

Advances in Computer Vision and Pattern Recognition



Massimo Tistarelli  
Christophe Champod *Editors*

# Handbook of Biometrics for Forensic Science

 Springer

# **Advances in Computer Vision and Pattern Recognition**

## **Founding editor**

Sameer Singh, Rail Vision, Castle Donington, UK

## **Series editor**

Sing Bing Kang, Microsoft Research, Redmond, WA, USA

## **Advisory Board**

Horst Bischof, Graz University of Technology, Austria

Richard Bowden, University of Surrey, Guildford, UK

Sven Dickinson, University of Toronto, ON, Canada

Jiaya Jia, The Chinese University of Hong Kong, Hong Kong

Kyoung Mu Lee, Seoul National University, South Korea

Yoichi Sato, The University of Tokyo, Japan

Bernt Schiele, Max Planck Institute for Computer Science, Saarbrücken, Germany

Stan Sclaroff, Boston University, MA, USA

More information about this series at <http://www.springer.com/series/4205>

Massimo Tistarelli · Christophe Champod  
Editors

# Handbook of Biometrics for Forensic Science



Springer

*Editors*

Massimo Tistarelli  
University of Sassari  
Alghero, Sassari  
Italy

Christophe Champod  
University of Lausanne  
Lausanne  
Switzerland

ISSN 2191-6586

ISSN 2191-6594 (electronic)

Advances in Computer Vision and Pattern Recognition

ISBN 978-3-319-50671-5

ISBN 978-3-319-50673-9 (eBook)

DOI 10.1007/978-3-319-50673-9

Library of Congress Control Number: 2016959546

© Springer International Publishing AG 2017

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

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

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer International Publishing AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

# Preface

Forensic Biometrics is a relatively novel discipline with a long-standing history. Even though this assertion may seem contradictory, it is true that biometric samples have been used in forensic examination since many decades. However, the systematic adoption of biometric technologies for criminal investigations and the incorporation in forensic processes is relatively new. In this regard, a considerable effort has been performed in the past 4 years by the European Union COST Action IC1106 “Integrating Biometrics and Forensics for the Digital Age.”

The COST IC1106 research consortium is composed of over 40 members, research institutions, and forensic laboratories mainly from Europe and also from Australia, China, and USA. Aim of the Action has been to promote synergies between the biometrics recognition and the forensics science communities. This was achieved by means of innovative networking and scientific exchange. Most of the chapters making this edited volume are the outcome of this scientific collaboration.

Today’s digital era is providing not only new computing solutions to assist forensics but also new threats and challenges, which cannot be solved with traditional approaches. These include identity-related scenarios such as attacks on security systems and the identification of abnormal/dangerous behaviors from remote cameras. New identification technologies and pattern recognition algorithms offer ways to provide proof of identity in these cases.

This book is the outcome of a strong new interdisciplinary community, which is establishing and disseminating good practice, and is stimulating novelty and interdisciplinarity in exploiting scientific possibilities.

While this trend deserves a growing attention, a strong impact is expected in many forensic scenarios, including identification at sensitive border crossing, analysis of video traces from surveillance cameras, and providing proof of evidence in court cases. This book presents a wide and in-depth view of the most advanced biometric technologies applied and positively developed to forensic cases.

A multiview approach is presented to the reader, with each chapter being designed to cover a different subject written by authors from different research institutions, and the objective of covering the subject from different perspectives.

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

Alghero, Italy  
June 2016

Massimo Tistarelli  
Christophe Champod

# Contents

<b>1 Biometric Technologies for Forensic Science and Policing: State of the Art . . . . .</b>	<b>1</b>
Christophe Champod and Massimo Tistarelli	
<b>Part I Analysis of Fingerprints and Fingermarks</b>	
<b>2 Capture and Analysis of Latent Marks . . . . .</b>	<b>19</b>
Mario Hildebrandt, Jana Dittmann and Claus Vielhauer	
<b>3 Automated Fingerprint Identification Systems: From Fingerprints to Fingermarks . . . . .</b>	<b>37</b>
Davide Maltoni, Raffaele Cappelli and Didier Meuwly	
<b>4 Challenges for Fingerprint Recognition—Spoofing, Skin Diseases, and Environmental Effects . . . . .</b>	<b>63</b>
Martin Drahanský, Ondřej Kanich and Eva Březinová	
<b>5 Altered Fingerprint Detection . . . . .</b>	<b>85</b>
John Ellingsgaard and Christoph Busch	
<b>Part II Face and Video Analysis</b>	
<b>6 Face Sketch Recognition via Data-Driven Synthesis . . . . .</b>	<b>127</b>
Nannan Wang, Shengchuan Zhang, Chunlei Peng, Jie Li and Xinbo Gao	
<b>7 Recent Developments in Video-Based Face Recognition . . . . .</b>	<b>149</b>
Jingxiao Zheng, Vishal M. Patel and Rama Chellappa	
<b>8 Face Recognition Technologies for Evidential Evaluation of Video Traces . . . . .</b>	<b>177</b>
Xingjie Wei and Chang-Tsun Li	

<b>9 Human Factors in Forensic Face Identification . . . . .</b>	<b>195</b>
David White, Kristin Norell, P. Jonathon Phillips and Alice J. O'Toole	
<b>Part III Human Motion, Speech and Behavioral Analysis</b>	
<b>10 Biometric Evidence in Forensic Automatic Speaker Recognition . . . . .</b>	<b>221</b>
Andrzej Drygajlo and Rudolf Haraksim	
<b>11 On Using Soft Biometrics in Forensic Investigation . . . . .</b>	<b>241</b>
Paulo Lobato Correia, Peter K. Larsen, Abdenour Hadid, Martin Sandau and Miguel Almeida	
<b>12 Locating People in Surveillance Video Using Soft Biometric Traits . . . . .</b>	<b>267</b>
Simon Denman, Michael Halstead, Clinton Fookes and Sridha Sridharan	
<b>13 Contact-Free Heartbeat Signal for Human Identification and Forensics . . . . .</b>	<b>289</b>
Kamal Nasrollahi, Mohammad A. Haque, Ramin Irani and Thomas B. Moeslund	
<b>Part IV Statistical Analysis of Forensic Biometric Data</b>	
<b>14 From Biometric Scores to Forensic Likelihood Ratios . . . . .</b>	<b>305</b>
Daniel Ramos, Ram P. Krish, Julian Fierrez and Didier Meuwly	
<b>15 Dynamic Signatures as Forensic Evidence: A New Expert Tool Including Population Statistics . . . . .</b>	<b>329</b>
Ruben Vera-Rodriguez, Julian Fierrez and Javier Ortega-Garcia	
<b>Part V Ethical and Legal Issues</b>	
<b>16 Ethics and Policy of Forensic Biometrics . . . . .</b>	<b>353</b>
Emilio Mordini	
<b>Index . . . . .</b>	<b>367</b>

## **Chapter 1**

# **Biometric Technologies for Forensic Science and Policing: State of the Art**

**Christophe Champod and Massimo Tistarelli**

**Abstract** In the last decades, biometric technologies have been applied in forensic investigations only to a limited extent of their possibilities. A number of factors have hindered the wider adoption of these technologies to operational scenarios. However, there have been a number of successful applications where biometric technologies were crucial to support investigation and to provide evidence in court. Given the great potential of biometric technologies for objective and quantitative evidence evaluation, it would be desirable to see a wider deployment of these technologies, in a standardized manner, among police forces and forensic institutes. In this chapter, after a review of the actual state of the art in forensic biometric systems, we try to identify some avenues to facilitate the application of advanced biometric technologies in forensic practice. Despite their impressive performance, some recent biometric technologies have never been applied to forensic evaluation. Other technologies will need adaptations to be ready for the forensic field. We postulate that there is a challenge to be faced with more advanced tools and testing on operational data. This will require a joint effort involving stakeholders and scientists from multiple disciplines as well as a greater involvement of forensic institutes and police forensic science departments.

### **1.1 A Short Historical Introduction and Forensic Context**

For all tasks of authentication or identification in the context of investigations (civil or criminal), be it for investigative purposes or to be used as evidence in a court of law, biometric techniques play a key role. The difficulties associated with the

---

C. Champod (✉)

School of Criminal Justice, University of Lausanne, Lausanne, Switzerland  
e-mail: christophe.champod@unil.ch

M. Tistarelli

PolComIng, Computer Vision Laboratory, University of Sassari,  
Viale Mancini 5, 07041 Sassari, Italy  
e-mail: tista@uniss.it

recognition of individuals on the basis of the memory of witnesses paved the way for recognition or identification techniques based on biometric features. As early as the turn of nineteenth century, Bertillon developed standardized descriptive techniques allowing the description and photographic documentation of faces [2]. Alongside, he implemented anthropometry, the standardized measurement of selected attributes of the human body [3]. That technique gained worldwide acceptance as a reliable method to help establishing identity when an unknown individual had to be confronted with a set of known persons that has been previously recorded. In the above historical context, the issue is the establishment of an identity of an unknown individual against a reference database of known individuals, either in verification mode (1 to 1) or in identification mode (1 to N). In addition, all measurements or photographed are acquired from living individuals in specified controlled conditions. Soon after Bertillon's anthropometry, the techniques taking advantage of fingerprints were introduced following the work by Galton, Faulds, Vucetich, and Henry (for a complete historical account refer to [19]). Fingerprints soon replaced anthropometry due to their ease of taking, classification and retrieval. But still, the technique is based on the pristine take of 10-print cards. Another major development, almost a side effect of the use of fingerprinting, is the capability to use marks, left inadvertently on objects potentially associated with criminal activities. These marks are left under uncontrolled conditions, may be distorted and smudged, visible or invisible to the naked eye. A range of techniques was developed to detect them. Even if marks are of significantly reduced quality compared to the quality of the controlled prints, they are an efficient mean for association with known prints. The first known cases of identification through the use of marks detected on crime scene are due to Vucetich in the 1892 Rojas case [16] and Bertillon investigated the 1902 Scheffer case [82]. In the first case, the mark was directly visible in blood, whereas in the second, Bertillon had to use white lead oxide powder to detect the marks on the glass pane. Through the years, the techniques used to visualize and enhance fingermarks (or other marks from friction ridge skin) progressed in terms of range of surfaces from which marks could successfully detected and in terms of sensitivity of the detection methods. Another key development was the organization of these methods in sequences starting from nondestructive optical methods towards more invasive but complementary physical or chemical techniques (see [15] for an overview of these methods).

Other biometric features left as marks have found some forensic applications since the early days before the major revolution i.e., DNA profiling. We note the use of barefoot impressions (either in soles of shoes or recovered on scenes [95], earmarks collected from doors or windows [92], bitemarks [27] or facial images [84]). But of course, the first type of behavioral biometric used in forensic science is handwriting and signature whose usage goes back a few centuries [38].

It is in 1986 in the Pitchfork case that Alec Jeffreys introduced the possibility to analyze DNA in forensic science with a view to singularize individuals. It gave new possibilities to test for kinship and also identity based on biological material, fluids, and traces left on crime scenes (for an extensive overview of the field of DNA

analysis refer to [10]). It was a major improvement compared to traditional serology testing that required substantial amount of material to obtain significant discrimination. Since its introduction in policing, DNA profiling gained in terms of sensitivity (capability to obtain a DNA profile from minute quantities of DNA, as low as a few picograms), and discrimination power. These progresses come with their lot of interpretative issues due to an increase of low level DNA mixtures with numerous contributors as it is now possible to obtain DNA from surface that have been in contact with multiple sources, as opposed to a large stain from on body fluid [11].

In forensic science, it all comes down to an attempt to compare material from unknown source(s) to material of known origin(s), in two typical scenarios, the first when the unknown material is taken under controlled conditions (such as for the identification of an unknown living individual) or under uncontrolled or unpredictable situations (such as for marks recovered from crime scenes or material from a dead body). This is one of the major contrasts between forensic science and other typical biometric systems used for access control or automatic border control. As soon as the gathered material is obtained under unconstrained conditions, the performance and the systems will suffer substantially. Another factor of complexity, that is shared with all biometric endeavor, is the within source variability, not only due to the acquisition conditions, but also to the source itself that may evolve to various degrees as a function of the age, time or intent.

Also, it is important to realize that the output of forensic work should not only be viewed in terms of provision of evaluative evidence in court, but it can take the form the production of information in the form of investigating leads. These two functions should be distinguished as the requirement for timeliness and reliability are different for both functions that we will refer to as “investigative” versus “evaluative” [44]. The use of DNA profiling will help to put this distinction into context. Most of national DNA databases are used to provide investigative leads in the form of the set of identifier for both known individuals and stains collected from investigated cases that have found correspondence (either in full, or partially suggesting close relatives) with a submitted unknown DNA profile. As such the information is not considered as evidence and will require being checked and interpreted by a DNA analyst. It is only the subsequent statement provided by the analyst (qualified with an appropriate measure of the weight to be attached to the findings) that will be presented as evidence in court. In some investigation, the fact that a DNA database has been used is not even disclosed in court (for example to avoid the court to be exposed to antecedents of the concerned individual). When a DNA database is searched, it is searched to minimize the risk of missed associations, as any association will be further challenged through the follow-up investigation and subsequent forensic analysis. However, when we ought to present the forensic findings in court as evaluative evidence, we wish to minimize the probability of a wrong association, or at the very least be transparent regarding this adverse probability.

## 1.2 Recent Developments of Biometric Technologies in Forensic Science

The development of automatic fingerprint identification systems (AFIS) has been steady since the mid-1960s [69]. Most of forensic systems were based on minutiae (either ridge endings or bifurcations of friction ridges) as the main features used for matching. From the outset of their development, these systems were not supposed to engage into any decision (as any other biometric system), but merely serve the purpose of speeding the search process when a fingerprint examiner is searching prints or marks against a database of already enrolled individuals or marks related to past cases. Indeed the decision as to whether the mark or the print belongs to a given known individual is left to the examiner that will apply the same process as if the suspect had been developed through another investigative means. The AFIS were used exclusively to bring appropriate candidates up in the list of candidates to consider one by one by the examiner as it represented a set of suspects to check against. The score, measuring the similarity between the query and the known is used only to rank the list of candidates. That being said, the accuracy of these systems is such that they allow bringing back the correct candidate in the considered list when a print or a mark is search among millions of cards and thousands of marks. These systems have progressed over the recent years to allow less interaction form the human examiner with the system. First, the images of high quality (often the case with good marks, well-taken prints) do not require much manual encoding but can benefit from auto-encoding algorithms that are very efficient. Second, when comparisons are undertaken on 2-print or 10-print cards, the accuracy allows setting automatic decision thresholds for identification. We are referring here about usage of AFIS systems in lights-out mode. The capacity to deal with marks (generally of lower quality than prints) is just a matter of time as the most advanced systems already allow to handle a fair amount of them (about 70%) in lights-out mode [65]. Third, the matching algorithms have progressed with the combined usage of both minutiae-based matching and image-based matching [41, 42].

The seminal work of Daugman [20–23] established the use of iris as an extremely strong biometric modality. It allows operating with extremely low probability of a false positive [33], used mainly for the identification of individual under supervised conditions (for immigration or border controls). For example, the wider system deployed for border control is in UAE. Iris is also one of the modalities retained alongside with fingerprints for the India's unique identification project Aadhaar (<https://uidai.gov.in/>). Progresses made in imaging sensors and at capturing irises at a distance allow using this modality in unconstrained conditions. Iris can be captured operationally with a 800 mm lens and a 850 nm infrared LED source at a standoff distance up to 8 m [93]. We can certainly see applications in forensic science but they will be limited given the current state of technology to some constrained environments with cooperative individuals. We are far from an application in covert surveillance modes.

Forensic speaker recognition (FSR) has a long forensic tradition. It takes advantage of aural-perceptual (structured listening by trained phoneticians), auditory-instrumental (structured listening by trained phoneticians helped with specific measures such as the fundamental frequency and formant analysis), and also automatic methods inspired by the techniques used in biometrics [72]. The forensic practice varies a lot internationally [29]. The difficulties that forensic scientists face with FSR are due to the unpredictable nature of the technical conditions and background conditions under which the recordings are obtained. It is observed not so much for the controlled utterances, but systemically for the questioned recordings. It can be for example session mismatch (e.g., GSM telephone versus microphone), variations in the distance between the speakers and the microphone or the conditions of speech (stress, emotion, physical activities, overlapping talkers). All these imponderable variables impact on the accuracy of FSR [13] and it is wise to carefully assess the situations in which FSR can be reasonably undertaken [83].

The next biometric modality that has been developed and improved enormously in the recent years is face. Facial recognition systems have progressed to allow accurate facial recognition in controlled conditions such as automatic border controls or searches through database of mugshots. Needless to say that performance is affected by the size of the reference database and the age difference between the query facial image and the corresponding facial image in the gallery [32]. Moving from controlled conditions to unconstrained facial recognition is challenging as shown by Jain and colleagues [46, 48]. This is due to the general poorer quality of the images, variations in illumination, pose and occlusion caused by other objects and people. In today's policing, the number of facial images of non-identified individuals of interest is increasing fast. Their sources go from surveillance cameras, cameras from automated teller machine (ATM), personal devices, etc.. Facial recognition, when the face is acquired under unconstrained conditions in terms of illumination or pose and with highly variable quality of images and background, is much more difficult but within the reach of current technology [4, 90] and certainly in the future. Some cases have been presented in the literature [53].

Recognition from police photo sketches is a niche application of biometry under mismatching conditions [52, 54]. It is an opportunity worth that every police force should be considering in serious cases.

Ear biometrics received attention as a means to improve on facial recognition under constrained conditions, apart from lightning [18, 40]. Performance has improved over the years to reach more than 95% accuracy [58]. To our knowledge, it has found limited applications in forensic science, mainly because the images of ears are not systematically acquired in adequate conditions during the booking process of individuals. Assessment of ear features while comparing two images of ears is currently carried out by forensic experts holistically on a case-by-case basis. This approach has shown limitations in terms of reliability [35]. The application of biometric techniques, in the sense of systematic and semi-automatic measurements on extracted features, on earmark collected from scenes and compared to earprints

has led to interesting results as well [49, 50]. Reference earprints are not collected systematically upon arrest and hence it is foreseen that this modality will not be used in identification mode (1:N), but more frequently in verification mode (1:1).

Facial information may not be the only information available in images helping towards identification and often, in forensic conditions, the face is obscured, masked, or simply not visible. In these cases additional features such as scars, tattoos, vein patterns or knuckle patterns may be the only attributes helping associate an image with a given individual. Tattoo image recognition is certainly the closest to concrete operational forensic applications [61]. Knuckle patterns started as a classic biometric development taking advantage knuckles patterns taken under constrained conditions, patterns available on hands, easy to acquire in a noninvasive way [57, 96, 97]. Both major and (secondary) minor knuckle patterns can be used in conjunction and early data suggests that these patterns are stable over time [55, 59]. The modality is still in research stage but a steady increase in accuracy has been achieved [1]. We do foresee forensic avenues even under unconstrained conditions as shown in a recent paper [56]. However, the case will probably not be an identification mode, as databases of knuckle prints of persons of interest are not acquired on a systematic basis, but will more be a one-to-one comparison once an individual of interest comes to police notice. Dorsal hand veins patterns (that can also be visible on images of forensic interest) received also research attention, but mainly in constrained conditions with specific acquisition techniques taking advantage IR cameras (e.g., [75]). Stability of these vein patterns over time is not fully researched yet. At this stage, it is difficult to envisage an application under unconstrained conditions based on query images acquired under forensic conditions.

The use of scars (or other features such as nevi) to help identify persons has not received much attention in the biometric community, but scars were recorded as identifying features from the early days of forensic science. They received renewed interest in forensic science with the proliferation of images showing limited identifying features such as in cases of pedo-pornographic material where only hand on individual may be seen [8]. Assessing these features based only on expert judgment only has shown to be difficult [85] and researchers are striving to acquire systematic data to allow assign an appropriate weight to these comparisons [6, 7, 43].

Scars, marks, and tattoos (SMT) are features known as soft biometrics traits that include also characteristics such as height, weight, body geometry, gender, etc. [73, 79]. They can be typically described using human understandable labels and measurements. They can be obtained under unconstrained conditions (typically at a distance) and hence offer opportunities in surveillance applications [79, 89]. Soft biometric features can extend to attributes of individuals such as garments [45].

Moving back to more macro traits, gait analysis is receiving an increased attention in forensic science due to the proliferation of CCTV images [26, 60, 63]. Experts may use computer-assisted tools to carry out specific measurements on images or correct for distortion and perspective, but the recognition process is essentially carried out visually. Though, they have shown good ability to identify individuals using that process [5]. The biometric community also investigated the

capabilities to detect and identify individuals based on gait features [74]. For a review of the latest development, refer to Makihara et al. [64]. In forensic science, one of the main challenge is related to the within source variability that may overlap the between sources variability, depending on the situation, the garments worn and fatigue [62]. Case-specific experiments are highly recommended.

### 1.3 Challenges

The challenges to deploy successfully biometric systems in forensic science have been described by a few authors already [24, 48, 68, 86]. For biometrics in general, we invite to refer to Jain et al. [47]. We will concentrate in this section on some key aspects.

One of the challenges posed by all forensic identification techniques described above lies in the fact that they are currently mostly based on the application of a holistic approach by trained experts. The decisions taken following the comparative examination are based on the informed judgment of the forensic analysts. Contrary to popular belief, all of these techniques, exception made of DNA profiling, do not find a systematic rooting in a statistical analysis, weighing the within-source variability against the between source variability. The case of fingerprint identification is covered in [15], but this state of affair runs across all the fields discussed above, apart from DNA. The paucity of systematic research in these areas has been highlighted by the 2009 report of the National Research Council [70].

Deploying biometric techniques in forensic science will require putting them in an adequate statistical reporting scheme. That scheme (see generally [14, 66]) is based on the concept of a likelihood ratio (LR) or Bayes factor. The LR is relative weighing of the probability to be assigned to findings given that the query and the known are from the same individual (hence allowing due consideration for the within source variability) against the probability of the findings given that the query and the known are from different individuals (hence giving due consideration for the between source variability in the relevant population). The LR provides the adequate measure of the weight to be assigned to the forensic observations and is recommended in Europe for reporting forensic evidence [94]. A specific guideline, also adopting a LR framework, has been developed for automatic and semi-automatic forensic speaker recognition as well [28].

Two challenges are then on the table when it comes to deploy biometric techniques in forensic science as evaluative evidence:

- (1) If the technique is used to provide a piece of evidence as a standalone computerized method, then the research should demonstrate that the method is properly characterized by standardized measures of forensic efficiency. Here, the work carried out in speaker recognition [30, 78] and further extended to other forensic disciplines [76, 77] represents a major advancement. It is hoped

that the recently published guideline will pave the way forward towards more standardization [67];

- (2) As mentioned before, forensic experts currently use most of the biometric techniques discussed above, applying their judgment informed training and experience developed through a large portfolio of cases. If biometric techniques bring new ways to assess the forensic findings, the combination of experts (traditional forensic and biometric experts) needs to be carefully researched. What will be the rules of engagement, how to deal with conflicting results, how to combine the respective information are not trivial questions and still have to find appropriate answers. We note some recent attempts with facial recognition [91] and hope to see more in future.

Not all biometric modalities used in forensic science will convey the same strength. For example, a fingermark displaying more than 12 minutiae deemed in correspondence with a print will on average bring a forensic weight (measured in terms of likelihood ratio) of above 1 billion [71], and so will a full 15-loci matching DNA profile [36]. In contrast, an earprints/marks correspondence will bring an average weight of 100 [50] or a few hundreds for voices under realistic forensic conditions [30]. It means that some biometric modalities will be well suited for both investigative and evaluative tasks, whereas others, such as face and ears, would be at first be used as investigative tools only. One of the challenges is to specify the appropriate framework of usage, so that the various stakeholders are very clear as how, when, and with what strength of information or evidence the modality will be used. An example on facial images is given in Dessimoz and Champod [25].

As 2D images and speech represent the main digital sensory data currently available to forensic scientists, we mainly concentrated on the processing of two-dimensional and one-dimensional signals. However, the development of 3D sensors will inevitably lead to an increased number of forensic cases where digital data will be available in 3D. On the other hand, researchers in biometric and computer vision technologies already demonstrated the feasibility of recognizing individuals from their 3D face shape, reconstructed from one or more 2D images. Inferring depth from one or more images has been studied for quite a long time in Computer Vision as *depth-from-X*, where “X” represents the visual cue exploited to infer the object 3D shape, i.e., stereo, motion, shading, etc. [31, 37, 81]. The main challenge has always been the accuracy required to reconstruct sufficient details in depth to allow the correct matching of individuals. The seminal work by Blanz and Vetter [9] introduced a novel parameterized model of face data, where the 3D facial appearance is directly related to both the face shape and the skin texture. The notable value of this research work is that the digital reconstruction of the 3D face shape is directly related to its 2D projection. In the last few years, several authors proposed different methods to reconstruct the 3D face shape from either a single image [34, 39, 51] or a sequence of images [17, 80]. The reconstructed face shape is generally used to produce a view-independent face image to match a reference face image captured at a given pose. This allows to reduce the errors due to the misalignment between the reference face, stored in a database, and the face captured

from a given individual, either at a security check point or from a camera casually found on a crime scene. Attempts to directly use the reconstructed 3D face shape to perform recognition did not yet achieve the current state of the art in 3D face recognition [87].

In order to determine the feasibility of 3D face recognition, several 3D databases are now available to the scientific community. Many of them consist of high resolution scans of hundreds different individuals acquired with different poses and expressions (see <http://www.face-rec.org/databases/> for an up to date list). This data allows to simulate real scenarios, where we need to take into account a high degree of variability. An algorithm devised to perform recognition from 3D face scans can be:

- completely automatic, operating in “lights out” mode as requiring human intervention may become prohibitive when dealing with thousands or millions of subjects;
- robust to changes in poses and facial expressions. In a real scenario, especially when dealing with images captured from surveillance cameras, it is impossible to select a given pose or facial expression.

Several algorithms for 3D face recognition have been proposed in the last decade. Recently an algorithm [12, 88] based on the extraction of facial features and the construction of a 3D signature manifold, based on 3D geometric invariants, demonstrated to be very effective in a number of very challenging test cases.

The results achieved so far, as well as other recent advances in biometrics research, do not ensure the applicability of the developed algorithms to forensic data, nor to solve every forensic case. However, given the great potential of these technologies in forensic investigations and for evidential evaluation, they are worth conducting further investigation to assess the applicability and performance which may be achieved on real forensic data.

## 1.4 Conclusions

As well exposed in this book, the identification of criminals from evidential traces or marks left on the crime scene has a long history in forensic science. As an example, in the late 50s the Federal Bureau of Investigations (FBI) in US regularly adopted fully manual procedures to match latent fingerprints with a large database of suspects on hundred thousands of print cards and started its work towards automation. Over the decades technology provided several instruments to automate the fingerprint matching process, also increasing the speed and accuracy. Since the late 90s, with the aid of the increased computation power and the reduced size of computing machines, the AFIS systems dramatically improved their performances allowing forensic expert to match and inspect hundreds of thousands fingerprint every day. As a result, the IAFIS (for Intergrated AFIS) currently working at FBI allows the forensic experts to match around 150,000 tenprints with a database of

over 100 million subjects every day. The current TAR is 99.6% with 0.103% FAR. Such a workload would not be possible with human inspectors alone.

Fingerprint matching is just one example of many success stories where several technologies allowed to facilitate the work of forensic examiners, or even to make it possible to produce evidential statement in court. Over the past decades, other biometric technologies have been successfully applied including gait recognition and face image analysis. However, there are many other “biometric traits” and associated technologies, which can be successfully exploited as well.

Innovation does not influence the speed and accuracy of the identification process alone, but also the possibility to:

- (1) handle different traces and their combination, including face pictures, action videos, speech, dna, iris, palmprints, footprints, skin marks, etc.
- (2) perform a fully automated identification of suspects in “lights out” mode.

Current biometric technology is offering a plethora of solutions to handle many different traces or marks. However, not all state of the art technologies are currently deployed in forensic investigations. It is envisaged that more efforts are required to define new technological solutions which can be applied to given operational scenarios under purposively tailored protocols.

This chapter analyzed the current state of the art in biometric technologies trying to define a path forward to the innovation of currently exploited technologies in forensic investigations. This is just a start of a path which is further described in the following chapters and it is hoped to eventually enrich the knowledge and wider adoption of biometrics in forensic cases.

**Acknowledgements** This research is based upon work supported by the European Commission under the project COST IC1106 “Biometrics and Forensics for the Digital Age” and H2020 MSCA RISE 690907 “IDENTITY”.

