They provide a differential voltage output according to pressure upon the floor tile and are digitised using a sample rate of 1024 Hz. The signals are then processed with an in situ micro-controller and stored on a desktop computer via a serial connection. To maximise data capture and to reduce the variance in walking

direction the instrumented floor tile is positioned in the doorway entrance of our research laboratory.

Due to the number of footsteps that are to be captured the provision for automatic labelling and rapid manual validation is deemed essential. A microphone situated a few steps ahead of the sensing area captures a four-digit spoken ID, if provided, whilst ensuring no disturbance in the natural walking process. The audio tokens facilitate automatic labelling with speaker recognition. Two video cameras capture images of the face and foot which can later be used for manual validation and to record metadata, i.e. to label different footwear, etc. Footstep data may be accessed by walker, date / time and other parametric details. Web-based administration allows viewing of footprint data in a graphical form and previews of video feeds ensuring a high confidence in the correct labelling of the data.

The work described here relates to a database comprised of 3174 footsteps collected from 41 persons who were each instructed to place their right foot over the centre of the instrumented floor tile. Two subsets have been identified: a client set of 17 persons with an average of 170 footsteps per person (2884 total footsteps) and an impostor set of 24 persons with an average of 15 footsteps per person (290 total footsteps). Each person in the client set provided footsteps with at least two different shoes.

The database has been further divided into independent development and evaluation data sets, and each of them is comprised of training and testing data sets. This is accomplished with random selection. The development set was used to set the different parameters and features of the recognition system, and two evaluation sets were used to test the established system with new unseen data.

Table 13.2 illustrates the distribution of the footsteps data into the different data sets. It is worth noting that there is no data overlap between the development set and the two evaluation sets. The development set is comprised of footsteps from clients P1 to P8, each contributing 40 footsteps for training and another 40 footsteps for testing. Evaluation set 1 is a balanced set comprised of footsteps from clients P1 to P17 where, for each client, there are 40 footsteps for training and another 40 for testing. Evaluation set 2 uses all the footsteps available in the database and is thus an unbalanced set in terms of the number of footsteps per person. It is comprised of footsteps from clients P1 to P17 with 45 footsteps per client for training and an average of 87 footsteps per client for testing, the range being 40–170 footsteps per client. Thus evaluation set 1 is a subset of evaluation set 2.

**Table 13.2** Distribution of footsteps in the data sets

	Development set		Evaluation set 1		Evaluation set 2	
	Training	Test	Training	Test	Training	Test
Clients	P1–P8	P1–P8	P1–P17	P1–P17	P1–P17	P1–P17
Footsteps per client	40	40	40	40	45	87
Impostors	P18–P41	–	P18–P41	–	P18–P41	–
Impostor footsteps	290	–	290	–	290	–
Subset data	610	320	970	680	1055	1479
Total set data	930		1650		2534	

As a part of the recognition system, the impostor footsteps are the same for all three data sets and come from persons P18 to P41 with a total number of 290 footsteps.

## 13.4 Experimental Work

In this section we present experiments carried out with the footstep database described. As an assessment protocol of the footstep recognition evaluation, index files were created to provide a list of the footstep signals to use in each one of the development and evaluation data sets following the structure utilised by the international NIST SRE [11].

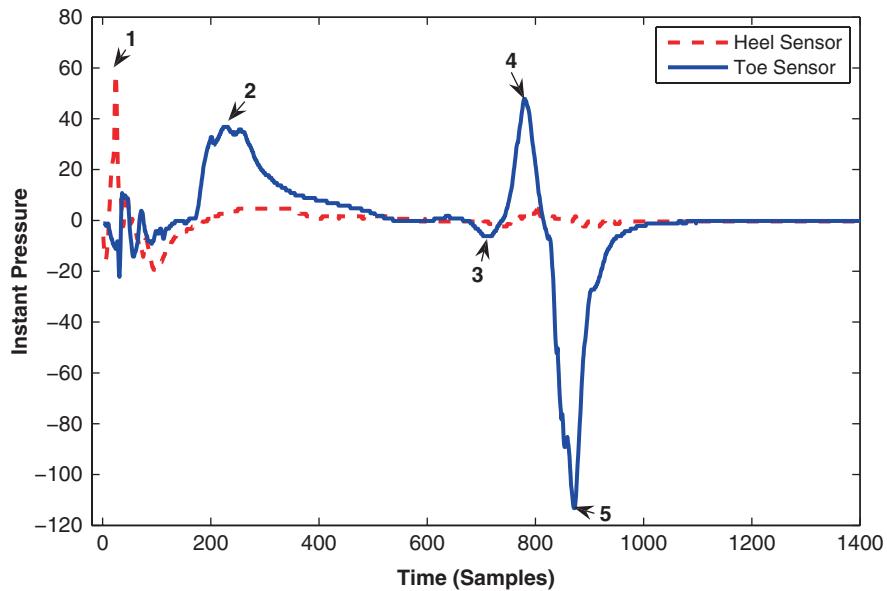
First we describe an optimisation of the geometric and holistic feature approaches followed and second the results of the footstep recognition evaluation. As regards the classification technique, a support vector machine (SVM) [12, 13] was used in all cases. A comparison between a nearest neighbour and an SVM classifiers is reported in [2] showing better performance for an SVM classifier as could be expected. The SVM is a statistical discriminative-based classifier that finds an optimal hyperplane which maximises the margin between in-class and out-of-class data. Different Kernel functions were tested having a better performance with a radial basis function (RBF) case used in all the experiments described above. Finally, results are presented with detection error trade-off (DET) [14, 15] curves as is popular with many biometric studies.

### 13.4.1 Feature Optimisation

In this section we present an optimisation of the features extracted from the footstep signals in order to improve performance with the SVM classifier. As described in [2], two different feature approaches, geometric and holistic, have been followed.

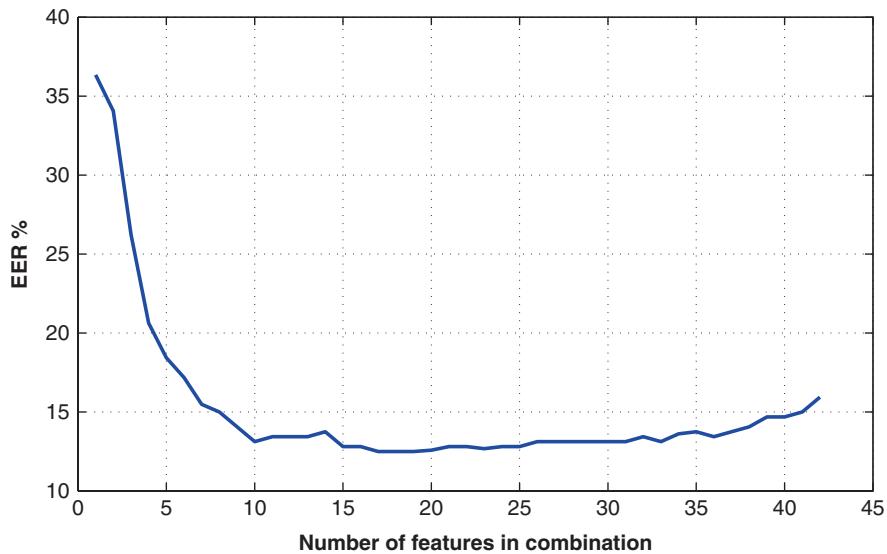
#### 13.4.1.1 Geometric Features

The signals that our system produces relate to the instantaneous pressure for each sensor along the footstep. Figure 13.1 shows a typical footstep waveform. The relevant points, shown by the numbers in Fig. 13.1, were chosen as an indication of the behaviour of the signals along time, similar to the work of [3, 4, 9]. These points coincide with some of the relative and absolute maxima and minima present in the footstep signals. Point 1 corresponds to the effect of heel pressure on the first sensor, the dashed profile in Fig. 13.2. Points 2–5 correspond to the second sensor, the solid profile in Fig. 13.1, and show the effect of the toe. Point 2 shows the initial pressure of the toe, point 5 shows the effect of the pushing off of the toe and points 3 and 4 mark the transition between points 2 and 5. The time and magnitude of these five points result in the first 10 features. Then, the inter-difference between each pair of points results in another 20 features (10 magnitude features and 10 time features). Finally, 12 additional features, the area, norm, mean, length and standard



**Fig. 13.1** Instant pressure against time. Relevant points for geometric feature extraction are indicated

deviation of both sensors and a relation for magnitude and time for the toe sensor, are concatenated to obtain a feature vector with a total of 42 geometric features for each footstep signal. These features were normalised with respect to the absolute maxima of the profile.

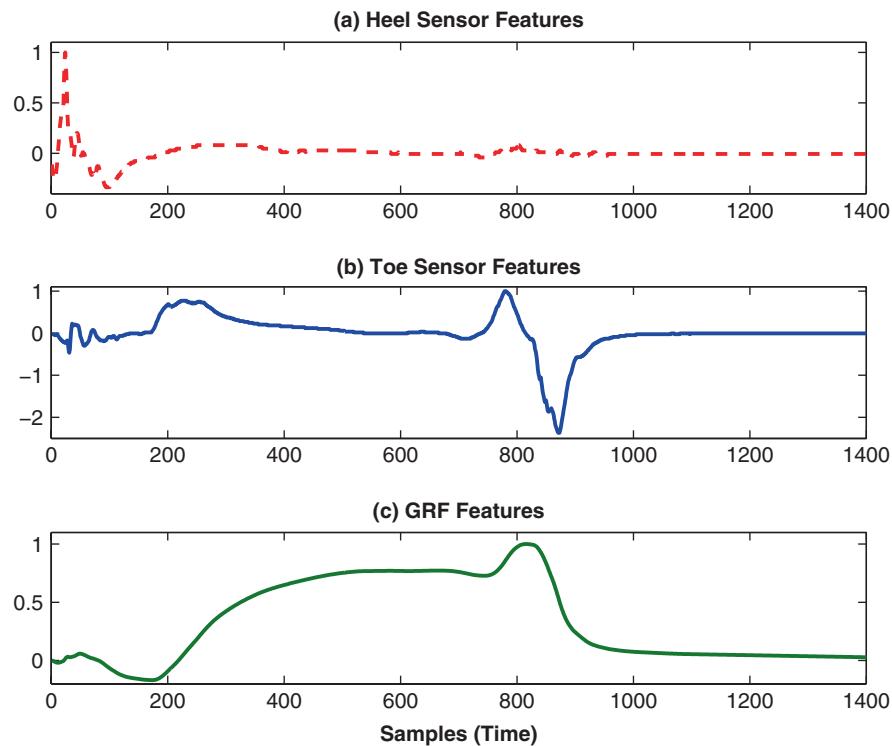


**Fig. 13.2** EER against number of features in combination for geometric features

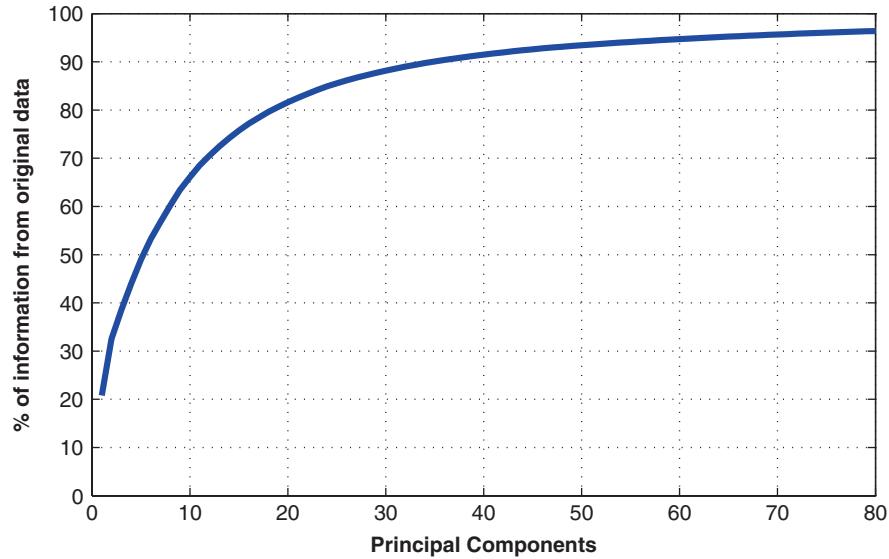
The optimisation of the geometric features was computed by an exhaustive search in order to find a combination of features which produces the minimum EER using the development set. Experiments were conducted using each one of the 42 geometric features separately to obtain a ranking in terms of performance. The feature with the minimum EER was identified and then a second set of experiments was conducted using the best feature together with each one of the remaining features to obtain another rank. This procedure was repeated until all 42 feature were used. Figure 13.2 shows the EER against the optimum combination of the features. As it is observed the set of the first 17 features produces an EER of 12.5% compared to the EER of 16% of the total combination of features. This equates to a relative improvement of 22% in terms of EER. This optimum combination of features is comprised of five features related to time, six related to magnitude and also the norm, area and deviation for both sensors.

#### 13.4.1.2 Holistic Features

Holistic features are comprised of the first 1400 samples (1.37 s) of the heel and toe sensors (as the example of Fig. 13.3(a) and (b)), and also the first 1400 samples of the GRF (as in Fig. 13.3(c)), calculated as the integration over time for these two



**Fig. 13.3** Holistic features used. (a) Heel sensor features. (b) Toe sensor features. (c) GRF features



**Fig. 13.4** Percentage of information from original data against number of principal components

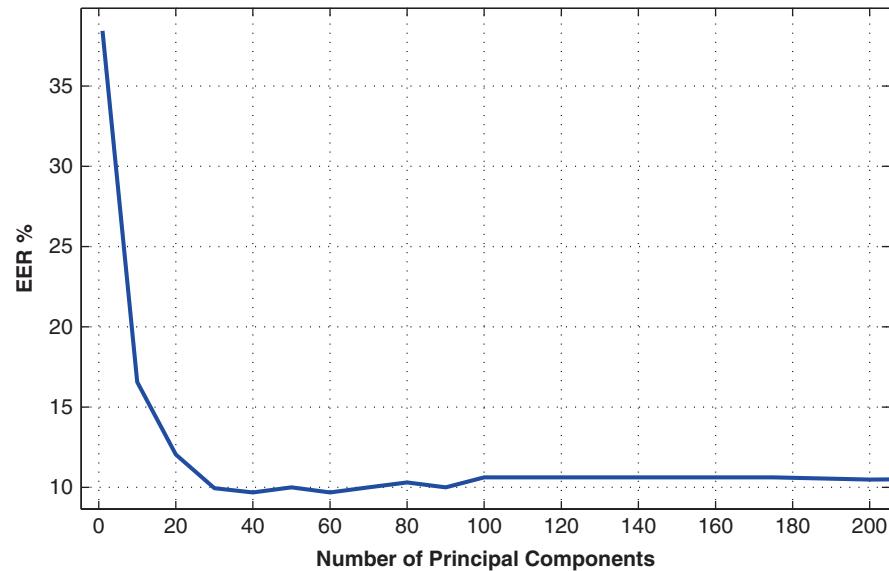
sensors. In total 4200 holistic features have been obtained after normalisation of each sensor and the GRF by its maxima.

Due to the high dimensionality of this holistic feature vector, principal component analysis (PCA) [16] was used to distil the information content. Thus, after PCA, a set of principal components are obtained, where each is a linear combination of the original feature set. Figure 13.4 shows the information contained in the principal components of the training data of development set. It is observed how using the first 80 principal components, more than 96% of the original information is retained whilst achieving a 98% reduction in dimensionality.

The purpose of an optimisation of the holistic features is to find the number of components of PCA with a minimum EER for the development set. For this experiment, the variation in EER is measured on the EER when adding more principal components to the SVM classifier. Figure 13.5 shows the EER against the variation in the number of principal components chosen as features to the SVM classifier. It is observed that a best EER of 9.5% is achieved when the first 60 principal components are used.

### 13.4.2 Footstep Recognition Evaluation

In this section we present the results of the footstep recognition evaluation. Figure 13.6 shows the DET curves result for the development, evaluation 1 and evaluation 2 data sets for the case of geometric and holistic features. It is observed

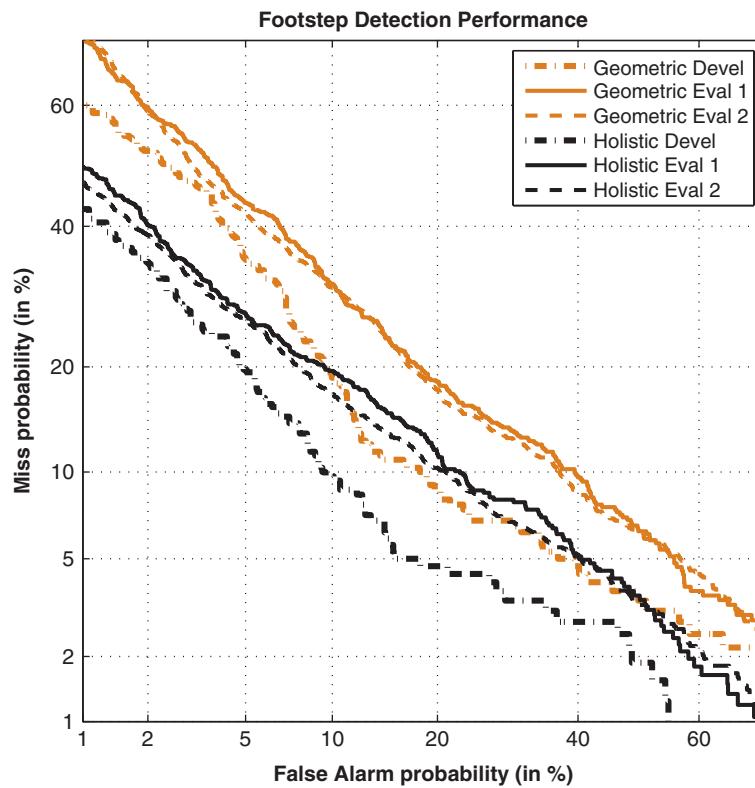


**Fig. 13.5** EER against number of principal components for holistic approach

that holistic features outperform geometric features in all cases. For the development set, EERs of 12.5% and 9.5% were achieved for the geometric and holistic features, respectively, as stated above. These are the best results as could be expected as the optimisation of the features has been carried out for the development set.

The purpose of the evaluation set is to test the footstep recognition system with new unseen data. For this experiment, all the parameters learnt from the development set like the PCA, scaling and normalising coefficients are applied to the evaluation sets. For evaluation set 1, an EER of 19% is achieved for geometric features. This contrasts with an EER of 15% obtained when the holistic features are applied to the same classifier, having a relative improvement of 21%. For evaluation set 2 the same trend is observed. An EER of 18.5% is achieved for geometric features compared to an EER of 13.5% for holistic, what equates to a relative improvement of 27%. It is observed that the DET curves for evaluation set 2 have a better performance than for evaluation 1 in general; this is because as it was stated in [2] more data were used to train the models as illustrated in Table 13.2.

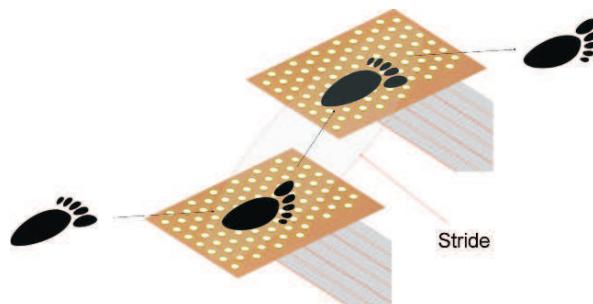
It is worth noting that results achieved with the optimisation of the geometric features are better than the results obtained in [2]. Only results for the evaluation sets with holistic features are marginally worse. This is due to the fact that previously published work used both training and testing data to evaluate the PCA. This makes the new experiments more realistic and statistically meaningful.



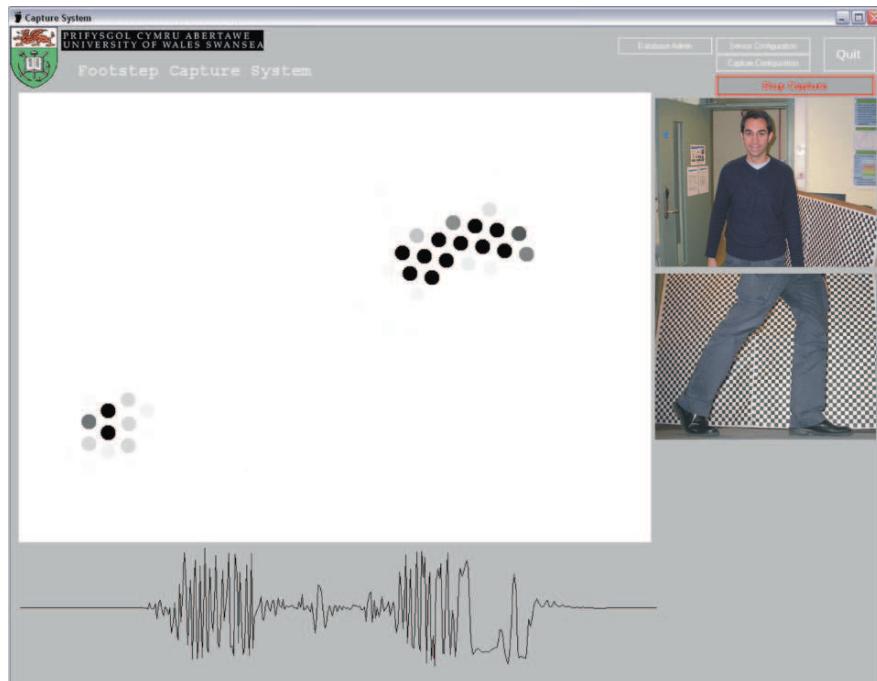
**Fig. 13.6** DET curves for geometric and holistic features for development set and evaluation sets 1 and 2

### 13.5 Current Work

Currently we are in the process of collecting a new database. We have developed a new footstep capture system which is comprised of two sensor mats each containing 88 piezoelectric sensors. The system captures two consecutive footstep signals, as illustrated in Fig. 13.7, and uses a sampling frequency of 1.6 kHz.



**Fig. 13.7** Spatial distribution of the piezoelectric sensors

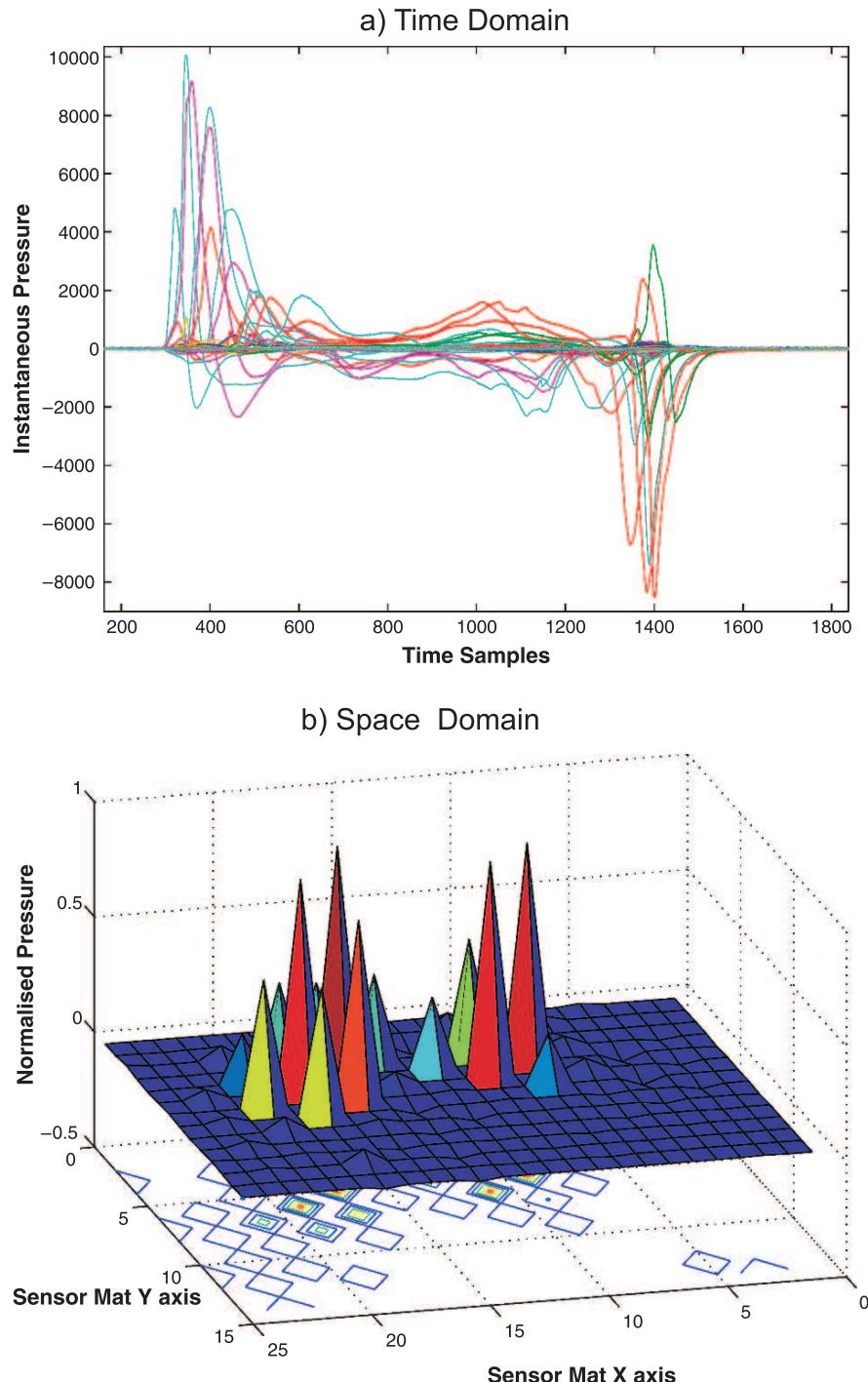


**Fig. 13.8** Screenshot of the footstep capture system software

Figure 13.8 shows a screenshot of the footstep capture system user interface. A distribution of the sensors activated by an example footstep stride is illustrated in the middle of the figure. Below that is the microphone output corresponding to the four-digit ID which is post-processed by the automatic speech recognition system for labelling purposes. The images to the right show frames from the videos that are captured during the footstep data collection. The top image shows the face and the bottom image shows the gait.

For the moment we have collected footstep data from more than 100 people. Data are collected in different sessions and with different conditions, namely with different footwear, including a barefoot condition, when the person carries a load (to examine weight variability effects), and also when they walk at different speeds. This will allow us to study, for the first time, how these conditions affect the performance of person verification using their footsteps.

The high sensor density of the new footstep capture system allows us to extract more information from the footstep signals compared to our previous capture system. Figure 13.9(a) shows a typical footstep signal in the time domain. In the order of 15 sensors are active for each footstep, thus giving a much more detailed account of the footstep dynamics. It is possible to create a 3D image of the footstep signals using the signals captured by the new system and an example is illustrated in Fig. 13.9(b). This gives us a more readily interpretable illustration of the footstep dynamics and illustrates how persons distribute their weight on the floor.



**Fig. 13.9** New possibilities on feature extraction. (a) Time domain profiles. (b) A 3D representation of the footstep signal in the space domain

A significant feature of the new system is the ability to capture two consecutive footsteps, i.e. stride data. The stride data allow the study of the differences between the right and left footsteps, as well as velocity, and angle between the feet, i.e. new features which have the potential to improve the discrimination of persons using their footsteps. These aspects are currently under investigation.

### 13.6 Conclusions

This chapter describes a semi-automatic system for capturing footsteps. A database comprised of more than 3000 footsteps has been gathered, allowing us to present more statistically meaningful results and potentially more reliable predictions of performance compared to related work. Also, this database is publicly available to the research community.

Experimental work has been conducted following best practice using independent development and evaluation sets. In addition, we report an optimisation of the two feature extraction approaches. Interestingly, holistic features show better performance with a relative improvement of around 27% in terms of EER compared to geometric features.

Finally a new footstep capture system has been presented. The new system has a high density of piezoelectric sensors. This facilitates the study of new footstep and stride-related features. The new database will also allow us to better investigate session variability and to study how different factors such as shoes, carried loads and speed affect recognition performance.

**Acknowledgments** The authors gratefully acknowledge the significant contributions of Richard P. Lewis on the development of the database capture system, central to this work. Also, we would like to acknowledge the support of the UK Engineering and Physical Science Research Council (EPSRC) grant and the European Social Funding (ESF).

### References

1. S.U.: Footstep recognition at Swansea University available at. <http://eeswan.swan.ac.uk>
2. Vera-Rodriguez, R., Evans, N.W.D., Lewis, R.P., Fauve, B., Mason, J.S.D.: An experimental study on the feasibility of footsteps as a biometric. In: Proceedings of 15th European Signal Processing Conference (EUSIPCO'07), Poznan, Poland (2007) 748–752
3. Addlesee, M.D., Jones, A., Livesey, F., Samaria, F.: The ORL active floor. IEEE Personal Communications. **4** (1997) 235–241
4. Orr, R.J., Abowd, G.D.: The smart floor: A mechanism for natural user identification and tracking. In: Proceedings of Conference on Human Factors in Computing Systems. (2000)
5. Cattin, C.: Biometric Authentication System Using Human Gait. Swiss Federal Institute of Technology, Zurich. PhD Thesis. (2002)
6. Yun, J.S., Lee, S.H., Woo, W.T., Ryu, J.H.: The user identification system using walking pattern over the ubiFloor. In: Proceedings of International Conference on Control, Automation, and Systems. (2003) 1046–1050

7. Suutala, J., Roning, J.: Combining classifiers with different footstep feature sets and multiple samples for person identification. In: Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP). **5** (2005) 357–360
8. Middleton, L., Buss, A.A., Bazin, A.I., Nixon, M.S.: A floor sensor system for gait recognition. In: Proceedings of Fourth IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05). (2005) 171–176
9. Gao, Y., Brennan, M.J., Mace, B.R., Muggleton, J.M.: Person recognition by measuring the ground reaction force due to a footstep. In: Proceedings of 9th International Conference on Recent Advances in Structural Dynamics. (2006)
10. Suutala, J., Roning, J.: Towards the adaptive identification of walkers: automated feature selection of footsteps using distinction-sensitive LVQ. In: Proceedings of International Workshop on Processing Sensory Information for Proactive Systems. (2004) 61–67
11. NIST: Speaker recognition evaluation website. <http://www.nist.gov/speech/tests/spk/index.htm>
12. Vapnik, V.N.: Statistical Learning Theory. Wiley, New York
13. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery. Springer. **2**(2) (1998) 1–47
14. Martin, A., Doddington, G., Kamm, T., Ordowski, M., Przybocki, M.: The DET curve in assessment of detection task performance. In: Eurospeech. **1** (1997) 1895–1898
15. NIST: Det-curve software for use with matlab. Available at <http://www.nist.gov/speech/tools>
16. Jolliffe, I.T.: Principal Component Analysis. Springer Series in Statistics. (2002)

# **Chapter 14**

## **Keystroke Dynamics-Based Credential Hardening Systems**

**Nick Bartlow and Bojan Cukic**

**Abstract** Keystroke dynamics are becoming a well-known method for strengthening username- and password-based credential sets. The familiarity and ease of use of these traditional authentication schemes combined with the increased trustworthiness associated with biometrics makes them prime candidates for application in many web-based scenarios. Our keystroke dynamics system uses Breiman's random forests algorithm to classify keystroke input sequences as genuine or imposter. The system is capable of operating at various points on a traditional ROC curve depending on application-specific security needs. As a username/password authentication scheme, our approach decreases the system penetration rate associated with compromised passwords up to 99.15%. Beyond presenting results demonstrating the credential hardening effect of our scheme, we look into the notion that a user's familiarity to components of a credential set can non-trivially impact error rates.

### **14.1 Introduction**

As the use of web-based applications involving e-commerce, banking, e-mail, word processing, and spreadsheets continues to proliferate, so does the need to further secure such applications. It is widely accepted that systems relying solely on username/password credential sets provide weak security in terms of their ability to deliver user authentication and access control [1, 17, 28]. To make matters worse, many reports [1, 8, 17, 18, 25] indicate that the amount of credential sets an average web user must maintain is growing rapidly. One 2007 large-scale study conducted by Microsoft estimates that an average web user has 25 accounts which require passwords to access [12]. In typical authentication environments requiring a high degree of security, we often turn to biometrics. The so-called "hard" biometrics such as fingerprint, face, and iris undoubtedly meet the need of increased security but often include obstacles such as cost, interoperability, or privacy standing in the

---

N. Bartlow (✉)  
West Virginia University, Morgantown, WV, USA  
e-mail: nick.barlow@mail.wvu.edu

way of deploying such systems in remote and unsupervised web-based environments. Keystroke dynamics, or the extraction and analysis of the patterns in which an individual types, does not incur the obstacles of cost, interoperability, or privacy. Although not as unique as “hard” biometrics, the achievable increase in security combined with their “ease of deployment” makes keystroke dynamics especially suited for application in web-based environments. At its core, keystroke dynamics is typically characterized by the combination of inter-key latencies and key hold times in typing sequences. Inter-key latency or delay can be defined as the length of time between two successive keystrokes (positive or negative in the case of overlapping strokes). Key hold times are simply the length of time in which a key is held down during a keystroke. In this chapter, we demonstrate the password hardening effect of keystroke dynamics using random forests with two different types of passwords. The first type is a set of English passwords (commonly found in password attack dictionaries) and the second type is a set of pseudo-random passwords which assuredly are not found in password dictionaries. Besides establishing the hardening effect of adding keystroke dynamics, we investigate the degree to which familiarity of credential set components affects keystroke dynamics performance. In other words, we attempt to quantify whether or not recognition performance is non-trivially affected when authentication is driven by input stimuli of varying familiarity (usernames, English words, random strings). Our findings are validated using a data set which contains almost 9,000 input sequences, a large collection with respect to those utilized in related studies. The remainder of this chapter is broken down as follows. Section 14.2 outlines the history of the field dating from the early 1980s through the state of the art. Section 14.3 describes the approach of the work. Section 14.4 provides background on the experimental method including the collection system, user interaction with the system, and data collection results. Section 14.5 examines the algorithm employed for classification and the features extracted from the data set. Section 14.6 presents the experimental analysis and results of the work. Section 14.7 provides information on assumptions and considerations of the work. Finally, Section 14.8 summarizes the contribution of the work.

## 14.2 Related Work

Keystroke dynamics as known today first came about in 1980 with the seminal work by Gaines and Lisowski [14]. Their work utilized statistical significance tests between 87 lowercase letter inter-key latencies (delays) to differentiate users and achieved impressive performance results. The field has continued to mature over the next three decades with contributions falling in four main categories: input features, classification algorithms, input requirements, and capture devices.

### 14.2.1 Input Features

In 1993, Brown and Rogers offered the next major contribution to the field by adding the duration of keystrokes (also known as hold time) to the feature space of keystroke dynamics [6]. Virtually all works after Obaidat and Sadoun’s 1997

contribution utilize both latency and hold time as the primary features used in classification [27]. Some derivative features have surfaced from time to time such as trigraph hold times found in Bergandano et al. [4]. Additionally, De Ru et al. investigated geometric distances between keys in [9].

### ***14.2.2 Classification Algorithms***

Traditionally, classification algorithms have compared the incoming samples to one or more reference samples in a template database through a distance metric. Popular distance metrics include Euclidean, Mahalanobis, Manhattan, Chebyshev, and Hamming. Gaines and Lisowski [14], Garcia [15], Young and Hammon [32], and Joyce and Gupta [20] all provide examples of algorithms that utilize one or more of these distance metrics as classification schemes. Beyond distance metrics, many machine learning approaches have been applied. Obaidat et al. [26, 27], Brown et al. [6], and Maisuria et al. [23] utilized neural networks. Cho and Yu have applied support vector machines (SVMs) extensively [31, 33]. Bartlow and Cukic explored the decision-tree-based approach of random forests [3]. Finally, evolutionary approaches such as genetic algorithms and particle swarm optimization have been investigated by Lee et al. [21].

### ***14.2.3 Input Requirements***

In [16], Gunetti and Picardi differentiate between keystroke dynamics based on fixed text and free text. By fixed text, they refer to analysis only at the time of authentication or login, whereas in free text, analysis occurs at the time of authentication and either periodically or continuously throughout the course of the session in question. The vast majority of the research in the field has focused on the fixed text segment. Although this dichotomy may contribute to the trend, a pattern of decreasing input requirements does seem to exist. In the 1980 seminal work by Gaines and Lisowski, 300–400 word passages were required for verification of users [14]. Garcia's 1986 system utilized individual's names as well as 1,000 common words [15]. In 1990, Joyce and Gupta started the decreasing input requirement trend, requiring only username, password, first name, and last name [20]. Most subsequent works have required only a combination of username, password, or both for verification [3, 6, 23, 24, 26, 27, 31, 33].

For obvious reasons it is not appropriate to compare the input requirements of fixed text systems with free text systems. Looking at free text works, no noticeable trend regarding input requirements seems to exist. Leggett et al. analyzed 537 character passages in their 1991 work [22]. Furnell et al. analyzed a continuous user authentication system using 2,200 characters collected twice in 1996 [13]. Although not clearly specified, Monroe and Rubin's 1997 work involved users entering a number of sentences and phrases. Two works by Dowland et al. involved continuous monitoring of Windows NT usage which naturally entailed

extremely large samples of text [10, 11]. Bergandano et al. conducted a work based on two passages of text each around 300 characters long [4]. Besides offering a summary of free text research from which most of this section originates, Gunetti and Picardi studied web-based Java input on the order of 700–900 characters [16]. Janakiraman et al. offer perhaps the largest data set considered in their 2007 study which involved 30,000–2,000,000 key events per user collected over 2 weeks of continuous Windows XP use [19].

#### **14.2.4 Capture Devices**

The overwhelming majority of work in the field has dealt with capturing keystroke input from standard keyboards as accompanied by desktop PCs, although, in the last 15 years, we have seen some studies begin to include the use of laptop keyboards. Since the inclusion of laptop keyboards, much speculation has gone into the degree to which changing keyboards affects one's unique typing signature. Although undisputed, this effect has not been thoroughly studied. Beyond that, some groups have investigated the potential of keystroke dynamics on mobile telephones, numeric keypads (such as ATMs and access control devices), etc. [7].

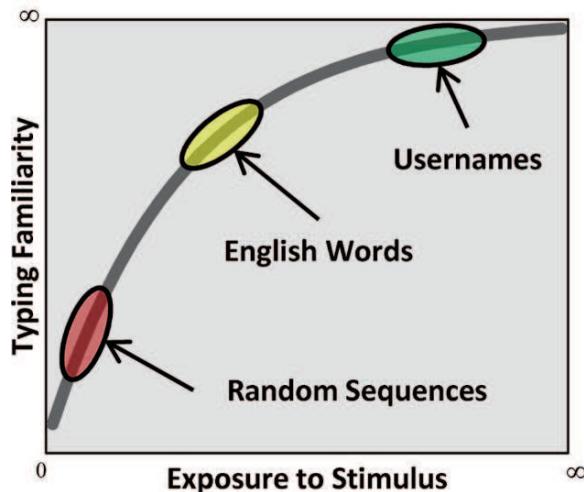
### **14.3 Approach**

Over the history of the field, a large number of keystroke dynamics studies have taken place in controlled environments. Whether subject to heavily supervised lab conditions, a variety of explicit rules, or standardized and uniform capture equipment, these experiments are not typically in line with our proposed application environments. Although stand-alone deployment in an environment with homogenous equipment is undoubtedly still a relevant area of application, current application demands are moving toward web-oriented environments in which choice of hardware, user supervision, and other external noise factors are not controllable. Although we admittedly do not cover all potential sources of noise in this type of application (i.e., mimicry), we attempt to simulate this “free” environment to the best of our ability by employing a completely unsupervised, web-based collection system with no hardware restrictions.