## References

1. Amraoui M, Abouchabaka J, Aroussi ME (2014) Finger knuckle print recognition based on multi-instance fusion of local feature sets. In: 2014 international conference on multimedia computing and systems (ICMCS), 14–16 April 2014, pp 87–92. doi:[10.1109/ICMCS.2014.6911188](https://doi.org/10.1109/ICMCS.2014.6911188)
2. Bertillon A (1890) La photographie judiciaire avec un appendice sur la classification et l'identification anthropométrique. Bibliothèque photographique. Gauthier-Villars et fils, Paris
3. Bertillon A (1893) Instructions signalétiques. Imprimerie administrative, Melun
4. Best-Rowden L, Han H, Otto C, Klare BF, Jain AK (2014) Unconstrained face recognition: identifying a person of interest from a media collection. IEEE Trans Inf Forensics Secur 9 (12):2144–2157. doi:[10.1109/TIFS.2014.2359577](https://doi.org/10.1109/TIFS.2014.2359577)
5. Birch I, Raymond L, Christou A, Fernando MA, Harrison N, Paul F (2013) The identification of individuals by observational gait analysis using closed circuit television footage. Sci Justice 53(3):339–342. doi:[10.1016/j.scijus.2013.04.005](https://doi.org/10.1016/j.scijus.2013.04.005)
6. Black S, MacDonald-McMillan B, Mallett X (2014) The incidence of scarring on the dorsum of the hand. Int J Legal Med 128(3):545–553. doi:[10.1007/s00414-013-0834-7](https://doi.org/10.1007/s00414-013-0834-7)

7. Black S, MacDonald-McMillan B, Mallett X, Rynn C, Jackson G (2014) The incidence and position of melanocytic nevi for the purposes of forensic image comparison. *Int J Legal Med* 128(3):535–543. doi:[10.1007/s00414-013-0821-z](https://doi.org/10.1007/s00414-013-0821-z)
8. Black SM, Mallett X, Rynn C, Duffield N (2009) Forensic hand image comparison as an aid for paedophile investigations. *Police Prof* 184:21–24
9. Blanz V, Vetter T (1999) A morphable model for the synthesis of 3D faces. In: Proceedings of the 26th annual conference on computer graphics and interactive techniques (SIGGRAPH'99). ACM Press/Addison-Wesley Publishing Co., New York, NY, USA, pp 187–194. doi:<http://dx.doi.org/10.1145/311535.311556>
10. Butler JM (2010) Fundamentals of forensic DNA typing. Elsevier, Academic Press, Amsterdam
11. Butler JM (2015) Advanced topics in forensic DNA typing: interpretation. Elsevier, Academic Press, Amsterdam
12. Cadoni M, Bicego M, Grossi E (2009) 3D face recognition using joint differential invariants, advances in biometrics. In: Proceedings of the 3rd international conference on biometrics, ICB 2009, pp 279–288. doi:[10.1007/978-3-642-01793-3\\_29](https://doi.org/10.1007/978-3-642-01793-3_29)
13. Campbell JP, Shen W, Campbell WM, Schwartz R, Bonastre JF, Matrouf D (2009) Forensic speaker recognition—a need for caution. *IEEE Signal Process Mag* 26(2):95–103. doi:[10.1109/MSP.2008.931100](https://doi.org/10.1109/MSP.2008.931100)
14. Champod C (2013) Overview and meaning of identification/individualization. In: Siegel JA, Saukko PJ (eds) Encyclopedia of forensic sciences. Academic Press, Waltham, pp 303–309. doi:[10.1016/B978-0-12-382165-2.00197-5](https://doi.org/10.1016/B978-0-12-382165-2.00197-5)
15. Champod C, Lennard CJ, Margot PA, Stoilovic M (2016) Fingerprints and other ridge skin impressions, 2nd edn. CRC Press, Boca Raton
16. Chapman CL (1992) Dr. Juan Vucetich: his contribution to the science of fingerprints. *J Forensic Ident* 42(4):286–294
17. Choi J, Medioni G, Lin Y, Silva L, Pereira Bellon OR, Pamplona M, Faltemier TC (2010) 3D Face reconstruction using a single or multiple views. In: 20th international conference on pattern recognition ICPR 2010, Istanbul, Turkey, 23–26 Aug 2010, pp 3963–3966. doi:[10.1109/ICPR.2010.963](https://doi.org/10.1109/ICPR.2010.963)
18. Choras M (2009) Ear biometrics. In: Li SZ, Jain A (eds) Encyclopedia of biometrics. Springer, New York. doi:[10.1007/978-0-387-73003-5\\_173](https://doi.org/10.1007/978-0-387-73003-5_173)
19. Cole S (2001) Suspect identities: a history of fingerprinting and criminal identification. Harvard University Press, Cambridge
20. Daugman J (2003) The importance of being random: statistical principles of iris recognition. *Pattern Recognit* 36(2):279–291. doi:[10.1016/S0031-3203\(02\)00030-4](https://doi.org/10.1016/S0031-3203(02)00030-4)
21. Daugman J (2006) Probing the uniqueness and randomness of IrisCodes: results from 200 billion iris pair comparisons. *Proc IEEE* 94(11):1927–1935. doi:[10.1109/JPROC.2006.884092](https://doi.org/10.1109/JPROC.2006.884092)
22. Daugman J, Downing C (2001) Epigenetic randomness, complexity, and singularity of human iris patterns. *Proc Royal Soc Biol Sci B* 268:1737–1740. doi:[10.1098/rspb.2001.1696](https://doi.org/10.1098/rspb.2001.1696)
23. Daugman JG (1993) High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans Pattern Anal Mach Intell* 15(11):1148–1161. doi:[10.1109/34.244676](https://doi.org/10.1109/34.244676)
24. Dessimoz D, Champod C (2007) Linkages between biometrics and forensic science. In: Flynn PJ, Jain AK, Ross A (eds) Handbook of biometrics. Springer, New York, pp 425–459. doi:[10.1007/978-0-387-71041-9\\_21](https://doi.org/10.1007/978-0-387-71041-9_21)
25. Dessimoz D, Champod C (2015) A dedicated framework for weak biometrics in forensic science for investigation and intelligence purposes: the case of facial information. *Secur J* (special issue on traceology) 29(4):603–617. doi:[10.1057/sj.2015.32](https://doi.org/10.1057/sj.2015.32)
26. DiMaggio JA, Vernon W (2011) Forensic gait analysis. In: Forensic podiatry: principles and methods. Humana Press, Totowa, NJ, pp 103–115. doi:[10.1007/978-1-61737-976-5\\_6](https://doi.org/10.1007/978-1-61737-976-5_6)
27. Dorion RBJ (2005) Bitemark evidence. Forensic science. Marcel Dekker, New York

28. Drygajlo A, Jessen M, Gfroerer S, Wagner I, Vermeulen J, Niemi T (2016) Methodological Guidelines for Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition. Verlag für Polizeiwissenschaften, Frankfurt
29. Gold E, Peter F (2011) International practices in forensic speaker comparison. *Int J Speech Lang Law* 18(2):293–307. doi:[10.1558/ijssl.v18i2.293](https://doi.org/10.1558/ijssl.v18i2.293)
30. Gonzalez-Rodriguez J, Rose P, Ramos D, Toledano DT, Ortega-Garcia J (2007) Emulating DNA: rigorous quantification of evidential weight in transparent and testable forensic speaker recognition. *IEEE Trans Audio Speech Lang Process* 15(7):2104–2115. doi:[10.1109/TASL.2007.902747](https://doi.org/10.1109/TASL.2007.902747)
31. Grimson WEL (1981) From images to surfaces: a computational study of the human early vision system. MIT Press
32. Grother PJ, Ngan ML (2014) Face recognition vendor test (FRVT)—performance of face identification algorithms, NIST interagency report 8009. National Institute for Standards and Technology, Gaithersburg
33. Grother PJ, Quinn GW, Matey JR, Ngan ML, Salamon WJ, Fiumara GP, Watson CI (2012) IREX III: performance of iris identification algorithms, NIST interagency report 7836. National Institute of Standard and Technology, Gaithersburg
34. Hassner T (2013) Viewing real-world faces in 3D. In: International conference on computer vision (ICCV), Sydney, Australia, pp 3607–3614
35. Hoogstrate AJ, van den Heuvel C, Huyben E (2001) Ear identification based on surveillance camera images. *Sci Justice* 41(3):167–172. doi:[10.1016/S1355-0306\(01\)71885-0](https://doi.org/10.1016/S1355-0306(01)71885-0)
36. Hopwood AJ, Puch-Solis R, Tucker VC, Curran JM, Skerrett J, Pope S, Tully G (2012) Consideration of the probative value of single donor 15-plex STR profiles in UK populations and its presentation in UK courts. *Sci Justice* 52(3):185–190. doi:[10.1016/j.scijus.2012.05.005](https://doi.org/10.1016/j.scijus.2012.05.005)
37. Horn BKP, Brooks MJ (eds) (1989) Shape from shading. MIT Press
38. Huber RA, Headrick AM (1999) Handwriting identification: facts and fundamentals. CRC Press, Boca Raton
39. Hu Y, Jiang D, Yan S, Zhang L (2004) Automatic 3D reconstruction for face recognition. In: Proceedings of 6th IEEE international conference on automatic face and gesture recognition, FRGC 2004. IEEE, pp 843–848. doi:[10.1109/AFGR.2004.1301639](https://doi.org/10.1109/AFGR.2004.1301639)
40. Hurley DJ, Nixon MS (2009) Physical analogies for ear recognition. In: Li SZ, Jain A (eds) Encyclopedia of biometrics. Springer, New York. doi:[10.1007/978-0-387-73003-5\\_172](https://doi.org/10.1007/978-0-387-73003-5_172)
41. Indovina M, Dvornychenko V, Hicklin RA, Kiebuzinski GI (2012) ELFT-EFS evaluation of latent fingerprint technologies: extended feature sets [Evaluation #2]. vol NISTIR 7859. National Institute of Standards and Technology, Gaithersburg
42. Indovina M, Hicklin RA, Kiebuzinski GI (2011) ELFT-EFS Evaluation of latent fingerprint technologies: extended feature [Sets Evaluation #1]. National Institute of Standards and Technology, Gaithersburg
43. Jackson G, Black S (2014) Use of data to inform expert evaluative opinion in the comparison of hand images—the importance of scars. *Int J Legal Med* 128(3):555–563. doi:[10.1007/s00414-013-0828-5](https://doi.org/10.1007/s00414-013-0828-5)
44. Jackson G, Jones S, Booth G, Champod C, Evett IW (2006) The nature of forensic science opinion—a possible framework to guide thinking and practice in investigations and in court proceedings. *Sci Justice* 46(1):33–44. doi:[10.1016/S1355-0306\(06\)71565-9](https://doi.org/10.1016/S1355-0306(06)71565-9)
45. Jaha ES, Nixon MS (2014) Soft biometrics for subject identification using clothing attributes. In: 2014 IEEE international joint conference on biometrics (IJCB), Sept 29 2014–Oct 2 2014, pp 1–6. doi:[10.1109/BTAS.2014.6996278](https://doi.org/10.1109/BTAS.2014.6996278)
46. Jain AK, Klare B, Park U (2012) Face matching and retrieval in forensics applications. *IEEE Multimed* 19(1):20. doi:[10.1109/MMUL.2012.4](https://doi.org/10.1109/MMUL.2012.4)
47. Jain AK, Nandakumar K, Ross A (2016) 50 years of biometric research: accomplishments, challenges, and opportunities. *Pattern Recog Lett* 79:80–105. doi:[10.1016/j.patrec.2015.12.013](https://doi.org/10.1016/j.patrec.2015.12.013)

48. Jain AK, Ross A (2015) Bridging the gap: from biometrics to forensics. *Philos Trans Royal Soc Lond B: Biol Sci* 370 (1674). doi:[10.1098/rstb.2014.0254](https://doi.org/10.1098/rstb.2014.0254)
49. Junod S, Champod C (2012) Earprint comparison: automated systems. In: Jamieson A, Moenssens AA (eds) *Wiley encyclopedia of forensic science*. John Wiley, Chichester. doi:[10.1002/9780470061589.fsa1033](https://doi.org/10.1002/9780470061589.fsa1033)
50. Junod S, Pasquier J, Champod C (2012) The development of an automatic recognition system for earmark and earprint comparisons. *Forensic Sci Int* 222(1–3):170–178. doi:[10.1016/j.forsciint.2012.05.021](https://doi.org/10.1016/j.forsciint.2012.05.021)
51. Kemelmacher-Shlizerman I, Basri R (2011) 3D face reconstruction from a single image using a single reference face shape. *IEEE TPAMI* 33(2):394–405. doi:[10.1109/TPAMI.2010.63](https://doi.org/10.1109/TPAMI.2010.63)
52. Klare B, Li Z, Jain AK (2011) Matching forensic sketches to mug shot photos. *IEEE Trans Pattern Anal Mach Intell* 33(3):639–646. doi:[10.1109/TPAMI.2010.180](https://doi.org/10.1109/TPAMI.2010.180)
53. Klontz JC, Jain AK (2013) A case study of automated face recognition: the Boston marathon bombings suspects. *IEEE Comput* 46(11):91–94 . doi:[10.1109/MC.2013.377](https://doi.org/10.1109/MC.2013.377)
54. Klum S, Han H, Jain AK, Klare B (2013) Sketch based face recognition: forensic vs. composite sketches. In: 2013 international conference on biometrics (ICB), 4–7 June 2013, pp 1–8. doi:[10.1109/ICB.2013.6612993](https://doi.org/10.1109/ICB.2013.6612993)
55. Kumar A (2012) Can we use minor finger knuckle images to identify humans? In: 2012 IEEE fifth international conference on biometrics: theory, applications and systems (BTAS), 23–27 Sept 2012, pp 55–60. doi:[10.1109/BTAS.2012.6374558](https://doi.org/10.1109/BTAS.2012.6374558)
56. Kumar A (2014) Importance of being unique from finger dorsal patterns: exploring minor finger knuckle patterns in verifying human identities. *IEEE Trans Inf Forensics Secur* 9 (8):1288–1298. doi:[10.1109/TIFS.2014.2328869](https://doi.org/10.1109/TIFS.2014.2328869)
57. Kumar A, Ravikanth C (2009) Personal authentication using finger knuckle surface. *IEEE Trans Inf Forensics Secur* 4(1):98–109. doi:[10.1109/TIFS.2008.2011089](https://doi.org/10.1109/TIFS.2008.2011089)
58. Kumar A, Wu C (2012) Automated human identification using ear imaging. *Pattern Recognit* 45(3):956–968. doi:[10.1016/j.patcog.2011.06.005](https://doi.org/10.1016/j.patcog.2011.06.005)
59. Kumar A, Xu Z (2014) Can we use second minor finger knuckle patterns to identify humans? In: 2014 IEEE conference on computer vision and pattern recognition workshops, 23–28 June 2014, pp 106–112. doi:[10.1109/CVPRW.2014.21](https://doi.org/10.1109/CVPRW.2014.21)
60. Larsen PK, Simonsen EB, Lynnerup N (2008) Gait analysis in forensic medicine. *J Forensic Sci* 53(5):1149–1153. doi:[10.1111/j.1556-4029.2008.00807.x](https://doi.org/10.1111/j.1556-4029.2008.00807.x)
61. Lee J, Jin R, Jain AJ, Tong W (2012) Image retrieval in forensics: tattoo image database application. *IEEE Multimed* 19(1):40–49. doi:[10.1109/MMUL.2011.59](https://doi.org/10.1109/MMUL.2011.59)
62. Ludwig O, Dillinger S, Marschall F (2016) Intra-individual gait pattern variability in specific situations: Implications for forensic gait analysis. *Forensic Sci Int* 264:15–23. doi:[10.1016/j.forsciint.2016.02.043](https://doi.org/10.1016/j.forsciint.2016.02.043)
63. Lynnerup N, Vedel J (2005) Person identification by gait analysis and photogrammetr. *J Forensic Sci* 50(1):1–7. doi:[10.1520/JFS2004054](https://doi.org/10.1520/JFS2004054)
64. Makihara Y, Matovski DS, Nixon MS, Carter JN, Yagi Y (2015) Gait recognition: databases, representations, and applications. *Wiley Encyclopedia of Electrical and Electronics Engineering*, pp 1–15. doi:[10.1002/047134608X.W8261](https://doi.org/10.1002/047134608X.W8261)
65. Meagher S, Dvornychenko V, Garris M (2014) Characterization of latent print “lights-out” modes for automated fingerprint identification systems. *J Forensic Ident* 64(3):255–284
66. Meuwly D (2006) Forensic individualisation from biometric data. *Sci Justice* 46(4):205–213. doi:[10.1016/S1355-0306\(06\)71600-8](https://doi.org/10.1016/S1355-0306(06)71600-8)
67. Meuwly D, Ramos D, Haraksim R (to appear) A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation. *Forensic Sci Int*. doi:[10.1016/j.forsciint.2016.03.048](https://doi.org/10.1016/j.forsciint.2016.03.048)
68. Meuwly D, Veldhuis R (2012) Forensic biometrics: from two communities to one discipline. In: Proceedings of the international conference of the biometrics special interest group (BIOSIG), 2012 BIOSIG, 6–7 Sept 2012, pp 1–12

69. Moses K (2011) Chapter 6: Automatic fingerprint identification systems (AFIS). In: McRoberts A (ed) *The fingerprint sourcebook*. National Institute of Justice, Washington DC, pp 6-1–6-33
70. National Research Council (2009) *Strengthening forensic science in the United States: a path forward*. The National Academies Press, Washington, DC
71. Neumann C, Evett IW, Skerrett J (2012) Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm. *J Roy Stat Soc Ser A (Stat Soc)* 175(Part 2):371–415 (with discussion)
72. Neustein A, Patil HA (eds) (2012) *Forensic speaker recognition*. Springer, New York. doi:[10.1007/978-1-4614-0263-3](https://doi.org/10.1007/978-1-4614-0263-3)
73. Nixon MS, Correia PL, Nasrollahi K, Moeslund TB, Hadid A, Tistarelli M (2015) On soft biometrics. *Pattern Recog Lett* 68(Part 2):218–230. doi:[10.1016/j.patrec.2015.08.006](https://doi.org/10.1016/j.patrec.2015.08.006)
74. Nixon MS, Tan TN, Chellappa R (2006) Human identification based on gait. *International series on biometrics*. Springer, New York. doi:[10.1007/978-0-387-29488-9](https://doi.org/10.1007/978-0-387-29488-9)
75. Raghavendra R, Surbirala J, Busch C (2015) Hand dorsal vein recognition: sensor, algorithms and evaluation. In: 2015 IEEE international conference on imaging systems and techniques (IST), 16–18 Sept. 2015, pp 1–6. doi:[10.1109/IST.2015.7294557](https://doi.org/10.1109/IST.2015.7294557)
76. Ramos D, Gonzalez-Rodriguez J (2013) Reliable support: measuring calibration of likelihood ratios. *Forensic Sci Int* 230(1–3):156–169. doi:[10.1016/j.forsciint.2013.04.014](https://doi.org/10.1016/j.forsciint.2013.04.014)
77. Ramos D, Gonzalez-Rodriguez J, Zadora G, Aitken C (2013) Information-theoretical assessment of the performance of likelihood ratio computation methods. *J Forensic Sci* 58(6):1503–1518. doi:[10.1111/1556-4029.12233](https://doi.org/10.1111/1556-4029.12233)
78. Ramos-Castro D, Gonzalez-Rodriguez J, Ortega-Garcia J (2006) Likelihood ratio calibration in a transparent and testable forensic speaker recognition framework. In: Speaker and language recognition workshop, 2006. IEEE Odyssey 2006: The, 28–30 June 2006, pp 1–8. doi:[10.1109/ODYSSEY.2006.248088](https://doi.org/10.1109/ODYSSEY.2006.248088)
79. Reid DA, Samangooei S, Chen C, Nixon M, Ross A (2013) Soft biometrics for surveillance: an overview. In: Govindaraju V, Rao CR (eds) *Handbook of statistics*, vol 31. Elsevier, North Holland, Amsterdam, pp 327–352. doi:[10.1016/B978-0-44-453859-8.00013-8](https://doi.org/10.1016/B978-0-44-453859-8.00013-8)
80. Roth J, Tong Y, Liu X (2015) Unconstrained 3D face reconstruction. In: The IEEE conference on computer vision and pattern recognition (CVPR), pp 2606–2615. doi:[10.1109/CVPR.2015.7298876](https://doi.org/10.1109/CVPR.2015.7298876)
81. Sandini G, Tistarelli M (1990) Active tracking strategy for monocular depth inference over multiple frames. *IEEE Trans PAMI* PAMI-11(12):13–27. doi:[10.1109/34.41380](https://doi.org/10.1109/34.41380)
82. Sannié C (1950) Alphonse Bertillon et la dactyloscopie. L'affaire Scheffer. *Revue internationale de police criminelle* 5(41):255–262
83. Schwartz R, Campbell JP, Shen W (2011) When to punt on speaker comparison? *J Acoust Soc Am* 130(4):2547. doi:[10.1121/1.3655180](https://doi.org/10.1121/1.3655180)
84. Spaun NA (2011) Face recognition in forensic science. In: Li SZ, Jain AK (eds) *Handbook of face recognition*. Springer International Publishing, pp 655–670. doi:[10.1007/978-0-85729-932-1\\_26](https://doi.org/10.1007/978-0-85729-932-1_26)
85. Stevenage SV, Walpole C, Neil GJ, Black SM (2015) Testing the reliability of hands and ears as biometrics: the importance of viewpoint. *Psychol Res* 79(6):989–999. doi:[10.1007/s00426-014-0625-x](https://doi.org/10.1007/s00426-014-0625-x)
86. Tistarelli M, Grosso E, Meuwly D (2014) Biometrics in forensic science: challenges, lessons and new technologies. In: Cantoni V, Dimov D, Tistarelli M (eds) *Biometric authentication: first international workshop, BIOMET 2014, Sofia, Bulgaria, June 23–24, 2014. Revised Selected Papers*. Springer International Publishing, Cham, pp 153–164. doi:[10.1007/978-3-319-13386-7\\_12](https://doi.org/10.1007/978-3-319-13386-7_12)
87. Tistarelli M, Cadoni M, Lagorio A (2016) Matching reconstructed 3D face shapes. Personal communication
88. Tistarelli M, Cadoni M, Lagorio A, Grosso E (2016) Blending 2D and 3D face recognition. In: Bourlai T (ed) *Face recognition across the imaging spectrum*. Springer International Publisher, pp 305–331. doi:[10.1007/978-3-319-28501-6\\_13](https://doi.org/10.1007/978-3-319-28501-6_13)

89. Tome P, Fierrez J, Vera-Rodriguez R, Nixon MS (2014) Soft biometrics and their application in person recognition at a distance. *IEEE Trans Inf Forensics Secur* 9(3):464–475. doi:[10.1109/TIFS.2014.2299975](https://doi.org/10.1109/TIFS.2014.2299975)
90. Tome P, Fierrez J, Vera-Rodriguez R, Ramos D (2013) Identification using face regions: application and assessment in forensic scenarios. *Forensic Sci Int* 233(1–3):75–83. doi:[10.1016/j.forsciint.2013.08.020](https://doi.org/10.1016/j.forsciint.2013.08.020)
91. Tome P, Vera-Rodriguez R, Fierrez J, Ortega-Garcia J (2015) Facial soft biometric features for forensic face recognition. *Forensic Sci Int* 257:271–284. doi:[10.1016/j.forsciint.2015.09.002](https://doi.org/10.1016/j.forsciint.2015.09.002)
92. Van der Lugt C (2001) Earprint identification. Elsevier's-Gravenhage
93. Venugopalan S, Prasad U, Harun K, Neblett K, Toomey D, Heyman J, Savvides M (2011) Long range iris acquisition system for stationary and mobile subjects. In: 2011 international joint conference on biometrics (IJCB), 11–13 Oct 2011, pp 1–8. doi:[10.1109/IJCB.2011.6117484](https://doi.org/10.1109/IJCB.2011.6117484)
94. Willis SM et al (2015) ENFSI guideline for evaluative reporting in forensic science. European Network of Forensic Science Institutes, Dublin. [http://enfsi.eu/sites/default/files/documents/external\\_publications/m1\\_guideline.pdf](http://enfsi.eu/sites/default/files/documents/external_publications/m1_guideline.pdf)
95. Yamashita BA, Kennedy RB (2009) Forensic barefoot comparison. In: Li SZ, Jain A (eds) Encyclopedia of biometrics. Springer, New York. doi:[10.1007/978-0-387-73003-5\\_181](https://doi.org/10.1007/978-0-387-73003-5_181)
96. Zhang L, Zhang L, Zhang D, Zhu H (2010) Online finger-knuckle-print verification for personal authentication. *Pattern Recognit* 43(7):1571–2560. doi:[10.1016/j.patcog.2010.01.020](https://doi.org/10.1016/j.patcog.2010.01.020)
97. Zhang L, Zhang L, Zhang D, Zhu H (2011) Ensemble of local and global information for finger-knuckle-print recognition. *Pattern Recognit* 44(9):1990–1998. doi:[10.1016/j.patcog.2010.06.007](https://doi.org/10.1016/j.patcog.2010.06.007)

**Part I**

**Analysis of Fingerprints  
and Fingermarks**

# **Chapter 2**

## **Capture and Analysis of Latent Marks**

**Mario Hildebrandt, Jana Dittmann and Claus Vielhauer**

**Abstract** The capture and analysis of latent marks in forensics has a history of over a century. The focus of this chapter is on the marks formed by fingerprint patterns. The process starts with the detection and acquisition of the marks using different physical, chemical and optical means. Additionally, experimental approaches for determining the age or exploiting the finger mark persistency, as well as digital enhancement techniques are described. Afterward, the analysis is performed in four steps. Here, features on three different levels are determined, evaluated and compared between two fingerprint patterns. The result of such a comparison, also known as dactyloscopy, is nowadays either an identification, exclusion or inconclusive. In the future those outcomes might be replaced with a likelihood ratio which allows for expressing uncertainties on a statistical foundation. In order to use new methods in court, particular requirements must be assessed. For this, the Daubert challenge, which includes the assessment of several factors by a judge, is briefly summarized.

---

M. Hildebrandt (✉) · J. Dittmann · C. Vielhauer

Department of Computer Science, Research Group Multimedia and Security,  
Otto-von-Guericke-University of Magdeburg, PO Box 4120, 39016 Magdeburg, Germany  
e-mail: mario.hildebrandt@iti.cs.uni-magdeburg.de

J. Dittmann  
e-mail: jana.dittmann@iti.cs.uni-magdeburg.de

C. Vielhauer  
e-mail: claus.vielhauer@iti.cs.uni-magdeburg.de

J. Dittmann  
The University of Buckingham, Buckingham, UK

C. Vielhauer  
Department of Informatics & Media, Brandenburg University of Applied Sciences,  
PO Box 2132, 14737 Brandenburg, Germany

## 2.1 Introduction

The acquisition and analysis of latent marks from the perspective of fingerprints is one of the oldest disciplines in forensic science which also paved the way for the utilization of the fingerprint as a biometric modality in a manifold of applications. Of course there are known differences in both domains such as briefly summarized as follows:

- (1) The fingerprint trace (finger mark) is formed by human secretions (e.g., invisible human sweat) left by a finger touching a surface (also called latent mark) or caused due to the removable, adherent or deformable nature of the a surface property (such as other liquids on the surface, e.g., visible blood or ink, surface dust—also called often patent mark, or ductile and plastic surface behavior also called plastic mark (see [1, pp. 105–106]). Depending on the visibility of the mark, the fingerprint characteristics might be invisible to the human eye. As the presence and location of fingerprint traces is not known a priori in forensics, the marks need to be searched for and found (detected) at first. Due to the small size of the marks and very often its invisible (latent) nature on the surface, the search needs to include mechanisms to turn the mark visible and/or accessible for further acquisition techniques. This can be done either by using a physical or chemical treatment, see details in the third section, allowing for acquiring camera photos, or with contact-less methods which exploit properties of the residue using special sensory such as optical, nanometer range sensing or special lighting conditions (UV or infrared, see also in the following forth section) further combined with an additional so-called digital treatment. Contact-less, non-destructive approaches are of course very valuable allowing trace processing several times without altering the trace itself.
- (2) For trace detection as well as for analysis after acquisition the trace needs to be separated or segmented from the surface to see if the finger mark contains a sufficient number and quality of forensically relevant characteristics (see Sect. 2.2).
- (3) The analysis is usually performed by a forensic expert (called dactyloscopy expert or latent print examiner) supported with further means (microscopes, digital processing, etc.) as described in Sect. 2.5.
- (4) Additional forensic challenges which occur in the case work are for example:
  - (a) **overlapping marks:** Detection of overlapping marks, determination of the kind of overlapping traces (e.g., when other traces such as fibers are expected) and/or number of overlapping marks, separation of the marks with its quality assessment (see e.g. [2, 3]).
  - (b) **trace age estimation:** Determination of absolute or relative time the trace was left at the crime scene, the trace age is valuable to exclude or include traces which are, e.g., older or younger at time of interest, as summarized in [4].

(c) **forgery detection:** As known and discussed in several publications such as Harper [5] or Wertheim [6] and more recently [7], crime scene trace forgeries are present and need to be detected. In [8] for example several artificial sweat printed fingerprints and corresponding forgery detection approaches are summarized and benchmarked.

In comparison to the forensic case, in the biometric case, the fingerprint is explicitly and intentionally presented to the biometric sensory, producing a so-called exemplar fingerprint. Therefore, the detection from the forensic case (1) is limited here to searching and locating the fingerprint in a particular limited sized sensor region. In respect to (2), depending on the kind of biometric sensory used, the surface separation and segmentation is fixed and well defined. The analysis from (3) is usually performed by a biometric algorithm.

In respect to further challenges (4):

(a) overlapping marks occur in case of contact-based sensing, where remaining secretions from a preceding touch remains on the sensory. Here it might be valuable that the sensory detects that there are disturbances and is doing or requesting a cleaning as solution in this case.

In biometrics, the mark age (trace age (4)(b)) can be determined explicitly during sensor acquisition with the sensor time. For both domains, the age of the subject might be also of interest, see for example discussion about challenges caused by aging of fingerprint in [9].

Forgery detection (c) is also relevant for biometric applications. The kind of forgeries differs in respect to the time and means, such as artificial sweat printings cannot be easily placed in front of a biometric sensory to forge it, but 3D fingers of course can be produced to forge both domains. Liveness detection or presentation attack detection is therefore a relevant feature in the biometric domain, while in forensic domain even marks and prints from dead subjects might be of interest in order to link them to a crime or to identify corpses.

In the forensic case the finger mark is evaluated in order to determine whether the mark's characteristics belong to a particular known subject by comparing it either with exemplar prints taken under controlled conditions directly from the subject or with other latent marks found at crime scenes where the subject's identity is not known, yet. Nowadays, potential outcomes of this comparison are an inconclusive result if an insufficient number of usable features is present and an exclusion for sufficiently different patterns or an identification if the mark has a sufficient number of characteristics in common with the pattern it has been compared to. In the future alternative measure likelihood ratios might be used (see e.g. [10]). In the biometric case, a verification or identification is performed by using samples from a template database which has to be created in advance in the enrollment stage (called enrollment sample(s), reference(s), template(s)). Known error rates are the false acceptance rate (FAR, also known as type 1 error) and false rejection rate (FRR, also known as type 2 error) as well as the failure to enroll rate (FTE) and failure to acquire rates (FTA), additionally the equal error rate (EER) is used to describe the performance of a biometric system [11, pp. 6–12].

An additional challenge here is the sensor dependency of template and test samples and cross-sensor evaluations are of interest (see e.g. [12]).

This chapter provides a brief overview over the state of the art of the acquisition and analysis of latent marks. In line with the literature, e.g. [1], the term finger mark is used for the latent trace at the crime scene, whereas the term fingerprint is used for an exemplar print captured directly from the finger e.g. within a biometric system. The term fingerprint pattern is used for the pattern of both, fingerprints as well as finger marks.

The remainder of this chapter is structured as follows: Sect. 2.2 summarizes the fingerprint characteristics as a foundation for the comparison. The current state of the art regarding conventional latent mark acquisition techniques is discussed in Sect. 2.3. An overview of contact-less, non-destructive acquisition techniques is presented in Sect. 2.4. In Sect. 2.5 the finger mark analysis process is discussed. Afterward, particular legal challenges for new techniques are discussed in Sect. 2.6. Subsequently, the chapter is summarized in Sect. 2.7.

## 2.2 Fingerprint Characteristics

The foundations for the analysis of fingerprints and marks are primarily established by Johann C. A. Mayer [13], Johann E. Purkinje [14], William Herschel [15], Henry Faulds [16], Francis Galton [17], Edward Henry [18], Edmond Locard and Salil Chatterjee whereas the usage of fingerprints in general dates back several centuries, see e.g. [19] for historical timeline of forensic sciences.

Mayer [13, pp. 5–6] described the uniqueness of the fingerprint pattern for the first time in his explanation of the second copper-plate: “*Although the wrinkles of two humans are never coincide with each other, nevertheless some humans have more similarities whereas others seem to differ more in their visual appearance. However, the peculiarities of all arrangement are similarly formed.*”

Purkinje [14] describes a classification of nine global patterns of the fingerprint: simple arch, tented arch, oblique stripe, oblique loop, almond, spiral, ellipse, circle and double whorl. Starting in 1858, Herschel [15] utilized the fingerprint patterns as a means of identification in India. The uniqueness of the skin furrows is also described by Henry Faulds [16] based on experiments with fingerprints from monkeys and humans. It is also one of the first public articles considering the usage of fingerprints for performing individualization at crime scenes.

Galton [17] describes different kinds of arches, loops and whorls as global patterns, as well as minutiae points and pores. He also discusses the evidential value of fingerprints, indexing methods and personal identification. Galton also performed experiments on the fingerprint pattern persistency.

Henry [18] developed a classification system for fingerprints in collaboration with Galton. He differentiates between the delta and core points of the global level one pattern. For his primary classification he differentiates between loops (including arches) and whorls for five pairs of two fingers [18, pp. 69–75].

Locard established the first rules towards a minimum number of minutiae points which are necessary for identification (see e.g. [20]) in 1914. Furthermore, he established the science of the poroscopy [21] in 1912. However, his most important contribution to forensic sciences is probably the formulation of his exchange principle (see e.g. [22, p. 44]), which is the foundation for many other forensic disciplines as well. It basically states that every offender inevitably leaves traces at the scene of crime and takes traces from it with him as well.

Chatterjee described the analysis of the edges of the ridge lines, known as edgeoscopy [21], in 1962.

Nowadays, as known, three different levels of features are used within the scope of fingerprints [1, pp. 15–20]. The first level describes the global pattern which is visible on the fingertip even to the bare eye. The second level describes local characteristics, known as minutiae points. Those particular features are primarily used for matching fingerprints. The third level of features describes microscopic details such as pores or the edges of papillary lines. In the following the characteristics for each feature level are described.

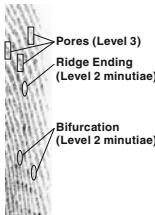
The first level of fingerprint features has already been used e.g. by Galton [17] and Henry [18] for their classification systems. In particular, usually the five different global pattern types left loop, right loop, whorl, (plain) arch, tented arch are used.

In forensic investigations those patterns can be used for a quick exclusion. However, a matching level 1 pattern is insufficient for a verification or identification of individuals. Besides their different visual appearance the patterns share some properties regarding the number of core and delta points. The delta points are characterized by a triangular-shaped region within the ridge flow. Thus, the ridge flow has three different orientations within close proximity of this point. The core point is a point where the ridge orientation significantly changes or, in other words, a point with a non-continuous ridge flow in its neighborhood.

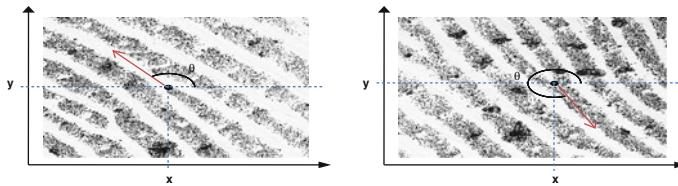
An arch has no core and delta points at all. Tented arches, left loops and right loops have one core and one delta point. Whorls have two core points and two delta points.

The level 2 patterns describe local characteristics of the ridge flow. The most common level 2 features are called minutiae points. Other features include warts, scars, creases, wrinkles, incipient ridges and subsidiary ridges [1, p. 17]. The two most common minutiae types are ridge endings (a termination of the papillary line) and bifurcations (a splitting point of one papillary line into two ridges) as illustrated in Fig. 2.1.

Usually four types of information are stored for each minutiae point: the x and y coordinates, the minutiae type and the orientation of the minutiae. Depending on the utilized format the origin of the coordinate system can be in the lower left or upper left corner of the image. Furthermore, the coordinates can be stored as pixel values or in metric units. Especially the latter has the advantage of achieving a resolution independent template. The orientation of the minutiae is determined as depicted in Fig. 2.2.