As noted previously, the relationship between the individual components of the credential sets is of specific interest to this work, in particular, the degree of familiarity that the user has to each component. Figure 14.1 presents the idea that the more times a user is exposed to a stimulus (component of a credential set), the more familiar the user becomes with typing the component. We hypothesize that the username is easily the most familiar of the three components considered, as most individuals are exposed (required to type it) multiple times on a daily basis. Following the username, an English password is likely the component with a lesser degree of typing familiarity as it may not be typed daily but it has probably been typed before if not multiple times. Finally, a random password containing characters of varying capitalization, digits, and special characters probably represents the lowest

**Fig. 14.1** Typing familiarity vs. exposure to stimulus

### Relationship Between Familiarity of Typing Input and Exposure



attainable familiarity as there is very little chance that the user has ever seen such a password, much less typed it. By assigning these three types of components to create credential sets, we are able to investigate the degree to which familiarity to input stimulus affects keystroke dynamics performance.

## 14.4 Experimental Method

### 14.4.1 Collection System

To achieve a completely web-based architecture, front-end client-side Java applets were developed in the NetBeans integrated development environment (IDE) and designed to run within standard web browsers (Mozilla Firefox, Internet Explorer, Netscape) using the Sun® Java Console. This offered both a tested open source platform as well as the bulk of the computation to occur in the client computers. On the server side, a MySQL® database (also an open source product) was used to house the data. Therefore, once client-side computation was finished, the input was entered into the database via the previously established client-server connection [2]. It should also be noted that the use of a web-based collection system does not preclude application of the techniques used in this work in non-web-based applications such as terminals and mail clients.

During the registration process, each user was given two sets of username/password credential sequences. The username remained the same across both sets of credentials and was of the form Firstname.Lastname with the first letter of each name capitalized. The first password was an eight letter lowercase English word taken from a cryptographic dictionary attack list [29]. Examples of such

passwords included computer and swimming. The second password consisted of 12 randomly generated characters in a consistent pattern. The format of the pattern was as follows:

SUUDLLLLDUUS

where S is a special symbol, U is an uppercase letter, L is a lowercase letter, and D is a digit. Examples of such passwords include +AL4lfav8TB= and \_UC8gkum5WH. This pattern was intended not to elicit any specific behavior but only to allow for easy interpretation of potentially ambiguous symbols. The extra length of the pseudo-random passwords was incorporated to arrive at what are considered to be cryptographically strong passwords.

Beyond the structure of the input, the behavioral nature of this biometric scheme required a slightly more involved data collection process than what is typical in conventional physiological biometric systems. Most notably, one cannot simply compare genuine input of one user to genuine input from another user in order to establish an instance of imposter input. In this study, the passwords were different for each and every individual. Therefore, the collection tool required the development of two different user interfaces. One interface requests users to input the genuine credentials provided to them in the registration phase. The second interface requests users to input the credentials assigned to another user, generating an imposter authentication attempt. Figure 14.2 shows the registration (a), genuine input (b), and imposter input (c) front ends.

Within the genuine input front end, users were asked to input each of their credentials (username + password1 and username + password2) five times every day for approximately 3 weeks. The imposter input front end was slightly different and can be seen in Fig. 14.2(c). In this front end, users were provided with credentials of a different registered user. To avoid inadvertently collecting genuine data in the imposter section, user always provides his/her username from the “My UserName” field. Upon selection of the username, the data collection system populates the “Imposter Credentials” fields, automatically selecting a pair of credentials that is short on imposter data. In this way, the number of imposter sequences was kept balanced over the set of all the enrolled users.

At this point, the user simply logged in the same way as if this were his/her own genuine credentials. Pending a successful collection step, the new username/password pair appears in the window and the process repeats itself. Similar to the genuine input screen, users were asked to input a total of 10 imposter sequences per day.

#### **14.4.2 User Supervision**

As mentioned previously, we attempted to minimize the supervision component of the data collection. To that effect, users were only provided with a basic series of instructions and short video clips (for those inclined) to explain the functionality and expected use of the system. Although a large number of the students were included in the study (from our lab and university classes), there was no requirement to offer

(a) Registration Front End.

(b) Genuine Input Front End.

(c) Imposter Input Front End.

**Fig. 14.2** The three main pages for data collection: (a) the registration front end, (b) the genuine input front end, (c) the imposter input front end

samples in the lab or in an on-campus setting. Additionally, many participants had no affiliation with the university and their data were collected from off-site locations throughout the internet. To that regard, no face-to-face guidance was provided to the participants.

#### 14.4.3 Collection Results

At the time of final analysis, the database had a total of 53 users with over 10,000 total input sequences. After applying a minimum number of 15 valid sequences of each type of password, a total of 41 users and 8,882 username/password sequences

**Table 14.1** Data included in the experiment. The abbreviations E, R, and T correspond to English password sequences, random password sequences, and total sequences. The final row includes the total number of users, average length for usernames and both types of passwords, and the total number of each type of input sequence collected

UserID	Username	English password (E)	Random password (R)	Genuine seq.			Imposter seq.		
				E	R	T	E	R	T
01	12	kathleen	@QZ4ozka1XE\$	120	110	230	46	46	92
03	13	williams	]RR4axpe0WA>	048	050	098	46	46	94
04	14	rosemary	:LC6nva9OO~	073	020	093	47	47	94
05	09	mitchell	>YH2avia0ER#	062	070	132	47	47	94
06	16	wolfgang	@WI7tjeb8WX}	117	108	225	47	47	94
07	13	aerobics	)YK2zquv9IQ+	083	076	159	47	47	94
08	10	firebird	.MS1suyf8MP^	053	052	105	47	47	94
09	11	fountain	(GC5idxx8TH{	051	051	102	47	47	94
10	18	caroline	^ZT7wyazz6JA[	016	017	033	47	47	94
11	12	zeppelin	!CN0srui6ZO=	122	119	241	46	46	92
12	19	bumbling	~XM6bywn6JL?	074	085	159	46	46	92
13	09	director	'VA0snuv1HA:	090	104	194	46	46	92
14	12	gonzales	&ZL7yfjj0GK*	059	061	120	46	46	92
15	16	password	?KK6cvuc1NK	059	088	147	46	46	92
17	14	business	;OI6vjog4QN>	053	058	111	46	46	92
18	12	fletcher	&UV1lkda5YH{	062	065	127	46	46	92
21	10	swimming	;KO3ovpt4QC>	046	021	067	46	46	92
23	17	wheeling	;LB3chtu2YX`	118	137	255	46	46	92
24	12	newcourt	:ZQ5grpX8VH;	027	028	055	46	46	92
25	15	snoopydog	,GG5ruft6IG+	052	045	097	46	46	92
26	07	colorado	\$ZZ9ilfg9RJ(	043	025	068	46	46	92
27	14	homebrew	*TY1drmj7CR\$	158	108	266	46	46	92
28	11	dolphins	[LO2uqam8UI+	047	041	088	46	46	92
29	15	plymouth	)VS0iaka5WW!	048	042	090	46	46	92
30	14	broadway	;PG3xue19LU}	054	049	103	46	46	92
31	13	woodwind	)EZ7mjjp4YM+	075	076	151	46	46	92
32	14	mountain	&BA1ishf0FC	054	041	095	46	46	92
33	10	strangle	=BP8duim7IF@	091	089	180	46	46	92
34	12	strangle	<WN1zegb5RS\$	077	074	151	46	46	92
35	13	princess	>GD0dgby6JU{	030	016	046	46	46	92
37	13	clusters	@UK8uudo5GS.	033	029	062	46	46	92
38	10	martinez	_UC8gkum5WH@	030	021	051	46	46	92
40	13	tacobell	[VB6jveb2PC~	015	017	032	46	46	92
43	12	baritone	&FO4ovcv0VK!	028	029	057	46	46	92
44	14	frighten	\$IP2ulld5QT@	053	051	104	46	46	92
46	12	starwars	\$SP3lhkt1YX{	031	039	070	46	46	92
47	15	thompson	#PO5dlfq0JW:	108	129	237	46	46	92
48	14	explorer	<JT5ocyi8TK=	099	099	198	46	46	92
49	15	elephant	&XQ5jwsp8KA]	080	080	160	46	46	92
51	13	springer	@SF5sjnd5EY/	017	015	032	46	46	92
53	12	sweatpea	^R08wxps1HI)	062	041	103	46	46	92
41	12.93	8	12	2,618	2,476	5,094	1,894	1,894	3,788

were used. Out of the 8,882 total sequences, 5,094 were of type genuine and 3,788 were of type imposter. A summarized breakdown of the data collected for each user can be found in Table 14.1. The demographics of the database represent a fairly diverse population in many regards. The gender split was approximately half, ages ranged from mid-teens to individuals in their early 60s, and there was also a relatively diverse racial makeup. Perhaps most importantly, the typing ability of the population was also very diverse, ranging from the most inept “hunt-and-peck” typists to individuals with professional training/experience. In that regard, the classification algorithms were required to differentiate not only between professional and amateur typists but also between the members of these two groups. Some of the users have been accustomed to working with multiple keyboard layouts. However, the type of keyboard used for data collection was not controlled. We know that a large percentage of subjects used desktop computers as well as laptops during the data collection. This is inevitably a source of noise but also demonstrates the ability of the technique to cope with variable hardware setups. The collection period lasted approximately 1 month.

## 14.5 Authentication Algorithm and Features

As mentioned in the related work section, a plethora of algorithms have been investigated as candidates for authentication mechanisms throughout the history of keystroke dynamics research. The scope ranges from simple distance metrics between probe and gallery templates, to complicated multi-layer neural networks. In this work, we arrive at feature vectors containing raw hold times and inter-key delays and use the decision-tree-based random forests algorithm developed by Breiman [5] to classify input sequences. It is important to note that we are assuming a verification framework where an identity is claimed. This assumption leads to a two-class problem; deciding whether an input sequence is genuine or imposter in nature. The following two sections provide expanded descriptions of the random forest algorithm (including the nature of the training and testing) and the feature set in which classification was based.

### 14.5.1 Random Forests and Training/Testing Framework

An elegant and powerful algorithm, random forests is named after two main characteristics. One, it is based on the development of a “forest” of decision-tree classifiers, each being similar to C5.0 decision trees developed by Quinlan [30]. Two, the method of generating the forests is based on the random sampling of features in the attribute space. The tree generation algorithm works as follows: each tree is grown based on a random sample selection of two-thirds of the instance population. In each decision tree that populates the forest, nodes, branches, and leaves are generated by continuously choosing the feature that yields the best split of the data based on  $m$  randomly selected features. Sub-tree generation continues to the

extent possible without pruning. Once all trees have been generated, new instances of feature vectors are passed through the trees of the forest and a voting process takes place to determine the classification result.

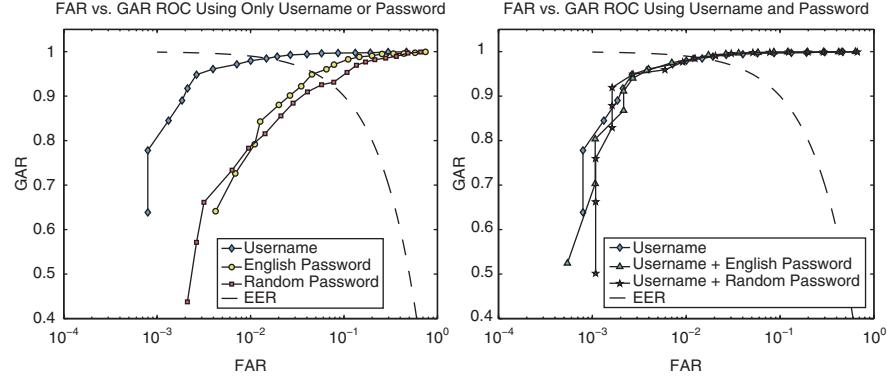
There are a number of attractive advantages random forests have over other machine learning algorithms. Our study pays particular attention to two of them. One, due to the two-thirds sampling used to train each tree, the remaining one-third so-called out-of-bag (OOB) sample is used to test the classification performance. Therefore, training and test sets do not need to be explicitly separated and the estimated error results are said to provide conservative estimates of future performance. Two, the ability to define varying voting schemes allows for generating forests tailored to specific matching applications. For instance, a 10–90% voting scheme for genuine and imposter classes places particular emphasis on minimizing the FRR, whereas a 90–10% scheme reverses the requirement, focusing attention on the FAR. This mechanism allows for the generation of receiver operating characteristic (ROC) curves which describe an entire range of achievable performance characteristics relative to FARs and FRRs. Other learners typically generate only a single operating point along a ROC curve. We generated 19 random forests for every user, each with a different voting scheme with voting increments of 0.05, ranging between 0.05–0.95 and 0.95–0.05. Furthermore, for each forest, 500 trees were generated and the default value of parameter  $m$  (features to consider at each node split) was used. In this case,  $m = \sqrt{X}$ , or the square root of the total number of attributes in the feature space, ( $X$ ), which was dependent on the particular input sequence. For instance, an 8-character English password might have 8 hold times and 7 delays for a total of 15 attributes. In this example  $m = 4 \approx \sqrt{15}$ .

#### **14.5.2 Feature Set**

Although we previously investigated using aggregate statistics based on the key-stroke hold times and delays of each input sequence such as averages and standard deviations in [3], this work utilizes the raw hold times and delays. In that regard, calculation of the feature vector for each sequence is arguably as simple as possible; no secondary calculation is necessary to arrive at the final feature vector. Given the requirement of shift key activity in the random passwords, it was possible that not every sequence had the same amount of hold times or delays. In this case, the longest feature vector was identified and all other sequences (both genuine and imposter) were padded with zeros to arrive at equal-length vectors for each user's data set.

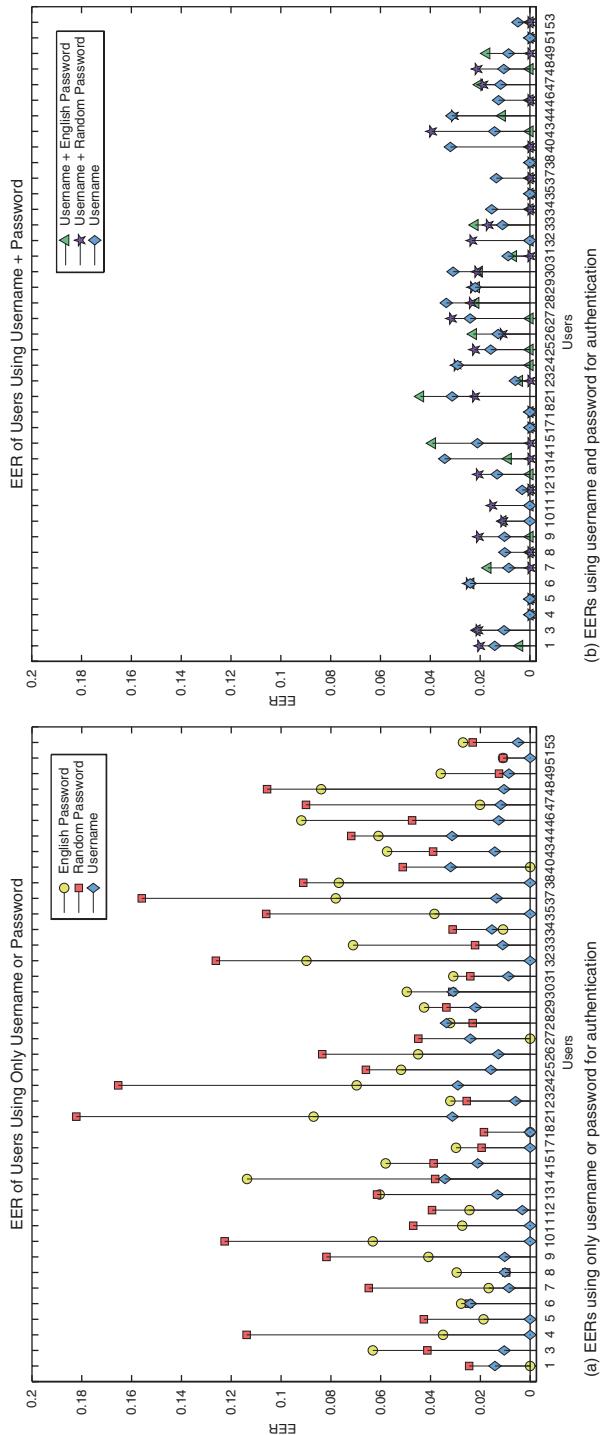
### **14.6 Experimental Analysis and Results**

In this section we look into three areas of interest: the performance of our keystroke dynamics system, the potential relationship between credential component familiarity and authentication results, and the credential hardening effect of applying



**Fig. 14.3** System ROC curves considering different credential set input components

keystroke dynamics to traditional username/password systems. We first demonstrate the performance of our keystroke dynamics system while varying the nature of algorithmic deployment and the input requirements in terms of credential components. First, to reiterate the behavioral nature of the system, genuine and imposter sequences were collected for each user in the system. Based on both types of sequences, 19 random forests were constructed for each user by varying the voting percentage required for classification. By doing so, we were able to calculate ROC curves and EERs for each user. Figure 14.3 shows the overall system ROC curves considering different credential components as input. These ROC curves were calculated by globally applying a random forest voting threshold to each user's dataset and averaging FARs and GARs across the 41 users. In the figure, plot (a) shows the three curves which result from considering only single credential components. In other words, only the hold times and delays for the username, English password, or random password were used for model creation and testing. Plot (b) shows the results when hold times and delays for both the username and the password components of the credential sets were used. As can be seen in (a), the English password outperforms the random password for most areas of the ROC curve while the username universally outperforms both passwords. This figure supports the hypothesis that performance is driven in part by familiarity of the credential set component used for authentication. In plot (b), there are select regions of the ROC curve where the addition of the random password to the username increases performance over using the username alone. However, in most regions sole use of the username results in good, if not better, performance, compared to the combination of both components of the credential set. Taking a more fine-grained look into the results, Fig. 14.4 displays the EERs for each of the 41 users in the study considering the same sets of credential components. Here we see the upper and lower bounds of the system's performance. The worst performing user for the English password and username is user #14 having EERs of 11.36% and 3.4%, respectively. User #21 has the worst random password performance with an EER of 18.22%. Approximately 25% of the users tested (10/41) achieve a 0% EER for usernames while 4 achieve this rate in English



**Fig. 14.4** Comparison of EERs across users considering different credential set input components

**Table 14.2** System performance using global and user-specific voting schemes

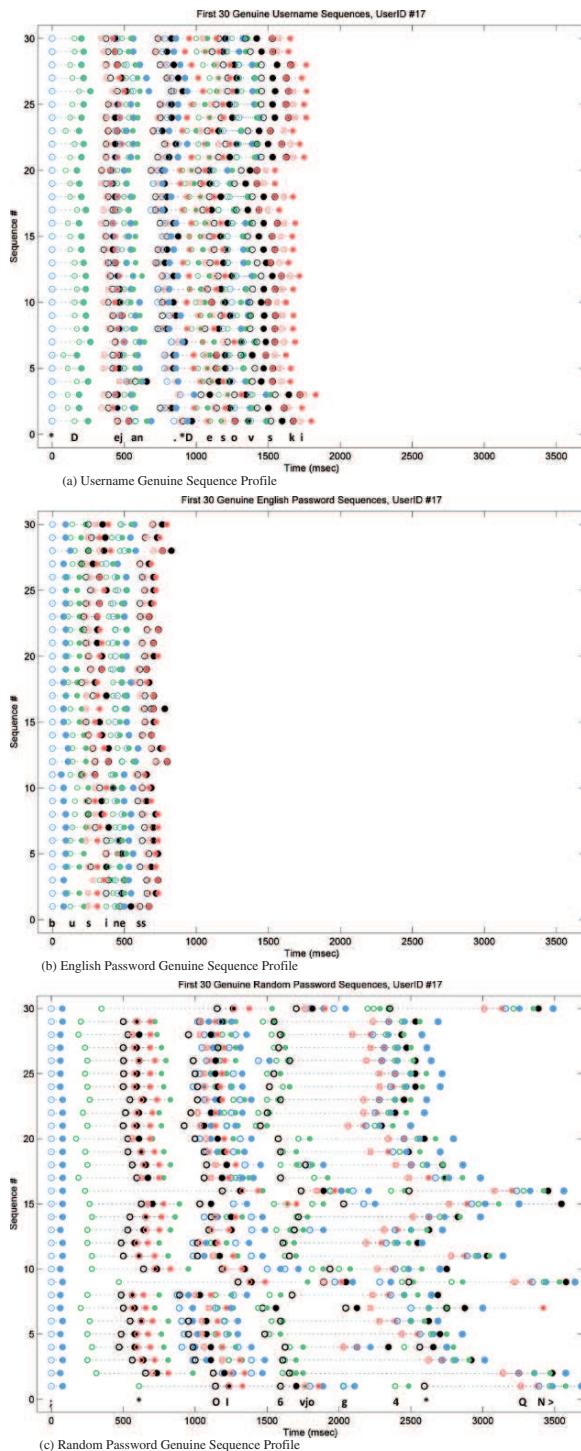
Credential set component(s)	EER global scheme (%)	EER user-specific scheme (%)	Difference (%)
Random password	6.796	5.507	1.289
English password	4.578	3.997	0.581
Username	1.511	1.284	0.226
Username + random password	4.581	1.173	3.408
Username + English password	3.138	0.852	2.286

passwords. The lower bound for performance on random passwords is 0.96% corresponding to user #8. If username and password components are combined, we find the maximum EER for both types of passwords improves substantially falling below 5%. Here user #21 has an EER of 4.4% for English passwords and user #43 has an EER of 3.9%. Additionally, more users achieve an EER of 0% with 23/41 for the combination of username and English passwords and 20/41 for username and random passwords.

As demonstrated earlier in Fig. 14.3, system-wide performance can be generalized by applying a global voting threshold across all users. Optionally, one may choose to apply user-specific voting schemes to optimize performance. Table 14.2 shows the difference in performance of applying global voting schemes vs. user-specific voting schemes across the different components of the credential sets. Naturally, if system-wide performance is characterized using voting schemes optimized to each user, error rates decrease. This notion is characterized by the last column in the table indicating the decrease in EER achieved from applying the global voting threshold to applying the user-specific voting schemes. Once again, we see the trend that credential set components that are more familiar to users perform better. Here, usernames outperform both the English passwords and the random passwords when only one component is considered. Furthermore, incorporating both the username and the English password yields the best performance result when looking at the user-specific EER of 0.852%.

In one final examination of the notion of familiarity, we look into the genuine keystroke profiles for one user over the three individual components tested: the username, English password, and random password. Intuitively, we would expect to see less variability from sequence to sequence in components a user is more familiar with than in those a user is less familiar with. Although space prohibits examining the profiles of every user for each sequence, Fig. 14.5 shows the first 40 sequence profiles for user #17's username, English password, and random password. In the three plots, keystrokes are defined by pairs of circles with matching colors. Here, open circles indicate a key down and filled circles correspond to the matching key up. Therefore, the time between the open circle and the closed circle of the same color represents the hold time. The time between two open circles represents the delay between two adjacent keystrokes. For clarity, the typed input corresponding to the first sequence is labeled in each plot with \*'s indicating a shift key was pressed. Not surprisingly, we see a greater deviation in the profile of each sequence in the random passwords than we do in the username and English password sequences.

**Fig. 14.5** The first 40 genuine input sequences of user #17 for username, English password, and random password components



**Table 14.3** Credential hardening effect with globally applied voting schemes

Credential set component(s)	FAR (%)			FRR (%)		
	Before	After	Difference	Before	After	Difference
Random password	100.00	6.80	↓ 93.20	0.00	6.80	↑ 6.80
English password	100.00	4.58	↓ 95.42	0.00	4.58	↑ 4.58
Username	100.00	1.51	↓ 98.49	0.00	1.51	↑ 1.51
Username + random password	100.00	4.58	↓ 95.42	0.00	4.58	↑ 4.58
Username + English password	100.00	3.14	↓ 96.86	0.00	3.14	↑ 3.14

This is potentially a reason for the difference in performances seen in Table 14.2. Should the first letter of each portion of the username not been capitalized, even less inter-sequence variation may have been found in the usernames, shown in plot (a).

To conclude the analysis of the results, we demonstrate the credential hardening effect of applying our keystroke dynamics system to a classic username/password authentication system. To do so, we assume that imposters attempting unauthorized entry have obtained the password of the user in which they are imposing. Based on this assumption, the FAR or penetration rate of the imposter, in absence of keystroke dynamics, will be 100%. In other words, if the imposter knows the targeted users' credentials he/she will always be able to type them correctly, thereby gaining access. On the other hand, when our keystroke dynamics biometric is used to augment the system, correct content of the password is only a partial requirement; the imposter must also type the credentials with the same keystroke dynamics signature. Table 14.3 demonstrates the effect of this additional requirement assuming operation at performance rates derived through global application of random forests voting schemes across all users.

The penetration rate into the system in terms of FAR decreases by 93.2%, 95.4%, and 98.49%, respectively, for random passwords, English passwords, and usernames taken as singleton components. These rates are calculated by simply subtracting the keystroke dynamics FAR from the assumed 100% penetration without the biometric augment. The higher security comes at a relatively low price as the associated increases in FRR are 6.8%, 4.58%, and 1.51% in random passwords, English passwords, and usernames, respectively. By applying user-specific voting schemes we can achieve a greater credential hardening effect through the addition of the keystroke dynamics augment. Table 14.4 shows the decrease in system penetration

**Table 14.4** Credential hardening effect with user-specific voting schemes

Credential set component(s)	FAR (%)			FRR (%)		
	Before	After	Difference	Before	After	Difference
Random password	100.00	5.51	↓ 94.49	0.00	5.51	↑ 5.51
English password	100.00	4.00	↓ 96.00	0.00	4.00	↑ 4.00
Username	100.00	1.28	↓ 98.72	0.00	1.28	↑ 1.28
Username + random password	100.00	1.17	↓ 98.83	0.00	1.17	↑ 1.17
Username + English password	100.00	0.85	↓ 99.15	0.00	0.85	↑ 0.85

in this case. Here, we see greater decreases in system penetration while also decreasing user inconvenience in terms of the FAR. Additionally, the improvement from applying keystroke dynamics on the username taken alone or combined with either type of password approaches the limits of performance with rates of penetration falling at 1.28%, 1.17%, and 0.85%, respectively. Similar to the globally applied voting schemes, this comes at an arguably low cost to user convenience.

## 14.7 Discussion

Despite the continuing maturity of the field, many open problems remain with keystroke dynamics and credential hardening systems. This work does not claim to answer issues such as rigid model training requirements of operational environments, imposter mimicry, or generation of imposter data sequences. Regarding issues specifically visited in this work, a number of assumptions may need further explanation. Since two individual components of the credential sets considered required shift-key behavior to modify characters, each extracted feature vector consisting of hold times and delays need not have the same number of attributes as every other vector in a data set. As mentioned earlier, each sequence was back-filled with zeros to ensure that the extracted feature vector for each sequence had the same length. Some may argue that this requires *posteriori* knowledge of the data, namely, the maximum length feature vector of all sequences in a data set. We would argue, however, that this knowledge is superficial as a maximum keystroke limit per credential set component could be imposed without having ill effect on the user's experience or the performance of the classification algorithm. To justify the latter, the random forest algorithm will automatically weed out attributes with little information gain (zero-valued for virtually all instances) during decision-tree generation.

Although this study admittedly does not consider user mimicry, we do emphasize that imposter keystroke “attacks” are considered “zero-effort” in that users were asked to type imposter data naturally. In other words, imposters made no effort to deviate from their normal typing patterns in an effort to more reliably emulate the targeted genuine sequence input. Similar to forgery in handwritten signatures, we assume that formal attempts beyond “zero-effort” attacks may result in reduction in credential hardening effects. This notion was not considered primarily based on the fact that it would be prohibitively difficult to do so given the completely remote and unsupervised nature of the data collection effort. We feel the importance of these characteristics coupled with the increased size of the data set outweighs the importance of gathering data incorporating mimicry.

Even though the results presented in this work support the hypothesis that user familiarity with credential set components may drive keystroke dynamics performance, statistical tests of sequence uniformity can be applied to further quantify differences in familiarity. This could be achieved by fitting distributions to individual keystroke hold times and delays across all genuine sequences of a data set and applying statistical tests of uniformity such as Rayleigh, Rao and Neyman. Should

our hypothesis hold, keystroke hold times and delays from usernames and English passwords would more closely resemble uniform distributions than those resulting from random passwords. That said, modification of the username to all lowercase letters may be necessary.

Additionally, the authors are aware that although the size of the data sets included in the work is similar to those found in related academic efforts, the strength of the evidence supporting the hypotheses presented could be increased through application on larger test beds. To that regard, continuing data collection efforts are in place to make larger scale studies possible.

## 14.8 Conclusion

In this work we further established the viability of keystroke dynamics as a method of hardening traditional username/password credential sets in remote, unsupervised environments that lack restrictions in hardware use (selection of keyboard). This hardening can be achieved by applying keystroke dynamics to individual components of a credential set or both simultaneously depending on the nature of the application environment and desired performance. Additionally, we investigated the notion that credential set component familiarity has an affect on keystroke dynamics performance. Namely, sequences that users are exposed to consistently and arguably familiar with such as usernames and English words offer better classification potential compared to undoubtedly unfamiliar sequences such as randomly generated strings. From a realistic application setting, this may be encouraging as despite their relative weakness, most users routinely select passwords with which they are familiar such as family names and English words. This phenomenon may be in part due to more consistent typing signatures in genuine sequences corresponding to components in which users are more familiar with.

## Proposed Questions and Exercises

- Unsupervised remote systems akin to the one described in this work are often subject to various sources of noise that can potentially make classification more challenging than in environments that are relatively “noise free.” Explain at least three different sources of noise that one might observe in such an unsupervised remote system.
- When testing keystroke dynamics systems, special considerations need to be taken when generating imposter data for training and testing. Explain the nature of these considerations, how imposter input is typically generated, and provide at least one example of a way to eliminate the need for this specialized generation of data.
- Considering the format of the usernames (`Firstname.Lastname`) and random passwords (`SUUDLLLLDUUS`) incorporated in this experiment, suggest one or more

derivative features that may aid in the process of classifying sequences as genuine or imposter.

- In the random forest algorithm, default settings rely on a scheme requiring a simple majority ( $> 50\%$ ) vote of trees to arrive at a classification decision. Why does varying the voting scheme beyond the default simple majority case allow one to change classification performance in terms of FAR and FRR?
- Outside of deviations in performance, why might a system choose to apply keystroke dynamics to only a single credential component (i.e., username or password) instead of the entire credential set?
- In the results section of the work, we notice an increase in performance when optimizing random forest voting schemes on a user-to-user basis as opposed to global application of voting schemes. What challenges may exist to deploying optimization of this nature?
- Examining Fig. 14.5 in Section 14.6, we see that the inter-sequence variation from the sequences in the username (shown in plot (a)) is similar to those found in the English password (shown in plot (b)) despite the fact that individuals are more familiar with their name than random English words. Provide one potential explanation why we do not see even less inter-sequence variation in the username sequences shown in plot (a).
- Aside from ease of deployment and lack of supervision, provide at least one more positive aspect of using the behavioral biometric of keystroke dynamics not listed in this work.

## References

1. Adams, A., Sasse, M.A.: Users are not the enemy. *Commun. ACM* **42**(12), 40–46 (1999). <http://doi.acm.org/10.1145/322796.322806>
2. Bartlow, N.: Username and Password Verification through Keystroke Dynamics. Master's Thesis. West Virginia University. 2005
3. Bartlow, N., Cukic, B.: Evaluating the Reliability of Credential Hardening Through Keystroke Dynamics. In: ISSRE, pp. 117–126 (2006)
4. Bergadano, F., Gunetti, D., Picardi, C.: User authentication through keystroke dynamics. *ACM Trans. Inf. Syst. Secur.* **5**(4) (2002)
5. Breiman, L.: Random forests. *Machine Learning* **45**(1), 5–32 (2001)
6. Brown, M. and Rogers, S.J.: User identification via keystroke characteristics of typed names using neural networks. *Int. J. Man-Mach. Stud.* **39**(6), 999–1014 (1993). <http://dx.doi.org/10.1006/imms.1993.1092>
7. Clarke, N.L., Furnell, S.: Authenticating mobile phone users using keystroke analysis. *Int. J. Inf. Sec.* **6**(1), 1–14 (2007)
8. Collins, D.: Irish computing users and the passwords they choose. National University of Ireland Master's Thesis (2006). <http://hdl.handle.net/10099/13>
9. De Ru, W., Eloff, J.: Enhanced password authentication through fuzzy logic. *IEEE Expert [see also IEEE Intelligent Systems and Their Applications]* **12**(6), 38–45 (Nov/Dec 1997). [10.1109/64.642960](https://doi.org/10.1109/64.642960)
10. Dowland, P., Furnell, S., Papadaki, M.: Keystroke analysis as a method of advanced user authentication and response. In: SEC, pp. 215–226 (2002)
11. Dowland, P., Singh, H., Furnell, S.: A preliminary investigation of user authentication using continuous keystroke analysis. In: In Proc. 8th IFIP Annual Working Conf. on Information Security Management and Small System Security (2001)

12. Florencio, D., Herley, C.: A large-scale study of web password habits. In: WWW '07: Proceedings of the 16th international conference on World Wide Web, pp. 657–666. ACM, New York, (2007). <http://doi.acm.org/10.1145/1242572.1242661>
13. Furnell, S., Morrissey, J.P., Sanders, P.W., Stockel, C.T.: Applications of keystroke analysis for improved login security and continuous user authentication. In: SEC, pp. 283–294 (1996)
14. Gaines, R., Lisowski, W., Press, W., Shapiro, S.: Authentication by keystroke timing: Some preliminary results. Rand Report R-256-NSF, The Rand Corporation, Santa Monica, CA (1980)
15. Garcia, J.: Personal identification apparatus. Patent 4,621,334, U.S. Patent and Trademark Office, Washington, D.C. (1986)
16. Gunetti, D., Picardi, C.: Keystroke analysis of free text. ACM Trans. Inf. Syst. Secur. **8**(3), 312–347 (2005). <http://doi.acm.org/10.1145/1085126.1085129>
17. Ives, B., Walsh, K.R., Schneider, H.: The domino effect of password reuse. Commun. ACM **47**(4), 75–78 (2004). <http://doi.acm.org/10.1145/975817.975820>
18. How to fix your life in 2004; simple ways to cut travel, college tabs, even waistlines. Wall Street Journal (Eastern Edition) p. D.1. 12/31/2003
19. Janakiraman, R., Sim, T.: Keystroke dynamics in a general setting. In: ICB, pp. 584–593 (2007)
20. Joyce, R., Gupta, G.: Identity authentication based on keystroke latencies. Commun. ACM **33**(2) (1990)
21. Lee, J.W., Choi, S.S., Moon, B.R.: An evolutionary keystroke authentication based on ellipsoidal hypothesis space. In: GECCO '07: Proceedings of the 9th annual conference on Genetic and evolutionary computation, pp. 2090–2097. ACM, New York (2007). <http://doi.acm.org/10.1145/1276958.1277365>
22. Leggett, J., Williams, G., Usnick, M., Longnecker, M.: Dynamic identity verification via keystroke characteristics. Int. J. Man-Mach. Stud. **35**(6), 859–870 (1991). [http://dx.doi.org/10.1016/S0020-7373\(05\)80165-8](http://dx.doi.org/10.1016/S0020-7373(05)80165-8)
23. Maisuria, L.K., Ong, C.S., Lai, W.K.: A comparison of artificial neural networks and cluster analysis for typing biometrics authentication. In: International Joint Conference on Neural Networks (IJCNN), **5**, 3295–3299 (1999)
24. Monroe, F., Reiter, M.K., Wetzel, S.: Password hardening based on keystroke dynamics. Int. J. Inf. Sec. **1**(2), 69–83 (2002)
25. Noguchi, Y.: Access denied. The Washington Post (2006). [http://www.washingtonpost.com/wp-dyn/content/article/2006/09/22/AR2006092201612\\_pf.html](http://www.washingtonpost.com/wp-dyn/content/article/2006/09/22/AR2006092201612_pf.html)
26. Obaidat, M.S., Macchiarolo, D.T.: An on-line neural network system for computer access security. IEEE Trans. Ind. Electron. **40**(2), 235–241 (1993)
27. Obaidat, M.S., Sadoun, B.: Verification of computer users using keystroke dynamics. IEEE Trans. Syst. Man Cybern. **27**(2), 261–269 (1997)
28. Schneier, B.: Applied Cryptography: Protocols, Algorithms, and Source Code in C, second edn. Wiley, New York (1996)
29. Shaffer, G.: Geodsoft good and bad passwords how-to: An example list of common and especially bad passwords (2004). <Http://geodsoft.com/howto/password/common.htm>
30. Software package c5.0 / see5 (2004). <Http://www.rulequest.com/see5-info.html>
31. Sung, K.S., Cho, S.: GA SVM Wrapper Ensemble for Keystroke Dynamics Authentication. In: ICB, pp. 654–660 (2006)
32. Young, J.R., Hammon, R.W.: Method and apparatus for verifying an individual's identity. Patent 4,805,222, U.S. Patent and Trademark Office, Washington, D.C. (1989)
33. Yu E., Cho S.: Keystroke dynamics identity verification - its problems and practical solutions. Comput. Secur. **23**(5), 428–440 (2004)

# Chapter 15

## Detection of Singularities in Fingerprint Images Using Linear Phase Portraits

Surinder Ram, Horst Bischof, and Josef Birchbauer

**Abstract** The performance of fingerprint recognition depends heavily on the reliable extraction of singularities. Common algorithms are based on a Poincaré Index estimation. These algorithms are only robust when certain heuristics and rules are applied. In this chapter we present a model-based approach for the detection of singular points. The presented method exploits the geometric nature of linear differential equation systems. Our method is robust against noise in the input image and is able to detect singularities even if they are partly occluded. The algorithm proceeds by fitting linear phase portraits at each location of a sliding window and then analyses its parameters. Using a well-established mathematical background, our algorithm is able to decide if a singular point is existent. Furthermore, the parameters can be used to classify the type of the singular point into whorls, deltas and loops.

### 15.1 Introduction

The application of fingerprint-based personal authentication and identification has rapidly increased in the recent years. The use of this technology can be seen in forensics, commercial industry and government agencies, to mention a few. Fingerprints are attractive for identification because they can characterize an individual uniquely and their configuration does not change throughout the life of individuals. The processing steps required for personal verification or identification based on fingerprints consist of acquisition, feature extraction, matching and a final decision [7].

Large volumes of fingerprints are collected and stored everyday in a wide range of applications. Automatic fingerprint recognition requires that the input fingerprint is matched with a large number of fingerprints stored in a database. Therefore, the first step in an identification system is the classification of a given fingerprint into five categories [7]. This reduces the amount of data to be searched for matches as

---

Surinder Ram (✉)  
Technical University of Graz, Austria Institute for Computer Graphics and Vision,  
e-mail: ram@icg.tugraz.at

the database can be partitioned into subsets. These five categories are arch, tented arch, left loop, right loop and whorl. Common algorithms extract singular points in fingerprint images and perform a classification based on the number and location of these singularities.