**Fig. 2.1** Illustration of minutiae and pores (re-sketched)



**Fig. 2.2** Minutiae orientation: ridge ending (*left*) and bifurcation (*right*) (re-sketched)

For ridge endings the angle of the minutiae is determined by measuring the angle  $\theta$  between the perpendicular line through the  $y$  coordinate and the prolonged ridge line. For bifurcations the angle is determined between the perpendicular line through the  $y$  coordinate and the prolonged valley between the two ridges.

On the third level of features, primarily small details are investigated. Such features include pores as illustrated in Fig. 2.1 and edges of the ridges. The pores are formed by the sweat glands on the ridges. Due to their nature, a high acquisition resolution of at least 1000 ppi is required in order to be able to extract such features reliably. Common biometric systems usually take no advantage of those features, see e.g. [23]. In contrast to that, especially for partial finger marks, in forensics such features can make a difference between an inconclusive comparison and an exclusion or identification of the mark.

### 2.3 Conventional Latent Mark Acquisition Techniques

Several latent mark detection and acquisition techniques exist and are applied in daily police work. Such methods are usually necessary to render the invisible residue from the fingertip visible. Some of those methods are also combined with special illumination techniques for further contrast enhancement.

The detection techniques are usually divided into methods for non-porous and porous substrates. On non-porous substrates the residue resides on top of the object with the finger mark. In contrast to that, the residue is absorbed over time into

porous substrates. A detailed description of the multitude of detection techniques is provided in [1, pp. 111–173] and [24]. Due to the nature of these techniques the latent mark is inevitably altered in terms of its chemical composition during the processing. Therefore, the forensic laboratories implement the ISO 17025 Standard to produce consistent results [24]. In the following, the most commonly applied known techniques for the latent mark development such as powdering, cyanoacrylate fuming, vacuum metal deposition, ninhydrin spraying, physical developer and multimetall deposition are summarized.

The oldest detection technique for smooth non-porous substrates is the powdering of the finger marks, see e.g. [1, pp. 136–137]. During this physical treatment powder particles are applied to the mark using brushes in order to detect the mark. The particles adhere to the greasy, sticky or humid substances in the residue and thus, render the mark visible. Afterward, the mark can be photographed and/or lifted using adhesive tape or gel lifters. Due to the low cost and low requirements regarding equipment and special training, this technique is still the most commonly applied method for latent marks detection at crime scenes. However, special substrate characteristics might require different powders in terms of color or other properties [1, pp. 136–137]. Other special forms of powdering include magnetic powders, which should cause less destruction by avoiding the brushing and wet powder, which can be applied on wet substrates in contrast to standard powdering techniques [1, p.137].

A chemical treatment for non-porous substrates is the cyanoacrylate fuming, also known as super glue fuming. This technique uses the effect of the polymerization of cyanoacrylate vapor at the fingerprint ridges. In particular, it reacts with eccrine and sebaceous components within the residue. The process is performed in special fuming chambers which provide a defined level of humidity and controlled heat sources for vaporizing the cyanoacrylate. It is the most commonly used technique for the latent mark development on non-porous substrates in laboratories [1, p. 138]. The result of this process is a hard white polymer on the ridges. Thus, after the fuming often an additional staining is applied to enhance the visibility of the developed marks.

The vacuum metal deposition technique is another method for visualizing latent marks on non-porous substrates [1, pp. 145–147]. The key aspect of this method is the effect that the finger mark residue hinders the deposition of metallic films. In this technique gold is evaporated under vacuum to form a thin layer of metal on the surface of an object. This layer is deposited across the surface and penetrates the residue which forms the mark. In a second step zinc is deposited in the same manner on the layer of gold. In contrast to the gold layer, this layer of zinc does not penetrate the fingerprint residue. Thus an increased contrast between the substrate and the mark is achieved. The result is usually a negative mark. An additional advantage of this technique is that it can be applied even after fuming the mark with cyanoacrylate.

On porous substrates the application of ninhydrin is one of the common detection techniques [1, pp. 114–124]. Ninhydrin reacts with amino acids, proteins and peptides within the residue resulting in a dark purple color. In the development

process the substrate is often dipped into a solution of the reagent. Alternatively, the solution can be applied by spraying or brushing. Afterward, the samples are dried and stored at room temperature for several hours for developing the purple color caused by the chemical reaction. Besides ninhydrin, several analogs can be applied as well; diazafluorene (DFO), see e.g. [1, pp. 128–131] is one example for such an analog reagent.

An alternative to the ninhydrin treatment is the utilization of a physical developer [1, pp. 131–133]. This particular technique is sensitive to water insoluble components of the residue and hence applicable for wet surfaces as well. The technique is based on the photographic physical developer. Once the substrate is placed in the reagent, silver is slowly depositing on it. The residue increases the amount of the deposited silver, resulting in darker areas on the developed surface.

The multimetal deposition (MMD) technique is a two-step process [1, pp. 133–134]. In the first step gold is deposited on the fingerprint residue within a colloidal gold solution. In the second step a modified physical developer solution is used to amplify the visibility of the treated mark. This technique can be applied on porous and non-porous substrates.

After the development of the latent mark, it is usually acquired by taking photographs [24, pp. 3–38, 289–321]. Those photos are afterward developed or printed for the following investigation steps. Depending on the characteristics of the developed mark special illumination techniques can be applied for an additional enhanced contrast.

## 2.4 Contact-Less Latent Mark Acquisition Techniques

The contact-less acquisition of the latent marks has the same purpose as conventional acquisition techniques—detecting the mark and making it usable for a forensic analysis by latent print examiners (LPEs). The main difference is the selection of the means of processing the mark. In particular, contact-less techniques exploit physical/optical properties and the chemical composition of the residue which can be sensed without a direct contact with the mark. However, not every contact-less acquisition method can be considered non-destructive, see e.g. [24, pp. 294–295]. Especially the utilization of UV radiation is known for its potential impact on DNA traces within the mark. Several known contact-less approaches are selected and summarized in the remainder of this section.

Early experiments with detecting the inherent fingerprint luminescence with lasers date back to 1977 [25]. However, the high performance could only be reproduced for contaminated residues. Without such a contamination, less than 20% of the marks allowed for the detection of fingerprints by laser excited luminescent [26]. Similar to the luminescence the fluorescence of fingerprints can be investigated as well [24]. However, the quality of the results for untreated marks also depends on particular contaminations.

Other approaches exploit the diffuse reflection of light on the residue. Such approaches can be applied to recover latent marks from smooth, non-porous, glossy substrates such as glass, plastic or polished metals [1, p.112] or for recovering marks in dust [24]. Such examples include oblique illumination setups, dark field illumination, coaxial illumination and the usage of polarized light (see [24, 27]). Those illumination techniques are already applied in forensic practice to enhance the contrast of latent and developed marks.

Various ultraviolet imaging techniques are also used in today's police practice. The most commonly used UV imaging technique is the reflected UV imaging system (RUVIS), which usually employs shortwave UV radiation [24, pp. 289–299]. This approach utilizes specific UV absorption characteristics of the substrate and the residue as well as the diffuse reflection from the residue [1, pp. 112–113]. The results depend on the substrate, the angle of the light source and the type of the light source.

Another spectrum is used by the EVISCAN system [28]. The system employs a heat source emitting long wavelength infrared radiation and a high-resolution thermal vision camera to capture the diffuse reflection of IR radiation from the fingerprint in an oblique top measurement setup.

For glossy, curved substrates, [29] employ a gloss-meter and a rotary stage to acquire the finger marks. Such systems have the advantage of compensating the perspective distortion which would occur in normal 2D imaging.

The technique of optical coherence tomography can be used to detect latent marks covered by a layer of dust [30]. The sensor is able to produce 3D volume data of light scattering material.

The approach described in [31] utilizes the chemical composition of the residue in conjunction with Fourier transform infrared spectroscopy (FTIR). This technique allows for acquiring latent marks from various porous and non-porous substrates. However, the sensory has significant limitations regarding the size of the samples.

Within the scope of the German Federal Ministry of Education and Research funded research project "DigiDak" reflection based sensors are employed as well. The experiments use three different measurement principles: hyperspectral imaging in the UV to NIR range [32], chromatic confocal sensing [33] and confocal laser scanning microscopy [34]. The latter two techniques additionally allow for capturing height information (topography) from the latent marks or other types of traces. Due to the perpendicular illumination and measurement, those sensors primarily sense an intensity reduction on the areas covered with residue caused by the previously mentioned diffuse reflection of the latent mark.

As an outcome of this research project, a data set of computer generated finger marks is available upon request [35]. This particular data set consists of fingerprint patterns generated using SFinGe [36] which have been afterward printed on an overhead foil using artificial sweat and standard inkjet printers. Subsequently, each of the 24 samples has been contact-less acquired using a FRT MicroProf 200 surface measurement device equipped with a FRT CWL 600 sensor with a resolution of 500 and 1000 ppi.

A specific advantage of the non-destructive acquisition is the possibility to observe a mark over an interval of time by acquiring a series of scheduled images of it, called time series in [4]. This is a foundation for estimating the age of the mark by determining its speed of degradation [4]. Additionally, such degradation can be used to determine the persistence of finger marks on the surface of specific substrates.

Estimating the age of latent marks is an old challenge which has not been solved, yet. If it is possible to determine the age of a mark, the evidential value would increase significantly because it would be possible to prove that an individual has been at the crime scene at the time of the crime. In [4] time series of latent marks covering the first 24 h after the placement of the mark are the foundation for extracting 17 features. The feature space consists of binary pixel features from intensity and topography images as well as statistical features. The individual aging speed is determined by calculating ten consecutive feature values for each sample. The experimental setup consists of numerous donors and various influence factors. In the experiments two disjunct time classes are defined in order to determine whether the mark is younger or older than 5 h with different classifiers. Here, the classification accuracy for single features varies between 79.29% in the best case and 30.02% in the worst case. However, the performance can be slightly improved up to 83.10% if non-deterministic aging tendencies are excluded from the classification. For the combined feature space the classification accuracy varies between 65.08 and 79.79% depending on the amount of excluded time series.

Besides the age estimation of finger marks, the degradation of the residue of the mark can be used to locate traces within a temporal feature space. This particular location approach is motivated by the persistence of fibers at crime scenes. The foundation for the temporal feature space for latent marks are spectral texture features which are observed over a series of consecutive scans. For comparison, this feature space is also applied in the spatial domain. The general approach of using temporal features is similar to the age estimation. However, there are some differences within the processing pipeline. In contrast to the age estimation where a set of features is extracted from a series of samples, each sample is separated into blocks of  $2 \times 2$  mm. Each block allows for determining the individual tendencies within the feature space in the region it is covering. In [37] the experiments are performed within low resolution scans (125 ppi) covering an area of  $20 \times 20$  cm of the three (semi-)porous substrates copying paper, photographic paper and a textured catalog paper. Each substrate is prepared with 60 latent prints from six test subjects. The low acquisition resolution is necessary in order to achieve a reasonable acquisition time for each scan of 2.1 h for copying paper and 1.1 h for reflective photographic and catalog papers. The results for the three investigated substrates in [37] show an improved performance of the temporal feature space in comparison to the spatial feature space. The largest performance gain of 6.7% points is achieved when eight consecutive images are used to determine the temporal features. However, the results in [37] also indicate that a large number of consecutive scans might lead to a deteriorated performance as well.

## 2.5 Latent Mark Analysis Process

After the acquisition of the mark, usually a digital preprocessing is applied for an additional enhancement of the fingerprint pattern. Depending on the substrate, various artifacts might interfere with the pattern of the mark and the contrast within the image. Thus, emphasizing the fingerprint pattern can be quite challenging because the fingerprint is not necessarily the dominating pattern within the image as summarized in [33]. Hence the digital preprocessing of latent marks often significantly differs from the enhancement of exemplar prints as used in biometric systems. Moreover, it is not possible to ask the subject to acquire a sample with a better quality. These particular challenges are addressed in [38] for conventionally captured latent marks within the NIST Special Database 27 [39] or [40] for contact-less acquired marks using image processing and pattern recognition approaches.

The approach in [38] consists of a manual markup of the ROI, core and delta points, the block-based computation of multiple dominant orientations utilizing the short-time Fourier transform, the orientation field estimation using R-RANSAC and subsequently the enhancement by employing Gabor filters. The hypothesis creation and evaluation for determining the orientation fields using R-RANSAC is necessary to detect a plausible ridge flow. Otherwise parts of the background orientations might be mixed with the fingerprint orientations which possibly alter minutiae points. The last step of utilizing Gabor filters is known from the processing of exemplar prints in biometric systems as well. Here, on the foundation of the local ridge frequency and orientation the fingerprint pattern can be emphasized.

The approach in [40] utilizes the whole amount of available data from the contact-less sensory, namely the intensity and topography image. In contrast to [38] no human interaction is necessary to process the image of the mark during its processing. The first step of this approach is the preprocessing of each input image using various Sobel operators, unsharp masking and various Gabor filters. Each preprocessed image is afterwards processed separately. The feature extraction for the pattern recognition based decision whether a block contains fingerprint residue or not is performed in blocks of  $50 \times 50 \mu\text{m}$ . The feature space consists of statistical, structural and fingerprint semantics features. Afterward, each block is classified using pre-trained models. Subsequently, a fingerprint image can be reconstructed based on the classifiers decisions. The primary challenge of this approach is the training of the classifiers. This step is time-consuming and usually requires human interaction in order to get a ground-truth for the supervised learning. The evaluation is performed for eight different substrates ranging from the rather cooperative white furniture surface to the very challenging blued metal and golden oak veneer. In the best case, a classification accuracy of 95.1% is achieved. On golden oak at least an accuracy of 81.1% is achieved. However, an automated biometric feature extraction and matching is only successful for fingerprints from three of the eight substrates.

Another challenge for the preprocessing is the handling of overlapping marks. Such marks frequently appear on locations which are often touched. Such places can be door handles, elevator buttons or touch screens. In current police practice such traces are usually discarded if the non-overlapped region of the mark does not contain a sufficient amount of minutiae points.

One of the first approaches addressing this challenge to separate two overlapping patterns is published in [2]. This approach is designed for conventionally captured marks. During the first step it is necessary to mark the region mask manually. Here, the two non-overlapped and the overlapped regions are defined by the user. Afterward, the initial orientation field is determined which contains two orientation labels within the overlapped region. The separation of the overlapped orientation field is performed using a relaxation labeling algorithm resulting in two component fingerprints which need to be constructed using compatibility coefficients based on the local neighborhood of each block or object. After this step, the separated orientation fields are merged with the orientation fields of the non-overlapped regions.

An extended separation approach for conventionally and contact-less acquired marks is proposed in [3]. Here, the context-based computation and optimization of parameters are introduced, e.g. for accounting for different acquisition resolutions. The biometric evaluation of the separation results show an equal error rate of 5.7% for contact-less acquired marks, and 17.9% for conventionally captured marks. These results show that the separation benefits from the increased acquisition resolution of 2540 ppi of the contact-less acquired latent marks.

For the analysis of the latent marks, usually the ACE-V process (see [41]) or variations of it, is applied. ACE-V is an abbreviation for the four processing stages: analysis, comparison, evaluation and verification.

During the first step, the analysis, the acquired latent mark is investigated regarding the overall quality and the clarity and number of usable features. Several influence factors, such as the substrate, the development medium, assumed deposition pressure or distortions, are taken into account in this step. If the latent print examiner comes to the conclusion that the quality is too low, e.g., due to a lack of sufficient features, the mark is discarded as insufficient without performing the remaining steps of the ACE-V process. Otherwise it is compared to a reference print in the following step.

The comparison step (2nd step) involves a side-by-side comparison between two fingerprint patterns—usually represented by the latent mark from a crime scene and a reference print with a known origin connected with an identity. During the first comparison step, the level 1 pattern is compared. The comparison can be aborted if those patterns do not match; if the patterns are identical or not visible in one of the samples, the comparison is continued by creating relationships between level 2 features. However, due to potential distortions and low-quality areas this is usually no physical measurement with a fixed scale and thus hard to automate. The features are usually matched by following ridge paths or by counting ridges between two feature points. The matching itself is a subjective process which requires extensive

training. Especially for poor quality marks the tolerance for variations is increased, which usually requires an increased number of features in order to decide about the result of the comparison in the next step.

The third step for the latent print examination is the evaluation. This step contains the formulation of the final conclusion based on the analysis and comparison steps of the samples. The examiner has to decide whether both patterns originate from the same source. This also requires the explanation of differences and variations found during the comparison. If the level 1, 2 and 3 features are sufficiently similar between the two prints the conclusion is called individualization. Likewise, if a number of different features are found and not explainable, e.g., by distortions, the conclusion is called exclusion which means that the patterns originate from different fingers. The result is marked as inconclusive if the examiner is unable to make a decision beyond reasonable doubt, e.g., due to a lack of matching and non-matching features.

The last (fourth) step of the ACE-V process is the verification. This step is performed because the comparison of the two fingerprint patterns is a subjective process. The verification consists of an independent analysis, a comparison and an evaluation of the samples by a second examiner. It is intended to increase the reliability of the final conclusion by reducing the subjectivity of the decision. This ideally means that the second investigator is unaware of the initial investigation results. If both examiners warrant the same conclusion, the examination of the latent mark is finished. Otherwise, a third examiner might repeat the investigation or the outcomes can be discussed and reevaluated by the two examiners in order to find an agreement.

The requirements for the decision-making are different in various countries. Currently two standards exist [42]: the numerical standard and the non-numerical standard.

The numerical standard defines a minimum number of matching level 2 feature points in order to draw the conclusion of identification. However, the threshold regarding the number of necessary features varies between 7 and 17 [42, p. 47]. This fixed threshold does not account for the rarity of particular feature points. Thus, the conclusion can be drawn from a specific number of rather common feature points or from rarer features as well. In order to account for this discrepancy in the evidential value of the features, some countries have switched to non-numerical standards [43]. This results in dynamic thresholds based on, e.g., the specificity of the features considering the rarity of feature points and the relationship to other points. In other words: a smaller number of matching rare features can suffice for drawing the conclusion of identification whereas a large number of matching common features might result in an inconclusive outcome. Thus, the non-numerical standard provides an increased flexibility in the decision making. However, on the other hand this results in an increased responsibility for the examiner and the requirement to provide sufficient statistical data and sophisticated knowledge of the state of the art to back the decision.

## 2.6 Legal Challenges of Applying New Techniques in the Latent Mark Processing

Each new procedure in forensics needs to be evaluated regarding their suitability and scientific foundation. In the US, a so-called Daubert challenge (see [44]) is known which can be used to assess the suitability of a scientific method prior to the admission of the evidence in court. In such a Daubert challenge at the Supreme Court, the judge has a role of a gatekeeper to screen scientific evidence in order to ensure that only relevant and reliable sources are admitted. In the original trial of *Daubert v. Merrell Dow Pharmaceuticals* in 1993 a list of five factors was provided, which a judge might consider during the assessment of the scientific validity of the theory or method [44]:

- *Whether it [the method] can be (and has been) tested,*
- *Whether it has been subjected to peer review and publication,*
- *The known or potential rate of error [of the method],*
- *The existence and maintenance of standards controlling the technique's operation,*
- *Whether it is generally accepted in the scientific community.*

This list has been extended after the initial Daubert decision by additional factors as summarized in [44, p. 38]. For new acquisition techniques this would require an extensive testing and the definition of particular standards. This would also require a comparison with existing techniques to show the validity of the results. Here, the contact-less acquisition techniques have the advantage that they would not interfere with any other conventional detection technique. Thus, it is possible to acquire the mark by non-destructive means in the first place and to verify the results using accepted standard procedures afterward.

However, in the context of the latent mark examination in general a critical review of court rulings is given in [45]. In the essence of [45] the fingerprint evidence merits its acceptance in court due to its long usage of over a century even if it would hardly withstand a careful evaluation of the Daubert factors. In [46] the rate of error is investigated in more detail. The author describes multiple cases of erroneous identifications indicating a non-zero error rate. However, he also states that the available data is inadequate to calculate a meaningful error rate. A solution to account for the non-zero error rate is the usage of known likelihood ratios (LR): instead of presenting the result of a binary decision, a quotient between the probabilities of two opposing hypotheses is given.

In the context of fingerprints [10], the two hypotheses are that the patterns originate from the same finger or from different fingers. Especially the variability of the pattern due to skin elasticity needs to be taken into account for determining the probabilities. In such an LR-based approach, the outcome of exclusion would relate to an LR of zero, whereas identification relates to an LR of infinity (due to the assumed error rate of zero). In practice the likelihood ratio is somewhere between those two extremes due to the variability of the patterns.

Although the first experiments in [10] look promising, they suffer from a challenge in a real-world application: it is almost impossible to calculate the probability for the two patterns originating from different fingers as the denominator for determining the LR since it would require knowing the patterns and properties of all fingers in the world. Thus, it is only possible to estimate this probability. Nevertheless, the application of LRs helps to express uncertainties in the decision process, see e.g. [47].

## 2.7 Summary

This chapter summarizes a selection of the state of the art of the latent mark acquisition and analysis. Even though the comparison of fingerprints has a long tradition in forensic investigations, several crucial questions remain unanswered to date. From a technical point of view, a broad variety of sensors allow for acquiring latent marks without the need for altering the trace composition by adding additional reagents. On the other hand, such techniques require an extensive testing in line with best practices of the police forces before they can replace the conventional techniques. Nevertheless, non-destructive sensors allow for new investigation techniques such as the age estimation or the observation of the fingerprint persistency.

The comparison of the fingerprint patterns bears several challenges as well. Especially the elasticity of the skin is a cause for uncertainty in the decision-making process. Here, it is necessary to establish a statistical foundation to determine likelihood ratios based on a common standard to be able to express and consider potential uncertainties while retaining a comparability of the resulting values.

With respect to the legal requirements there is currently no common ground. Some countries employ the numerical standard with thresholds requiring between 7 and 17 matching features [42, p. 47]. Others use non-numeric standards which account for the specificity of particular feature points.

## References

1. Champod C, Lennard C, Margot P, Stoilovic M (2004) Fingerprints and other ridge skin impressions. CRC Press
2. Chen F, Feng J, Jain AK, Zhou J, Zhang J (2011) Separating overlapped fingerprints. IEEE Trans Inf Forensics Secur 6(2):346–359
3. Qian K, Schott M, Zheng W, Dittmann J (2014) A context-based approach of separating contactless captured high-resolution overlapped fingerprints. IET Biom 3(2):101–112
4. Merkel R, Gruhn S, Dittmann J, Vielhauer C, Bräutigam A (2012) On non-invasive 2D and 3D chromatic white light image sensors for age determination of latent fingerprints. Forensic Sci Int 222(1–3):52–70

5. Harper W (1937) Fingerprint forgery—transferred latent fingerprints. *J Criminol* 28(4): 573–580
6. Wertheim PA (1994) Latent fingerprint fabrication, website request 12/02/2016, see also in Wertheim, Pat A., Detection of Forged and Fabricated Fingerprints. *J Forensic Ident*, 44(6), pp 652–681. <http://www.iowaiai.org/latent-fingerprint-fabrication/>
7. Champod C, Espinoza M (2014) Forgeries of fingerprints in forensic science, In: Marcel S, Nixon MS, Li SZ (eds) *Handbook of biometric anti-spoofing*, ser. *advances in computer vision and pattern recognition*. Springer, London, pp 13–34
8. Hildebrandt M, Dittmann J (2015) StirTraceV2.0: enhanced benchmarking and tuning of printed fingerprint detection. *IEEE Trans Inf Forensics Secur* 10(4):833–848
9. Modi SK, Elliott SJ, Whetsone J, Kim H (2007) Impact of age groups on fingerprint recognition performance. In: IEEE workshop on automatic identification advanced technologies, pp 19–23
10. Neumann C, Champod C, Puch-Solis R, Egli N, Anthonioz A, Bromage-Griffiths A (2007) Computation of likelihood ratios in fingerprint identification for configurations of any number of minutiae. *J Forensic Sci* 52(1):54–64
11. Jain AK, Flynn P, Ross A (2008) *Handbook of biometrics*. Springer, US
12. Ross A, Jain AK (2004) Biometric sensor interoperability: a case study in fingerprints. In: International workshop on biometric authentication: ECCV 2004, pp 134–145
13. Mayer JCA (1788) *Anatomische Kupfertafeln: nebst dazu gehörigen Erklärungen* (Band 4): Eilf Kupfertafeln von den Sinnwerkzeugen und den Brüsten. Berlin, Leipzig, Universitätsbibliothek Heidelberg. <http://digi.ub.uni-heidelberg.de/diglit/mayer1788bd4/0006>
14. Purkinje JE (1823) *Commentatio de examine physiologico organi visus et systematis cutanei: quam pro loco in gratioso medicorum ordine rite obtinendo die XXII. decembris MDCCCXXIII, H.X.L.C. Vratislaviae: Typus Universitatis*. <https://archive.org/details/68020950R.nlm.nih.gov>
15. Herschel W J (1916) The origin of finger-printing. Oxford University Press. <http://galton.org/fingerprints/books/herschel/herschel-1916-origins-1up.pdf>
16. Faulds H (1880) On the skin-furrows of the hand. *Nature* (Oct. 28 1880), p. 605
17. Galton F (1892) Finger prints. Macmillan and Co
18. Henry ER (1900) Classification and uses of finger prints. George Routledge and Sons <https://archive.org/details/ClassificationAndUsesOfFingerPrints>
19. Rudin N, Inman K (2016) The forensic science timeline. <http://www.forensiccdna.com/timeline.html>, Accessed 02 Dec 2016
20. Kingston CR, Kirk PL (1965) Historical development and evaluation of the ‘12 point rule’ in fingerprint identification. *Int Crim Police Rev*
21. Ashbaugh DR (1999) Ridgeology: modern evaluative friction ridge identification. Royal Canadian Mounted Police, Forensic Identification Support Section
22. Inman K, Rudin N (2000) *Principles and practice of criminalistics: the profession of forensic science (Protocols in forensic science)*. CRC Press
23. Zhao Q, Jain AK (2010) On the utility of extended fingerprint features: a study on pores. In: Computer vision and pattern recognition workshops (CVPRW), pp 9–16
24. Bleay SM, Sears VG, Bandey HL, Gibson AP, Bowman VJ, Downham R, Fitzgerald L, Ciukszta T, Ramadani J, Selway C (2012) *Fingerprint source book*. Home Office CAST, United Kingdom
25. Dalrymple BE, Duff JM, Menzel ER (1977) Inherent fingerprint luminescence—detection by laser. *J Forensic Sci* 22:106–115
26. Salares VR, Eves CR, Carey PR (1979) On the detection of fingerprints by laser excited luminescence. *Forensic Sci Int* 14:229–237
27. Lin S-S, Yemelyanov KM, Pugh EN Jr, Engheta N (2006) Polarization-based and specular-reflection-based noncontact latent fingerprint imaging and lifting. *J Opt Soc Am A* 23(9):2137–2153
28. German eForensics GmbH, EVISCAN contactless evidence detection. <https://www.eviscan.com/files/71/eviscan-brochure-lowres.pdf>, Accessed 29 Dec 2015

29. Kuivalainen K, Peiponen K-E, Myller K (2009) Application of a diffractive element-based sensor for detection of latent fingerprints from a curved smooth surface. *Meas Sci Technol* 20(7):077002
30. Dubey SK, Mehta DS, Anand A, Shakher C (2008) Simultaneous topography and tomography of latent fingerprints using full-field swept-source optical coherence tomography. *J Opt A: Pure Appl Opt*, 10(1), 015 307–015 315
31. Crane NJ, Bartick EG, Perlman RS, Huffman S (2007) Infrared spectroscopic imaging for noninvasive detection of latent fingerprints. *J Forensic Sci* 52(1):48–53
32. Makrushin A, Scheidat T, Vielhauer C (2015) Capturing latent fingerprints from metallic painted surfaces using UV-VIS spectroscope. In: Proceedings of SPIE 9409: media watermarking, security, and forensics 2015
33. Hildebrandt M, Merkel R, Leich M, Kiltz S, Dittmann J, Vielhauer C (2011) Benchmarking contact-less surface measurement devices for fingerprint acquisition in forensic investigations: results for a differential scan approach with a chromatic white light sensor. In: Proceedings of DSP'11, pp 1–6
34. Kirst S, Vielhauer C (2015) Detection of latent fingerprints using high-resolution 3D confocal microscopy in non-planar acquisition scenarios. In: Proceedings of SPIE 9409: media watermarking, security, and forensics 2015
35. Hildebrandt M, Sturm J, Dittmann J, Vielhauer C (2013) Creation of a public corpus of contact-less acquired latent fingerprints without privacy implications. *Commun Multimedia Secur*, LNCS 8099:204–206
36. Cappelli R (2015) Fingerprint sample synthesis. In: Li SZ Jain AK (eds) Encyclopedia of biometrics, 2nd edn. Springer
37. Merkel R, Dittmann J, Hildebrandt M (2014) Latent fingerprint persistence: a new temporal feature space for forensic trace evidence analysis. In: IEEE international conference on image processing (ICIP), Paris pp. 4952–4956. doi: [10.1109/ICIP.2014.7026003](https://doi.org/10.1109/ICIP.2014.7026003)
38. Yoon S, Feng J, Jain AK (2011) Latent fingerprint enhancement via robust orientation field estimation. In: 2011 international joint conference on biometrics (IJCB), pp 1–8
39. National Institute of Standards and Technology NIST special database 27 [Online]. <https://www.nist.gov/srd/nist-special-database-27>
40. Hildebrandt M, Kiltz S, Dittmann J, Vielhauer C (2014) An enhanced feature set for pattern recognition based contrast enhancement of contact-less captured latent fingerprints in digitized crime scene forensics. In: Proceedings of SPIE 9028: media watermarking, security, and forensics 2014
41. Holder EH, Robinson LO, Laub JH (eds.) (2011) The fingerprint sourcebook, US Department of Justice, Office of Justice Programs, National Institute of Justice, NCJ 225320
42. Polski J, Smith R, Garrett R, et al (2011) The report of the international association for identification, standardization II committee. Document No. 233980. <https://www.ncjrs.gov/pdffiles1/nij/grants/233980.pdf>, Accessed 30 Dec 2015
43. BBC News (2006) Court fingerprint system scrapped. [http://news.bbc.co.uk/2/hi/uk\\_news/scotland/5310246.stm](http://news.bbc.co.uk/2/hi/uk_news/scotland/5310246.stm), Accessed 30 Dec 2015
44. Dixon L, Gill B (2001) Changes in the standards for admitting expert evidence in federal civil cases since the daubert decision. rand institute for civil justice
45. Benedict N (2004) Fingerprints and the daubert standard for admission of scientific evidence: why fingerprints fail and a proposed remedy. *Ariz Law Rev* 46(3):519–549
46. Cole SA (2005) More than zero: accounting for error in latent fingerprint identification. *J Crim Law Criminol* 95(3):985–1078
47. Martire KA, Kemp RI, Sayle M, Newell BR (2014) On the interpretation of likelihood ratios in forensic science evidence: presentation formats and the weak evidence effect. In: *Forensic science international*, vol 240, pp 61–68

# Chapter 3

## Automated Fingerprint Identification Systems: From Fingerprints to Fingermarks

Davide Maltoni, Raffaele Cappelli and Didier Meuwly

**Abstract** The aim of this chapter is to present the automated fingerprint recognition technology and its use for forensic applications. After a brief historical review, we provide an introduction to modern Automated Fingerprint Identification Systems (AFIS) by discussing their functionalities and accuracy. The topic then becomes more technical and goes through some of the recently introduced approaches for fingerprint recognition (both for fingerprint and fingermarks). Forensic applications exploiting the recognition of fingerprints (identity verification and identification) and fingermarks (forensic intelligence, investigation and evaluation) are then described. Finally, a discussion about the current topics and foreseeable challenges in terms of technology and application concludes the chapter.

### 3.1 Introduction

#### 3.1.1 History

In the early twentieth century, fingerprint recognition was formally accepted as a valid personal identification method and became a standard routine in forensic science [1]. Fingerprint identification agencies were set up worldwide and criminal fingerprint databases were established [1]. Various fingerprint recognition techniques, including fingermark<sup>1</sup> processing, fingerprint classification, and finger-

---

<sup>1</sup>The finger dermatoglyphics and their standard rolled or flat inked or scanned impressions are named *fingerprints*, whereas the recovered or lifted traces are named *fingermarks* (latent fingerprints is a popular but imprecise synonym for fingermarks) [2].