In order to ‘match’ two fingerprints it is necessary to extract minutiae, which are special points in fingerprints where ridges end or bifurcate. Two fingerprints can be reported as equal if a certain number of minutiae positions are identical in both fingerprints. In general, matching of fingerprint images is a difficult task [5], mainly due to the large variability in different impressions of the same finger (i.e. displacement, rotation, distortion, noise, etc.). One way to relax the problem in terms of performance and runtime is to use certain ‘landmarks’ in the image in order to apply a pose transformation. Since singular points are unique landmarks in fingerprints, they are used as reference points for matching [8].

### **15.1.1 Methods for Extraction of Singularities**

Many approaches are described for singular point detection in literature. Karu and Jain [6] referred to a Poincaré Index method. However, there are principle weaknesses adhered to this method. Many rules and heuristics have been proposed by various authors (e.g. [20]) in order to make the method robust against noise and minor occlusions. Due to its simplicity and more than adequate performance in most images, this method enjoys high popularity in fingerprint recognition systems.

Another method described in [11] exploits the fact that partitioning the orientation image in regions, characterized by homogeneous orientations, implicitly reveals the position of singularities. The borderline between two adjacent regions is called a fault line. By noting that fault lines converge towards loop singularities and diverge from deltas, the authors define a geometrical method for determining the convergence and divergence points.

Nilson et al. [14] identify singular points by their symmetry properties. In particular this is done using complex filters, which are applied to the orientation field in multiple resolution scales. The detection of possible singularities is done by analysing the response created by these complex filters.

### **15.1.2 Model-Based Detection of Singularities**

In [16] Rao et al. proposed a novel algorithm for singular point detection in flow fields. Their main idea is to locally approximate a flow pattern by a two-dimensional linear differential equation. This allows a parametric representation of different types of phase portraits, and their classification is possible based on the extracted parameters.

In this chapter, we present a novel method for the detection of singularities based on the work of Rao et al. In comparison, our method is robust against noise in

the input image and is able to detect singularities even if they are partly occluded. Additionally, we present methods for detection and recognition of all types of singularities in fingerprint images, whereas Rao et al. presented a model for vortices only. This model-based attempt is new to the field of fingerprint singularity detection.

### 15.1.3 Outline

In Section 15.2 an explanation of linear phase portraits is given. Furthermore, the fitting of the parameters is explained in detail. We analyse the weaknesses of existing algorithms and propose a robust parameter fitting method.

In Section 15.3 we explain how this algorithm can be applied to fingerprint images.

Section 15.4 shows the conducted experiments. In the first part, we demonstrate the noise and occlusion robustness of our algorithm. Furthermore, we have tested our algorithm on  $2 \times 280$  hand-labelled images in order to demonstrate the singular point detection capabilities.

Finally, in the last section a summary of the proposed method is given.

## 15.2 Two-Dimensional Linear Phase Portraits

Phase portraits are a powerful mathematical model for describing oriented textures and, therefore, have been applied by many authors [3, 16, 19]. Linear phase portraits can be expressed by the following differential equation system:

$$\frac{dx}{dt} = \dot{x} = p(x, y) = cx + dy + f \quad \frac{dy}{dt} = \dot{y} = q(x, y) = ax + cy + e \quad (15.1)$$

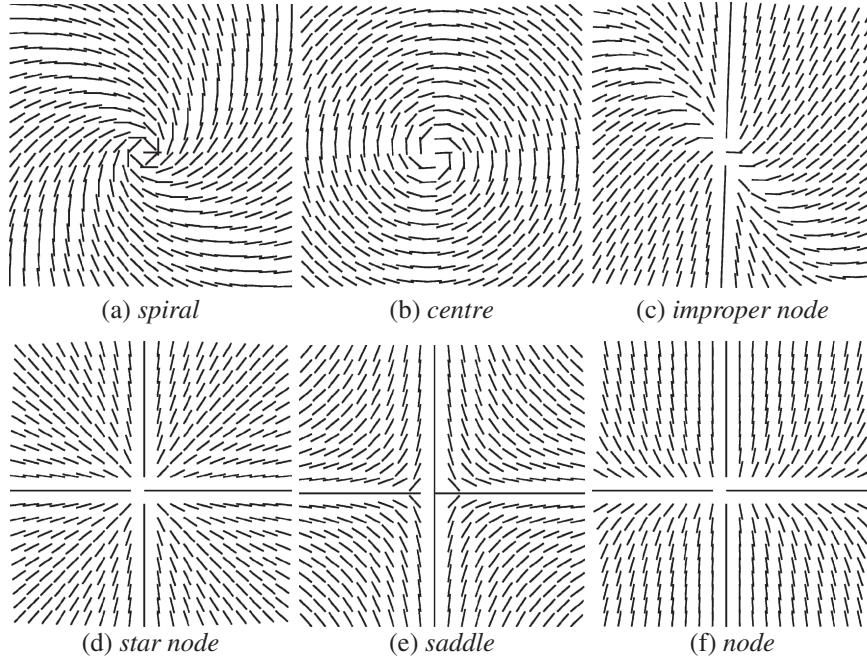
By varying the parameters of these equations we can describe a set of oriented textures comprising saddles, star nodes, nodes, improper nodes, centres and spirals [9] (examples are given in Fig. 15.1). The orientation of these fields is given by

$$\phi(x, y) = \text{atan} \left( \frac{dy}{dx} \right) = \text{atan} \left( \frac{\dot{y}}{\dot{x}} \right) = \text{atan} \left( \frac{ax + by + e}{cx + dy + f} \right) \quad (15.2)$$

Equation (15.1) can further be represented in a more convenient matrix notation as

$$\dot{\mathbf{X}} = \mathbf{A} * \mathbf{X} + \mathbf{B} \quad \text{where } \mathbf{X} = \begin{bmatrix} x \\ y \end{bmatrix}, \mathbf{A} = \begin{bmatrix} c & d \\ a & b \end{bmatrix}, \mathbf{B} = \begin{bmatrix} f \\ e \end{bmatrix} \quad (15.3)$$

where  $\mathbf{A}$  is called the *characteristic matrix* of the system. A point at which  $\dot{x}$  and  $\dot{y}$  are zero is called a *critical point*  $(x_0, y_0)$  [9]. The elements of the characteristic matrix are used to determine between six flow patterns [15]. The type of the flow pattern is determined by the eigenvalues of the characteristic matrix.



**Fig. 15.1** A classification of different phase portraits based on the characteristic matrix  $\mathbf{A}$  [15]. Complex eigenvalues result in a spiral (a) or a centre (b)-type pattern, differentiated from each other only by the real part of the eigenvalues. If the eigenvalues are real and both equal, the pattern can be classified into a start node (d) or into an improper node (c). Miscellaneous real-valued eigenvalues result in a saddle (e) or a node (f), only distinguished by their signs

### 15.2.1 Parameter Estimation

Rao and Jain [15] presented an algorithm for parameter estimation of linear phase portraits. However, the non-linear least squares computation required in their original algorithm is computationally expensive and prone to local minima. In [18], Shu et al. presented a linear formulation of an algorithm which computes the critical points and parameters for a two-dimensional phase portrait. Because their approach is linear, there exists a closed form solution. Recent applications of the algorithm can be seen in [10, 21]. Throughout this chapter we refer to this algorithm as the linear least squares algorithm. In the following section we give a brief introduction of the algorithm presented by Shu et al. in [18, 19].

To solve the problem one can apply a least squares algorithm. Equation (15.2) can be expressed as

$$p(x_i, y_i) - \tan\phi_i * q(x_i, y_i) = 0 \quad (15.4)$$

We can directly estimate the parameters by using the triplet data  $(x_i, y_i, \tan\phi_i)$  and (15.4), where  $(x_i, y_i)$  is the coordinate of a pixel and  $\tan\phi_i$  the observed data.

Let  $\tan\phi_i = \zeta$ ; The optimal weighted least square estimator is one that minimizes the following cost function:

$$\sum_{i=0}^n \omega_i^2 \cdot [p(x_i, y_i) - \zeta_i \cdot q(x_i, y_i)]^2 \quad (15.5)$$

which can be rewritten as

$$\sum_{i=0}^n \omega_i^2 \cdot [ax_i + by_i - \zeta_i cx_i - \zeta_i dy_i + e - \zeta_i f]^2 \quad (15.6)$$

which is subject to the constraint  $\sqrt{a^2 + b^2 + c^2 + d^2} = 1$ . Here  $\omega_i = \cos\phi_i$  and is applied because the tangent function is not uniformly sensitive to noise, so each observed data has to be weighted by the inverse of the derivative of the tangent function.  $n$  is the total number of triplet data used to estimate the parameter set  $(a, b, c, d, e, f)$ .

Let

$$L_4 = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}, \quad L_2 = \begin{bmatrix} e \\ f \end{bmatrix}, \quad \Omega_2 = \begin{bmatrix} \omega_0 - \zeta_0 \omega_0 \\ \omega_1 - \zeta_1 \omega_1 \\ \omega_2 - \zeta_2 \omega_2 \\ \vdots \\ \omega_n - \zeta_n \omega_n \end{bmatrix} \quad (15.7)$$

and

$$\Omega_4 = \begin{bmatrix} x_0 \omega_0 & x_0 \omega_0 & \zeta_0 x_0 \omega_0 & \zeta_0 y_0 \omega_0 \\ x_1 \omega_1 & x_1 \omega_1 & \zeta_1 x_1 \omega_1 & \zeta_1 y_1 \omega_1 \\ x_2 \omega_2 & x_2 \omega_2 & \zeta_2 x_2 \omega_2 & \zeta_2 y_2 \omega_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_n \omega_n & x_n \omega_n & \zeta_n x_n \omega_n & \zeta_n y_n \omega_n \end{bmatrix} \quad (15.8)$$

Now we can express the previous constrained optimization as minimizing the cost function:

$$C = (\Omega_4 L_4 + \Omega_2 L_2)^T (\Omega_4 L_4 + \Omega_2 L_2) + \lambda(L_4^T L_4 - 1) \quad (15.9)$$

Differentiating  $C$  with respect to  $L_4$ ,  $L_2$  and to the Lagrangian multiplier  $\lambda$  and setting the derivatives to zero, we obtain

$$\begin{aligned} \frac{\partial C}{\partial L_4} &= 2\Omega_4^T \Omega_4 L_4 + 2\Omega_4^T \Omega_2 L_2 + 2\lambda L_4 = 0 \\ \frac{\partial C}{\partial L_2} &= 2\Omega_2^T \Omega_2 L_2 + \Omega_2^T \Omega_4 L_4 = 0 \\ \frac{\partial C}{\partial \lambda} &= L_4^T L_4 - 1 = 0 \end{aligned}$$

which yields

$$L_4^T L_4 = 1$$

$$L_2 = -(\Omega_2^T \Omega_2)^{-1} \Omega_2^T \Omega_4 L_4 \quad (15.10)$$

$$\psi L_4 = \lambda L_4 \quad (15.11)$$

where

$$\psi = -\Omega_4^T \Omega_4 + \Omega_4^T \Omega_2 (\Omega_2^T \Omega_2)^{-1} (\Omega_2^T \Omega_4). \quad (15.12)$$

$L_4$  is an eigenvector of the symmetric matrix  $\psi$  and  $\lambda$  is its eigenvalue. Therefore, the eigenvector with the smallest absolute eigenvalue gives the best estimation of  $L_4$ . We can further compute  $L_2$  by using (15.10).

### 15.2.2 Algorithm Analysis

In [19], Shu et al. presented a detailed analysis of their algorithm. From this analysis and our own experiments (see Section 15.4) two conclusions can be drawn:

1. The presented algorithm works well in the case of Gaussian distributed noise. In the presence of occlusions, the algorithm may fail to extract the correct parameters.
2. The method has non-uniform sensitivity to noise, depending on the position of the point. The sensitivity in regions close to the singular point is low, whereas the sensitivity in regions away from the singular point is increased.

### 15.2.3 RANSAC-Based Approach

Although the roots of the linear phase portrait estimation algorithm can be tracked back to the year 1990 [18], only recently several authors applied this algorithm in their work. For example in [10], the authors applied this method in order to extract a high level description of fingerprint singularities and direction fields thereof. As mentioned above, there are conceptional weaknesses adhered to this algorithm. In order to improve the robustness of the original algorithm, we propose a random sample consensus (RANSAC) [2]-based approach for parameter fitting. The RANSAC algorithm is an iterative method to estimate parameters of a mathematical model from a set of observed data which contain outliers. Random sampling and consensus has been applied to a wide range of problems including fundamental matrix estimation, trifocal tensor estimation, camera pose estimation and structure from motion [4]. This algorithm has proven to give better performance than various other robust estimators.

The application of this robust algorithm to fitting parameters of linear phase portraits can be described as in the following:

1. Randomly select six triplet data points  $(x, y, \zeta)$  from the oriented texture and compute the model parameters using the linear least squares algorithm (as described in Section 15.2.1).
2. Verify the computed model by using a voting procedure. Every pixel lying within a user, given threshold  $t$ , increases the vote.
3. If the vote is high enough, accept fit and exit with success.
4. Repeat 1–3 for  $n$  times.

The number of iterations  $n$  can be computed using the following formula [2]:

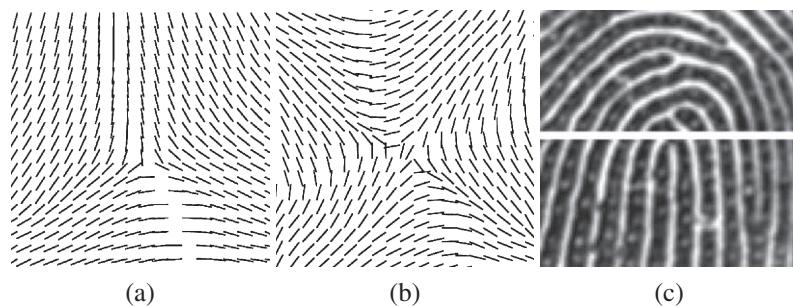
$$n = \frac{\log(1 - z)}{\log[1 - (1 - \epsilon)^m]} \quad (15.13)$$

where  $z$  is the confidence level,  $m$  is the number of parameters to be estimated and  $\epsilon$  is the outlier proportion.

### 15.3 Application to Fingerprint Singularities

Most approaches classify singular points into two types, namely in cores and deltas. The approach presented in this chapter distinguishes between three types of singular points: whorls, deltas and loops.

Describing whorl-type singularities using linear phase portraits is straightforward. Delta- and loop-type singularities need a special treatment in order to be modelled using linear phase portraits. Our approach for modelling deltas is to double the orientation and then to fit a saddle-type pattern to it. Loops are modelled in two parts. The upper part of the loop is modelled as a whorl. The second part of the loop consists of a homogeneous region (see Fig. 15.2(c), for an example). In order to compute the number of inliers for this region we compute the median of the doubled angle orientation. Every orientation which lies within a user, given threshold  $t$ , is counted as inlier.



**Fig. 15.2** The orientation fields of a delta (a). The doubled angle orientation fields can be seen in (b). The delta is modelled in the doubled angle orientation field as a saddle-type singularity. Loops (c) are modelled in two parts, the first half whorl type and the second a homogeneous region

### 15.3.1 Singularity Detection Using Sliding Window Approach

By using linear phase portraits, it is possible to model the local area around a singular point. Furthermore linear phase portraits can only describe one singular point, while a complete fingerprint impression contains at least two singular points. In order to detect multiple singular points in fingerprint images, we approach the use of a sliding window approach. The orientation in the window area is locally approximated by linear phase portraits. The size of the sliding window depends on the sensor resolution. Values for the FVC2000 database 1 and FVC2004 database 2 are  $50 \times 50$  pixels for deltas,  $80 \times 80$  pixels for whorls and  $100 \times 70$  pixel for loop-type singularities. For the sliding step we found values of  $1/3$  of the window size to be sufficient to detect every singular point. Too fine steps slow down the search process. At each position of the sliding window, parameters are fitted by the RANSAC algorithm as described in Section 15.2.3. A detection of a singular point is accepted only if the number of inliers exceeds 65%. This value is low enough in order to allow a certain number of outliers, which usually are present in noisy fingerprint images. On the other hand it is high enough to prevent random parameters to be fitted to a region and therefore causing spurious detections. The results of singularity detection algorithm are shown in Fig. 15.3.

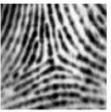
Once parameters for a given subwindow have been found, these parameters must be inspected. This inspection is done using the eigenvalues of the characteristic matrix  $\mathbf{A}$ . In general, the ratio of the two eigenvalues  $\lambda_1/\lambda_2$  expresses the aspect ratio of an oriented pattern around a given singularity. In order to prevent physically impossible parameters to be fitted, we introduce a threshold for this ratio. In Table 15.1 an overview of the classification and the thresholds is given.

Also note that because of the sliding window approach in certain cases a singularity may be detected multiple times. For combining these multiple detections to



**Fig. 15.3** The search procedure: A window of fixed size slides over the whole image in order to search for singularities. At each position of the window, parameters of a linear phase portrait are fitted. Based on these parameters it is decided if a singularity is present. Furthermore it is possible to classify the different types of singular points

**Table 15.1 Classification schema for fingerprint singularities.** The classification is done based on the extracted parameters of the linear phase portraits. Whorls are detected in the original orientation field. Detection of deltas is done in the doubled angle orientation field. In case of the loop, first a half centre is detected, followed by the detection of a homogeneous region. Thresholds are introduced in order to prevent the fitting of physically impossible parameters

Appearance	Eigenvalues	Thresholds
Whorl 	Complex eigenvalues $\lambda_1 = \Re + j\Im$ $\lambda_2 = \Re - j\Im$	$\frac{1}{3} < \frac{\lambda_1}{\lambda_2} < 3$
Delta 	Real distinct eigenvalues $\lambda_1$ and $\lambda_2$ with opposite sign	$\frac{1}{4} < \frac{\lambda_1}{\lambda_2} < 4$
Loop 	Upper part only: complex eigenvalues $\lambda_1 = \Re + j\Im$ $\lambda_2 = \Re - j\Im$	$\frac{1}{3} < \frac{\lambda_1}{\lambda_2} < 3$ $\Re < 0.2$

one single detection, we use the mean shift algorithm as described in [1]. The mean shift algorithm is a non-parametric technique that locates density extrema or modes of a given distribution by an iterative procedure.

## 15.4 Experimental Results

### 15.4.1 Parameter Fitting Examples

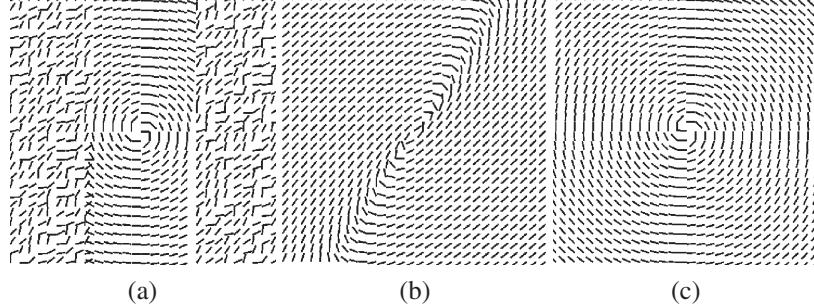
In the following section, we give an illustrative overview of the parameter fitting capability of our algorithm. Therefore, we use synthetic orientation fields (Fig. 15.4) and patches taken from real fingerprint images (Fig. 15.5).

The illustration in Fig. 15.4 shows that our robust model-based algorithm can extract the correct parameters even in the case of noisy data. In Fig. 15.5 we want to emphasize on the uniform sensitivity of our algorithm. As the aim is to extract the singular point position as accurately as possible, this property of the algorithm is essential.

The estimation of the orientation field is accomplished by a gradient-based algorithm as described in [17].