---

D. Maltoni (✉) · R. Cappelli

Ingegneria e Scienze Informatiche, Università di Bologna, Via Sacchi 3, 47521, Cesena, Forlì-Cesena, Italy

e-mail: maltoni@csr.unibo.it

D. Meuwly

Services, Cybersecurity and Safety, Netherlands Forensic Institute, University of Twente, Drienerlolaan 5, 7522, NB Enschede, The Netherlands

comparison were developed. For example, in the US, the FBI fingerprint identification division was set up in 1924 with a database of 810,000 fingerprint cards (see [3, 4]).

With the rapid expansion of fingerprint recognition in forensic science, operational fingerprint databases became so huge that manual fingerprint identification turned to be infeasible. For example, the total number of fingerprint cards (each card contains one impression for each of the 10 fingers of a person) in the FBI fingerprint database now stands well over 200 million from its original number of 810,000 and is growing continuously. With thousands of requests being received daily, even a team of more than 1300 fingerprint experts were not able to provide timely responses to these requests [1]. Starting in the early 1960s, the FBI, Home Office in the UK, Paris Police Department in France, and the National Police Agency of Japan, began to invest a large amount of effort in developing automatic fingerprint identification systems [1]. In 1963, FBI Special Agent Carl Voelker asked the assistance of NIST engineers Raymond Moore and Joe Wegstein to study the first automated solution for fingerprint identification. Moore and Wegstein analyzed the manual methods used by human fingerprint experts, mainly based on comparing minutiae (i.e. ridge endings and ridge bifurcations) on fingerprint ridges [5]. If the minutiae from two fingerprints were determined to be topologically equivalent, the two fingerprints were declared to be from the same finger of the same person. Based on the above observations, three major problems in designing AFISs were identified: digital fingerprint acquisition, minutiae extraction, and minutiae comparison. During 15 years of work on this topic, Wegstein developed and improved some minutiae comparison algorithms [6–13], including the well-known M82 comparison algorithm [13]. In the same period, some companies and organizations started developing AFIS components and prototypes, responding to requests issued by government agencies. By the early 1990s, four major vendors leaded the AFIS market: Printrak and Cogent (US), Morpho (France), and NEC (Japan). AFIS rapidly evolved from initial vendor-specific prototypes to efficient and interoperable systems able to process millions identification requests per second. Today almost every law enforcement agency in the developed and most emerging countries uses an AFIS. Large-scale civil projects (such as India UIDAI) also make use of these technologies. These systems have greatly improved the operational productivity of law enforcement agencies and reduced the cost of hiring and training human fingerprint experts.

### 3.1.2 AFIS Functionalities

The basic functions of a modern AFIS are:

- *Single/multiprint identification*—identifying a single fingerprint (or a set of fingerprints) from an individual to determine where (s)he has an existing record in the system. Depending on the application and the required identification

accuracy, a variable number of fingerprints of the same individual can be used (typically 1, 2, 4, 8 or 10).

- *Fingermark search*—attempting to link fingermarks to other fingermarks or to single/multiprint cards.

In both cases the result provided by the system is a candidate list of a given length, sorted according to a comparison score. The relationship between a score and the probability of correct identification is still debated, as discussed in Sect. 3.3. Possible enhanced functions provided by recent AFIS include:

- *Palmpoint/mark search*—adding palmpoint records to the database to allow search of pampprints/marks.
- *Multimodal identification systems*—combining the recognition of several biometric modalities (fingerprint, palmpoint, iris, and face).

In forensic science these AFIS functionalities are exploited inside several processes [2]:

- **Identity verification**, that is a *decision* about the identity of a person based on the comparison of her/his fingerprints;
- **Forensic intelligence**, that is an *association* of criminal cases through fingermarks and palmmarks;
- **Forensic investigation**, that is a *selection* of candidates on the basis of the comparison of finger- and palmmarks recovered on a crime scene with a database of finger- and palmpoints;
- **Forensic evaluation**, that is a *description* of the strength of evidence resulting from the study, using a comparative approach of the similarity and the distinctiveness of a fingermark recovered on a crime scene and a fingerprint of a candidate selected during the forensic investigation process. It seeks individualization but remains probabilistic and inductive in essence [14].

### 3.1.3 *Fingerprint Identification Accuracy*

Identification accuracy in AFIS open-set scenario<sup>2</sup> is typically denoted by the value of FNIR when the system is tuned to operate at a given FPIR. Let  $t$  be a security threshold, then:

- False negative identification rate (FNIR) is the fraction of the mated searches where the enrolled mate is outside the candidate list, or is inside the list but its score is less than  $t$ ;

---

<sup>2</sup>In open-set scenarios, some of the searched users have not a record in the database (*non-mated search*), while in closed-set scenarios it is assumed to search only users with at least one record in the database (*mated search*).

- The false positive identification rate (FPIR) is the fraction of the non-mated searches where one or more enrolled identities are returned in the candidate list with score higher than  $t$ .

Another common way to characterize AFIS accuracy is Rank- $n$  metrics, denoting the percentage of mated searches where the correct identity is returned in the first  $n$  positions of the candidate list.

For more details refer to Sect. 6.1 of [15].

Several independent evaluation campaigns have been organized on fingerprint recognition since the year 2000 (see Sect. 4.7.2 of [16]). Referring to identification on large databases, one of the most comprehensive tests is FPVTE 2012 [15]. State-of-the-art algorithms from leading AFIS vendors have been submitted and evaluated on several tasks. The most accurate fingerprint identification algorithms achieved false negative identification rates (FNIR, or “miss rates”) of:

- 1.9% for single index fingers,
- 0.27% for two index fingers,
- 0.45% for four-finger identification flats (IDFlats),
- 0.15% for eight-finger IDFlats,
- 0.05% for ten-finger IDFlats,
- 0.1% for ten-finger rolled-to-rolled,
- 0.13% for ten-finger plain-to-plain,
- 0.11% for ten-finger plain-to-rolled.

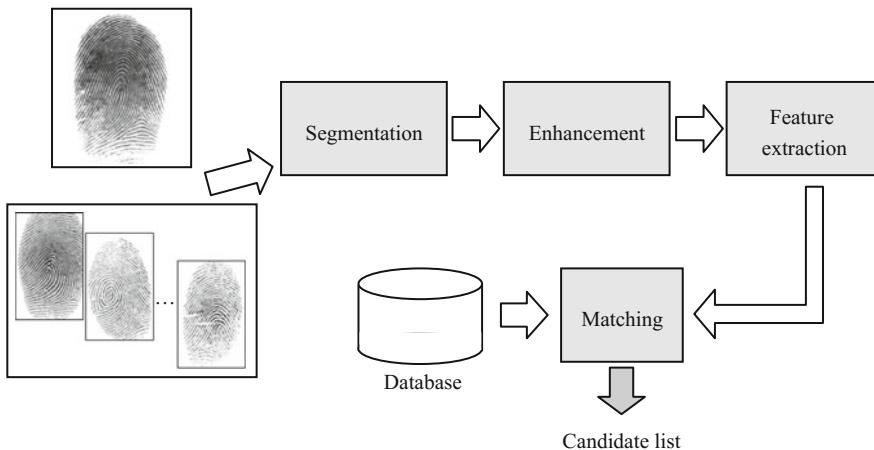
These numbers are reported at a false positive identification rate (FPIR) of 0.1%. 30,000 search subjects were used for these results (10,000 mates and 20,000 nonmates). The number of enrolled subjects used for single index fingers was 100,000, 1.6 million for two index fingers, 3 million for IDFlats, and 5 million for ten-finger plains and rolled. A larger test with a search set of 50,000 mates and 300,000 nonmates has been scheduled so that accuracy at a smaller FPIR can be reported.

## 3.2 Automated Fingerprint/Mark Technology

### 3.2.1 Fingerprints

The block diagram of a typical fingerprint-based recognition system is depicted in Fig. 3.1.

Historically, in law enforcement applications, the acquisition of fingerprint images was performed by using the so-called “ink-technique”: the subject’s fingers were covered with black ink and pressed or rolled on a paper card; the card was then scanned by using a general-purpose scanner, producing a digital image. This kind of acquisition process is referred to as off-line fingerprint acquisition. Nowadays, most AFIS accept live-scan digital images acquired by directly sensing



**Fig. 3.1** Block diagram of a fingerprint-based recognition system, where the input can be a single fingerprint or a set of fingerprints

the finger surface with an electronic fingerprint scanner. Appropriate image quality requirements for AFIS fingerprint scanners were defined by the FBI in Appendix F of the Electronic Fingerprint Transmission Specification (EFTS) [17]. Depending on the operating modality (single or multiprint identification), the input consists of one or more input raw images. Single fingerprints can be acquired as *plain* (or *flat*) representations or in *rolled* representation, which is characterized by an unwrapped nail-to-nail pattern.

In the following, the main processing steps applied to each raw image are summarized. The raw image is first passed to an image enhancement module, whose goal is to improve the clarity of the ridge pattern, especially in noisy regions, to simplify the subsequent feature extraction. Special digital filtering techniques, known as contextual filtering [16], are usually adopted at this stage. The feature extraction module further processes the enhanced image and extracts a set of features from it. This feature set often includes minutiae but, depending on the comparison algorithm, other features (e.g. local orientation, local frequency, singularities, ridge shapes, ridge counts, parts of the enhanced image, sweat pores, etc.) can be extracted too. It has been recently proved that using extended features [18, 19] in combination with minutiae can lead to better accuracy [20, 21].

Finally, the matching module compares the extracted features against all records in the database. Comparing two fingerprints typically requires finding the spatial correspondence of a minimum number of minutiae; this is not a simple task because of the large variations (e.g. displacement, rotation, skin condition, distortion, noise, etc.) that can characterize two fingerprints acquired from the same finger at different times. Earlier comparison algorithms try to find a suitable alignment by maximizing minutiae pairing at global level (i.e. throughout the whole fingerprint pattern). Nowadays the most effective algorithms first compare local arrangements of

minutiae and then consolidate them at global level [16]. This improves efficiency and better tolerates skin distortion.

A recently proposed effective algorithm is based on the Minutia Cylinder-Code (MCC) representation [22]. MCC builds a local descriptor for each minutia  $m_i = (x_i, y_i, \theta_i)$ , where  $(x_i, y_i)$  is the minutia location and  $\theta_i$  is the minutia direction (in the range  $[0, 2\pi]$ ). The descriptor encodes spatial and directional relationships between the minutia and its neighbourhood of radius  $R$ , and can be conveniently represented as a cylinder, whose base and height are related to the spatial and directional information, respectively (see Fig. 3.2a, b). The cylinder is divided into  $N_D$  sections: each section corresponds to a directional difference in the range  $[-\pi, \pi]$ ; sections are discretized into cells ( $N_S$  is the number of cells along the section diameter). During the cylinder creation, a numerical value is calculated for each cell: it represents the likelihood of finding minutiae that are close to the cell and whose directional difference with respect to  $m_i$  is similar to a given value. Figure 3.2c, d shows the cylinder associated to a minutia with six minutiae in its neighbourhood. Once a cylinder is built from a minutia  $m_i$ , it can be simply treated as a single feature vector, obtained by linearizing the cell values. With a negligible loss of accuracy (see [23, 24]), each element of the feature vector can be stored as a bit (Fig. 3.2e): in the following,  $\mathbf{v}_i \in \{0, 1\}^n$  denotes an MCC bit-vector obtained from minutia  $m_i$ , and  $T = \{c_i\}$ , denotes an *MCC template* obtained from a finger-print, where each  $c_i = (\mathbf{v}_i, x_i, y_i, \theta_i)$  is a tuple containing a bit-vector and its associated minutia information. Note that, although strictly speaking the term “cylinder” corresponds to  $\mathbf{v}_i$  [22], in the following, for simplicity, it denotes the whole tuple  $c_i$ .

Each bit-vector  $\mathbf{v}_i$  is a fixed-length local descriptor:

- invariant for translation and rotation, since (i) it only encodes distances and directional differences between minutiae, and (ii) its base is rotated according to the corresponding minutia angle;
- robust against skin distortion (which is small at a local level) and against small feature extraction errors, thanks to the smoothed nature of the functions defining the contribution of each minutia.

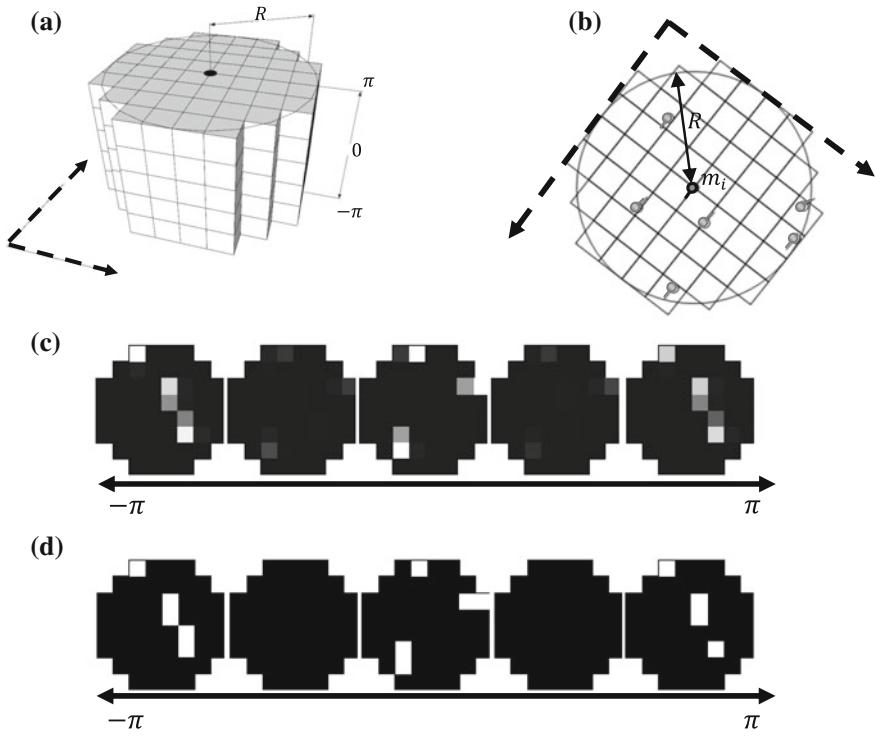
As described in [22], a simple but effective similarity measure between two cylinders  $c_i = (\mathbf{v}_i, x_i, y_i, \theta_i)$  and  $c_j = (\mathbf{v}_j, x_j, y_j, \theta_j)$  is:

$$s_L(c_i, c_j) = \begin{cases} 1 - \frac{\|\mathbf{v}_i \oplus \mathbf{v}_j\|}{\|\mathbf{v}_i\| + \|\mathbf{v}_j\|} & \text{if } d_\phi(\theta_i, \theta_j) \leq \delta_\theta \\ 0 & \text{otherwise} \end{cases}. \quad (1)$$

where

$\oplus$  Denotes the bitwise XOR operator;

$\|\cdot\|$  Denotes the Euclidean norm;



**Fig. 3.2** **a** a graphical representation of the local descriptor associated to a minutia in the MCC representation, with  $N_S = 8$  and  $N_D = 5$ ; **b** minutiae involved in a cylinder; **c** cell values in the  $N_D$  sections (*lighter areas* represent higher values) of the cylinder built over the minutiae in (b); **d** binarized cell values stored as bits. Note that cylinder sections in (c) and (d) are rotated according to the direction of minutia  $m_i$

$d_\phi(\theta_i, \theta_j)$  is the difference between the two angles;

$\delta_\theta$  is a parameter controlling the maximum rotation allowed between two fingerprints

Note that (1) is a *local* similarity measure between two cylinders: in order to compare two fingerprint templates  $T_A$  and  $T_B$ , a single value (*global score*), denoting the overall fingerprint similarity, has to be obtained from the pairwise (local) cylinder similarities. Various global similarity measures have been proposed for MCC [22, 23, 25]; the simplest and most efficient one is the *Local Similarity Sort* (LSS), which is calculated as the average of the top local similarity scores between cylinders of the two templates. A more accurate, but less efficient, similarity measure is the *Local Similarity Sort with Distortion-Tolerant Relaxation* (LSS-DTR) [25]. LSS-DTR adds a consolidation step to LSS, in order to obtain a score that reflects to what extent the local similarities hold at global level. LSS-DTR is based on similarity measures between candidate pairs of corresponding minutiae

**Table 3.1** Published parallel fingerprint identification methods on various hardware architecture: the throughput is expressed in million fingerprint comparisons per second, each speedup is measured with respect to a corresponding baseline sequential CPU implementation

Method	Hardware	Throughput	Speedup
Lindoso et al. [69]	FPGA Xilinx Virtex-4 LX	0.007	24
Jiang and Crookes [70]	FPGA Xilinx Virtex-E	1.220	47
Peralta et al. [71]	Cluster of 12 PC	0.813	181
Gutierrez et al. [72]	One Tesla M2090 GPU	0.050	31
	Two Tesla M2090 GPUs	0.098	61
Cappelli et al. [26]	One Tesla C2075 GPU	9.305	514
	Four Tesla C2075 GPUs	35.221	1 946

and uses spatial and directional features that are invariant for rotation/translation and tolerate skin distortion.

Owing to its bitwise fixed-length features and simple similarity metrics, MCC allows very fast fingerprint identification methods to be developed. In [26], an extremely fast MCC-based algorithm for graphic processing unit (GPU) is introduced. Thanks to a careful design of the algorithm, ad hoc data structures and lookup tables, special sorting methods and other optimizations, this algorithm is able to compare more than nine million fingerprint per second on a single GPU. Table 3.1 reports results published in the scientific literature for recent parallel fingerprint identification algorithms on various hardware architectures; it is well evident that the MCC-based algorithm described in [26] overcomes all the other approaches, both in terms of absolute performance and of relative speedup. In particular, on a single PC with four GPUs, it achieved a throughput of more than 35 million fingerprint comparisons per second; with such a throughput, less than 0.3 s are required to perform ten queries on a database with one million fingerprints, thus enabling real-time fingerprint identification on large databases with low-cost hardware.

### 3.2.2 Fingermarks

Fingermarks are partial impressions of the finger left on the surface of objects when they are touched. When invisible (latent), they are detected with optical methods and/or developed chemical/physical techniques before to be either captured photographically or lifted from objects. They are of critical value in forensic applications because they are usually encountered at crime scenes and serve as crucial evidence in a court of law.

However, compared to fingerprint comparison, fingermark comparison accuracy is significantly lower due to complex background noise, poor ridge quality and small area. As an example, in recent NIST evaluations, while the best-performing fingerprint algorithms achieved a rank-1 identification rate of about 99% on a

background database of 100,000 fingerprints [15], the best-performing commercial fingermark comparison algorithm achieved a rank-1 identification rate of 67.2% when searching 1,066 fingermarks against a background database of 100,000 exemplar prints, without using manually annotated features [19].

In the current generation of AFIS, to make fingermark identification more reliable, manual markup of various features (e.g. region of interest, singular points and minutiae) is typically necessary. To reduce this markup cost and to improve the consistency, fully automatic (“lights-out”) and highly accurate fingermark comparison algorithms are needed: “lights-out” capability for fingermark identification is one the major objectives of FBI’s Next Generation Identification (NGI) program. In the last five years, research has been very active in fingermark processing and relevant steps have been done toward the practical implementation of systems operating in lights-out mode. In the Netherlands, for example, more than 60% of the pairing of fingermarks and reference fingerprints is obtained using automatic encoding only [2].

Referring to the general schema of Fig. 3.1, what makes fingermark identification different from traditional fingerprint identification is not the comparison step but the segmentation, enhancement and feature extraction stages. In fact, in principle, the comparison algorithm used for fingerprint identification can be used for fingermarks search too; actually, by noting that fingermark minutiae are typically confined to a small area, comparing such a compact set against a full minutiae set characterizing a fingerprint can be viewed as a specific point pattern matching problem, for which optimized approaches can be designed (see for example the method proposed in [27] where the discriminant power of MCC descriptors is exploited to improve a Hough transform-based rigid matcher).

### 3.3 Segmentation

Separating the print area from the background is very critical in fingermarks, and ad hoc methods are being proposed also according to the background type (see for example [28], where a method is proposed to deal with fingermarks on banknotes).

State-of-the-art fingermark segmentation approaches are based on local total variation [29–31]. For example the model in [30] decomposes a fingermark image into two layers: cartoon and texture. The cartoon layer, consisting of smooth components, contains unwanted components (e.g. structured noise) while the texture layer, which contains the oscillatory or textured component, mainly consists of the fingerprint. This cartoon-texture decomposition facilitates the process of segmentation, as the region of interest can be easily detected from the texture layer using traditional segmentation methods.

A particularly complex case is when two or more fingermarks are partially overlapped and, besides isolating the composite pattern form the background, it is also necessary to separate the pattern into its components [32–34].

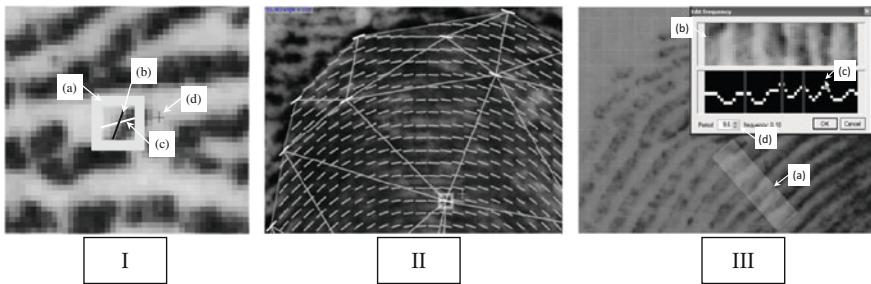
### 3.4 Enhancement

Once the foreground has been segmented, before minutiae can be automatically extracted, the pattern quality needs to be improved. The most effective approaches rely on contextual filtering [16], where the filter characteristics are locally tuned according to the context (i.e. local ridge orientation and local ridge frequency); however when dealing with fingermarks, the context estimation is very critical and inaccurate estimations can compromise the efficacy of the enhancement step.

In [35] a tool for semi-automatic context estimation is proposed. This approach reduces the markup efforts (since the user's input is limited to a few points, see Fig. 3.3), while making the fingermark enhancement quite accurate. However if a fully automatic processing is required, more sophisticated techniques have to be designed. In fact, typical bottom-up context estimation methods (i.e. gradient based approaches for orientations) proved to be ineffective because the noise level can be locally higher than the underlying signal.

Quite recently, some interesting techniques have been introduced based on prior knowledge of fingerprint structure. An off-line learning technique is adopted to determine the space (or dictionary) of feasible context patches, to which each patch in the current fingermark has to be conducted:

- In [36] prior knowledge on ridge-line orientations is represented by means of a dictionary of reference orientation patches. Figure 3.4 shows a flowchart of the proposed algorithm. In the off-line stage, a set of good quality fingerprints is used to construct the dictionary; in the online stage, given a fingermark, its orientations are first estimated using a traditional technique, then corrected by



**Fig. 3.3** Examples of the user interface proposed in [35] for the markup of the local orientations and frequencies. From *left to right*, **I** manual adjustment of a single local orientation element: (a) the selected element, (b) the initial orientation proposed by the software, (c) the orientation selected by the user moving the mouse cursor (d); **II** a larger portion of a fingermark image shows some local estimations made by the user (white segments) and all the local orientations interpolated by the software (grey segments); **III** markup of the local frequencies: (a) the user clicks on a given location, (a) the corresponding oriented window is highlighted on the fingermark image and (b) displayed on a popup window, (c) the grey-level profile is shown and the users selects some consecutive maxima using the mouse, (d) the estimated *ridge-line* frequency and the corresponding period are shown

means of dictionary lookup of orientation patches and compatibility analysis between neighbouring patches.

- [36] proved that the use of prior knowledge (in the form of orientation patch dictionary) is helpful for correcting many errors. However, since the positions of the patches have not been restrained, some orientation patches may occur at impossible locations. For example, the orientation field estimated by this algorithm for the fingermark in Fig. 3.5 contains a wrong delta singularity in the top region, which is not likely to occur in real fingerprints. In [37], Yang et al. propose a robust fingerprint orientation field estimation algorithm, which is based on localized dictionaries of orientation patches. A drawback of localized patches is the need to register the fingerprint with respect to a universal coordinate system; in [37] this is accomplished by a robust registration algorithm

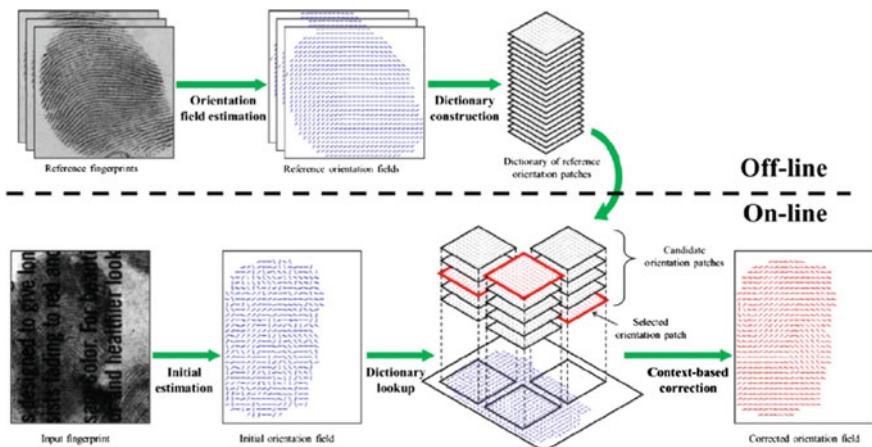


Fig. 3.4 A flowchart of the algorithm proposed in [36]

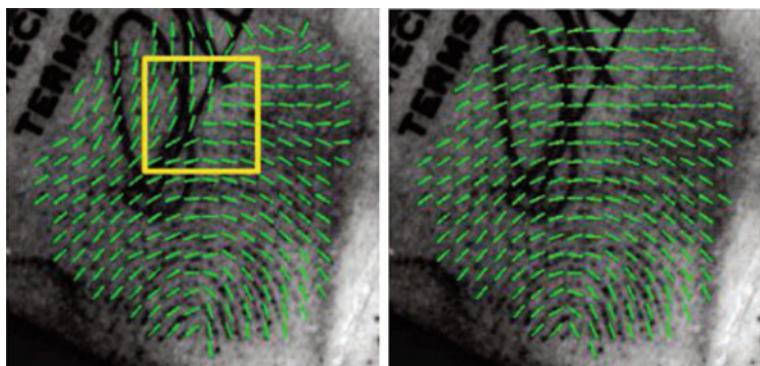
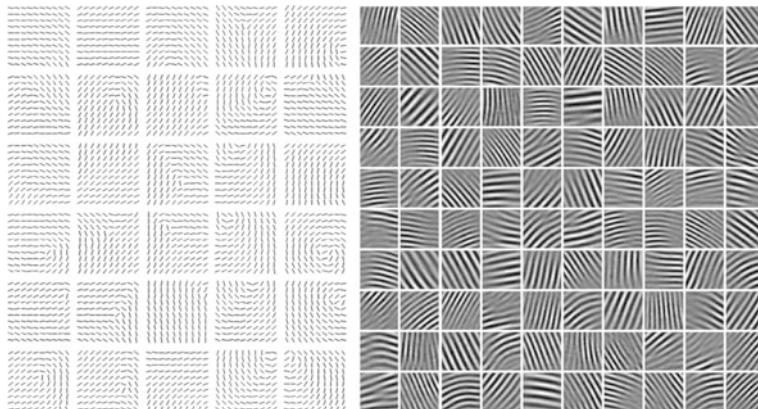


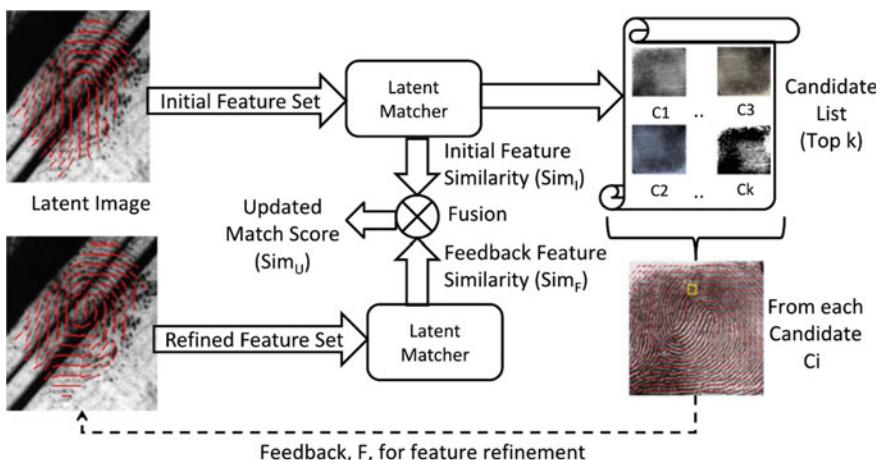
Fig. 3.5 Orientation extraction from a fingermark using the approach proposed in [36] (on the left) and in [37] (on the right)

based on probabilistic voting of all local orientation patches. The local dictionary approach outperformed the global one in a number of experiments [37] and is currently the best-performing approach in the FvcOnGoing FOE (Fingerprint Orientation Estimation) benchmark [38].

- A limitation of orientation dictionaries is that they cannot provide information about frequency. In [29] a ridge structure dictionary, which is learnt from a set of high quality ridge patches, is then used to restore ridge structure in these patches. Figure 3.6 shows a subset from a dictionary of orientation patches and a subset from a dictionary of ridge structure patches. Experimental results on two fingermark databases (i.e. NIST SD27 and WVU DB) show that the proposed algorithm outperforms previous segmentation and enhancement



**Fig. 3.6** On the left A portion of a dictionary of orientation patches; on the right a portion of a dictionary of ridge structure patches



**Fig. 3.7** The main steps of the feedback-based feature refinement described in [39]

algorithms and improves the performance of a state-of-the-art commercial fingermark comparison algorithm: in particular, at a rejection rate of 20%, the rank-1 identification rate accuracy increases by 7.43% and 7.23%, respectively, for the NIST SD27 and WVU DB with respect to the baseline performance of a commercial comparison algorithm.

A different approach to get contextual information in a fully automatic manner [39] is relying on the features of the fingerprints in the candidate list to guide the fingermark enhancement and then (re)compare the improved pattern against the candidate list exemplars to consolidate a more reliable score (Fig. 3.7).

## 3.5 Forensic Applications

This section describes the forensic use of fingerprints and fingermarks within the field of forensic dactyloscopy and details the role of biometric fingerprint recognition technology for both of them.

### 3.5.1 *Applications Using fingerprints*

#### 3.5.1.1 Identity Management Within Criminal Justice Systems

Criminal Justice Systems have been forerunners in the adoption of fingerprint recognition technology for the identification and the verification of the identity of criminals, victims and missing persons. AFIS are commonplace in law enforcement since the 1980s to collect and compare fingerprint, and later fingermark specimens [40]. Fingerprint data were first and still are used within Criminal Justice Systems in an identity management perspective, for identity verification as well as closed and open-set identification. It has to be stressed that the reliability of any forensic fingerprint application relies on the integrity of the identity management deployed by the criminal justice systems [41]. The identity infrastructure in place needs to be able to create, challenge and end biometric identities reliably. For example, before the use of biometric solutions in the Dutch prisons, some individuals were serving sentences, substituting themselves to the convicted criminals. Nowadays the Netherlands have implemented, in case of serious crime, a system combining the fingerprint and face modalities to identify and verify that the person behind a claimed identity remains the same along the whole criminal justice chain, from the arrest to the custody to serve the sentence [42].

For the administration of justice it is relevant to reach formal decisions of identity verification or identification. This is only possible if the prior probability of the hypothesis of common source for the test and reference fingerprint specimens can be assigned reliably and if a set of high quality fingerprints (not fingermarks)

combined to an accurate biometric technology can provide an overwhelming strength of evidence to support this hypothesis. The availability of secure identity management infrastructures based on the integrity of the fingerprint data and metadata, accurate fingerprint recognition technology and valid inference frameworks are a prerequisite to the forensic use fingerprints and fingermarks within Criminal Justice Systems.

### 3.5.1.2 Forensic Identification of Missing Persons

The identification of missing persons from a mass disaster depends on the form of this disaster, closed or open. A closed disaster relates to a known number of individuals from a defined group, like an aircraft crash with a passengers list. Open disasters like traffic accidents, natural disasters, technical accidents (fires, explosions), terrorist attacks and events occurring within the context of war relates to an unknown number of individuals from an undefined group [43]. Combinations of these two forms are also conceivable (e.g. aircraft crash in a residential area) [44]. When the prior probabilities can be assigned, the evidential value of the biometric features can be assessed and decision thresholds can be determined. Closed-set identification (1 to N) and open-set identification (1 to N + 1) frameworks apply, respectively, to these two types of disaster. When the prior probabilities cannot be assigned and the decision thresholds cannot be determined, the likelihood ratio inference model (defined hereunder) applies to assess the strength of evidence of the fingermarks and fingerprints [43, 45, 46].

## 3.5.2 Application Using Fingermarks

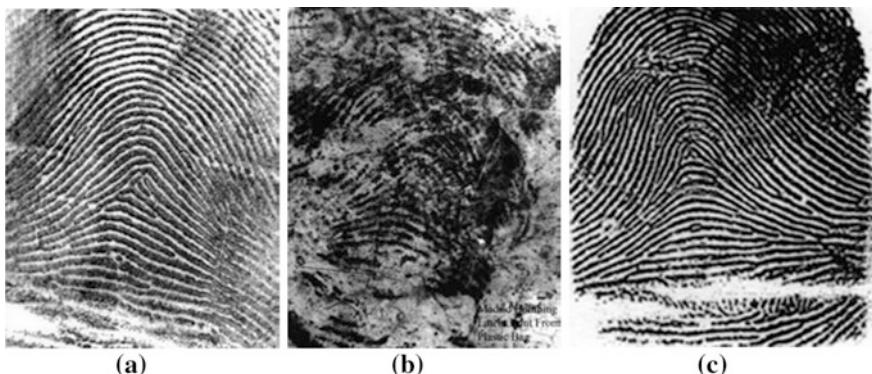
### 3.5.2.1 Forensic Intelligence

Repeat offenders do not only repeat offences but also generally operate according to a similar *modus operandi*, leaving similar traces at different crime scenes. Fingerprint recognition technology can be used to associate fingermarks from different cases to uncover crime phenomena and produce forensic intelligence. For example fingermarks recovered in separate cases can be compared pairwise using fingerprint recognition technology to produce a rank list of N scores, from which a subset of M relevant links from N will be selected (M from N). Comparing fingermarks with fingermarks is scientifically the most challenging application of fingerprint recognition technology to forensic science, due to the limited quality of the specimens compared [47]. For instance the Europol Information System (EIS) is in the process to integrate forensic intelligence capabilities for fingermarks to support the operational activities of the Agency like the fight against illicit drugs, trafficking of human beings or terrorism [43, 48].

### 3.5.2.2 Forensic Investigation

Even if offenders are aware that they are leaving traces during their criminal activity, large amounts of them are still recovered on crime scenes. That is first of all because it is impossible for a criminal to act without leaving a trace, considering the intensity of the criminal activity as Locard first observed [49]. Traces cannot be avoided due to the physical intensity of criminal activities such as breaking in or killing, but also due to the emotional intensity of the criminal activity. Offenders tend to underestimate, neglect or even forget about the traces they leave, even in meticulously organized and prepared actions like terrorist attacks. For example, how else can we explain the presence of a blue plastic bag containing detonating devices found near the Madrid Atocha train station on 11 March, 2004? It is on this blue plastic bag that one of the most (in)famous fingermarks within the field of forensic dactyloscopy was recovered, developed and submitted for identification. The use of this fingermark to establish who could be a possible suspect in this terrorist attack has demonstrated first during the forensic investigation phase the impressive sensitivity of the AFISs, optimized to minimize the false rejection rate to levels approaching zero, but it has also provided an example of the misuse of this information later during the phase of forensic evaluation.

Because of the large amount of distortion present in the fingermark found on the blue plastic bag and also because of the uncertainty whether it was the result of a single or double imposition (Fig. 3.8b), this fingermark had been initially considered as unusable by the Spanish National Police. This first assessment was later revised and the fingermark was compared with the fingerprint references stored in the Spanish AFIS and shared through Interpol for a search against the 10 billions of fingerprint references (1 billion ten-print cards) stored in the national and immigration fingerprint databases worldwide. And even with a mark of such extreme poor quality, modern systems were able to select a few reference specimens with a surprisingly high degree of similarity with the mark from immense databases. For example, in this case the US Federal Bureau of Investigations (FBI) compared the fingermark with the 700 million reference fingerprints of the criminal master file of the United States Integrated Automated Fingerprint Identification System (US-IAFIS) and established a shortlist of 20 candidates. But the forensic evaluation phase was not performed according to the ACE-V protocol (Analysis, Comparison, Evaluation, Verification), considered as a best practice procedure [50]. The initial examiner failed to conduct a complete analysis of the fingermark before conducting the AFIS search. As a result, three fingermark examiners, two from the FBI and one independent, initially attributed the fingermark developed on the blue plastic bag to an American citizen, Brandon Bieri Mayfield (Fig. 3.8a). The case took an unexpected turn when the Spanish National Police attributed the same mark to an Algerian citizen, Ouhnane Daoud (Fig. 3.8c). Although in casework the ground truth about the origin of a fingermark remains unknown, in this case the FBI withdrew their identification of Mayfield and attributed the fingermark to Ouhnane Daoud. In 2006, the Office of the Inspector General of the US Department of Justice



**Fig. 3.8** **a** Reference fingerprint of the left index of Brandon Bieri Mayfield [51], **b** fingermark found on a plastic bag containing detonating devices near the Atocha train station, March 11th 2004 [51], **c** reference fingerprint of the right middle finger of Ouhnane Daoud [51]

published a review explaining the reasons for what is considered as an erroneous decision of identification [51].

In their evaluation, the FBI examiners clearly failed to consider some key parameters that govern the search of a poor quality fingermark in an immense database of reference fingerprints during the forensic investigation phase. First, the probability to select reference fingerprints with a very high degree of similarity and minimal differences with the fingermark (and sometimes between them) increases with the size of the database. The forensic practitioners name such pairs of biometric specimens “look alikes”. These observations provide the empirical demonstration that the distinctiveness between two specimens (trace and reference) is inherently limited. Second, any forensic investigation procedure leads to the production of a shortlist of reference specimens, but the extreme low quality of a trace increases the probability to shortlist reference specimens with only a limited degree of similarity with the trace. Finally the probability to detect differences between a trace and a reference specimen decreases with their quality, to the point where even qualified and experienced practitioners cannot detect differences of origin.

In these circumstances an erroneous decision of identification may occur, as in the Mayfield case, if the question of the inference of the source of the fingermark is envisaged as a decision based on the criterion of uniqueness. On the other hand, a logical answer can be provided to the court if the answer to the question of the source of the fingermark is envisaged as a measure of distinctiveness leading to the description of the strength of evidence, combined with the explicit estimation of the order of magnitude of the relevant population in the case [52].

Applying such an approach in the Mayfield case, the FBI examiners would have noticed that the size of the relevant population that can be estimated (considering that the fingermark has been recovered in Madrid, Spain and the suspected person was living in Portland, Oregon USA) is superior by several orders of magnitude to the strength of evidence that can be assigned on the basis of the observations made

on the fingermark (Fig. 3.8b) and the fingerprint (Fig. 3.8a), making their decision of identification unlikely to be correct.

### 3.5.2.3 Forensic Evaluation

The example of the Mayfield case demonstrates that the crucial challenge of forensic evaluation consists in assigning or computing the most correct (and not the biggest) strength of evidence, on basis of the observations made on a pair of fingermark and fingerprint specimens. Human-based and computer-assisted methods using fingerprint recognition technology have been developed to assign or compute the strength of evidence. Both approaches rely on the Bayes theorem of conditional probabilities to express the strength of evidence in the form a likelihood ratio (LR) in a uniform framework [53]. This framework is considered as logical and balanced and the LR can be considered as the metric describing the strength of evidence [54].

At source level, the LR is defined as the ratio of the probability of the observations (evidence E) in the light of a pair of mutually exclusive alternative hypotheses as follows: hypothesis H1—the two specimens originate from a common source and hypothesis H2—the two specimens originate from different sources. Generally, hypothesis H1, stating that the trace material originates from the suspected person, is supported by the prosecution. Hypothesis H2, stating that the trace material originates from another individual randomly chosen within the relevant population of potential sources of the trace, is supported by the defence.

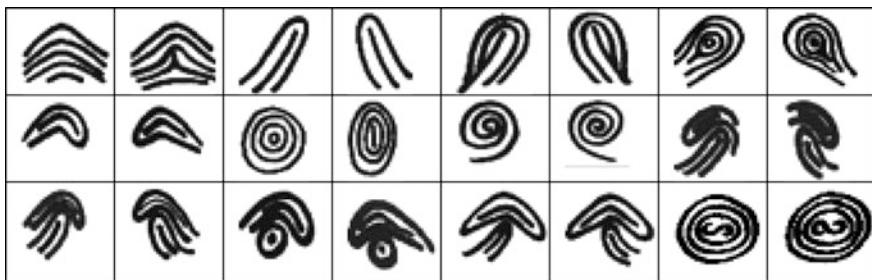
The strength of evidence is assigned or computed as the ratio of two probabilities: the probability of the evidence when the prosecution hypothesis is true when divided by the probability of the evidence when the defence hypothesis is true. In the Eq. (2), I represents the relevant background information about the case, for instance the selection process of the suspected person and the nature of the relevant population [52]. The result, expressed as a likelihood ratio, is calculated as follows:

$$\frac{P(H_1|E, I)}{P(H_2|E, I)} = \frac{P(E|H_1, I)}{P(E|H_2, I)} * \frac{P(H_1, I)}{P(H_2, I)} \quad (2)$$

Posterior probability ratio	Likelihood ratio	Prior probability ratio
-----------------------------------	---------------------	-------------------------------

The posterior probability ratio is calculated as the multiplication of the prior probability ratio by the likelihood ratio. The role of the forensic practitioner is limited to the assessment of the likelihood ratio. The duty to provide the prior probability ratio and to make decisions on basis of the posterior probability ratio remains to the trier of fact [43].

Human-based methods rely on forensic practitioners using knowledge and experience to assign the LR on basis of their personal probabilities. On the other hand, computer-assisted methods and LR-based systems rely on statistical



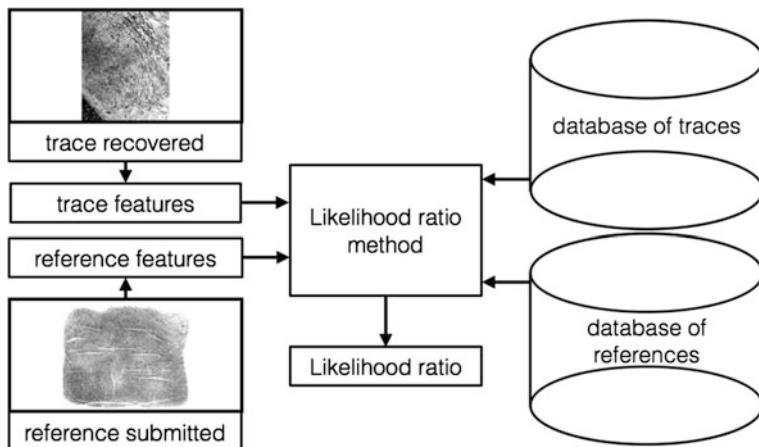
**Fig. 3.9** Extended set of fingerprint general patterns (Courtesy from Arent de Jongh and Anko Lubach, Netherlands Forensic Institute)

probabilities obtained from the combination of databases and fingerprint recognition technology to compute the LR. The strength of a LR-based system is to provide statistical probabilities on the set of distinctive features that can be extracted automatically. The strength of human beings is to also be able to consider features that cannot be handled by biometric technology yet, like an extended set of general pattern in fingermarks (Fig. 3.9). Statistical probabilities are considered as more objective and personal probabilities more subjective.

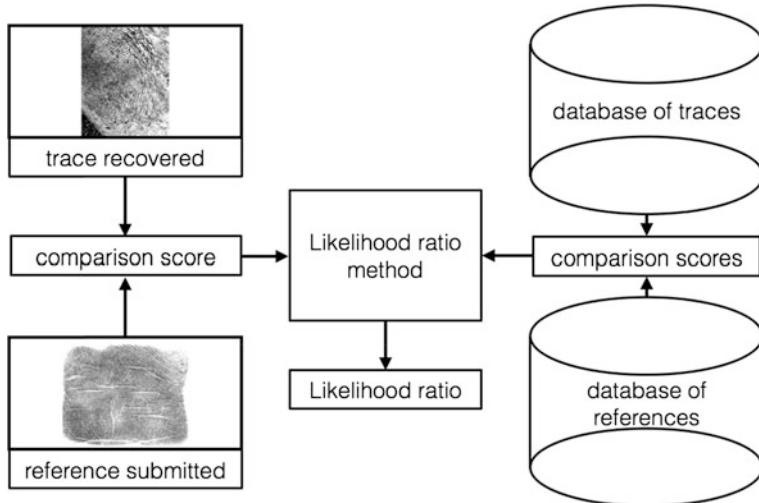
The classical “forensic identification” disciplines relying mainly on personal probabilities for the assessment of the evidence are being increasingly challenged [55], especially because of the development of evidence based on DNA profiles governed by statistical data and the evolving requirements for the admissibility of evidence following the Daubert decision by the Supreme Court of the USA [52]. The quality of the inference provided by computer-assisted LR-based systems strongly depends on the quantity and properties of the data used to estimate the intravariability/within-source variability and the intervariability/between-source variability [56]. The LR approach has been first implemented for the DNA modality [53]. Since then numerous computer-assisted methods have been proposed to compute LRs for different biometric modalities, and among them fingermarks [57].

The current methods described in the literature can be classified as feature-based (Fig. 3.10) [58] or score-based approaches (Fig. 3.11) [59]. When a limited number of features are extracted from each specimen, the model parameters can be estimated with a limited amount of data [60]. Such an approach is feature-based; it involves statistical modelling at the level of the features, using for example probability density functions under the two alternative hypotheses to produce the LR values. When a large number of features are extracted from each specimen, score-based methods are frequently used as a way to reduce the multidimensionality of the feature space and to estimate the model parameters in one dimension [61]. The LR values computed from comparison scores are typically the result of a comparison performed by pattern recognition and machine-learning algorithms [62].

Ongoing discussions are taking place within the forensic community about the scientific merits and flaws of these approaches and about the challenges that represent their implementation in practical systems [63–65].



**Fig. 3.10** Framework for a feature-based LR computation



**Fig. 3.11** Framework for a score-based LR computation

### 3.5.3 Current Challenges

#### 3.5.3.1 Automation and Transparency

In forensic applications, the feature extraction from poor quality fingerprint and fingermark images remains a computer-assisted process, combining the performance of the subjective and somehow inconsistent ability of humans to the more objective but also more limited performance of computers for some complex pattern

recognition tasks. Both the human inconsistencies and the limits of the computer affect the performance of the feature extraction and, as a consequence, the performance of the forensic processes based on the use of fingermarks and fingerprints. In the future, a feature extraction process at once reliable and completely automated (lights-out-mode) is desirable. Steps in this direction can be done, not only by refining the minutiae extraction process, but also by enriching the feature vector with other available, measurable, discriminatory, permanent and robust features, for e.g. the extended set of fingerprint general patterns (see above) or the generalization of the use of the number of ridges between the minutiae [40].

Fingerprint practitioners need a better understanding of the functioning of the fingerprint recognition technology implemented in the AFISs in order to operate AFIS processes in a more transparent manner. In absence of an authoritative source to help them interpret the AFIS results, they can only speculate about how the core algorithms are designed and what could influence their output. This is not without raising difficulties for the practitioners to be accountable for results driven by the technology [40, 66].

### 3.5.3.2 Scalability and Interoperability

The scalability of the paper-based ten-print collections was limited by the necessary trade-off between selectivity and reliability imposed by manual classification systems, and the coexistence of several paper-based systems around the world limited their interoperability at an international level. The implementation of Automated Fingerprint Identification Systems solved the problem of scalability for the national fingerprint databases, but the interoperability problem remains as the first generations of AFIS incorporated feature vectors that are encoded using proprietary formats [40]. Currently, the problem of interoperability between different types of AFISs is on the way to be solved partially by the widespread use of the ANSI-NIST biometric exchange file format, but the predominant use of proprietary formats for the feature vectors remains. This improvement opens a new opportunity in terms of scalability, with the possibility to extend the interoperability of the AFISs at a global level and to integrate the AFIS functionality into a broader platform for multimodal biometric identification, for example including facial recognition, speech and DNA [40]. But this opportunity also raises the challenge of merging the results from the different biometric modalities using fusion strategies adapted to the different forensic applications: forensic identification (decision), forensic investigation and intelligence (selection) and forensic evaluation (description) [43].

### 3.5.3.3 Forensic Fingermark Processes

In forensic investigation and intelligence processes, the short lists of potential sources still predominantly have a fixed size based on the rank information provided by the AFISs. The efficiency of these processes would largely benefit from

lists of variable size, combining the information of the rank and score value to output a description of the strength of the link made between the fingermark and the reference fingerprints [40].

Forensic evaluation of fingermarks and fingerprints consists mainly in the inference of identity of the source of a fingermark and a fingerprint. Currently, this task remains mainly performed by fingermark examiners. They apply the Analysis, Comparison, Evaluation, Verification (ACE-V) procedure and, in some countries, a numerical standard, to substantiate three types of qualitative opinions: identification, exclusion or inconclusive. As their evaluation is deterministic, fingermark examiners also make an implicit use of their own subjective probabilities of the rarity of the features used for identification. They refine these subjective probabilities through training and experience [40, 50].

In forensic research, the inference of identity of the source of a fingermark and a fingerprint is also envisaged, combining statistical models, digitized fingerprint and fingermark databases as well as a scientific methodology; namely the likelihood ratio approach based on Bayes' theorem. This approach aims to offer a uniform framework and a transparent methodology to fingermark examiners, and to assist them in producing a logical, testable and quantitative evaluation of fingerprint evidence [67]. Prototypes for forensic fingermark evaluation exist and some rely on fingerprint biometric technology. But their validation, including criteria about the robustness of the underlying assumptions and criteria about the precision, the accuracy and the calibration of the results is a critical step prior to their acceptance by the fingerprint scientific and legal communities [68].

### 3.6 Conclusion

This chapter introduced the use of AFIS for ID verification, identification, forensic investigation and intelligence. These systems have greatly improved the operational productivity of law enforcement agencies and reduced the cost of hiring and training human fingerprint experts. In about fifty years they evolved from initial prototypes to powerful systems, able to process millions identification requests per second. Scientific research in fingerprint recognition and fingermark processing is still very active; among the various recent interesting developments, it is worth mentioning:

- an extremely efficient fingerprint comparison algorithm, which exploits GPUs to achieve a throughput of millions of comparisons per second on a simple PC;
- the availability of tools for semi-automatic feature markup for the enhancement of fingermarks;
- learning-based techniques for fingermark processing, which exploits prior knowledge of fingerprint structure to improve the quality of automatically extracted features.

All the forensic applications have benefited from the improvements offered by the fingerprint biometric technology in terms automation and performance. Automatic fingerprint identification systems are currently an integral part of the different forensic processes: identification, intelligence, investigation and evaluation. But if the contribution of the fingerprint biometric technology is essential in the good execution of the operational forensic processes, several challenges remain:

- First, the extraction of features from poor quality fingerprint and fingermark images is still a computer-assisted process, when full automation (lights-out-mode) is desirable.
- Second, increasing the transparency of the fingerprint biometric technology would help the practitioners to interpret the AFIS results.
- Third, cooperation, harmonization and standardization have enabled for an improvement gap of the fingerprint biometric technology. This technological improvement still has to reverberate operationally in order for the forensic processes to become fully scalable and interoperable in practice.
- Finally the fingerprint biometric technology has opened the door to render the fingermark evaluation, in particular, more objective, but a long research, validation and education effort is still necessary to achieve its implementation and to gain the acceptance of the fingerprint scientific and legal communities.

## References

1. Lee HC, Gaenslen RE (2001) Advances in fingerprint technology. Elsevier Publishing, New York
2. Meuwly D (2015) Forensic use of fingermarks and fingerprints. In: Stan Z., Jain, AL (eds) Encyclopedia of biometrics, 2<sup>nd</sup> edn. Springer
3. Federal Bureau of Investigation (1984) The Science of fingerprints: classification and uses. Federal Bureau of Investigation, Government Publication, Washington, DC, US
4. Federal Bureau of Investigation (1991) The FBI fingerprint identification automation program: issues and options. Federal Bureau of Investigation, Government Publication, Washington, DC, US
5. Moses KR, Higgins P, McCabe M, Prabhakar S, Swann S (2011) Automated fingerprint identification system (AFIS). In: SWGFAST—the fingerprint sourcebook, scientific working group on friction ridge analysis study and technology and national institute of Justice, pp 1–33
6. Wegstein JH, Rafferty JF, Pencak WJ (1968) Matching fingerprints by computer, National Bureau of Standards, U.S. Department of Commerce, Washington, DC, Technical Note 466
7. Wegstein JH (1969) A computer oriented single-fingerprint identification system, National Bureau of Standards, U.S. Department of Commerce, Washington, DC, Technical Note 443 1969
8. Wegstein JH (1970) Automated fingerprint identification, National Bureau of Standards, U.S. Department of Commerce, Washington, DC, Technical Note 538
9. Wegstein JH (1972) Manual and automated fingerprint registration, National Bureau of Standards, U.S. Department of Commerce, Washington, DC, NBS Technical Note 730
10. Wegstein JH (1972) The M40 fingerprint matcher, National Bureau of Standards, U.S. Department of Commerce, Washington, DC, Technical Note 878

11. Wegstein JH, Rafferty JF (1978) The LX39 latent fingerprint matcher, National Bureau of Standards, U.S. Department of Commerce, Washington, DC, Special Publication 500–536
12. Wegstein JH, Rafferty JF (1979) The automated identification of fingerprints. In: Dermatoglyphics—fifty years later, Washington, DC
13. Wegstein JH (1982) An automated fingerprint identification system, National Bureau of Standards, U.S. Department of Commerce, Washington, DC, NBS Special Publication 500–589
14. Stoney DA (1991) What made us ever think we could individualize using statistics. *J Foren Sci Soc* 31(2)
15. Watson CI et al (2015) Fingerprint vendor technology evaluation 2012, NIST, NIST Interagency/Internal Report (NISTIR)—8034
16. Maltoni D, Maio D, Jain AK, Prabhakar S (2009) Handbook of fingerprint recognition, 2nd edn. Springer, New York, NJ, USA
17. FBI—CJIS division (1999) Electronic fingerprint transmission specification, FBI, CJIS-RS-0010 (V7)
18. Indovina M, Hicklin RA, Kiebzinski GI (2011) ELFT-EFS evaluation of latent fingerprint technologies: extended feature sets (Evaluation 1), National Institute of Standards and Technology, US Department of Commerce, NISTIR 7775
19. Indovina MD, Dvornychenko V, Hicklin RA, Kiebzinski GI (2012) ELFT-EFS evaluation of latent fingerprint technologies: extended feature sets (Evaluation 2), National Institute of Standards and Technology, US Department of Commerce, NISTIR 7859
20. Jain AK, Feng J (2011) Latent fingerprint matching. *IEEE Trans Pattern Anal Mach Intell* 33 (1):88–100
21. Zhao Q, Jain AK (2010) On the utility of extended fingerprint features: a study on pores. In: CVPR workshop on biometrics, San Francisco
22. Cappelli R, Ferrara M, Maltoni D (2010) Minutia Cylinder-Code: a new representation and matching technique for fingerprint recognition. *IEEE Trans Pattern Anal Mach Intell* 32 (12):2128–2141
23. Cappelli R, Ferrara M, Maio D (2012) A fast and accurate palmprint recognition system based on minutiae. *IEEE Trans Syst Man Cybern Part B* 42(3):956–962
24. Cappelli R, Ferrara M, Maltoni D (2011) Fingerprint Indexing based on minutia cylinder code. *IEEE Trans Pattern Anal Mach Intell* 33(5):1051–1057
25. Cappelli R, Ferrara M, Maltoni D, Tistarelli M (2010) MCC: a baseline algorithm for fingerprint verification in FVC-onGoing. In: Proceedings 11th international conference on control, automation, robotics and vision (ICARCV), Singapore
26. Cappelli R, Ferrara M, Maltoni D (2015) Large-scale fingerprint identification on GPU. *Inf Sci* 306:1–20
27. Paulino AA, Feng J, Jain AK (2013) Latent fingerprint matching using descriptor-based hough transform. *IEEE Trans Inf Forensics Secur* 8(1):31–45
28. Si X, Feng J, Zhou J (2014) Enhancing latent fingerprints on banknotes. In: IEEE international joint conference on biometrics, Clearwater, FL, USA, pp 1–8
29. Cao K, Liu E, Jain AK (2014) Segmentation and enhancement of latent fingerprints: a coarse to fine ridge structure dictionary. *IEEE Trans Pattern Anal Mach Intell* 36(9):1847–1859
30. Zhang J, Lai R, Kuo C-CJ (2013) Adaptive directional total-variation model for latent fingerprint segmentation. *IEEE Trans Inf Forensics Secur* 8(8):1261–1273
31. Choi H, Boaventura M, Boaventura IAG, Jain AK (2012) Automatic segmentation of latent fingerprints. In: IEEE fifth international conference on biometrics: theory, applications, Arlington, VA, USA, pp 303–310
32. Zhao Q, Jain AK (2012) Model based separation of overlapping latent fingerprints. *IEEE Trans Inf Forensics Secur* 7(3):904–918
33. Zhang N, Zang Y, Yang X, Jia X, Tian J (2014) Adaptive orientation model fitting for latent overlapped fingerprints separation. *IEEE Trans Inf Forensics Secur* 9(10):1547–1556
34. Feng J, Shi Y, Zhou J (2012) Robust and efficient algorithms for separating latent overlapped fingerprints. *IEEE Trans Inf Forensics Secur* 7(5):1498–1510

35. Cappelli R, Maio D, Maltoni D (2009) Semi-automatic enhancement of very low quality fingerprint. In: 6th international symposium on image and signal processing and analysis (ISPA09), Salzburg, pp 678–683
36. Feng J, Zhou J, Jain AK (2013) Orientation field estimation for latent fingerprint enhancement. *IEEE Trans Pattern Anal Mach Intell* 35(4):925–940
37. Yang X, Feng J, Zhou J (2014) Localized dictionaries based orientation field estimation for latent fingerprints. *IEEE Trans Pattern Anal Mach Intell* 36(5):955–969
38. BioLab. (2015) FVC-onGoing web site. <http://biolab.csr.unibo.it/fvcongoing>
39. Arora SS, Liu E, Cao K, Jain AK (2014) Latent fingerprint matching: performance gain via feedback from exemplar prints. *IEEE Trans Pattern Anal Mach Intell* 36(12):2452–2465
40. Meuwly D (2014) Friction ridge skin—AFIS. In: Jamieson A, Moenssens A (eds) Encyclopedia of forensic science, Chichester, UK. Wiley
41. Meuwly D (2010) ID management in 2020, ID.academy. The Hague
42. Plomp MGA, Grijpink JHAM (2011) Combating identity fraud in the public domain: information strategies for healthcare and criminal justice. In Proceedings of the 11th European conference on e-government, Ljubljana, Slovenia
43. Meuwly D, Veldhuis R (2012) Forensic biometrics: from two communities to one discipline. In: 2012 BIOSIG-proceedings of the international conference of the biometrics special interest group (BIOSIG), pp 207–218
44. Interpol (2009) Disaster victims identification guide. Interpol, Lyon
45. Biedermann A, Taroni F, Margot P (2012) Reply to Budowle, Ge, Chakraborty and Gill-King: use of prior odds for missing persons identifications. *Investig Genet* 3:1–2
46. Budowle B, Ge J, Chakraborty R, Gill-King H (2011) Use of prior odds for missing persons identifications. *Investig Genet* 2:1–6
47. Ribaux O, Walsh SJ, Margot P (2006) The contribution of forensic science to crime analysis and investigation: forensic intelligence. *Forensic Sci Int* 156:171–181
48. Europol (2011) Europol information management: products and services. Europol, The Hague
49. Locard E (1920) L'enquête criminelle et les méthodes scientifiques. Ernst Flammarion, Paris
50. Langenburg GM (2012) A critical analysis and study of the ACE-V process, University of Lausanne, Switzerland, PhD thesis
51. Fine GE (2006) A review of the FBI's handling of the Brandon Mayfield case, Office of the Inspector General, U.S. Department of Justice
52. Dessimoz D, Champod C (2008) Linkages between biometrics and forensic science. In: Handbook of biometrics. Springer, pp. 425–459
53. Evett I (1998) Towards a uniform framework for reporting opinions in forensic science casework. *Sci Justice* 38(3):198–202
54. Good II (1991) Weight of evidence and the Bayesian likelihood ratio. In: Aitken CGG, Stoney DA (eds) The use of statistics in forensic science. Ellis Horwood, Chichester UK, pp 85–106
55. Saks M, Koehler J (2005) The coming paradigm shift in forensic identification science. *Science* 309(5736):892–895
56. Meuwly D (2006) Forensic individualisation from biometric data. *Sci Justice* 46(4):205–213
57. Neumann C et al (2006) Computation of likelihood ratios in fingerprint identification for configurations of three minutiae. *J Forensic Sci* 51(6):1255–1266
58. Lindley DV (1977) A problem in forensic science. *Biometrika* 64(2):207–213
59. Van Leeuwen DA, Brümmer N (2007) An introduction to application-independent evaluation of speaker recognition systems. In: Speaker classification I. Springer, pp 330–353
60. Bolck A, Weyermann C, Dujourdy L, Esseiva P, van den Berg J (2009) Different likelihood ratio approaches to evaluate the strength of evidence of MDMA tablet comparisons. *Forensic Sci Int* 191(1):42–51
61. Gonzalez-Rodriguez J, Drygajlo A, Ramos-Castro D, Garcia-Gomar M, Ortega-Garcia J (2006) Robust estimation, interpretation and assessment of likelihood ratios in forensic speaker recognition. *Comput Speech Lang* 20(2):331–355

62. Jain A, Ross A (2015) Bridging the gap: from biometrics to forensics. In: *Philosoph Trans Roy Soc B Biol Sci* 370(1674)
63. Alberink I, de Jongh A (2015) Authors' Response. *J Forensic Sci* 60(1):257–258
64. Alberink I, de Jongh A, Rodriguez C (2014) Fingermark evidence evaluation based on automated fingerprint identification system matching scores: the effect of different types of conditioning on likelihood ratios. *J Forensic Sci* 59(1):70–81
65. Neumann C, Saunders CP (2014) Commentary on: Alberink I, de Jongh A, Rodriguez C. Fingermark evidence evaluation based on automated fingerprint identification system matching scores: the effect of different types of conditioning on likelihood ratios. *J Forensic Sci* 59(1):70–81
66. Smith ML, Noorman ME, Martin AK (2010) Automating the public sector and organizing accountabilities. *Commun Assoc Inf Syst* 26(1)
67. Neumann C, Evett IW, Skerrett J (2012) Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm. *J Roy Stat Soc: Ser A (Stat Soc)* 175(2):371–415
68. Neumann C (2012) Statistics and probabilities as a means to support fingerprint examination. In: Ramotowski R (ed) Lee and Gaenslen's advances in fingerprint technology. CRC Press, pp 419–466
69. Lindoso A, Entrrena L, Izquierdo J (2007) FPGA-based acceleration of fingerprint minutiae matching. In: 2007 3rd southern conference on programmable logic, 2007. SPL'07, Mar del Plata, Argentina, pp 81–86
70. Jiang RM, Crookes D (2008) FPGA-based minutia matching for biometric fingerprint image database retrieval. *J Real-Time Image Proc* 3(3):177–182
71. Peralta D, Triguero I, Sanchez-Reillo R, Herrera F, Benitez JM (2014) Fast fingerprint identification for large databases. *Pattern Recogn* 47(2):588–602
72. Gutierrez PD, Lastra M, Herrera F, Benitez JM (2014) A high performance fingerprint matching system for large databases based on GPU. *IEEE Trans Inf Forensics Secur* 9(1): 62–71

## **Chapter 4**

# **Challenges for Fingerprint Recognition—Spoofing, Skin Diseases, and Environmental Effects**

## **Is Fingerprint Recognition Really so Reliable and Secure?**

**Martin Drahanský, Ondřej Kanich and Eva Březinová**

**Abstract** This chapter tries to find answers to the questions whether the fingerprint recognition is really so reliable and secure. The most biometric systems based on fingerprint recognition have very low error rates, but are these error rates really telling us everything about the quality of such a biometric system? What happens when we use spoofs to deceive the biometric system? What happens when the genuine user has any kind of skin disease on his fingertips? And could we acquire a fingerprint with acceptable quality if there are some distortions on a finger or there are some environmental effects influencing the scanning technology? Reading this chapter brings you an introduction of preparation of finger fakes (spoofs), spoof detection methods, summarization of skin diseases and their influence on papillary lines, and finally the environmental effects are discussed at the end.