### 15.4.2 Comparison with Shu et al.

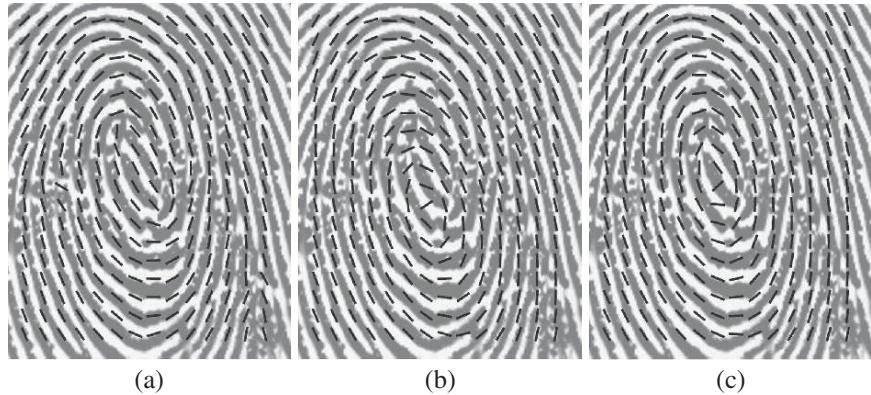
In order to compare the robustness of our method with the algorithm of Shu et al. [18], we have extracted 100 image patches from the FVC2004 database 2a [13]. These patches have a size of  $150 \times 150$  pixels (as shown in Fig. 15.6, with the singular point centred) and represent whorl-type singularities. All singular point positions



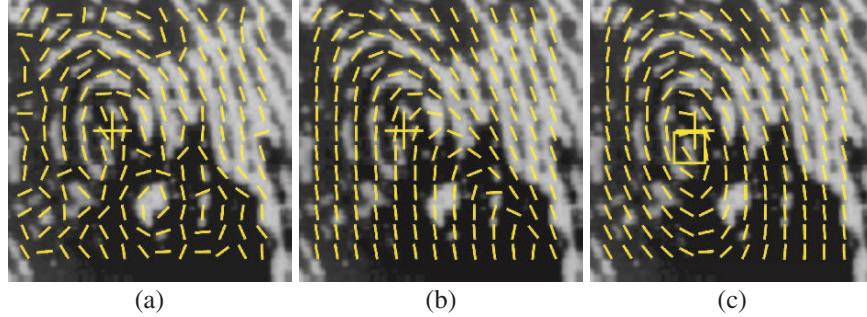
**Fig. 15.4 Occlusion.** In this example we illustrate the robustness of the two algorithms against occlusions. First an orientation field of centre type has been created. On both sides of the orientation field an occlusion has been simulated by replacing the original values by random values (a). While the original algorithm of Shu et al. [18] fails (b), our new approach (c) extracts the parameters precisely

have been labelled manually. While the proposed algorithm classified every patch correctly, the original algorithm of Shu et al. failed in one case.

The average distance between labelled and detected singular points is 11.3 pixels for our algorithm and 19.2 pixels for the method of Shu et al. The standard deviation for our algorithm is 4.2 pixels, while being 5.6 pixels for the algorithm of Shu et al. From this experiment, it can clearly be seen that the algorithm as proposed by Shu et al. has a major drawback concerning precision. This is due to the lower sensitivity of this algorithm around the singular point and a higher sensitivity away from the singular point.



**Fig. 15.5 Uniform sensitivity.** (a) The orientation has been extracted using a gradient-based method. (b) One can see the fitting result of the linear least squares algorithm as proposed by Shu et al. and applied in many recent papers. (c) The fitting results of our proposed method. The uniform sensitivity property of our algorithm is essential for the precise localization of singular points



**Fig. 15.6** A comparison of 100 image patches using our robust approach and the original algorithm of Shu et al. This experiment shows that our algorithm is not only more robust, but also more precise in locating the correct position of the singular point (manually labelled position shown as ‘+’ and detected as ‘□’). (c) The orientation field has been extracted using a gradient-based method. Figure (b) The fitting results of the linear least square algorithm. (a) The results of our robust approach

### 15.4.3 Singular Point Detection in Fingerprint Images

In this section we present results of the detection and recognition algorithm. For performance evaluation we use the first 280 images from the FVC2000 database 1a [12] and the FVC2004 database 2a [13]. The true positions of the singular points have been annotated manually. Using this ground truth data, the detections of the algorithm can be categorized into three types. These types are true positive (TP), false positive (FP) and false negative (FN). The maximum distance for classifying a detection as TP was set to a value of 25 pixels. Detections which are closer than 10 pixels to the segmentation border are discarded. The iterations for the RANSAC procedure are 150 and the threshold for counting an orientation as inlier is  $8^\circ$ . From the above manual classification, we can compute the following figures:

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Recall} = \frac{TP}{TP + FN} \quad (15.14)$$

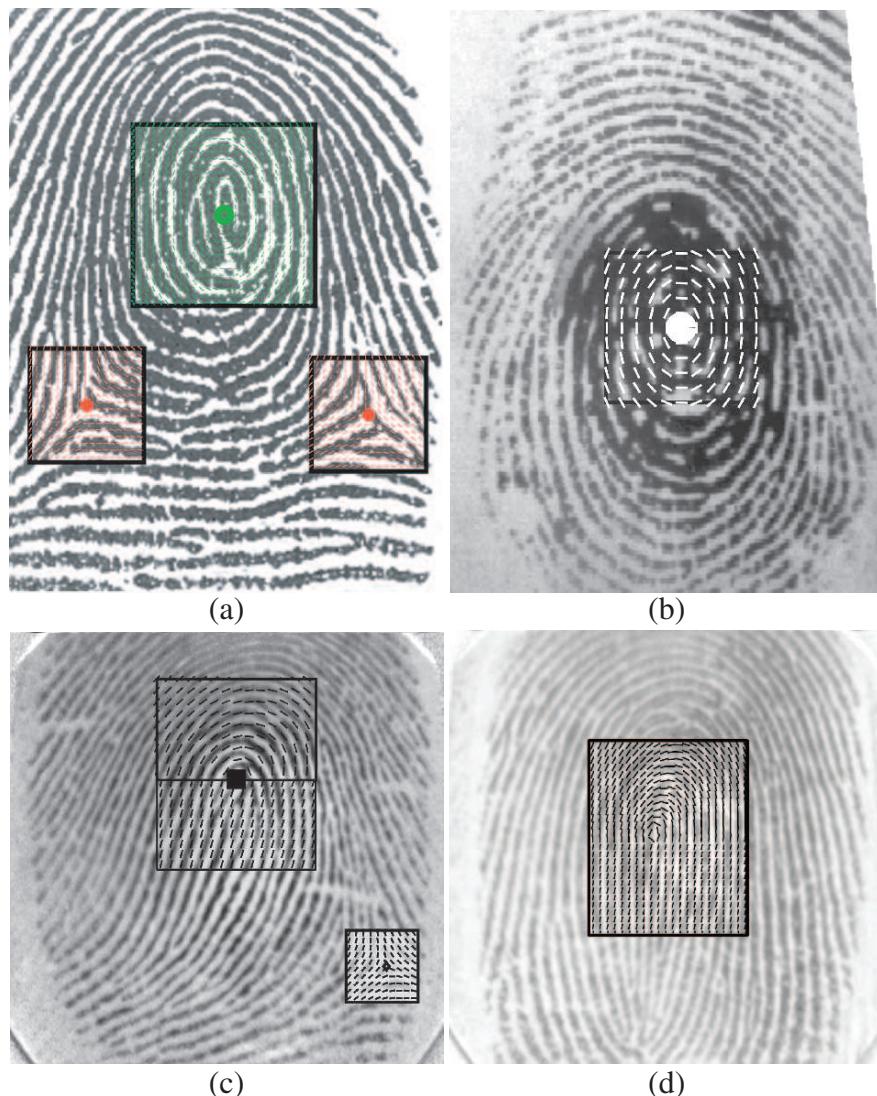
For the FVC2000 database 1a, we achieve a precision of 99.4% and a recall figure of 96.0%. See Table 15.2 for more details on the different types of singularities. For the FVC2004 database 2a, the algorithm achieves a precision rate of 95.4% and a recall rate of 96.4%. More details are given in Table 15.3. Detected singular points and corresponding sliding windows for different types of fingerprints are shown in Fig. 15.7.

**Table 15.2** Results on the FVC2000 database 1a using the first 280 images (fingers 1–35, eight impressions per finger)

Type	True positive	False positive	False negative	Precision (%)	Recall (%)
Delta	94	1	7	98.9	93.0
Whorl	78	0	1	100	98.7
Loop	193	1	7	99.5	96.5
Total	365	2	15	99.4	96.0

**Table 15.3** Results on the FVC2004 database 2a using the first 280 images (fingers 1–35, eight impressions per finger)

Type	True positive	False positive	False negative	Precision (%)	Recall (%)
Delta	89	1	6	98.8	93.6
Whorl	55	7	1	88.7	98.2
Loop	233	10	7	98.8	93.7
Total	377	18	14	95.4	96.4



**Fig. 15.7** The detected singular points with the corresponding sliding windows. The orientation has been reconstructed using linear phase Portraits. **(a)** A whorl-type fingerprint with the detected whorl- and delta-type singularities. A noisy whorl type can be seen in **(b)**, while in **(c)** it is shown how the sliding windows fit a loop-type singularity. **(d)** A wrong detection is shown. The displayed fingerprint is of whorl type, while the algorithm detected a loop-type singularity

## 15.5 Conclusion

We presented a model-based method for singular point detection in fingerprint images. Our proposed method is robust to noise and occlusions in the input image. The algorithm proceeds by fitting linear phase portraits at each location of a sliding window and then analyses its parameters. Using a well-established mathematical background, our algorithm is able to decide if a singular point is existent. Furthermore, the parameters can be used to classify the type of the singular point into whorls, deltas and loops.

We performed several tests on synthetic and natural images in order to point out the mentioned capabilities of our algorithm. We also showed how the algorithm is able to reconstruct the orientations near singular points. The final evaluation of our algorithm is done on a data set of  $2 \times 280$  images from publicly available fingerprint images. This evaluation attests our algorithm as an excellent singular point detection capability.

Future work includes merging our model-based approach with existing numerical methods.

**Acknowledgments** This work has been funded by the Biometrics Center of Siemens IT Solutions and Services, Siemens, Austria.

## References

1. D. Comaniciu and P. Meer. Mean shift analysis and applications. *IEEE Int. Conf. Computer Vision (ICCV'99), Kerkyra, Greece*, 1197–1203. Published by IEEE Computer Society Press. 2: 1197–1203, 1999.
2. M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, June 1981.
3. R. M. Ford and R. N. Strickland. Nonlinear phase portrait models for oriented textures. *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR '93., 1993 IEEE Computer Society Conference on*, pages 644–645, 1993.
4. R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004.
5. A. K. Jain, L. Hong, S. Pankanti, and R. Bolle. An identity-authentication system using fingerprints. *Proceedings of the IEEE*, 85(9):1365–1388, 1997.
6. A. K. Jain and K. Karu. Learning texture discrimination masks. *IEEE Transactions on Pattern Analysis Machine Intelligence*, 18(2):195–205, 1996.
7. A. K. Jain and D. Maltoni. *Handbook of Fingerprint Recognition*. Springer-Verlag, New York, Inc., Secaucus, NJ, USA, 2003.
8. X. Jiang, M. Liu, and A. C. Kot. Reference point detection for fingerprint recognition. *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 1:540–543, 2004.
9. E. Kreyszig. *Advanced Engineering Mathematics*. 6th edition, Wiley, New York, 1988.
10. J. Li, W.-Y. Yau, and H. Wang. Constrained nonlinear models of fingerprint orientations with prediction. *Pattern Recognition*, 39(1):102–114, January 2006.
11. D. Maio and D. Maltoni. A structural approach to fingerprint classification. *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, 3:578–585, 1996.

12. D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. Fvc2002: Second finger-print verification competition. *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, 3:811–814, 2002.
13. D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. Fvc2004: Third finger-print verification competition. In D. Zhang and A. K. Jain, editors, *ICBA, Lecture Notes in Computer Science*, 3072: 1–7. Springer, New York 2004.
14. K. Nilsson and J. Bigun. Localization of corresponding points in fingerprints by complex filtering. *Pattern Recognition Letters*, 24(13):2135–2144, September 2003.
15. A. R. Rao and R. Jain. Analyzing oriented textures through phase portraits. *Pattern Recognition, 1990. Proceedings of the 10th International Conference on*, 1:336–340, 1990.
16. A. R. Rao and R. C. Jain. Computerized flow field analysis: oriented texture fields. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(7):693–709, 1992.
17. A. R. Rao and B. G. Schunck. Computing oriented texture fields. *Computer Vision and Pattern Recognition, 1989. Proceedings CVPR '89., IEEE Computer Society Conference on*, pages 61–68, 1989.
18. C. F. Shu, R. Jain, and F. Quek. A linear algorithm for computing the phase portraits of oriented textures. *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, pages 352–357, 1991.
19. C.-F. Shu and R. C. Jain. Direct estimation and error analysis for oriented patterns. *CVGIP: Image Underst.*, 58(3):383–398, November 1993.
20. S. Wei, C. Xia, and J. Shen. Robust detection of singular points for fingerprint recognition. *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, 2:439–442, 2003.
21. W.-Y. Yau, J. Li, and H. Wang. Nonlinear phase portrait modeling of fingerprint orientation. *Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th*, 2:1262–1267, 2004.

# Chapter 16

## Frequency-Based Fingerprint Recognition

Gualberto Aguilar, Gabriel Sánchez, Karina Toscano, and Héctor Pérez

**Abstract** Fingerprint recognition is one of the most popular methods used for identification with greater success degree. Fingerprint has unique characteristics called minutiae, which are points where a curve track ends, intersects, or branches off. In this chapter a fingerprint recognition method is proposed in which a combination of fast Fourier transform (FFT) and Gabor filters is used for image enhancement. A novel recognition stage using local features for recognition is also proposed. Also a verification stage is introduced to be used when the system output has more than one person.

### 16.1 Introduction

The biometry or biometrics refers to the automatic identification (or verification) of an individual (or a claimed identity) by using certain physiological or behavioral traits associated with the person. Traditionally, passwords (knowledge-based security) and ID cards (token-based security) have been used to moderate access to restricted systems. However, security can be easily breached in these systems when a password is divulged to an unauthorized user or when an impostor steals a card. Fingerprints are fully formed at about 7 months of fetus development and finger ridge configurations do not change throughout the life of an individual except due to accidents such as bruises and cuts on the fingertips [1]. Unimodal biometric systems perform person recognition based on a single source of biometric information. Such systems are often affected by the following problems: noisy sensor data, non-universality, lack of individuality, lack of invariant representation, and susceptibility to circumvention. In fact, at present some investigators are studying the vulnerability of biometric systems [2]. In this chapter, we have decided to make a system with more than a biometric characteristic. The results show that the multibiometric systems are more reliable than the biometric systems (only with one characteristic) [3]. There are different architectures in a multibiometric system [4]. This chapter proposes a system with multiple units (i.e., two fingerprints).

---

G. Aguilar (✉)  
SEPI ESIME Culhuacan, Instituto Politécnico Nacional, Av. Santa Ana 1000, México D.F.  
e-mail: gualberto@calmecac.esimecu.ipn.mx

## 16.2 Proposed System

The proposed system uses two fingerprints, that of the right thumb and the left thumb. Each fingerprint is separately processed and in the last stage, the individual results are compared to give a single result. The recognition process is divided into two stages. The first stage is a combination of two algorithms for image enhancement using FFT and Gabor filters to reconstruct the image's information. The second stage performs the fingerprint recognition process; however, sometimes two or three images result, therefore, a verification stage is required.

### 16.2.1 Data Acquisition

The fingerprint acquisition is made using an optical device UareU 4000 from Digital Persona Inc. with a USB 2.0 interface.

The images were captured with a 512 DPI resolution and a size of 340 by 340 pixels in gray scale. In this research, a database with 1000 fingerprints from 50 different people was used, using 10 images per thumb.

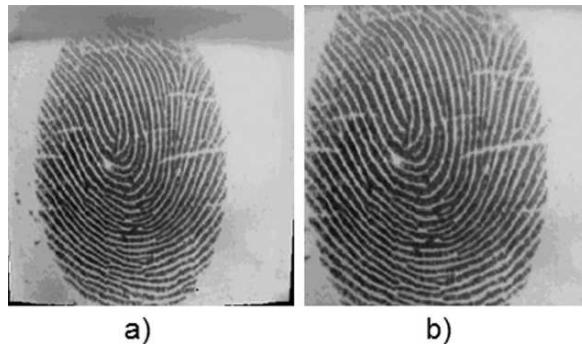
### 16.2.2 Preprocessing Stage

Most of the fingerprint images are distorted in the zones near to the image borders. This distortion can be caused by different factors like finger movement during the capture time or little pressure in the lateral areas from the scanner. This distortion must be eliminated to assure that only useful information will be processed at the



**Fig. 16.1** (a) Optical reader used. (b) Images obtained with optical reader. (c) Images from FVC2002

**Fig. 16.2** (a) Original image.  
(b) Same image in (a) after distortion correction



time of minutiae's extraction. In case these distortions are not eliminated, the algorithm could detect false minutiae. Therefore, about 10% of the image was cut from each side ensuring that useful information from the fingerprint was not removed. Figure 16.2 provides an example of this operation.

### 16.2.3 Fingerprint Enhancement

The performance of minutiae extraction algorithms and other fingerprint recognition techniques relies heavily on the quality of the input fingerprint images. In an ideal fingerprint image, ridges and valleys alternate and flow in a locally constant direction. In such situations, the ridges can be easily detected and minutiae can be precisely located in the image. However, in practice, due to skin conditions, sensor noise, incorrect finger pressure, and a significant percentage of fingerprint images are of low quality. The goal of an enhancement algorithm is to improve the clarity of the ridge structures in the recoverable regions and mark the unrecoverable regions as too noisy for further processing. The filters may be defined in the spatial or in the Fourier domain. In this chapter a combination of filters in the two domains is used.

*Spatial Domain Filtering:* O'Gorman et al. [5] proposed the use of contextual filters for fingerprint image enhancement for the first time. They used an anisotropic smoothening kernel whose major axis is oriented parallel to the ridges. For efficiency, they recomputed the filter in 16 directions. The filter increases the contrast in a direction perpendicular to the ridges while performing smoothening in the direction of the ridges. Gabor filters have important signal properties such as optimal joint space frequency resolution [6]. Gabor elementary functions form a very intuitive representation of fingerprint images since they capture the periodic, yet non-stationary nature of the fingerprint regions. The even symmetric Gabor functions have the following general form (16.1):

$$G(x, y) = \exp \left\{ -\frac{1}{2} \left[ \frac{x^2}{\delta_x^2} + \frac{y^2}{\delta_y^2} \right] \right\} \cos(2\pi f x) \quad (16.1)$$

Here  $f$  represents the ridge frequency and the choice of  $\delta_x^2$  and  $\delta_y^2$  determines the shape of the filter envelope and also the trade-off between enhancement and spurious artifacts.

*Fourier Domain Filtering:* Sherlock et al. [7] performed contextual filtering completely in the Fourier domain. Each image was convolved with precomputed filters of the same size as the image. Watson proposed another approach for performing enhancement completely in the Fourier domain. Here the image is divided into overlapping blocks and in each block, the image is obtained by

$$I_{enh}(x, y) = FFT^{-1} \{ F(u, v) |F(u, v)|^k \} \quad (16.2)$$

$$F(u, v) = FFT(I(x, y)) \quad (16.3)$$

From the previous paragraphs, it is clear that both approaches present desirable features that can be combined to obtain better image enhancement results. The combination strategy is illustrated in Fig. 16.3.