### **4.1 Spoofing and Anti-spoofing**

The first subchapter starts with spoofing and anti-spoofing techniques for fingerprint recognition systems. It means that we try to use any kind of finger fake or dead real finger to get an unauthorized access to a biometric system. When a genuine user has already registered his finger in a fingerprint recognition system, there are still several ways how to deceive the biometric system. In order to deceive the

---

M. Drahanský (✉) · O. Kanich  
Brno University of Technology, Faculty of Information Technology,  
Brno, Czech Republic  
e-mail: drahan@fit.vutbr.cz

E. Březinová  
1st Department of Dermatovenereology, St. Anne's University Hospital,  
Faculty of Medicine & Masaryk University, Brno, Czech Republic

fingerprint system, an attacker may put the following objects on the fingerprint scanner [1, 3]:

- *Registered (enrolled) finger.* The highest risk is that a legitimate user is forced, e.g., by an armed criminal, to put his live finger on the scanner under duress. Another risk is that a genuine user is compelled to fall asleep with a sleeping drug in order to make free use of his live finger.
- *Unregistered finger* (an impostor's attempt). An attack against a biometric system by an impostor (not necessarily an attacker—this could be only an inquisitive person) with his own biometric characteristic is referred as a non-effort forgery.
- *Severed fingerprint of enrolled finger.* A reprehensible attack may be performed using the finger severed from the hand of a genuine user, registered in the biometric system. This kind of attack could be done on a living user or dead user. In both cases, the finger could be used for a limited time only. After several hours, the fingerprint is very often in so bad condition that no papillary lines could be acquired.
- *Genetic clone of enrolled finger.* It should be stated that monozygotic twins do not have the same fingerprint, and the same will be very probably true for clones [10]. The reason is that fingerprints are not entirely determined only genetically but rather by the pattern of nerve growth in the skin. Furthermore, different intrauterine pressures in mother's uterus play an important role in creation of the fingerprint global structure (class) of the fingerprint. It means that such pattern is not exactly the same even for identical twins.
- *Artificial clone of enrolled finger.* More probable attacks against fingerprint systems may use an artificial finger. An artificial finger can be produced from a printed fingerprint made by a copy machine or a graphical processing technique in the same way as forged documents. If an attacker can make then a mold of the enrolled finger by directly modeling it, he can finally also make an artificial finger from a suitable material (see examples in this subchapter). He may also make a mold of the enrolled finger by making a 3D model based on its residual fingerprint.
- *Others.* In some fingerprint systems, an error in authentication may be caused by making noise or flashing a light against the fingerprint scanner, or by heating up, cooling down, humidifying, impacting on, or vibrating the scanner outside its environmental tolerances. We will discuss some of these factors in the third subchapter.

There are possible attacks on algorithms, data transport, and hardware, but these kinds of attacks are out of scope of this subchapter. We will discuss attack possibilities on the input device—fingerprint scanner. One of the simplest possibilities is to produce an artificial finger(print) using soft silicon, gummy and plastic material, or similar substances [10, 18]—see the later part of this subchapter. The fingerprint of a person enrolled in a database is easy to acquire, even without the user's cooperation. Latent fingerprints on daily-use products or on sensors of the access

control system itself may be used as templates. To discourage potential attackers from presenting a fake finger (i.e., an imitation of the fingertip and the papillary lines) or, even worse, to hurt a person to gain access, the system must be augmented by an anti-spoofing component [3]. To prevent false acceptance we have to recognize whether the finger on the plate of the fingerprint sensor (also referred to as fingerprint scanner) is alive or not. First of all, we will introduce anti-spoofing techniques (often called liveness detection methods), which are based on various principles. Their purpose is to detect whether the presented finger is alive (and simultaneously the fingerprint is acquired) or any kind of finger(print) spoof is used.

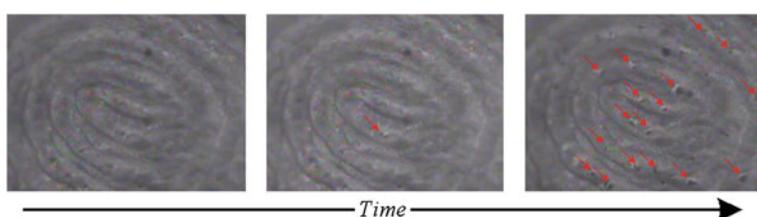
#### 4.1.1 Perspiration

This software-based method processes the information already acquired by a fingerprint scanner—the principle of this technique is the detection of perspiration as an indication of liveness (see Fig. 4.1) [17]. It is worth noting that the outmost layer of the human skin contains around 600 till 1,000 sweat glands per square inch [17]—this amount changes according to the position of skin on the body. These sweat glands diffuse sweat (a dilute sodium chloride solution) on to the surface of skin through pores. The position of skin pores does not change over time and their pore-to-pore distance is approximately 0.5 mm over fingertips [17].

The perspiration method is based on a high difference in the dielectric constant and electrical conductivity between the drier lipids that constitute the outer layer of the skin and the moister sweaty areas near the perspiring pores. The dielectric constant of sweat is around 30 times higher than the lipid.

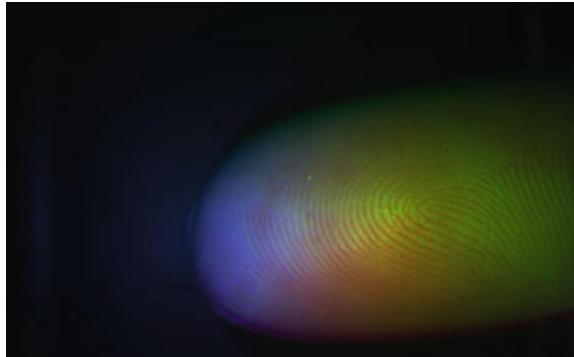
#### 4.1.2 Spectroscopic Characteristics

This hardware-based method may be regarded not only as an anti-spoofing mechanism but also as an individual biometric system with an inherent anti-spoofing capability.



**Fig. 4.1** Ascent of sweat from sweat pores on a fingertip (4 × zoomed)

**Fig. 4.2** Merge of original fingerprint images illuminated by various wavelengths in RGB [13]



Living human skin has certain unique optical characteristics due to its chemical composition, which predominately affects optical absorbance properties, as well as its multilayered structure, which has a significant effect on the resulting scattering properties [16]. When collecting images generated from different illumination wavelengths sent into the skin, different subsurface skin features may be measured and used to ensure that the material is a living human skin. When such a multispectral sensor is combined with a conventional fingerprint reader, the resulting sensing system can provide a high level of certainty that the fingerprint originates from a living finger. The principle of this technique lies in passing light of different wavelengths through a sample and measuring the light returned, which is affected by the structural and chemical properties of the sample. Figure 4.2 shows the color image obtained from a finger illuminated by light with various wavelengths and by merging the original images in RGB color model to one final image.

#### 4.1.3 Ultrasonic Technology

General ultrasonic method [10] uses a transmitter, which emits acoustic signals toward the fingerprint, and a receiver, which detects the echo signals affected by the interaction with the fingerprint—very often transmitter and receiver are combined into one unit called transceiver. A receiver utilizes the fact that ridges (skin) and valleys (air) have difference in acoustic impedance, therefore the echo signals are reflected and diffracted differently in the contact area. This approach with inherent anti-spoof testing capability among its foremost principles uses the fact that sound waves are not only reflected and diffracted, but are also subject to some additional scattering and transformation. This phenomenon is called contact scattering [10] and it was discovered that this scattering is, to a significant extent, affected by the subsurface structure of the acquired object. Hence, the class corresponding to the live tissue could be modeled and whenever the received acoustic waves are inconsistent with this class, they are rejected. The main problem here is not to

obtain clear signals, but to analyze and to make a reconstruction of internal structures from signals, which are very difficult to interpret.

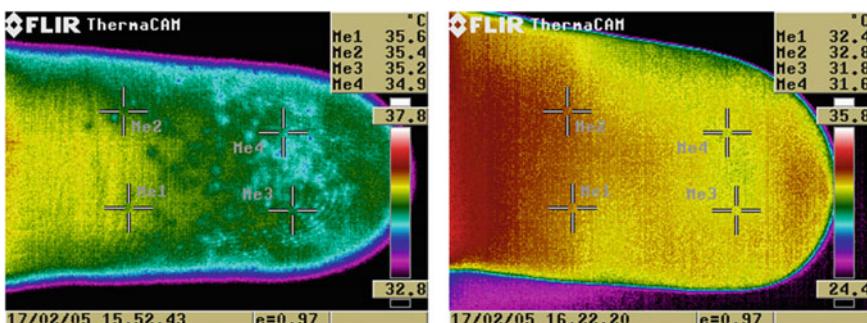
#### 4.1.4 Physical Characteristics: Temperature

This simple method measures the temperature of epidermis during a fingerprint acquisition. The temperature of human epidermis of a finger lies in the range of approximately 25–37°C (see examples in Fig. 4.3). However, this range usually has to be wider to make the system usable under different conditions. In addition, there are many people who have distortions in blood circulation, i.e., a fact which leads to deviations in the body temperature and hence to wrong anti-spoof module decision.

#### 4.1.5 Physical Characteristics: Hot and Cold Stimulus

This technique is based on the fact that the human finger reacts differently to thermal stimuli compared with other artificial (nonliving) material.

The designed anti-spoofing testing module [19] works as follows. A stimulus-giving section gives a stimulus (it may cover cool and hot stimulus) to the finger by a contact plate with which the finger makes contact. Next, typical information could be measured by an organism information-measuring section, which is produced by the live finger in response to the stimulus. Concretely, the amount of the fluctuation for the flow rate of the blood flowing in the peripheral vascular tracts varies according to the stimuli. Hence, as peripheral vascular tracts of the fingertip are extended or contracted, the amplitude value of the blood flow is measured and processed by an organism information-measuring section. Under hot stimulus the amplitude of the blood flow increases, while it decreases under cold stimulus. Moreover, according to the autonomic nervous system, the amplitude is delayed a



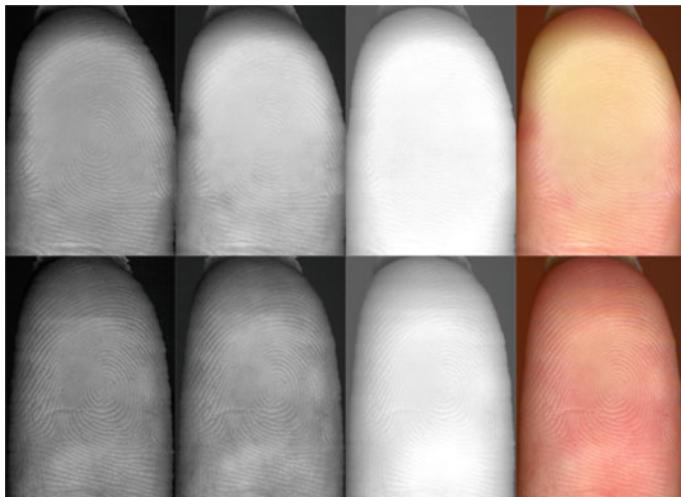
**Fig. 4.3** Thermo-scans of fingertips acquired using a thermo-camera FLIR

little with respect to the application of the stimulus. Since these facts are typically observed when the live fingers are measured, they could be employed to distinguish live among artificial and dead samples.

#### 4.1.6 Physical Characteristics: Pressure Stimulus

The principle of this method lies in concrete changes in characteristics of the live skin, which are realized due to pressure applied to the finger [11]. Since the structure and the characteristics of artificial and dead samples are different, when compared with a live finger, this phenomenon could not be seen if such samples were used.

The color of the live skin of the finger without pressure is usually reddish, but becomes whitish when pressure is applied to the skin of the finger. Hence, for the purposes of the device it is suitable to measure the spectral reflectance in the blue and green spectral range (see Fig. 4.4). The difference (in RGB model) between average R values is approximately 11, in G approximately 42, and in B approximately 20 [11].



**Fig. 4.4** Comparison of pressed finger (*1st row*) and non-pressed finger (*2nd row*). In the *1st column* (from the *left*), there is only the R-channel, the G-channel is in the *2nd column*, the B-channel is in the *3rd column* and in the *4th channel* there is the finger in all RGB colors [11]

#### 4.1.7 Physical Characteristics: Electrical Properties

Some anti-spoofing methods are based on the fact that the live human skin has different electrical properties compared with other materials [10, 14]. The suitable fingerprint recognition system could be extended by an electrode system and an electrical evaluation unit.

The *conductivity* [10] of the human skin is based on humidity, which is dependent on people's biological characteristics and environmental conditions: some people have dry fingers and others have sweaty (wet) ones; also during different seasons, climatic, and environmental conditions, humidity differs significantly. As a result, the span of permissible resistance levels has to be big enough to make the system usable. In such a situation it is quite easy for an impostor to fool the system.

The *relative dielectric constant* (RDC) [10] of a specific material reflects the extent to which it concentrates the electrostatic lines of flux. Many advocates claim that the RDC has the ability to distinguish between real and artificial samples. However, the RDC is highly dependent on the humidity of the sample, so the same situation as in the case of conductivity arises. To fool this method, an attacker can simply use an artificial sample and dip it into a compound of 90% alcohol and 10% water.

*Bio-impedance* [14] describes the passive electrical properties of biological materials and serves as an indirect transducing mechanism for physiological events, often in cases where no specific transducer for that event does exist. It is an elegantly simple technique that requires only the application of two or more electrodes (this is a big disadvantage of this method). The impedance between the electrodes may reflect “seasonal variations in blood flow, cardiac activity, respiration volume, bladder, blood and kidney volumes, uterine contractions, nervous activity, the galvanic skin reflex, the volume of blood cells, clotting, blood pressure and salivation” [14].

#### 4.1.8 Physical Characteristics: Pulse

Scanners based on this technique try to detect whether the scanned object exhibits characteristics of the pulse and blood flow consistent with a live human being [4, 10]. It is not very difficult to determine whether the object indicates some kind of pulse and blood flow, but it is very difficult to decide whether the acquired characteristics are coincident with a live sample. It is difficult to create an acceptance range of the sensor, which would lead to small error rates. The main problem is that the pulse of a human user varies from person to person—it depends on the emotional state of the person and also on the physical activities performed before the scanning procedure. In addition, the pulse and blood flow of the attacker's finger may be detected and accepted when a wafer-thin artificial sample is used.

The sensor usually detects variation in the levels of the reflected light energy from the scanned object as evidence of the pulse and blood flow [3, 4]. First, the light source illuminates in infrared the object and then a photodetector measures the light energy reflected from the object. Finally, there is the processing instrument (this also controls the light source) which processes the output from the photodetector. Since there are some ways how to simulate pulse and blood flow characteristics (e.g., by flashing the light or by motion of the scanned object), scanners should have a deception detection unit [10].

#### ***4.1.9 Physiological Basics of Heart Activity***

Heart activity measurements are well known as electrocardiogram (ECG) in medicine. In [4], two approaches for measuring of fine movements of papillary lines, based on optical principles, are suggested. The first solution is based on a close-up view of the fingertip acquired with a camera; the second one is distance measurement with a laser sensor. It should be noted that adding the proposed anti-spoof detection solution (either camera or laser based) to a fingerprint recognition system may significantly influence the hardware requirements imposed on the complete system. Both solutions measure the fine movements of skin on a fingertip caused by heart activity (blood circulation in a cardiovascular system), i.e., volume changes in arteries and veins.

#### ***4.1.10 Physical Characteristics: Blood Oxygenation***

Sensors which measure blood oxygenation [10] are mainly used in medicine (oximeters) and have also been proposed for use in anti-spoof testing modules. The technology involves two physical principles. First, the absorption of light having two different wavelengths by hemoglobin differs depending on the degree of hemoglobin oxygenation. The sensor for the measurement of this physical characteristic contains two LEDs: one emits visible red light (660 nm) and the other infrared light (940 nm). When passing through the tissue, the emitted light is partially absorbed by blood depending on the concentration of oxygen bound on hemoglobin. Second, as the volume of arterial blood changes with each pulse, the light signal obtained by a photodetector has a pulsatile component which can be exploited for the measurement of pulse rate.

### 4.1.11 *Fingerprint Spoof Preparation*

Nowadays, it is well known that there do exist many various materials, which could be used for production of fingerprint fakes [3].

The whole process of creation of the fingerprint fakes can be divided into several categories according to input data that are available. Usually, we do not have the possibility of cooperation with the user and simultaneously we have to create a mold indirectly using another information. This method is very popularized by films but we need a little bit of “cooperation” from a genuine user.

Other methods suppose that we have an access to a sensor that we want to deceive. It is possible to either use fingerprint reactivation or fingerprint synthesis [8]. The first method is based on reactivation of a fingerprint, which remains on the sensor, using for example our breath. The second method gets fingerprint image by reconstruction from enrolled template in the biometric database.

The planned procedure of making fakes using manufactured mold and chosen material is described below. First of all the mold (we used a printed circuit board) has to be cleaned—the best way appeared to be a common toothbrush and tooth paste. After cleaning the CH14 separator is applied. When the applied separator becomes dry (approx. after 10 min, i.e., the white film can be observed), it is possible to continue. A sufficient amount of silicon is placed into a measuring glass. Optionally, a small amount of body tone paint or grated graphite powder is added. In that case the silicon mixture has to be stirred thoroughly. This mixture is then spread on the mold and pushed with spatula in order to get rid of as many air bubbles as possible. When the silicon mixture gets dry, we can remove it from the mold. The removing is performed from one chosen mold side by slowly pulling. Due to the usage of CH14 separator, the dried silicon mixture does not tear apart and it is not glued to the mold (see Fig. 4.5).

Generally, we are using over 30 various materials for preparation of a fingerprint spoof, i.e., gelatin, aspic, gummy bears, aquarium silicone (black, white, skin color), epoxy resin (CHS 1200, L285, Hobbyking), latex, etc. Some of these

**Fig. 4.5** Removing dried mixture from the mold



materials are mixed with skin color or graphite, as mentioned before, because we need skin color for optical sensor or conductivity for capacitive sensors. Using these materials in combination with other advanced methods, we are able to deceive many of fingerprint sensors, incl. those ones with built-in anti-spoof mechanism.

## 4.2 Skin Diseases

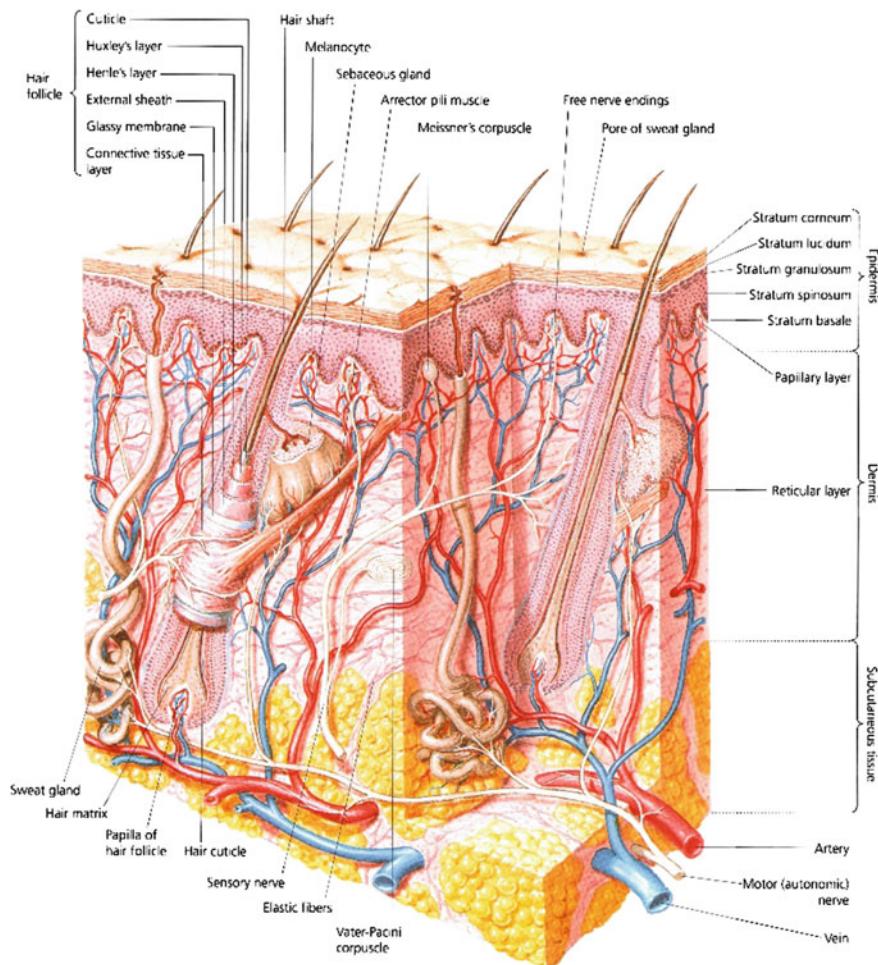
The skin is one of the largest organs in the body, having a surface area of  $1.8 \text{ m}^2$  and making up to 12–15% of an adult's body weight. It consists of three layers (see Fig. 4.6) [7]: *epidermis* (the outer layer), *dermis* ("true skin") and *subcutaneous* (fatty) *layer*. Structure and thickness of the skin vary with site (e.g., thick epidermis on palms and soles due to mechanical protection—up to 1.4 mm).

Skin diseases represent very important, but often neglected factor of the fingerprint acquisition. It is not possible to say in general how many people suffer from skin diseases, because there are so many various skin diseases [8, 20]. In a general, medical practice about 20–25% of patients with skin complaints are referred. When discussing whether the fingerprint recognition technology is a perfect solution capable to resolve all our security problems, we should always keep in mind those potential users who suffer from some skin disease.

The situation after successful recovery of a potential user from such skin diseases is, however, very important for the possible further use of fingerprint recognition devices. If the disease has attacked and destroyed the structure of papillary lines in the epidermis and underlying dermis (so-called dermoepidermal junction—connection of the top two layers of the skin), the papillary lines will not grow in the same form as before (if at all) and therefore this user could be restricted in his future life by being excluded from the use of fingerprint recognition systems, though his fingers do not have any symptoms of the skin disease anymore.

Skin is constantly being regenerated. A keratinocyte ("skin cell") starts its life at the lower layer of epidermis (the basal layer), which is nourished by blood vessels and is supplied with nerve endings from dermis. The cell migrates upward from basal layer to stratum corneum (the outermost skin layer). During 4 weeks the cell undergoes a series of changes, gradually flattening out and moving toward the surface. Then it dies and is shed. This physiological process can be negatively affected in many diseases of the skin. The epidermis is not supplied with blood vessels, but has nerve endings. The shape of dermoepidermal junction basically forms the structure of papillary lines.

In the most cases of dermatological disorders we find a lot of changes in the ultrastructure of the skin, including epidermis and dermis. There is often inflammation (inflammatory cells), atrophy or hypertrophy, fibrotisation and many other changes visible in the microscope. These differences result in changes of color (optical characteristics), changes of dermal vessels and capillaries (blood perfusion), changes of elasticity, and thickness of the skin (optical characteristics after pressure change).



**Fig. 4.6** Skin structure [7]

### Diseases Causing Histopathological Changes of Epidermis and Dermis

These diseases usually cause problems for all kinds of fingerprint scanners, because they can influence either color or internal structure of the skin.

The most common representatives of this group are [7, 20]: *Hand and fingertip eczema, Dyshidrosis, Tinea, Pyoderma, Pitted keratolysis, Pyogenic granuloma, Systemic sclerosis, or Raynaud's phenomenon.*

### Diseases Causing Skin Discoloration

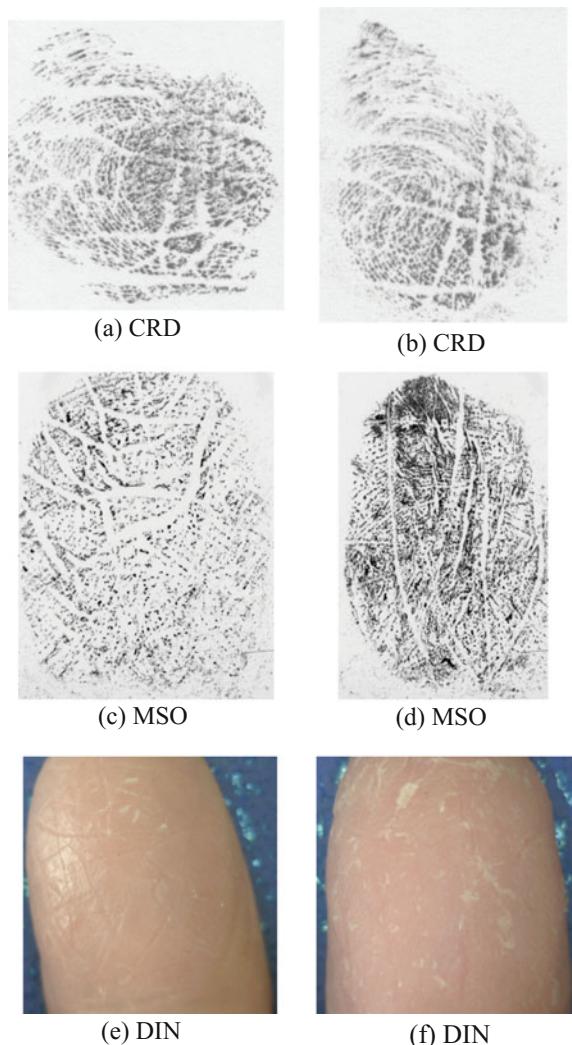
These diseases may cause problems for optical fingerprint scanners and also for scanners which use a fingerprint anti-spoof detection check based on the color or spectral analysis of the human skin.

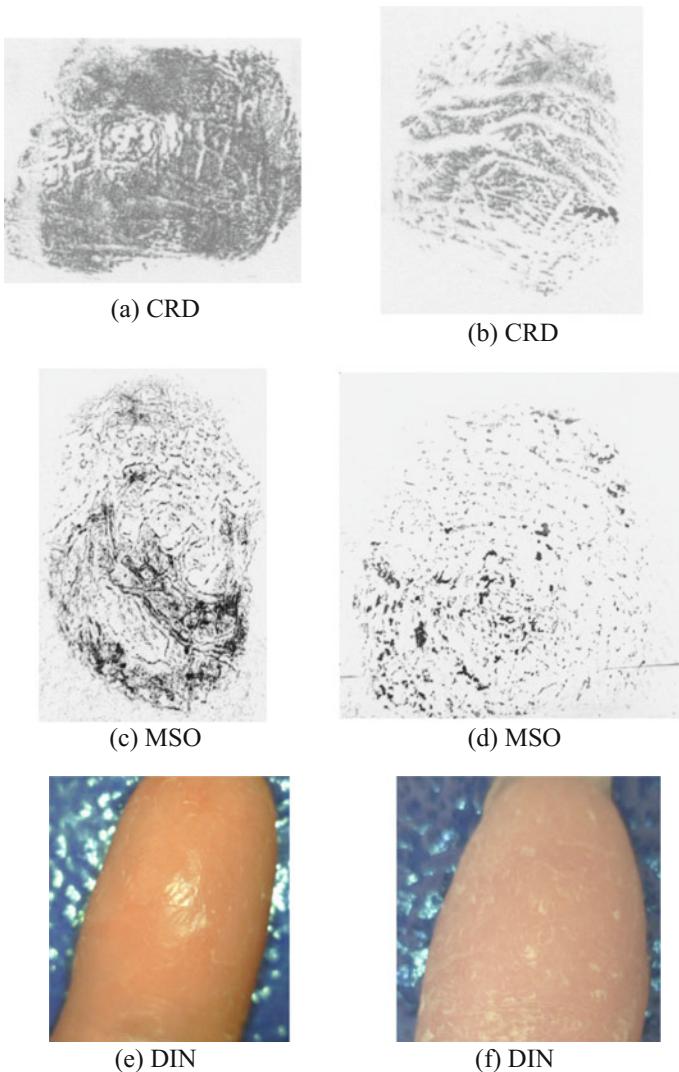
Typical representatives are [7, 20]: *Macular drug eruptions and rashes in infectious diseases (Hand, foot and mouth disease, Scarlet fever, Secondary syphilis, Kawasaki's disease), Pitted keratolysis, Raynaud's phenomenon, Xanthomas, Carotenosis, or Hereditary hemorrhagic telangiectasia.*

### Diseases Causing Histopathological Changes in Junction of Epidermis and Dermis

These diseases could cause structure changes underneath the skin in the junction between dermis and epidermis, i.e., in the area from that ultrasonic fingerprint

**Fig. 4.7** Fingertip eczema—a severe form

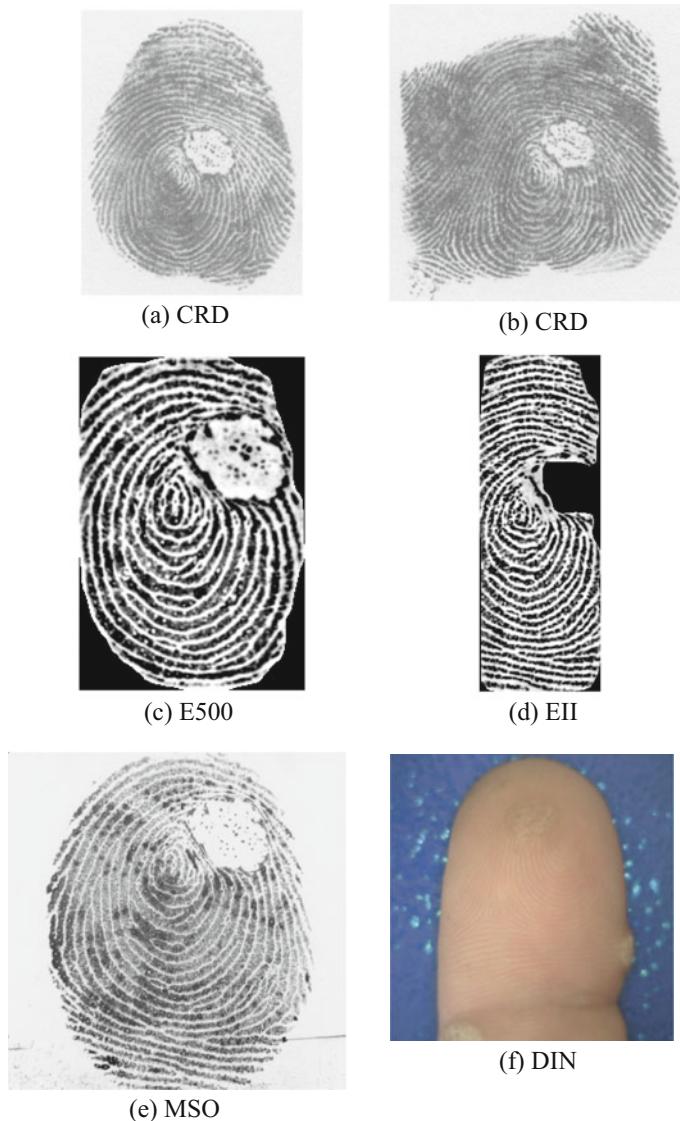




**Fig. 4.8** Psoriasis—a full seizure

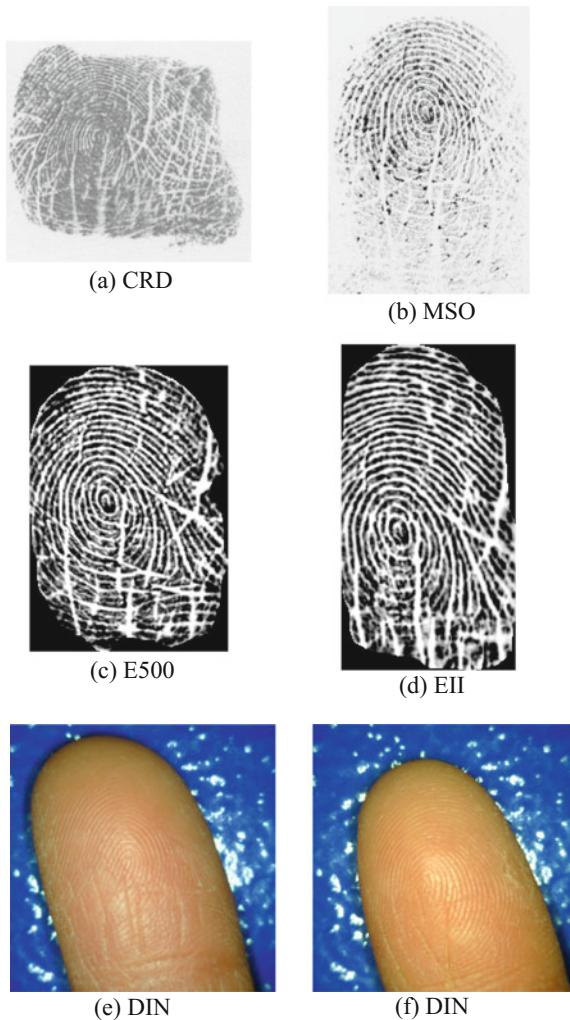
scanners acquire fingerprint pattern images. Typical representatives are [7, 20]: *Hand eczema*, *Verruca vulgaris* (warts), *Psoriasis*, or *Epidermolysis bullosa*.

In the following section, examples (Figs. 4.7, 4.8, 4.9 and 4.10) of results of disease-affected fingerprint data collection are shown. The name of the disease and description are given in each figure heading. The codename of applied capturing



**Fig. 4.9** Verruca vulgaris (wart)

principle is stated under each subfigure. We use the following codenames: CRD (dactyloscopic card), MSO (Sagem MSO 300), E500 (UPEK EikonTouch500), EII (UPEK Eikon II), TBS (TBS 3D Enroll 2011), and DIN (Dinolite).

**Fig. 4.10** Collagenosis

### 4.3 Environmental Distortions

When we finally place finger in the acquisition area there are still some difficulties. Environment around us is not ideal like in laboratory—so there are some influencing factors. We can divide them into three categories: *finger condition*, *sensor condition*, and *environment itself*. Some of these factors can be prevented but only with full support of users and thorough care of the sensor and specialized room. As you can see it is not possible to fully prevent all these factors. Consequently, sensors and recognition algorithms have to be prepared for this situation. The only way how to prepare them is by using large test dataset of damaged

fingerprints. Unfortunately such data is not usually available. But we can create this data artificially. These topics are covered in the following subsections, first factors that influence resulting fingerprint are discussed, followed by creation of artificial fingerprints.

### 4.3.1 *Phenomena Influencing Fingerprint Acquisition*

As was mentioned before we can divide these factors into three groups: *user condition*, *sensor condition*, and *environment*—these factors will be discussed next. Each of these factors is described with example and sensor technologies that are influenced more or technologies that are not influenced at all.

First user condition is a *dirt on the finger*—it could be a small particle, a few grains of dust, or just a greasy finger. Conductive materials and liquids are usually the most problematic types of dirt. Only ultrasonic, contactless, and e-field technologies are resistant against this type of damage. *Dry or moist finger* is one of the most typical cases of damage done to a fingerprint. This is caused because we wash our hands or we are nervous and our fingers are getting sweat or on the other hand we have very dry hands because of some lotion, our skin resistance can increase or decrease ten times to the normal value. This plays usually a huge role in the recognition by optical, capacitive, and e-field sensors. *Physical damage* of a finger like cuts or abrasions is obviously damaging the fingerprint. If it is not a deep injury that influences papillary lines forever, the ultrasonic and e-field technologies scan the finger in the deeper dermis layer where the fingerprint keeps undamaged. Another factor is *skin diseases* which were thoroughly described in the previous subchapter. *Pressure and skin plasticity* can turn the fingerprint into a big black oval. Only contactless sensors are fully immune to the damage that the pressure can make. The change of pressure, very big or very low pressure or moving of the finger is also considered being part of the next category that is *non-cooperative behavior*. The user usually uses an unexpected pressure, moves when the device is scanning and/or places the finger in a wrong place or with a wrong rotation. None of the technologies is fully resistant to these types of behavior [3, 9].

The second group of factors is connected to the sensor. *Dirt on the surface* has the same effects like the dirt on the finger. The problem is that it is affecting everyone who is using this device. It means that in the registration phase this factor can create a common error for every user and there is a danger that these users will not be able to be identified after cleaning up the device. In addition to fingers there are more types of dirt than can pollute the sensor area: for example metallic dust, wooden dust, earth dust, fine sand, excrements (in outdoor use). In addition to ultrasonic and e-field technologies, every sweep sensor is also more resistant to this type of damage. *Latent fingerprint* is closely related to the previous topic. It is in some way a type of dirt on the surface of the sensor. More than damaging a new fingerprint there is a security hazard. The technologies, which are resistant to latent fingerprint, are the same like those in the previous topic. *Sensor technology* itself

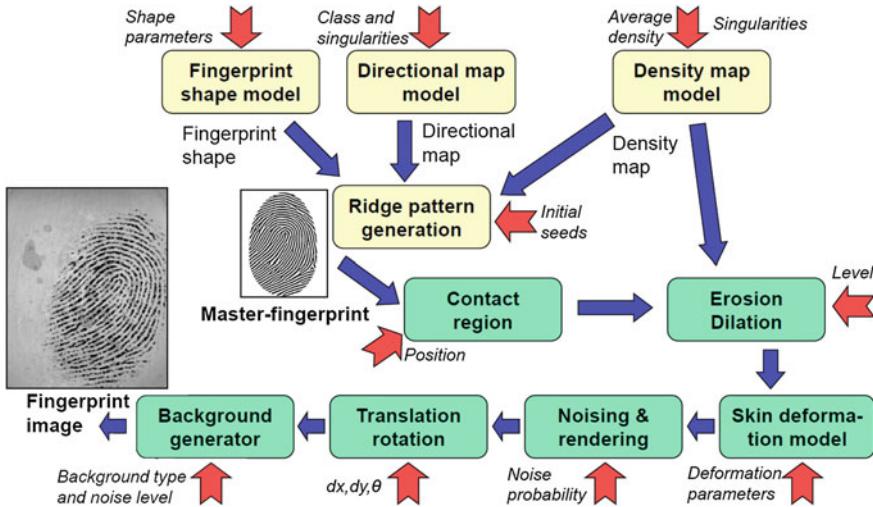
determines how the fingerprint will look like. Some technologies support only binary images or use specific pallet of colors to represent ridges and valleys, some can create 3D maps. *Physical damage* is an extreme but a possible influencing factor of the resulting fingerprint. There is no easy way to prevent the sensor from damaging. The damage of the sensor will have various effects on every technology [3, 9].

Finally, the surrounding environment itself will be discussed. *Vibration* can break down the sensor and slightly change the position of a finger. Only sensors using the sweep technology are, to a certain degree, resistant to this type of damage. *Temperature* can be different for the sensor, the finger or the environment. Typically, there are no problems with the exception of the thermal technology. But when we think about extreme temperatures, we have to deal with very dry or very moist fingers which can affect the resulting image. Also it is known that the ultrasonic technology does not operate properly in extremely low temperatures. *Surrounding light* is only affecting optical and electrooptical technologies because they have a light sensing unit. When the sensor area is larger or the finger of the user is smaller or the contactless technology is used, the influence of the surrounding light can be huge. *Electromagnetic radiation* is an influencing factor which affects every technology. The device as a whole can be influenced by electromagnetic radiation. Wires inside or outside connecting it to other parts of a biometric system and all electronic components can be influenced [3, 9].

#### 4.3.2 Methods for Generation of Synthetic Fingerprints

The synthetic fingerprint generation is an inverse biometrics problem. According to input variables we basically do the fingerprint recognition process from the end to the start. There are several methods how to generate a synthetic fingerprint. When we thoroughly study them, we can find that they are all based on the same principle. The method used by the SFinGe seems to be the oldest one and also the most commonly known so it will be described as a pattern for others.

For better understanding we can look at the upper part of Fig. 4.11 to see the process of the generation. Firstly, the fingerprint shape is determined. The basic shape is oval and each elliptical segment can be changed to create the required shape. The second step is the directional field model. In this step, the fingerprint class is chosen and together with that the position of cores and deltas. This step is using the Sherlock and Monroe ridge [12] flow model to generate a consistent directional field. The third step creates the density map. When we look at a fingerprint, we can find that the density of papillary lines is not the same throughout the whole area. After examining several real fingerprints some heuristic criteria could be made. These criteria are based on the position of singularities (cores and deltas) and according to them the density map is generated. The last step is ridge pattern generating. This phase uses all previous steps and some initial seeds. Iteratively, the image with initial seeds is refined with the Gabor filter. The filter



**Fig. 4.11** SFinGe process of artificial fingerprint generation

orientation and frequency is adjusted according to the directional field and density map. Minutiae are automatically generated at random places with random types that are not only ridge ending and bifurcation, but also more complex ones. After that phase, the fingerprint is done [9, 12, 21].

As we can see, the SFinGe<sup>1</sup> generating process is not exactly an inverted recognition process. When we strictly follow this process, we do the so-called fingerprint reconstruction. These are the methods that focus on the creation of the whole fingerprint only from minutiae saved as a template in fingerprint recognition. Another method is somewhere between these two. It says that fingerprint features are dependent on each other. It is following the same scheme but with dependencies on other steps. The orientation field is influenced by singular points. The minutiae density is higher around singularities and also their appearance is not random but it is statistically driven. The minutiae direction is also dependent on their types and on the orientation of ridges around. This method firstly determines singular points, after that the orientation field and lastly the minutiae. Each step is dependent on the previous one. After all the steps the fingerprint is made with the use of the AM-FM method [21].

The last method uses minutiae as an input. The creation of the whole fingerprint is based only on these minutiae. The biggest difference is that the orientation field is generated from minutiae and not from classes or singular points as it was in the previous methods. It is generated from the minutiae direction and each minutia has a weight based on the distance of it from the point where we are determining the orientation field. The disadvantage of this method is that the final fingerprint could

<sup>1</sup><http://biolab.csr.unibo.it/sfinge.html>.

have a class that does not exist in the real world. The density map can be manually changed in this method. After that, using a similar method of Gabor filter like in SFinGe, master fingerprint is generated. Note that instead of initial seeds, this method uses minutiae as these seeds and the generation start with them so the precisely defined minutiae don't change in the process of generation [9].

To sum it up, we can create these artificial fingerprints from ISO template, based on statistical values, using random seeds and random minutiae points or using fixed minutiae points. For testing random fingerprints is this usually fully sufficient. All these methods end with creation of the artificial fingerprint which is called master fingerprint. This master fingerprint is what was mentioned at the start of this chapter that is perfect finger perfectly scanned with no influence of the environment.

To create more realistic fingerprint we have to add these influential factors. It is time to look at the lower part of the Fig. 4.11. There are certain damage simulation methods. The first step is the selection of the contact region. To simulate the different placements of the finger on the sensor area a random translation of the ridge pattern is made. The next step is the variation in ridge thickness. The ridge thickness is modified to simulate various skin dampness and finger pressure. The next phase is the fingerprint distortion. In this phase, the skin deformation according to different finger placements over the sensor is simulated. The skin plasticity (compression or stretching) and a different force applied on each part of the finger creates a nonlinear distortion. The next step is noising and rendering. Another phase is the global translation or rotation. This phase simulates the not perfectly placed finger on the sensor. So it slightly translates and/or rotates the whole image. The last step is the generation of a realistic background. The background is generated randomly from a set of background images. At the end of that step, the fingerprint impression is made [9, 12, 21].

Please note that SFinGe system, which was used here as an example of generation and damaging of the artificial fingerprint, lies the focus in realistic looking fingerprints without exact simulation of the damage done to the fingerprint. Some steps clearly simulate specific damage, others are used as an appropriation of several factors, which is to some extent sufficient but when we want to create fingerprints acquired in some extreme environment or one which is often and more severely influenced by some phenomena we have to use precise damage simulations. We work on these damage simulations (incl. skin diseases) at the moment.

## 4.4 Conclusion

At the beginning of this chapter, there are described anti-spoofing methods, which are used for liveness detection on fingers in general. Some relevant advantages and disadvantages of these methods are discussed as well. Furthermore, the second part includes description of skin diseases on fingers influencing the fingerprint recognition process. It has to be noted that the anti-spoofing methods have big troubles with diseased fingers, because these fingers are very often falsely evaluated as

nonliving (fake fingers). Therefore, it is very important to implement appropriate methods for anti-spoofing, which can handle diseased fingers. The last part of this chapter is devoted to other environmental influencing factors, which can cause troubles in the process of fingerprint recognition. These factors have similar impact as diseased fingers, i.e. there could be seen comparable impact to the fingerprint recognition process. As mentioned before, it is very important to find a good and reliable method for anti-spoofing, which can correctly treat diseased fingers or acquired fingerprints by various influencing factors coming from the environment.

**Acknowledgments** This work was supported by The Ministry of Education, Youth and Sports of the Czech Republic from the National Programme of Sustainability (NPU II); project “IT4Innovations excellence in science”—LQ1602; “New solutions for multimodal biometrics—enhancement of security and reliability of biometric technologies”—COST LD14013 (CZ); “Reliability and Security in IT”—internal Brno University of Technology project FIT-S-14-2486 (CZ).

## References

1. Ambalakat P (2005) Security of biometric authentication systems. In: 21st Computer Science Seminar, SA1-T1–1 (2005), p 7
2. Daugman J (2001) Biometric Decision Landscapes, University of Cambridge, p 13
3. Drahanský M (2011) Fingerprint Recognition Technology—Related Topics. Saarbrücken, DE, LAP, 2011, p 172. ISBN 978-3-8443-3007-6
4. Drahanský M, Funk W, Nötzel R (2006) Liveness detection based on fine movements of the fingertip surface. In: IEEE—The West point workshop, West Point, New York, USA, pp 42–47. ISBN 1-4244-0130-5
5. Drahanský M, Hejtmánková D (2010) New experiments with optical liveness testing methods. *J Inf Hiding Multimedia Signal Process* 1(4):301–309. ISSN 2073-4212
6. Habif TP (2004) Clinical dermatology, 4th edn. Mosby, China, p 1004. ISBN 978-0-323-01319-2
7. Habif TP (2004) Clinical dermatology, 4th edn. Mosby, China, p 1004. ISBN 978-0-323-01319-2
8. Jain AK, Flynn P, Ross AA (2008) Handbook of biometrics. Springer, p 556. ISBN 978-0-387-71040-2
9. Kanich O (2014) Fingerprint damage simulation—A simulation of fingerprint distortion, damaged sensor, pressure and moisture, LAP LAMBERT Academic Publishing GmbH & Co. KG, p 57. ISBN 978-3-659-63942-5
10. Kluz M (2005) Liveness testing in biometric systems. Master Thesis, Faculty of Informatics, Masaryk University Brno, CZ, p 57
11. Lodrová D (2013) Security of biometric systems. Dissertation thesis, FIT BUT, Brno (CZ), p 152
12. Maltoni D, Maio D, Jain AK, Prabhakar S (2009) Handbook of fingerprint recognition, 2nd edn. Springer, p 494. ISBN 978-1-84882-253-5
13. Malý T (2013) Detekce živosti prstu pomocí osvětlení různé délky (Detection of Finger Liveness using Illumination with Different Wavelengths), Diploma thesis, FIT BUT, Brno (CZ), p 59
14. Martinsen ØG, Grimnes S, Haug E (1999) Measuring depth depends on frequency in electrical skin impedance measurements. In: Skin research and technology, No. 5, pp 179–181. ISSN 0909-752X

15. Matsumoto T, Matsumoto H, Yamada K, Hoshino S (2005) Impact of artificial “Gummy” fingers on fingerprint systems. In: Proceedings of SPIE, Optical Security and Counterfeit Deterrence Techniques IV, vol 4677, p 11
16. Rowe RK (2008) Spoof Detection. Summer school for advanced studies on biometrics for secure authentication. Italy, Alghero, p 43
17. Schuckers S, Hornak L, Norman T, Derakhshani R, Parthasaradhi S (2003) Issues for liveness detection in biometrics. West Virginia University, Presentation, CITeR, p 25
18. Tan B, Lewicke A, Schuckers S (2008) Novel methods for fingerprint image analysis detect fake fingers. In: SPIE, p 3. doi:[10.1117/2.1200805.1171](https://doi.org/10.1117/2.1200805.1171)
19. Organism Identifying Method and Device. US Patent 6,314,195, Nov 2001
20. Wolff K, Johnson RA, Suurmond D (2005) Fitzpatrick’s Color Atlas and Synopsis of Clinical Dermatology, 5th Edition. McGraw-Hill, USA, p 1085. ISBN 0-07-144019-4
21. Zhao Q, Jain AK, Paulte NG, Taylor M (2012) Fingerprint image synthesis based on statistical feature models. In: 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS). IEEE, pp 23–30

# Chapter 5

## Altered Fingerprint Detection

John Ellingsgaard and Christoph Busch

**Abstract** The success of Automated Fingerprint Identification Systems (AFIS) has lead to an increased number of incidents where individuals alter their fingerprints in order to evade identification. This is especially seen at border crossings where fingerprints are subject to comparison against a watch list. This chapter discusses methods for automatically detecting altered fingerprints. The methods are based on analyses of two different local characteristics of a fingerprint image. The first analysis identifies irregularities in the pixel-wise orientations which share similar characteristics to singular point. The second analysis compares minutia orientations covering a local, but larger area than the first analysis. A global density map is created in each of the analysis in order to identify the distribution of the analyzed discrepancies. Experimental results suggest that the method yields performance fully comparable to the current state-of-the-art method. Further improvements can be achieved by combining the most efficient analysis of the two methods. The promising results achieved in this study are attractive for further investigations. Especially, studies into the possibility of introducing alteration detection into standard quality measures of fingerprints which would improve AFIS and contribute to the fight against fraud.

### 5.1 Introduction

Identification using fingerprints is the most matured and widespread biometric technique that currently exists. Fingerprints have a long history as a tool for identification and forensic purposes. Technological advancement has lead to the development

---

J. Ellingsgaard  
Technical University of Denmark, Anker Engelunds Vej 1, 2800  
Kongens Lyngby, Denmark  
e-mail: jellingsgaard@hotmail.com

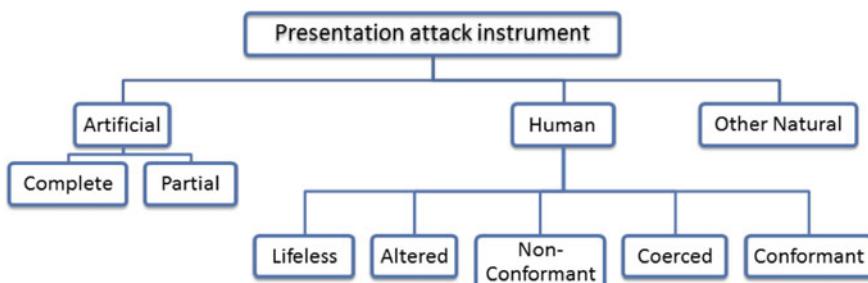
C. Busch (✉)  
Norwegian Biometrics Laboratory, Norwegian University of Science  
and Technology (NTNU), P.O. Box 191, 2802 Gjøvik, Norway  
e-mail: christoph.busch@ntnu.no

of Automated Fingerprint Identification Systems (AFIS), which are used by border control and law enforcement agencies for identification purposes.

One such application is the European Visa Information System (VIS). The system consists of distributed national interfaces (NI-VIS) that are linked together with a central information system (CS-VIS) [8]. The system contains alphanumeric data as well as biometric data in the form of fingerprints and photographs, for identification and verification purposes. One method used to avoid identification of such a system is to alter one's fingerprints e.g. by obfuscating ridge flows by scraping, cutting or burning, or even in extreme measures using plastic surgery [43].

Altered fingerprints on a real finger are not necessarily easy to spot by a quick glance on the fingers of a data subject. Changes can be subtle to the naked eye and would require officers to do a closer inspection of every finger to positively identify alterations. The international standard ISO/IEC 30107-1 [18] defines a presentation attack instrument (PAI) as the biometric characteristic or object used in a sensor-based presentation attack. Artificial (fake) or human-based characteristics are the two main categories; a third category covering natural cases such as animal- and plant-based PAI is also included for completeness. Figure 5.1 gives an overview of presentation attack types while examples of each of the types belonging to the two main categories are shown in Table 5.1.

This chapter will deal with the aspect of detecting altered fingerprints. The objective is not to identify the actual identity of an individual that has altered fingerprints,



**Fig. 5.1** Types of presentation attacks. *Source* [ISO-IEC 30107-1]

**Table 5.1** Artificial and human attack presentation characteristics [18]

Main	Characteristic	Example
Artificial	Complete	Gummy finger
	Partial	Glue on finger
Human	Lifeless	Cadaver part, severed finger/hand
	Altered	Mutilation, surgical switching of fingerprints
	Non-Conformant	Tip or side of finger
	Coerced	Unconscious, under duress
	Conformant	Zero effort impostor attempt

but instead to detect and raise an alarm if a fingerprint is considered to be altered. The chapter is structurally divided into four parts. The first part introduces the history of altered fingerprints. The second part describes related work and state-of-the-art algorithms for detecting altered fingerprints. The third part is the contribution of a proposed algorithm for detecting altered fingerprints. Finally the results are evaluated and discussed.

## 5.2 Background of Fingerprint Alterations

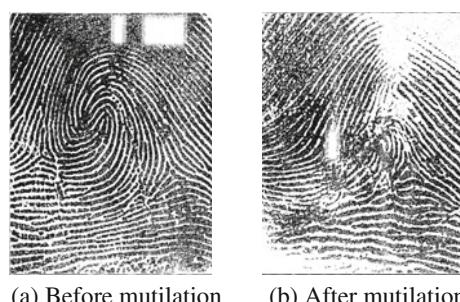
Fingerprints have a long history of being used for forensics and other identification purposes. As the importance of the usage of fingerprints has grown through time and identification techniques have improved, the instances of individuals trying to deceive the system and avoid being identified have become more common. Already back in 1935 H. Cummins [9] published information on three criminal cases involving altered fingerprints. The cases were among others the following:

- John Dillinger applied acid to the finger tips in order to burn and permanently change the fingerprints. After his death it was determined that careful examination of the remaining undamaged areas of the fingerprints would be enough to positively identify him solely on the fingerprints.
- Gus Winkler mutilated four of his fingerprints on the left hand, excluding the thumb, possibly by the combination of slashing and deeply scraping. He was actually successful in changing his pattern type from double loop to left loop (see Fig. 5.2).

Other incidents demonstrate individuals using more advanced and inventive techniques for masquerading their identity.

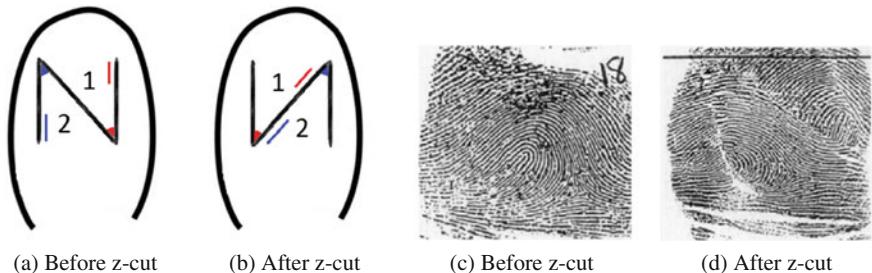
- Robert J. Philipps (1941) attempted to completely erase his fingerprints by transplanting skin grafted from the side of his chest onto the fingertips [16].
- Jose Izquierdo (1997) cut a “Z” shaped cut (see Fig. 5.3) on his fingertip and exchanged the two flaps of skin. After manually reconstructing his real fingerprint images officials managed to reveal his true identity; this came with a large cost of approximately 170 h of manual and computer searching [41].

**Fig. 5.2** Gus Winkler changed the pattern type from double loop to left loop. *Source* [9]



(a) Before mutilation

(b) After mutilation



**Fig. 5.3** Illustrations of how two flaps of skin can be exchanged within a “Z” shaped cut. The numbers and colors are merely for illustration purposes to show skin positions before and after the surgery. Altered fingerprints using a “Z” shaped cut in the skin [41]

**Fig. 5.4** Images of altered fingertips [22]

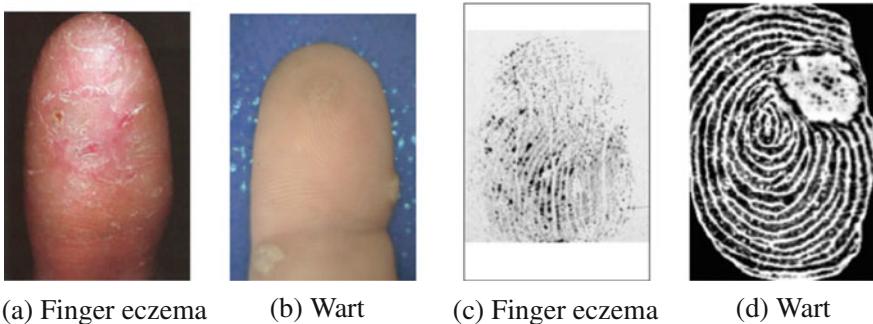


The above mentioned examples are examples from forensic investigations. However, border crossings are seeing an increased amount of asylum seekers and migrants with mutilated fingerprints who try to avoid being identified. The *MailOnline* reported on migrants mutilating their fingertips to hide their identity [29]. Images that show actual fingers with altered fingertips can be seen in Fig. 5.4.

Based on the observations by Feng [13] altered fingerprints are classified into three categories based on the changes in ridge pattern due to alteration [43]. The classification types are based on the fingerprint image and not on the actual alteration process [44]. The following will describe these categories, analyze discrepancies, and special features of fingerprints in these categories, which will serve as the basis for understanding the structure of common alterations.

### 5.2.1 Obliteration

Probably the most common form of alteration is *obliteration*, which relates to diminishing or even destroying the quality of the friction ridge patterns on the fingertips in order to make it problematic to match with the enroled reference. Obliteration can be performed by incision, scraping, burning, applying acids or transplanting smooth



**Fig. 5.5** Diseases can obliterate a fingerprint. Images and subsequent fingerprint images do not belong to the same subjects. *Source* [10]

skin [13]. John Dillinger for instance mutilated his fingertips with acids. Obliteration can be perceived as a natural extension to the problem of identifying low quality fingerprints [34]. The quality of unaltered fingerprints can vary depending on different factors such as the quality of the actual scan or damages such as scars. Also skin diseases such as eczema or warts can have a degrading impact on the quality of the friction ridge patterns (see Fig. 5.5). Fingerprint quality assessment software such as NFIQ2.0 [33] could in many cases be used to deny enrolling or comparing a heavily obliterated fingerprint, since the quality would simply be deemed too low for comparison.

### 5.2.2 Distortion

Distortion is the reshaping of the original patterns of the friction ridges. This can be done by removing and reorganizing portions of skin from fingertips or by transplanting other skin with friction ridge patterns unto the fingertip. The resulting fingerprints on the fingertips will have unnatural ridge patterns. Jose Izquierdo distorted his fingerprints by exchanging two portions of skin on the fingertip by a “Z” shaped cut (see Fig. 5.3). Fingertips that have been successfully distorted may reach a high sample quality score [17, 32], since they will have clearly visible friction ridge patterns throughout the whole fingerprints; possibly even preserving ridge properties such as width and frequency over the entire fingerprint area. A closer look at a distorted fingerprint will however show clear irregularities. There will typically be sudden changes in the orientation of friction ridges along the scars where different skin patches are joined together. Also, distortion can result in unnatural distribution of singular points (Table 5.2).

Unaltered fingerprints normally have a smooth orientation field throughout the whole fingerprint except in the vicinity of singular points.

**Table 5.2** Characteristics of distortion

	Distortion
Definition	Misrepresentation, misshape, a change in perception so that it does not correspond to reality (psychology)
Performed by	Removing and reorganizing portions of skin from fingertips or by transplanting skin with friction ridge patterns
Characteristics	Unnatural ridge patterns and scarred areas

### 5.2.3 *Imitation*

The most advanced category of altered fingerprints is imitated fingerprints. This is not referring to spoofing or artefacts as presentation attack instrument but rather to the required skills and the quality of the alteration. Imitated fingerprints have friction ridge patterns that both preserve ridge properties, e.g., width and frequency, while also containing the typical smooth orientation field pattern found in unaltered fingerprints. A typical imitation technique includes transplantation of a large area of friction ridge skin. An example is to remove a portion of friction ridge skin and thereafter join together the remaining skin. For this to be a success, friction ridges on each side of the scar must principally avoid abrupt changes in orientations. Gus Winkler was successful in this technique, even changing the type of his finger pattern in the process. The main difference between *distortion* and *imitation* is the fact that imitated fingerprints maintain the smooth orientation field characteristics of an unaltered fingerprint.

The problem with imitated fingerprints is that they contain so many properties of an unaltered fingerprint and are of such a good quality that they will successfully pass fingerprint quality assessment software. Well executed imitation can even be hard to spot even with a close inspection of the fingertips by the human expert.

The main focus of the following section is on distorted fingerprints. Obliterated fingerprints will, in most cases, already be processed correctly based on the area and amount of obliteration. Either fingerprint quality assessment software will evaluate that the fingerprint quality is too low or it will be processed correctly in the biometric identification system. Distorted fingerprints can have a high quality level and share many properties with unaltered fingerprints. However, they have clearly identifiable properties, such as irregular and abrupt changes in the orientation of friction ridges.

## 5.3 Related Work

Relatively limited research has been done in the field of automatically detecting altered fingerprints. Yoon [43] proposed a technique based on analyzing discontinuity and abnormality in the flow of the friction ridges along with analyzing the spatial distribution of minutiae. The algorithm is based on two different analyses.

- Analysis of the friction ridge orientations. Fingerprints generally have a smooth ridge flow except in the vicinity of a *singular point*. Altered fingerprints will typically result in irregular and abrupt changes in the ridge flow in some areas of the fingerprint. This approach tries to identify regions of unnatural ridge flow. The algorithm is called *Orientation Field Analysis (OFA)* in this chapter.
- Analysis of *minutia* distribution. A minutia point is located at local discontinuities in the fingerprint pattern where friction ridges begin, terminate or bifurcate. This analysis will be named *Minutiae Distribution Analysis (MDA)*.

Feature vectors are constructed from each of the analyses, fused into one larger feature vector and fed into a Support Vector Machine (*SVM*) for classification. The following subsections will describe how the feature vectors are constructed.

### 5.3.1 Orientation Field Analysis

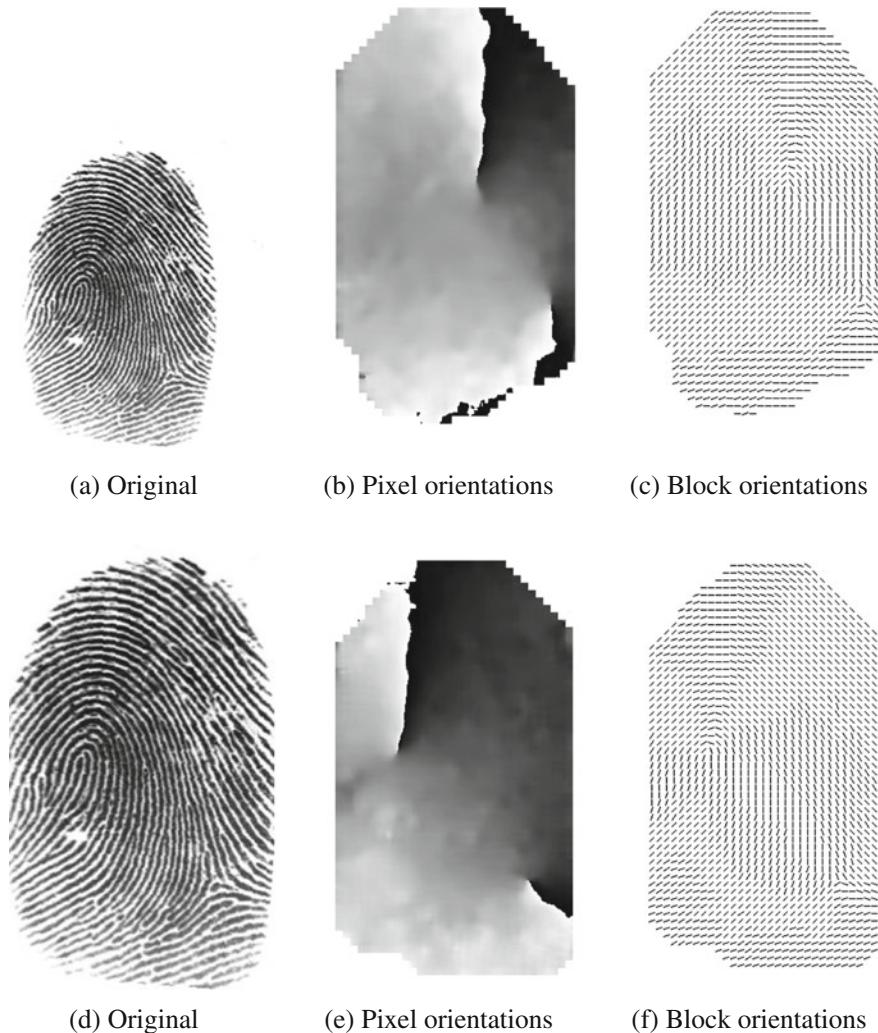
The *Orientation Field Analysis (OFA)* uses a mathematical model for constructing an approximation of an estimated ridge flow of the fingerprint. The analysis identifies discontinuities based on differences of the ridge flow approximation and estimation, e.g. areas where the approximation is unable to correctly simulate the actual fingerprint image. The orientation of friction ridges, typically called *orientation field* and denoted  $\theta$ , is defined as an image where  $\theta(x, y)$  holds the estimated orientation at pixel  $(x, y)$ . The steps of the analysis are the following:

1. Normalization of the image to have a common rotation and size.
2. Segmentation. The foreground of the fingerprint is separated from the background of the image in order to be able to only analyze the actual fingerprint.
3. Orientation field estimation. The  $n \times n$  block-wise averaged orientation field,  $\theta(x, y)$ , is computed using the gradient-based method.
4. Orientation field approximation. A polynomial model is used to approximate the orientation field  $\theta(x, y)$  to obtain  $\hat{\theta}(x, y)$ .
5. Orientation error map. The absolute difference between the orientation field  $\theta(x, y)$  and the approximation  $\hat{\theta}(x, y)$  is computed to yield an error map,  $e(x, y)$ .