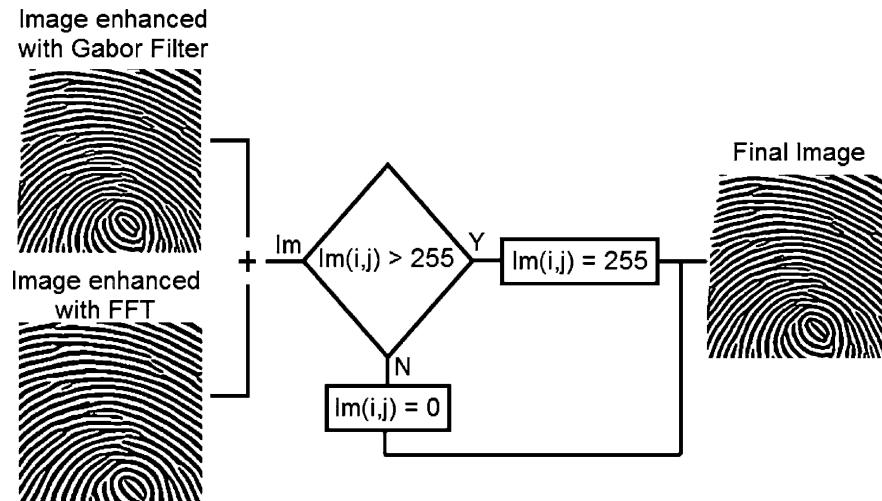


Fig. 16.3 Combination process

#### 16.2.4 Thinning

Before minutiae extraction stage, a thinning process [8] is applied, that is, an algorithm where the result is an image with minimum possible thickness lines. In order to understand the algorithm better, it is necessary to know some definitions. Let us remember that after the binarization process the image contains binary values 1 and 0, where a 1 means a white pixel and a 0 a black pixel. A pixel  $O(x, y)$  is internal, if its four neighbors  $(x + 1, y)$ ,  $(x - 1, y)$ ,  $(x, y + 1)$ , and  $(x, y - 1)$  are 0 (black pixel). A limit pixel is defined using its eight connections. A pixel is a pixel limit if it is not an internal pixel and at least one of its eight neighbors is a

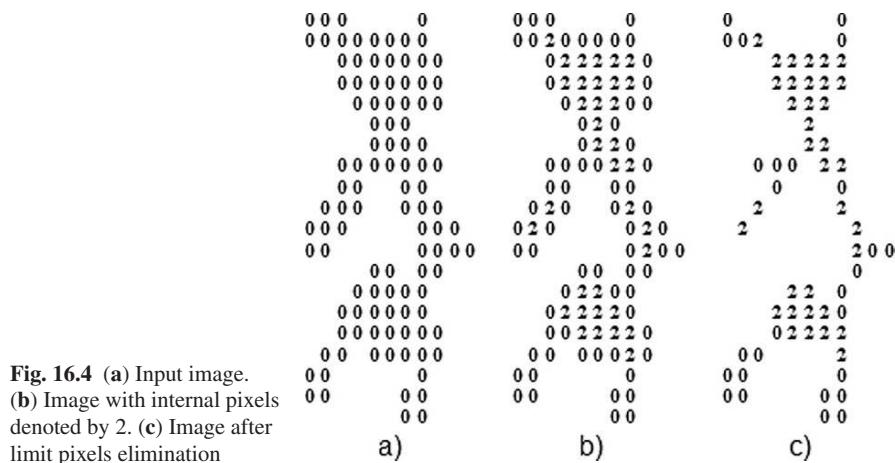
1. A connection pixel is one that cannot be eliminated in a 3 by 3 matrix without disconnecting a line made with at least 3 pixels. This process is shown in Fig. 16.1. The thinning algorithm consists of finding internal pixels in the image and later eliminating the pixel limit. This process is carried out until it is not possible to find more internal pixels and it is explained in greater detail. The first algorithm stage consists to find the total internal pixels that exist in the image. Later, all limit pixels that are not connected pixels are removed. This algorithm is repeated until no more internal pixels are found. After thinning the image and not finding more internal pixels, the algorithm is applied again but in this occasion with a small change. This change consists in finding internal pixels only with three neighbor pixels.

The last step is the repetition of the algorithm again but in this occasion to find internal pixels with only two neighbors. Considering the elimination of an internal pixel it is not possible to eliminate some neighbor pixel. The results of these operations are illustrated in Figs. 16.4 and 16.5.

### 16.2.5 Minutiae Detection

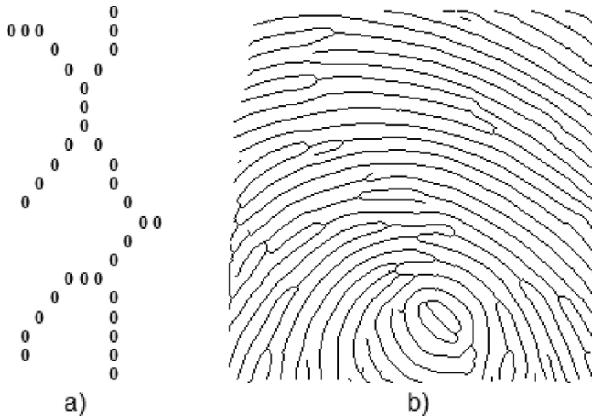
After thinning process, minutiae detection algorithm is applied which consists in calculating the pixel number crossing the center pixel ( $P_c$ ). This process is carried out at full binary image applying 3 by 3 window as follows:

$$P_c = \sum_{i=1}^8 p(i) \quad \text{if} \begin{cases} P_c = 7 & \text{TERMINATION MINUTIAE} \\ P_c = 6 & \text{BLOCK WITHOUT MINUTIAE} \\ P_c \leq 5 & \text{BLOCK WITH BIFURCATION} \end{cases} \quad (16.4)$$

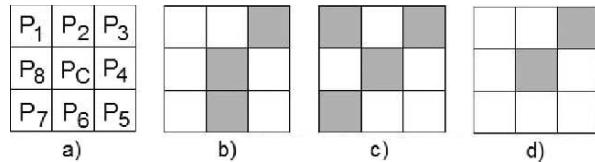


**Fig. 16.4** (a) Input image.  
 (b) Image with internal pixels denoted by 2. (c) Image after limit pixels elimination

**Fig. 16.5** Image obtained using thinning process. (a) Example. (b) Original fingerprint



**Fig. 16.6** (a) Window of  $3 \times 3$  used to find minutiae. (b) Block without minutiae. (c) Block with bifurcation. (d) Block with ending



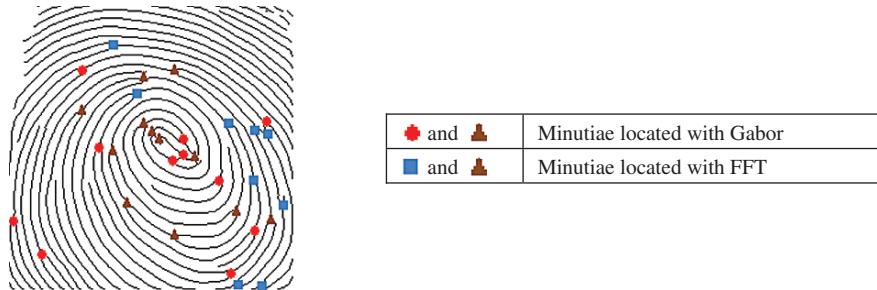
where  $P_1$  to  $P_8$  is an ordered sequence of pixels that define the block of eight neighbors of the center pixel.

In Fig. 16.6a the configuration of the used window to locate bifurcations and ending is shown. Figure 16.6b–d shows the possible configurations that we can find. A  $P_C=7$  means that we are on a window with ending. A  $P_C=6$  means that there is no bifurcation or ending.  $P_C \leq 5$  means that a bifurcation is found. This process is applied to the hold image. The result of this process is a vector with the characteristic points that will be used later in the recognition or verification stage. This process is shown in Fig. 16.7.

Figure 16.8 shows that the combination of Gabor filters with fast Fourier transform helps to detect a greater amount of minutiae. This is the reason why this chapter



**Fig. 16.7** Minutiae extracted by applying the technique based on (a) Gabor filters; (b) fast Fourier transform; (c) combination of FFT and GF



**Fig. 16.8** Final image with the total of minutiae

proposed a combination of two stages in image enhancement, i.e., to avoid some minutiae being eliminated during the enhancement process.

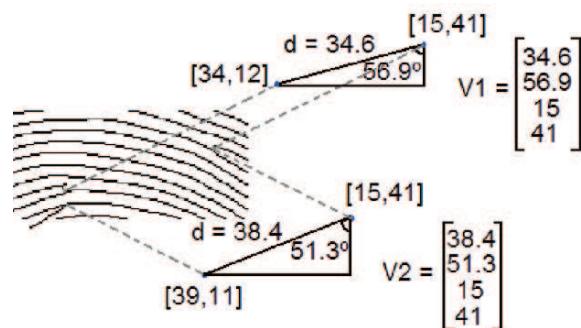
### 16.2.6 Identification from Fingerprints

The recognition process is carried out using three important characteristics: distances, angles, and coordinates. The fingerprint feature is a 500 by 4 matrix (500 vectors of  $1 \times 4$ ). Each vector has four values: minutiae coordinates ( $x, y$ ), distance to nearest minutiae, and the angle with respect to  $y$ -axis. These values are shown in Fig. 16.9.

To perform recognition, the input image is re-arranged in a  $500 \times 4$  elements array. This matrix is then compared with all arrays stored in the database.

First, equal distances are located and only those with the same angle are retained. Second, the arrays whose coordinates are very different with respect to the array under analysis are eliminated to assure a better recognition. After several tests, it was decided that the coordinate values can change around 10 pixels. With this process, it was found that 15 equal vectors give a good recognition; this means that recognition exists only when the input array contains more than 15 equal vectors to the array stored in the database.

This process is repeated with each thumb, that is, the process is made twice.



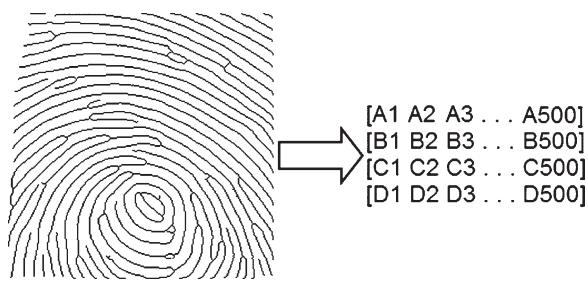
**Fig. 16.9** Minutiae feature vector

### 16.2.7 Two-Stage Recognition

In the first stage a high recognition percentage was obtained; however, the values of coordinates, distances, and angles between minutiae for some arrays in the database were 15 in more than one stored image. Therefore, for some tests the result showed more than one recognized image and the percentage of false acceptance was high. Due to this, a second test was carried out consisting of verifying the resulting images.

With the previous stage similar arrays are eliminated and only true array is accepted. In [9] a similar approach is presented. So, the verification stage is done with statistical parameters. The principal disadvantage is the processing of the complete thinned image. Therefore, the system became slow and less efficient. The verification stage proposed in this chapter is faster, is efficient, and has small complexity. This verification stage consists in analyzing the direction and the pixel numbers that were moved in the input image. When two images from the same person are compared and one of these images is moved, all minutiae from that fingerprint are moved in the same direction and by the same pixel numbers. Figure 16.10 shows this example. In this example, the input image was moved 11 pixels on the left and 33 pixels upward with respect to the storage image; therefore, all minutiae in the input image were moved by the same pixel numbers in the same directions.

When two fingerprint images of different people are compared and one of these images is moved, the pixels numbers and the direction change. In Fig. 16.11 there are two images from different people; the first minutia was moved 4 pixels on the right and 5 pixels above but the second minutia was moved 10 pixels on the right and 0 pixels on the y-axis.



A = Distances between minutiae

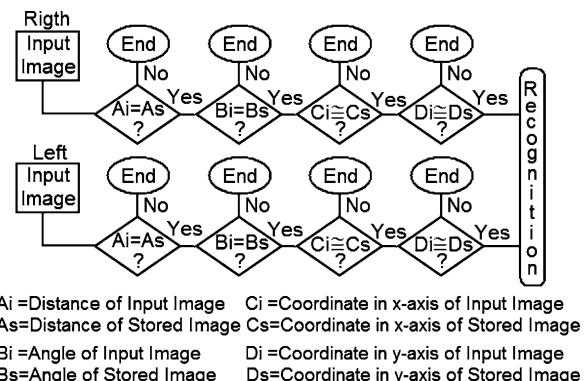
B = Angles between minutiae

C = X-axis value

D = Y-axis value

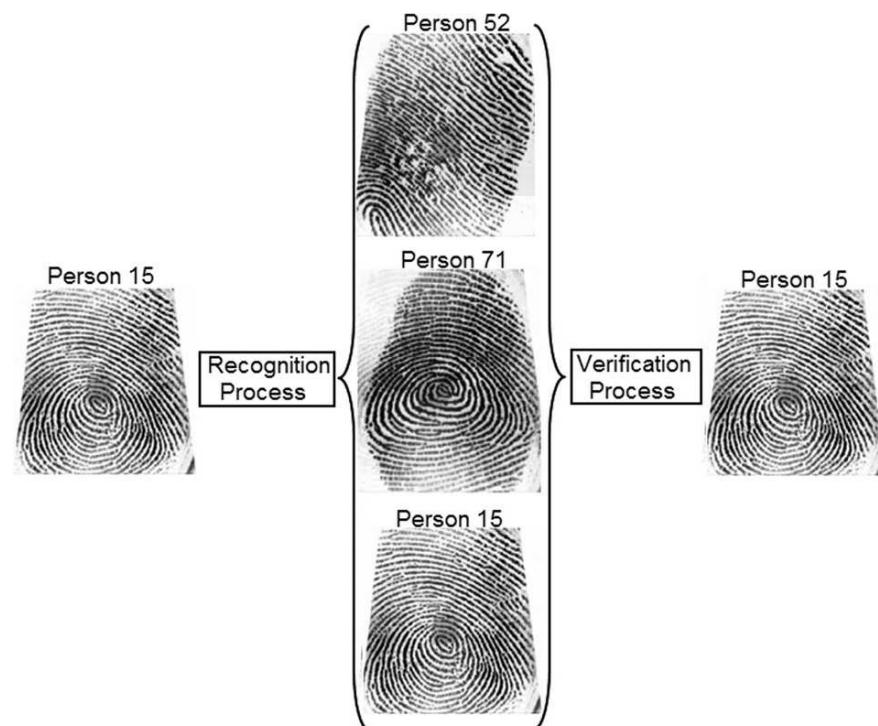
**Fig. 16.10** Matrix input image

**Fig. 16.11** Recognition process



### 16.3 Experimental Results

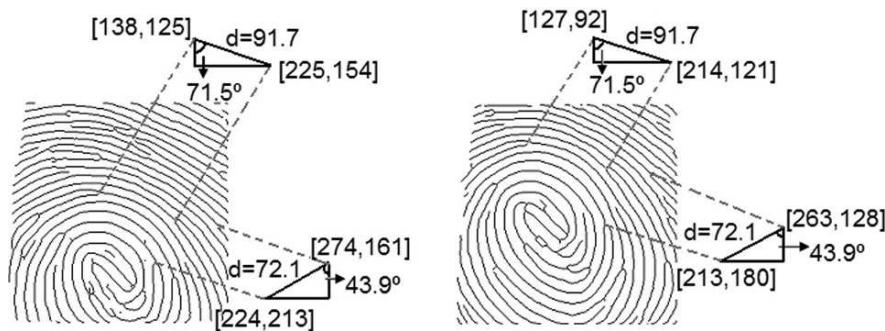
Table 16.1 shows the identification results obtained from a database of 50 subjects matching either one or two fingerprints the steps involved in identification process are shown in Fig. 16.12. In the third row the total percentage is shown. There will be a true recognition only when the recognition in the two previous stages is from



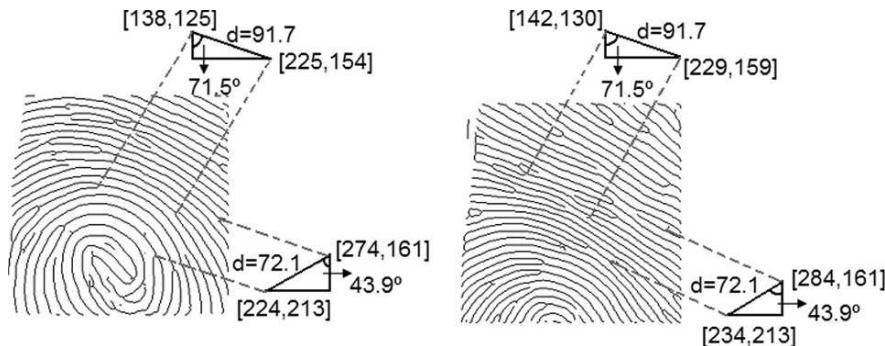
**Fig. 16.12** Identity verification process

**Table 16.1** Identification results

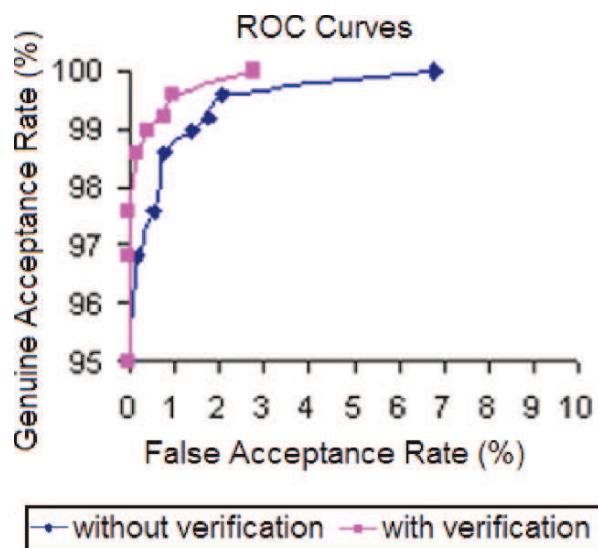
	Stage without verification		Stage with verification	
	Recognition (%)	False recognition (%)	Recognition (%)	False recognition (%)
Right thumb	97.8	1.6	98.8	0.6
Left thumb	98.0	1.4	98.8	0.6
Both thumbs	97.0	0.8	97.8	0.0
Both thumbs	97.4	1.2	98.2	0.4
Both thumbs	93.2	0.6	93.8	0.0

**Fig. 16.13** Two fingerprints from the same subject

the same person as illustrated in Fig. 16.13. The threshold used in the recognition stage was 15. In other words, at least 15 equal values are required to say that the image is from the person under analysis. Later, we made some modifications to the threshold. The fourth row shows the results of the total percentage with an acceptance threshold of 10. The fifth row shows the results of the total percentage with an acceptance threshold of 20. The distribution of minutiae in fingerprints taken from two different persons is shown in Fig. 16.14. Figure 16.15 shows the corresponding ROC curves.

**Fig. 16.14** Two fingerprints from different subjects

**Fig. 16.15** ROC curves from a database of 50 subjects



**Acknowledgments** We thank the National Science and Technology Council of Mexico for the financial support to the development of this work.

### Proposed Questions and Exercises

- What kinds of sensors do you know to capture fingerprints?
- What are the advantages and disadvantages of using one optical sensor to capture fingerprints?
- In your opinion, why are biometric systems based on fingerprints the most used in the world?
- Why is the enhancement stage in a fingerprints recognition system important?
- How many kinds of minutiae do you know?
- In your opinion, is it important to use all different kinds of minutiae in a recognition system?
- What is the most important stage in a fingerprints recognition system?
- How many classifiers of fingerprints do you know?
- What is the most used fingerprints characteristic to match the fingerprints?
- What kind of false fingerprints do you know?
- What are the advantages and disadvantages of using two different fingerprints in a recognition system?
- In your opinion do you think that the thinning process is useful to achieve a better recognition performance?

## References

1. M. Kuchen, C. Newell, A Model for fingerprint formation, *Europhysics Letters*, 68(1):141–147, 2004.
2. T. Matsumoto, H. Matsumoto, K. Yamada, S. Hoshino, Impact of artificial gummy fingers on fingerprint systems, SPIE. 4677, 2002.
3. A. K. Jain and A. Ross, Multibiometric systems, *communications of the ACM*, Special Issue on Multimodal Interfaces, 47(1):34–40, January 2004.
4. R. Brunelli, D. Falavigna, Person identification using multiple cues, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:955–966, October 1995.
5. L. O’Gorman, J. V. Nickerson, An approach to fingerprint filter design, *Pattern Recognition* 22(1):29–38, 1989.
6. S. Qian, D. Chen, Joint time-frequency analysis, *Methods and Applications*, Prentice Hall, Upper Saddle River, 1996.
7. B. G. Sherlock, D. M. Monro, K. Millard, Fingerprint enhancement by directional Fourier filtering, *Visual Image Signal Processing*, 141:87–94, 1994.
8. D. Eberly, *Skeletonization of 2D Binary Images*, Geometric Tools, Inc., June 2001.
9. G. Aguilar, G. Sanchez, K. Toscano, M. Nakano, H. Perez, Multimodal biometric system using fingerprint, International Conference on Intelligence and Advance Systems, November 2007.

# Index

## A

Access control, 13  
Active vision, 118  
Agamben, Giorgio, 301  
Aliveness detection, 297  
Ambient intelligence, 11, 20  
Applications, 11, 221  
Atmospheric conditions, 181  
Automatic Identification and Data Capture Technologies - AIDC, 304

## B

Bhattacharyya error, 285  
Binary feature selection, 253  
Bioethics, 293  
Biological diversities, 306  
Biometric sensor  
    2D face camera, 169  
    3D face, 219  
    3D fingerprint, 89  
    3D touchless fingerprinting, 89  
    at a distance, 5  
    camera calibration, 89  
    CCD retina, 119  
    contact, 5  
    contact-less, 5, 83  
    Flashscan3D fingerprint, 91  
    footsteps, 314  
    Frustrated Total Internal Reflection (FTIR), 86  
    interlace, 158  
    interoperability, 92  
    led array, 89  
    Log-polar imaging, 119  
    long-range acquisition, 170  
    low resolution, 156  
    multi-sensor systems, 278  
    multi-vision system, 89  
    out of focus, 157

Reflection-based Touchless Finger Imaging (RTFI), 85

Space-variant imaging, 119  
structured light illumination (SLI), 91  
Surround Imager, 91, 104, 106  
TBS device, 89  
Touchless Finger Imaging(TTFI), 85  
Transmission-based Touchless Finger Imaging (TTFT), 87

Vulnerability of touchless imaging, 105

Biometric sensors

    keystroke, 332  
Biometrics  
    behavioral, 313, 334  
    Data Exchange Format, 92  
    DNA, 297  
    Ear, 77  
    ethics, 298, 301, 307, 308  
    Face, 111, 155, 169, 194, 213  
    Fingerprint, 83, 349, 363  
    footsteps recognition, 313  
    Gait, 61  
    Iris, 23  
    keystroke dynamics, 329  
    Multimodal, 273  
    outdoor, 181  
    policy, 294, 295, 301  
    Soft traits, 281  
    Touchless fingerprint, 83

## C

Camera  
    Airy disk, 174  
    CCD, 173  
    CCTV, 8, 11, 13, 16  
    CMOS, 178  
    Depth of Field, 175  
    effective resolution, 174  
    Field of View, 175

- illumination, 170
- inter pupil distance (IPD), 175
- interlace artifacts, 177
- lens geometry, 174
- long-range acquisition, 170
- luminous emittance, 170
- Modulation Transfer Function (MTF), 174
- motion blur, 158, 178
- Near infrared (NIR), 12
- optimal optics, 174
- outdoor, 181
- Photo-heads, 179
- rolling shutter artifacts, 179
- set up, 170
- structured light, 219
- Capacity of a template, 275
- Cauchy-Binet kernel, 208, 216
- Chernoff error, 285
- Client-server architecture, 15
- CMC, 186
- Condensation, Kalman filter, 197
- Contextual filtering, 366
- Curvature descriptors, 228
  
- D**
- Data fusion, 9, 10, 194
- Data minimization, 295
- Data quality, 39, 169
  - measurement, 45, 184
  - quality-based fusion, 281
  - signal-to-noise ratio, 38, 184
- Databases, 16, 17
  - 3D face
    - ASU, 232
    - BJUT-3D, 232
    - Bosphorus, 232
    - BU-3DFE, 232
    - CASIA, 232
    - Extended M2VTS, 232
    - FRAV3D, 232
    - FRGC v2.0, 226, 232
    - GavabDB, 232
    - MIT-CBCL, 232
    - ND2006, 232
    - UND, 232
  - Face, 193, 194, 211
    - Banca database, 212, 214–216
    - Boğaziçi University database, 213
    - FERET, 155, 165
    - FRGC, 155, 165
    - Honda/UCSD video database, 211, 212
    - Moving faces and people (UTD), 196, 212
- Photo-heads, 180
- Fingerprint
  - FVC2000, 359
  - FVC2004, 357
  - TBS, 103
  - Touchless, 106
- Footsteps
  - EMFi Floor, 315
  - ETH Zurich, 315
  - ORL Active Floor, 315
  - Smart Floor, 315
  - Swansea University, 315
  - Ubifloor, 315
- Gait
  - CASIA, 72, 74
  - CMU, 68
  - Motion of Body (MoBo), 75
  - NIST, 68
  - University of Southampton
    - Multi-Biometric Tunnel, 76
- Iris
  - NETBCS, 29
  - U.S. Naval Academy, 23, 48
- Large scale, 16
- Density estimation, Kernel, 210, 211
  
- E**
- E-passports, 14
- Ear recognition
  - University of Southampton
    - Multi-Biometric Tunnel, 77
- Emotional state, 296
  
- F**
- Face detection, 249
- Face from video, 111, 136, 156, 195, 214, 216
  - Cauchy-Binet kernel, 208, 216
  - Condensation, 197, 204
  - Grassmann manifold, Grassmannian, 209, 210, 216
  - Hidden Markov models, Baum-Welch
    - Re-estimation, Forward-Backward procedure, 205, 215
  - Identity model, 201
  - interlace, 158
  - Kalman filter, 196, 197
  - kernel methods, 210
  - Linear Dynamic System, 207
  - low resolution, 156
  - Martin metric, 208
  - Motion model, 200
  - N4SID, 207, 216

- observability matrix, 209
  - out of focus, 157
  - Particle Filtering, 196, 198
  - PCA-ID, 207, 208
  - PH-HMM, 205, 206
  - Procrustes distance, 210, 211
  - projection matrix, 210
  - sampling, Importance Sampling, Sequential Importance Sampling, 197
  - Subspace angles, 208, 211, 214
  - tracking, 195, 196, 198–202, 211–213
  - Face recognition
    - 3D, 217
      - average face model - AFM, 225
      - challenges, 233
      - curvature descriptors, 228
      - deformable model, 229
      - feature extraction, 227
      - fusion, 228
      - Gaussian curvature, 229
      - geodesic distances, 229
      - ICP, 226
      - non-rigid registration, 226
      - part-based, 237
      - performance analysis, 230
      - registration, 225
      - subspace analysis, 227
    - Active vision, 118
    - AdaBoost, 161
    - Appearance Manifold, Shape-Illumination manifold, 214
    - Appearance model, 195, 203
    - at a distance
      - evaluation, 165
      - high quality imaging, 160
      - surveillance, 163
    - at a distance (FRAD), 111, 141, 155, 169, 195, 208, 213
    - average face, 124
    - Beijing 2008 Olympic Games, 161
    - CBSR-AuthenMetric, 161
    - cooperative system, 159
    - dynamics, 194, 196, 198, 204, 207, 215
    - eco-centric vs ego-centric representations, 128
    - eigenfaces, 249
    - elastic bunch graph matching, 132
    - expression analysis, 195
    - face acquisition at 100m, 175
    - face detection, 249, 259
    - face landmarks, 131, 223
    - face representation, 135
    - facial features, 131
  - feature reduction, 139
  - Foveal vs Peripheral processing, 118
  - frontal face imaging, 160
  - fusion of human and machine performance, 144
  - gestures, 193, 194, 206, 207
  - graph matching, 139
  - Hidden Markov models, 193, 194, 202, 214
  - human skills, 111
  - Human Visual System (HVS), 111
  - Human vs Machine performance, 141
  - illumination, 194, 195, 198–200, 211, 213
  - illumination compensation, 134
  - image acquisition, 169
  - impact of weather, 181
  - Importance Sampling, 197
  - interlace artifacts, 177
  - learning from examples, 249
  - Local Binary Pattern (LBP), 161
  - Log-polar imaging, 119
  - Machine learning
    - Lasso scheme, 252
    - learning features, 252, 258
  - Modulation Transfer Function (MTF), 174
  - morphing, 124
  - motion, 193, 194, 196, 198, 201, 202, 205, 212, 213, 215
  - motion blur, 158, 178
  - near infrared (NIR), 12, 14
  - near-distance, 155
  - neurophysiology, 111
  - neuroscience, 111
  - non negative matrix factorization, 129
  - Photo-heads, 179
  - quality, 15, 169
  - rolling shutter artifacts, 179
  - sampling, Importance Sampling, Sequential Importance Sampling, 197
  - shading, 193
  - SIFT features, 138
  - Space-variant imaging, 119
  - stereo, 193
  - stereo acquisition, 219
  - subspace methods, 128
  - Subspaces, 195, 209, 210
  - system, 159
  - video-based, 111, 155, 156, 169, 193–196, 202, 204, 205
  - viewing perspective, 159
  - Visual attention, 121
- Failure analysis, 188
- Fast Fourier Transform, 363, 365, 368
- Fingerprint recognition, 83, 349, 363

- 3D non-parametric virtual rolling, 94  
 3D parametric models, 93  
 3D parametric virtual rolling, 93  
 3D reconstruction, 89  
 3D touchless, 89  
 3D unwrapping methods, 92  
 3D virtual rolling, 92  
 camera calibration, 89  
 Camera lenses design, 87  
 contactless, touchless, 84  
 contextual filters, 365  
 fault lines, 350  
 FFT, 363, 364, 366, 368  
 finger distortion, 96  
 fingerprint categories, 350  
 fingerprint enhancement, 99, 365  
 Flashscan3D, 91  
 Frustrated Total Internal Reflection (FTIR), 86  
 Gabor filters, 363–365, 368  
 image quality, 101  
 interoperability, 92  
 latents, tenprints, 84  
 led array, 89  
 linear phase portraits, 351  
 matcher, 103  
 Minutiae detection, 367  
 multi-vision system, 89  
 noise, 354  
 occlusion, 354  
 RANSAC parameter fitting, 354  
 Reflection-based Touchless Finger Imaging (RTFI), 85  
 ridge shape, 89  
 singularities detection, 350  
 skin reflection, 85  
 skin surface, 85  
 structured light illumination (SLI), 91  
 Surround Imager, 91, 104, 106  
 sweating activity, 106  
 TBS device, 89  
 thinning, 366–368, 373  
 Touchless Finger Imaging(TTFI), 85  
 Transmission-based Touchless Finger Imaging (TTFT), 87  
 video analysis, 106  
 Vulnerability of touchless imaging, 105  
 Footstep dynamics, 313  
 Footsteps recognition, 313  
     classification, 315  
     data acquisition, 317  
     feature optimization, 318  
     ground reaction force - GRF, 314  
     Hidden Markov Model, 314  
     holistic features, 320  
     neural networks, 315  
     performance evaluation, 317, 321  
 FPFAR, 186  
 FPMDR, 186  
 FPROC, 186  
 Frobenius norm, 266  
 Function creep, 293–296, 298, 299, 301, 307
- G**
- Gabor filtering, 365  
 Gait recognition, 61  
     3D, 74  
     3D Visual Hull, 75  
     3D model, 74  
     3D skeletal model, 75  
     3D stereo, 76  
     appearance model, 73  
     binary silhouette, 68  
     camera calibration, 64, 67  
     data acquisition, 76  
     deformable model, 76  
     epipolar trajectory, 63  
     markerless estimation, 65  
     Mean Correlation Coefficient, 64  
     optical flow, 67  
     parameters, 64  
     pendulum model, 64  
     pose-based methods, 62  
     pose-free methods, 68  
     Procrustes shape analysis, 72  
     state of the art, 62  
     Units of Self-Similarity, 71  
     University of Southampton  
         Multi-Biometric Tunnel, 76  
         view invariant, 62  
 Gaussian curvature, 229  
 Geodesic distance, 229  
 Globalization, 305  
 Grassmann manifold, Grassmannian, 209, 210, 216
- H**
- Health information, 297  
 Holistic features, 320  
 Human body information, 299, 301, 302, 308  
 Human face perception, 112  
     brain activation, 112  
     brain lesions, 113  
     caricatures, 124  
     Changeable feature processing, 118  
     dynamic identity signature, 126

- EEG, 114  
electrophysiological recordings, 112  
expression processing, 113  
eye gaze direction, 117  
eye movements, 119  
face representation, 122  
Faces in motion, 117  
facial gestures, 126  
familiar vs unfamiliar faces, 126  
Foveal vs Peripheral processing, 118  
Functional Magnetic Resonance Imaging (fMRI), 114  
Functional neuroimaging, 112  
    brain reading, 115, 117  
    pattern classification, 115, 117  
    preference method, 115  
    repetition priming/suppression, 115  
    voxel labeling, 117  
Fusiform Face Area (FFA), 116  
identity processing, 113  
illusions, 112, 123  
Invariant feature processing, 118  
Lateral inferior occipital gyrus, 118  
memory, 112  
motion, 112, 126  
Neural architectures, 116  
    core system, 116  
    modular representation, 116  
Neurobiology, 113  
Neuropsychological case studies, 113  
Occipital Face Area (OFA), 117  
other race effect, 125  
PET, 114  
Prosopagnosia, 112, 113  
prototype face, 124  
representation, 112  
selective processing, 121  
Superior Temporal Sulcus (pSTS), 117  
Thatcher illusion, 123  
Visual attention, 121  
Visual fixation, 119  
Visual saccade, 121  
Visual scan-path, 119  
Human rights, 307, 308  
Human visual system, 111  
    EEG, 114  
    eye gaze direction, 117  
    eye movements, 119  
    face representation, 122  
    Functional Magnetic Resonance Imaging (fMRI), 114  
IT cortex, 114  
PET, 114  
receptive fields, 114  
superior temporal sulcus (STS), 114  
Visual attention, 121  
visual pathways, 114  
Visual saccade, 121  
Hybrid systems, 279
- I**
- Identification, 4  
Identity management, 17, 18, 308  
Informational intrusiveness, 296  
Informatization of the body, 293, 294, 301, 302, 307, 308  
Iris recognition, 22, 23  
    algorithms, 40, 47, 48  
    anatomy, 24  
    At a distance, 42, 43, 47, 49–51  
        state of the art, 49, 55  
    data acquisition, 44, 46, 47, 50  
    History, 28  
    illumination, 34–36, 40, 45, 49–51  
    image quality, 33, 39, 40, 44, 46, 47  
    iridology, 27  
    Iris pattern, 25  
    light diffraction, 33, 44–46  
    matching, 32, 33, 40  
    optics, 23, 36, 37, 50, 51, 56  
    pathologies, 26  
    photometrics, 34  
    Safety, 38, 39, 57  
    sensors, 34, 35, 37  
    Sharbat Gula, 31, 32, 55  
    signal-to-noise ratio, 38  
    state of the art, 39  
    template, 40–42, 48, 49  
    United Arab Emirates Expellee Program, 29
- Iterative closest point - ICP, 226  
Iterative feature selection, 253
- K**
- Keystroke dynamics, 329  
    behavioral, 334  
    capture devices, 332  
    classification, 331  
    distance metrics, 331  
    features, 330, 338  
    input requirements, 331  
    inter-key latency, 330  
    interface, 334  
    Random Forests algorithm, 337  
    user authentication, 337  
    user supervision, 334

**L**

- Landweber scheme, 253
- Lasso regression, 253
- Levels of fusion, 279
  - decision, 281
  - feature, 280
  - rank, 280
  - score, 280
  - sensor, 279
- Local Binary Pattern (LBP), 161
- Log-polar transform, 119

**M**

- Machine learning, 247
  - binary classification, 252
  - binary feature selection, 253
  - face features, 258
  - Frobenius norm, 266
  - iterative feature selection, 253
  - Landweber scheme, 253
  - Lasso regression, 253
  - Lasso scheme, 252
  - learning shared representations, 255
  - linear Support Vector Machine, 255
  - multi-task feature learning, 256
  - performance evaluation, 265
  - regularization parameter, 265
  - ridge regression, 252
  - Tikhonov regularization, 252
- Martin metric, 208
- Minutiae detection, 367
- Motion blur, 158, 178
- Multi-algorithm systems, 278
- Multi-Classifiers, 195
- Multi-instance systems, 278
- Multi-sample systems, 278
- Multi-sensor systems, 278
- Multi-task feature learning, 256
- Multibiometrics, 9, 228, 273
  - accuracy, 275
  - advantages, 276
  - challenges, 274
  - correlated experts, 287
  - data noise, 275
  - expert selection, 285, 287
  - fusion architecture, 287
  - fusion with soft biometric, 281
  - hybrid systems, 279
  - levels of fusion, 279
    - decision, 281
    - feature, 280
    - rank, 280
    - score, 280

**sensor, 279**

- multi-algorithm systems, 278
- multi-instance systems, 278
- multi-sample systems, 278
- multi-sensor systems, 278
- multimodal systems, 279
- multiple biometric experts, 282
- non-universality, 275
- quality-based fusion, 281
- spoof attacks, 275
- taxonomy, 277

**Multiclass problems, 247**

- Multimodal Biometrics, 273
- Multimodal systems, 279
- Multiple biometric experts, 282
- Multiple sensors, 10, 11

**N**

- Near infrared (NIR), 12

**O**

- Observability matrix, 209

**P**

- Particle Filtering, 196, 198, 202
- Password hardening, 330
- Performance analysis, 11, 141, 186
  - 3D face, 230
  - CMC, FPROC, ROC, 186
  - failure analysis from Similarity Surface theory, 188
  - footstep recognition, 317, 321
  - FPFAR, FPMDR, 186
  - FRGC, 230
  - FRVT 2006, 232
  - multi-task learning, 265
  - similarity surface theorem, 188
- Performance factors, 6
- Phase portraits, 351
- Privacy, 17–19
- Procrustes distance, 210, 211
- Projection matrix, 210
- Pseudospeciation, 306

**R**

- Random Forests algorithm, 337
- Random Sample Consensus - RANSAC, 354
- Recognition
  - one-to-many, 13
  - one-to-one, 4, 13
- Remote Biometric
  - Face, 111, 141, 155, 169, 193, 221
    - image acquisition at 100m, 175
    - impact of weather, 181

- interlace artifacts, 177
- motion blur, 158, 178
- Photo-heads, 179
- rolling shutter artifacts, 179
- Fingerprint, 83
- Gait, 61
- Iris, 23
- outdoor, 181
- RFID, 7, 13, 14, 221
- Ridge regression, 252
- ROC, 186
- S**
- 3D scanner, 219
- Security, 294, 295, 298, 307–309
- Sensing from a distance, 7
- Sensors
  - 3D scanner, 219
  - CCD, EMCCD, 173
  - CMOS, 178
  - correlation, 9, 10
  - Electro Mechanical Film - EMFi, 315
  - list of 3D scanners, 219
  - multi-sensor systems, 278
  - piezo force, 314
  - switch, 315
- T**
- Technology challenges, 6
- Template
  - reverse engineering, 298
- Terminology, 28
- Tikhonov regularization, 252
- V**
- Verification, 3