Since the orientation field serves as a central part of this method and also in the upcoming proposed algorithm, a technique for constructing such an orientation field will be described in more detail.

#### 5.3.1.1 Orientation Field Estimation

The *orientation field* is the local orientation of the friction ridges. It serves as an essential part in all stages of analyzing a fingerprint, such as preprocessing and feature extraction. The orientation field basically holds information on the local orientations of friction ridges. Orientations are typically defined in the range  $[0, \pi)$ . Depending on the context pixel-wise and block-wise orientation fields are used. Instead of



**Fig. 5.6** Orientation of fingerprints. **a, d** show the original fingerprint images [6], **b, e** contain pixel-wise orientations in greyscale, and **c, f** show block-wise orientations

using local ridge orientation at each pixel, it is common to partition the image into smaller blocks. The block-wise orientations are derived by simply averaging the orientations within each block. Figure 5.6 shows the orientation field of two fingerprint images; the pixel-wise orientations are illustrated in greyscale while block-wise orientations use lines to represent orientations within each block.

There are two common approaches to compute the orientation field of a fingerprint: filter bank-based approaches and gradient-based approaches. An example of a filter bank-based approach is a method proposed by Kamei and Mizoguchi [24] using

**Fig. 5.7** Sobel's two  $3 \times 3$  gradient kernels

$$S_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad S_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

directional filters in the frequency domain. According to Gu and Zhou [15] filter bank-based approaches are more resistant to noise than gradient-based approaches, but computationally expensive. Gradient-based methods seems to be the most common approach for extracting local ridge orientation; probably since it is the *simplest and most natural approach* [30].

### 5.3.1.2 Pixel Orientation

A natural approach for extracting ridge orientation is based on computation of gradients in the fingerprint image. The first step is determining the gradient components  $\delta_x$  and  $\delta_y$  for each pixel in the image. This implementation uses a *Sobel* operator to define the pixel gradient components. The Sobel operator uses two  $3 \times 3$  gradient filters—one for calculating the horizontal changes in the image and the other for vertical changes. The two kernels are illustrated in Fig. 5.7 where  $S_x$  is the kernel for the horizontal direction and  $S_y$  is the vertical. For each pixel,  $(i, j)$ , in the image two two-dimensional convolutions with the Sobel kernels are computed, yielding the gradient components  $\delta_x(i, j)$  and  $\delta_y(i, j)$ .

For each gradient the magnitude and vector angle can be calculated using Eqs. (5.1) and (5.2). However, pixel-wise orientation is very sensitive to noise in the fingerprint image and therefore is too detailed and somewhat inaccurate. The solution is to calculate block-wise averages of the pixel gradients.

$$G = \sqrt{\delta_x^2 + \delta_y^2} \quad (5.1)$$

$$\theta(i, j) = \frac{\pi}{2} + \arctan\left(\sqrt{\frac{\delta_x(i, j)}{\delta_y(i, j)}}\right) \quad (5.2)$$

### 5.3.1.3 Block-Wise Ridge Orientation

Block-wise averages of gradients have multiple purposes when processing fingerprint images. Typically the orientation (or gradients) of each pixel is first smoothed using an averaging filter from a larger area of the image before assigning block-wise orientation averages. The same averaging technique is used in both cases. Yoon et al. [43] use a  $16 \times 16$  averaging filter to smoothen the pixel-wise orientations prior to computing the block-wise orientations. The equations for calculating the block-

wise orientations for each block are given where pixel  $(i, j)$  is the center of the block being calculated. Equations (5.3) and (5.4) show the two components,  $V_x$  and  $V_y$ , of the doubled local ridge orientation vector [37].  $W$  is the block size, Yoon here uses a  $8 \times 8$  pixel block [43]. The dominant ridge flow is then computed according Eq. (5.5).

$$V_x(i, j) = \sum_{u=i-\frac{W}{2}}^{i+\frac{W}{2}} \sum_{v=j-\frac{W}{2}}^{j+\frac{W}{2}} 2\delta_x(u, v) \cdot \delta_y(u, v) \quad (5.3)$$

$$V_y(i, j) = \sum_{u=i-\frac{W}{2}}^{i+\frac{W}{2}} \sum_{v=j-\frac{W}{2}}^{j+\frac{W}{2}} (\delta_x(u, v)^2 - \delta_y(u, v)^2) \quad (5.4)$$

$$\theta(i, j) = \frac{\pi}{2} + \frac{1}{2} \arctan 2(V_y(i, j), V_x(i, j)) \quad (5.5)$$

### 5.3.1.4 Orientation Field Approximation

The global orientation field,  $\theta(x, y)$ , is approximated by a polynomial model to obtain  $\hat{\theta}(x, y)$ . The cosine and sine components of the doubled orientation at  $(x, y)$  can be represented by bivariate polynomials of order  $n$ :

$$g_c^n(x, y) \stackrel{A}{=} \cos 2\theta(x, y) = \sum_{i=0}^n \sum_{j=0}^i a_{i,j} x^j y^{i-j} \quad (5.6)$$

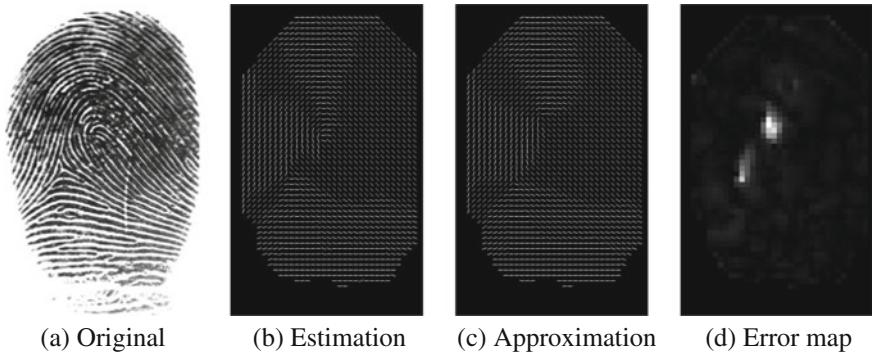
$$g_s^n(x, y) \stackrel{A}{=} \sin 2\theta(x, y) = \sum_{i=0}^n \sum_{j=0}^i b_{i,j} x^j y^{i-j} \quad (5.7)$$

where  $a_{i,j}$  and  $b_{i,j}$  are the polynomial coefficients for  $g_c^n(x, y)$  and  $g_s^n(x, y)$ , respectively [43]. The order of the polynomial model,  $n$ , is selected to be 6. Coefficients  $a_{i,j}$  and  $b_{i,j}$  for the approximated polynomials  $\hat{g}_c(x, y)$  and  $\hat{g}_s(x, y)$ , respectively, can be estimated by the least squares method. The approximated orientation field,  $\hat{\theta}(x, y)$ , is constructed by

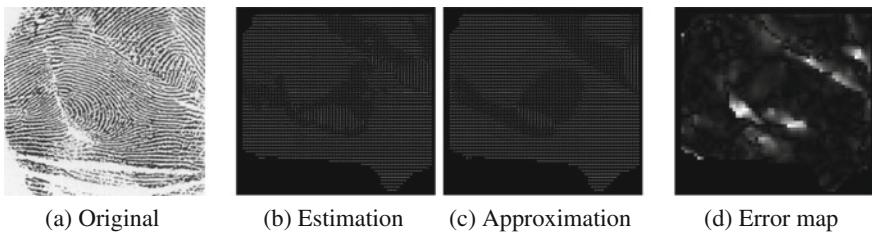
$$\hat{\theta}(x, y) = \frac{1}{2} \tan^{-1} \left( \frac{\hat{g}_s(x, y)}{\hat{g}_c(x, y)} \right) \quad (5.8)$$

### 5.3.1.5 Orientation Error Map

Altered areas in a fingerprint, e.g. around scars and obliterated areas, can result in discontinuous or unnatural changes in the orientation field. The approximated orientation field will not be able to accurately represent these abrupt and irregular changes



**Fig. 5.8** OFA of unaltered fingerprint. The error map has high values around singularities. Source of original fingerprint image: [6]



**Fig. 5.9** OFA of altered fingerprint. The approximation is unable to correctly model abrupt changes around distorted areas. The error map therefore has high values around singularities and in some scarred regions. Source of original fingerprint image: [41]

caused by alterations. An error map,  $\epsilon(x, y)$ , is therefore computed as the absolute difference between  $\theta(x, y)$  and  $\hat{\theta}(x, y)$ .

$$\epsilon(x, y) = \min(|\theta(x, y) - \hat{\theta}(x, y)|, \pi - |\theta(x, y) - \hat{\theta}(x, y)|)/(\pi/2) \quad (5.9)$$

The error map shows how precise the approximation is to the estimation. Abrupt changes and discontinuities in the ridge flow will result in high values in the error map. Unaltered fingerprints of good quality will therefore only have small errors around singular points, whereas altered fingerprints can additionally have errors in scarred or mutilated areas. Figures 5.8 and 5.9 illustrate the resulting orientation fields and error map of an unaltered fingerprint and an altered fingerprint, respectively.

A feature vector can be constructed from the error map in the following manner:

1. Two columns of blocks are removed from each side of the error map which results in an error map of size  $60 \times 60$  blocks.
2. The error map is divided in  $3 \times 3$  cells. Each cell is therefore  $20 \times 20$  blocks.
3. Histograms in 21 bins in the range  $[0,1]$  are computed for each of the nine cells.
4. The nine histograms are concatenated into a 189-dimensional feature vector.

### 5.3.2 Minutiae Distribution Analysis

In the *Minutiae Distribution Analysis* (*MDA*) the minutiae extractor *MINDTCT* in NBIS [39] is used to extract minutiae from a fingerprint. The analysis is based on the observation that the minutiae distribution of altered fingerprints often differs from that of natural fingerprints [43]. The analysis constructs a density map of the minutiae points by using the Parzen window method with uniform kernel function. Let  $\mathbf{S}_m$  be the set of minutiae of the fingerprint, i.e.,

$$\mathbf{S}_m = \{\mathbf{x} \mid \mathbf{x} = (x, y) \text{ is the position of minutia}\}. \quad (5.10)$$

The density map of the minutia is constructed as follows:

1. The initial minutia density map,  $M'_d(\mathbf{x})$ , is obtained by

$$M'_d(\mathbf{x}) = \sum_{\mathbf{x}_0 \in \mathbf{S}_m} K_r(\mathbf{x} - \mathbf{x}_0), \quad (5.11)$$

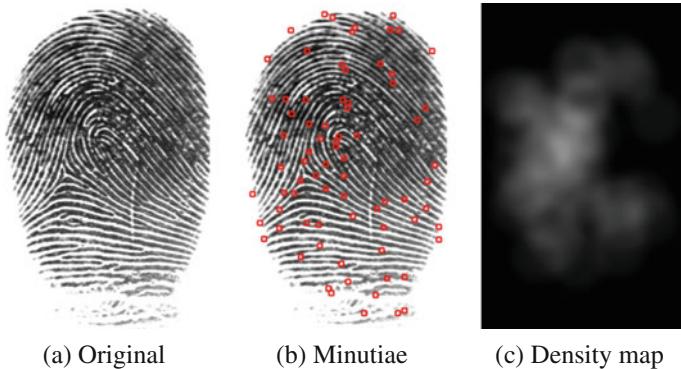
where  $K_r(\mathbf{x} - \mathbf{x}_0)$  is a uniform kernel function centered at  $\mathbf{x}_0$  with radius,  $r$ . Yoon [43] set the radius to 40 pixels. However, the implementation generated for study uses  $r = 30$ , since this gave better results.

2. The initial density map,  $M'_d(x, y)$  is smoothed by a Gaussian filter ( $30 \times 30$  pixels) with a standard deviation of 10 pixels.
3.  $M_d(x, y)$  is transformed to lie in the interval  $[0, 1]$  by

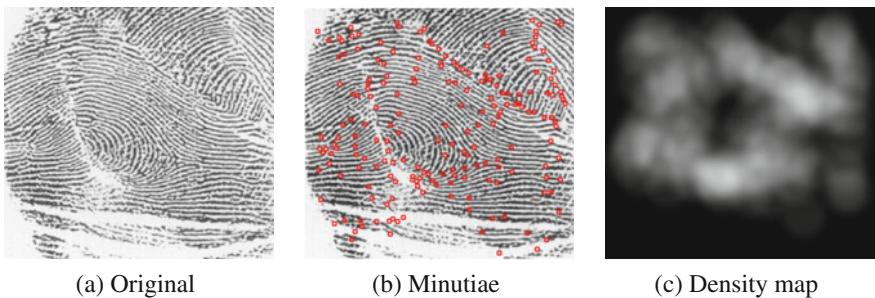
$$M_d(x, y) = \begin{cases} M'_d(x, y)/T, & \text{if } M'_d(x, y) \leq T, \\ 1, & \text{otherwise} \end{cases} \quad (5.12)$$

where  $T$  is a predetermined threshold ( $T$  is set to 6.9 in this specific implementation).

Figures 5.10 and 5.11 show the density maps of an unaltered and altered fingerprint, respectively. Alterations will cause ridge discontinuities which will result in many spurious minutiae. A feature vector is also constructed from the density map in the same fashion as for the *OFA*. 16 columns of pixels are removed from each side of the density map, resulting in an image of size  $480 \times 480$  pixels. The density map is divided into  $3 \times 3$  cells, where each cell is  $160 \times 160$  pixels. Histograms of each cell of the density map are computed in 21 bins in the range  $[0, 1]$ . The nine histograms are concatenated to construct a 189-dimensional feature vector. The feature vector from the *OFA* is concatenated to the feature vector from the *MDA*. This results in a feature vector of 378 dimensions, which is fed into a *SVM* for classification.



**Fig. 5.10** Minutiae density map of an unaltered fingerprint. Source of FP image: [6]



**Fig. 5.11** Minutiae density map of an altered fingerprint. Source of FP image: [41]

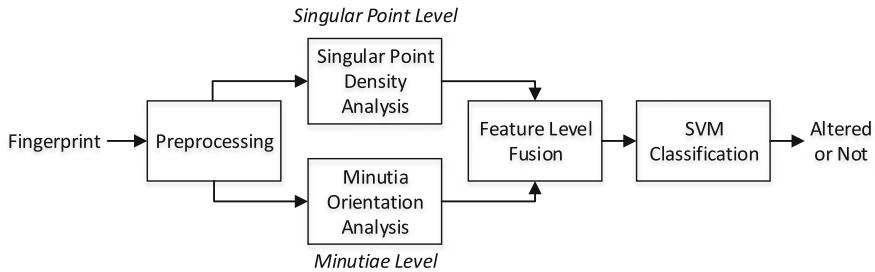
## 5.4 Recent Algorithms for Fingerprint Alteration Detection

Recently a new approach to detect fingerprint alterations has been proposed [12]. A rough overview of the approach is given in Fig. 5.12. The following subsections will give an introduction to the main steps of the method.

### 5.4.1 Preprocessing

The preprocessing pipeline is used to identify foreground and background of the image, to reduce noise and increase the contrast between ridges and valleys, and to transform the image into a common and *invariant* format. The output should thus be a fingerprint image with properties and enhancements specifically designed for the following analysis process. The pipeline consists of the following five steps:

1. Cropping. Locating and isolating the fingerprint image.
2. Segmentation. Separating the foreground from the background.



**Fig. 5.12** Flowchart of the proposed algorithm

3. Rotation. Align fingerprint image along the longitudinal direction.
4. Resize. The size of the image is changed to fit a specified size.
5. Enhancement. Improve the clarity of friction ridges and minimize noise.

#### 5.4.1.1 Cropping

The NIST NBIS package [39] includes a fingerprint segmentation algorithm, *Nfseg*, for cropping a rectangular region of an input fingerprint. As it does not align the fingerprint along the longitudinal direction, further rotation techniques will be used later in the pipeline.

#### 5.4.1.2 Segmentation

An important image preprocessing operation is that of separating the fingerprint image ridge area (i.e., the *ROI*) from the image background. The input fingerprint image,  $I$ , is intensity normalized to have zero mean along with unit standard deviation. This is done by the following pixel-wise function:

$$\forall x \in \{1..R\}, y \in \{1..C\} : I_n(x, y) = \frac{I(x, y) - \text{avg}(I)}{\text{std}(I)} \quad (5.13)$$

where  $\text{avg}(I)$  is the average pixel intensity of the input image and  $\text{std}(I)$  is the standard deviation.  $I(x, y)$  is the pixel intensity at pixel  $(x, y)$  of the input image. From  $I_n$  it is possible to generate a binary image,  $I_{mask}$ , known as the *mask* of the fingerprint where ones belong to the image *ROI* and zeros belong to the background. The normalized image is divided into blocks of size  $8 \times 8$ . If the standard deviation of a block is above a threshold,  $T$ , then the block is regarded as being part of the actual fingerprint, i.e., the foreground. This is a block-wise process; the function for creating the mask is the following:

$$Ib_{mask}(x, y) = \begin{cases} 0 & \text{if } std(Ib(x, y)) \leq T, \\ 1 & \text{otherwise} \end{cases} \quad (5.14)$$

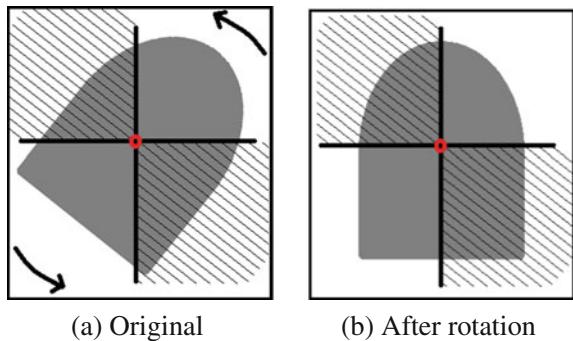
where  $Ib$  contains the blocks of size  $8 \times 8$  pixels. The threshold,  $T$ , is set to 0.1. Morphological operations are run on the block-wise mask image,  $Ib_{mask}$ , for filling holes and removing isolated blocks yielding  $Ib'_{mask}$ . Two different structures are used for the morphological operations: a square rotated  $45^\circ$  with a diagonal length of five blocks and a square with a length of three blocks. The series of open operations are conducted by alternating between the two shapes, starting with the rotated square. The close operations are executed in the same fashion, however starting with the *unrotated square*. The shapes have been set empirically. The final pixel-wise fingerprint mask,  $I_{mask}$ , is the up-scaled version of the final block-wise mask,  $Ib'_{mask}$ . This segmentation is specifically for our analysis. The strength of the approach is that it removes unclear borders which might be identified as altered while keeping low quality areas that reside within the *ROI*. Likewise, this segmentation is not ideal for fingerprint recognition purposes in general since too many foreground blocks would be removed possibly erasing important minutiae. Low quality areas within the fingerprint can also add unwanted minutia or features that can compromise the comparison algorithm whereas they will play an important part in determining if the image has been altered or not.

#### 5.4.1.3 Rotation

Landmarks of the fingerprint image are normally used as a reference point for rotation purposes. Different approaches have been proposed for rotating a fingerprint image, such as computing the image orientation using *singular point* as reference points [28] or using the fingerprint Center Point Location (*CPL*) as a reference point [31]. Using minutia or singular points as reference points requires significant analysis of the fingerprint image and can be quite complex. The approach taken in this work is based on [31] since it is a simple and efficient approach which does not require complex computations. The approach is built on the assumption that most fingerprints have an ellipsoidal shape. The rotation method therefore uses the *mask* of the fingerprint which was constructed in the previous step. The image is shifted so that the centroid of the *ROI* is at point  $(0, 0)$ .

Consider the image being placed in a Cartesian coordinate system, see Fig. 5.13. Two areas,  $a_1$  and  $a_2$  are defined where  $a_1$  is composed of quadrant 2 (upper left) and 4 (bottom right) while  $a_2$  is composed of the remaining diagonal quadrants. The idea is to rotate the foreground image around the center so that the amount of foreground pixels is balanced between  $a_1$  and  $a_2$ . This is an iterative process.

**Fig. 5.13** The fingerprint mask is placed in a coordinate system and rotated. The sums of the fingerprint foreground of the quadrants with the patterned background are compared to the sums of the quadrants with a clear background



#### 5.4.1.4 Resizing

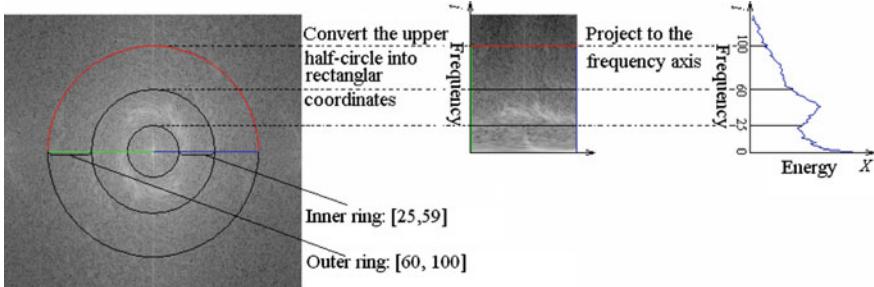
Before resizing, the image must be cropped again since rotation might have added additional background to the sides of the image. There are many ways that the image can be re-cropped, e.g. by using the *Nfseg* algorithm or removing columns and rows that contain only zeros. The image is resized to the closest fit for  $512 \times 480$  pixels and thereafter segmented afresh.

#### 5.4.1.5 Enhancement

The final step of the preprocessing pipeline is enhancement of the fingerprint image. The goal of fingerprint enhancement techniques are traditionally to improve the clarity of friction ridges and remove unwanted noise in order to assist the following analyses or feature extraction algorithms. Numerous fingerprint enhancement techniques have been proposed. Two processes were conducted on the image to slightly enhance it: histogram equalization in the spatial domain and a simple enhancement in the frequency domain.

Histogram equalization is a common method for enhancing the contrast of a image. The method defines a mapping of grey levels  $p$  into grey levels  $q$  which attempts to uniformly distribute the grey levels  $q$  [20]. A cumulative histogram of the enhanced image would show a relatively linear curve and the ideal mean would be right in the center of the density value. Histogram equalization is described in Eq. (5.15) where  $k$  is the greyscale level of the original image,  $n_j$  is the number of times pixel value  $j$  appears in the image,  $n$  is the total number of pixels and  $L$  is the number of grey levels (for example 256). The contrast of grey levels are stretched near the histogram maxima using histogram equalization. This improves the detectability of many image features [26].

$$\forall i \in \{1..R\}, j \in \{1..C\} : G(i,j) = H(I(i,j)) = H(k) = \sum_{j=0}^k \frac{n_j}{n}(L-1) \quad (5.15)$$

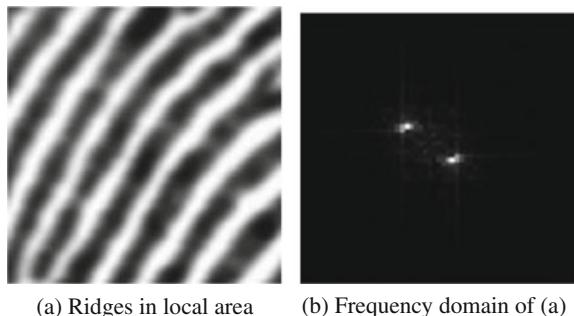


**Fig. 5.14** Fingerprint images produce a ring pattern in the frequency domain. Note the spectral energy in the *inner ring*. *Source* [23]

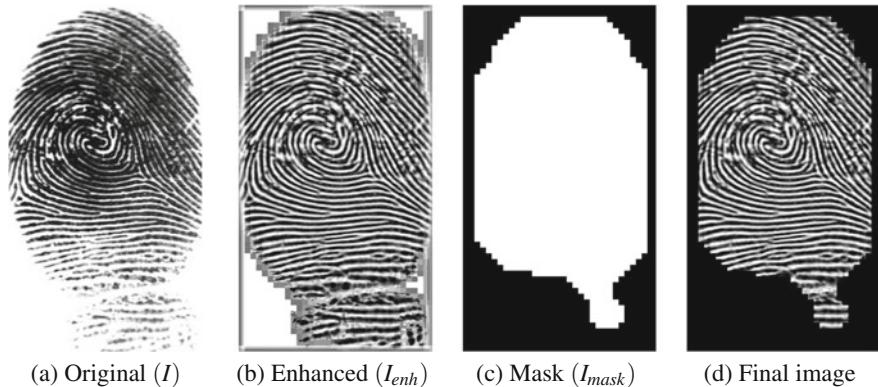
Fingerprints can also be enhanced in the frequency domain. Ridges can be locally approximated by single sine waves. The orientation of the ridges gives frequencies in all directions. A good quality fingerprint image will therefore produce a clear ring pattern around the center in the frequency domain (see Fig. 5.14). The clearness of the ring is based on the consistency of the ridge characteristics.

The radial power spectrum is an established method to determine a quality score of the fingerprint [33]. The quality score is based on the peak of maximum energy in the *inner ring*. This can also be seen in Fig. 5.14. One method of improving the clarity of the ridge patterns could be to apply a filter which enhances the frequencies in the inner ring while minimizing or even eliminating frequencies outside of the inner ring. Altered fingerprints will have unnatural ridge properties on a global scale and—depending on the type of alteration—be of a less quality. It is desired to also enhance discrepancies since the impending analyses will be on these inconsistencies. Watson et al. [40] presented a simple enhancement method based on enhancing dominant frequencies in the frequency domain. This method concentrates on local properties and is performed using overlaying blocks. This approach will be used to enhance fingerprint images in this study (Fig. 5.15).

**Fig. 5.15** Orientation and frequency of ridges in local area. *Source* [7]



(a) Ridges in local area      (b) Frequency domain of (a)



**Fig. 5.16** Enhancement of an unaltered fingerprint. *Source a* [6]

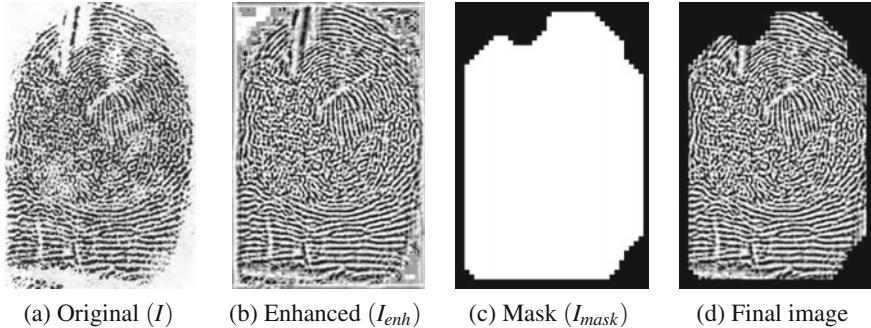
This method is quite efficient in enhancing the image for the specific purpose, as it fills up small holes in ridges and also otherwise enhances the appearance of the friction ridges.

The fingerprint image is divided into blocks of size  $8 \times 8$  pixel. An additional overlapping border of  $8 \times 8$  pixels is added around each block such that the actual size of each block is  $24 \times 24$  pixels. The *FFT* of each block is multiplied by an exponentiation of its magnitude; the exponent factor  $k$  is set to 0.45. The enhanced block  $B'(x, y)$  based on the original block  $B(x, y)$  is done accordingly:

$$B'(x, y) = \text{FFT}^{-1}(\text{FFT}(B(x, y)) \cdot |\text{FFT}(B(x, y))|^k) \quad (5.16)$$

Notice that in Eq. 5.16 blocks  $B(x, y)$  and  $B'(x, y)$  are  $24 \times 24$  pixels, i.e. they are extended by the borders. The enhanced image,  $I_{enh}$ , is combined by the center  $8 \times 8$  pixels in each block of  $B'$ . Examples of the resulting enhancements on altered and unaltered fingerprint images can be seen in Figs. 5.16 and 5.17, respectively. The reason for using overlapped blocks is to minimize the border effect of the block-wise *FFT*. The idea of the method is that dominant frequencies of each block correspond to the ridges, amplifying these dominant frequencies increases the ratio of ridge information to nonridge noise [40].

It is important that the choice of enhancement algorithm complements the alteration detection algorithms. As opposed to traditional minutia extraction methods used for fingerprint recognition purposes the selected algorithm should retain some local inconsistencies.



**Fig. 5.17** Enhancement of an altered fingerprint. *Source a* [38]

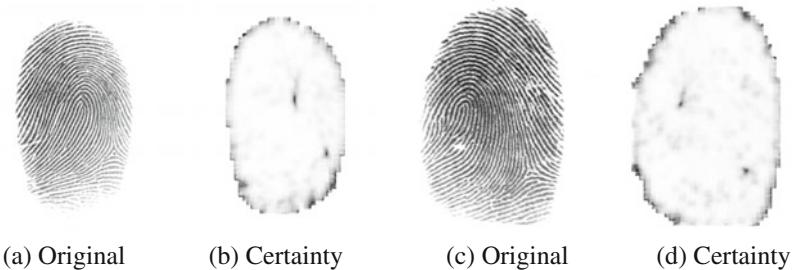
### 5.4.2 Singular Point Density Analysis

The proposed method, Singular Point Density Analysis (SPDA) is based on local analysis of the orientation field using the Poincaré index to detect noisy friction ridge areas. An altered fingerprint will have a higher density of such areas than an unaltered. Singular points namely *core* and *delta*, act as control points around which the ridge lines are wrapped [30]. The suggested approach is partly inspired from research done by Petrovici [35] and calculates amplitudes of singular points based on the reliability of their orientation. Before describing the actual method, a brief analysis of the characteristics of the orientation field will be conducted on altered and unaltered images.

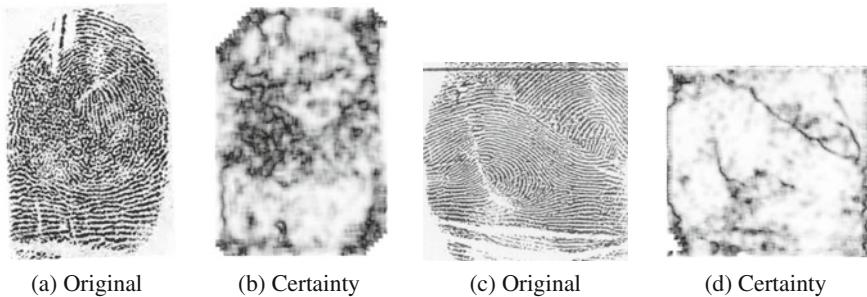
#### 5.4.2.1 Orientation Certainty Level

The Orientation Certainty Level (*OCL*) is a common analysis used as a quality measure to determine the certainty of local orientations. It measures the energy concentration along the direction of ridges [42]. In Sect. 5.3.1.2 the two intensity gradients,  $\delta_x$  and  $\delta_y$ , were found for each pixel using a Sobel operator. The orientation certainty can be calculated as the ratio of the two eigenvalues of the covariance matrix of the gradient vector. The covariance matrix  $C$  of the gradient vector for an  $N$  points image block is given by:

$$C = \frac{1}{N} \sum_{i=1}^N \left\{ \begin{bmatrix} dx_i \\ dy_i \end{bmatrix} \begin{bmatrix} dx_i & dy_i \end{bmatrix} \right\} = \begin{bmatrix} \frac{1}{N} \sum_{i=1}^N dx_i^2 & \frac{1}{N} \sum_{i=1}^N dx_i dy_i \\ \frac{1}{N} \sum_{i=1}^N dy_i dx_i & \frac{1}{N} \sum_{i=1}^N dy_i^2 \end{bmatrix} = \begin{vmatrix} a & c \\ c & b \end{vmatrix} \quad (5.17)$$



**Fig. 5.18** Two unaltered images and their corresponding orientation certainty. Since the images are of relatively good quality, low certainty is in areas of high curvature. *Source a, c [6]*



**Fig. 5.19** Two altered images and their corresponding orientation certainty. Areas with low friction ridge quality caused by scars and obliteration cause low orientation certainty. *Source a [38], c [41]*

where  $dx = \delta_x$  and  $dy = \delta_y$  are the gradient densities at each pixel. The eigenvalues,  $\lambda_{max}$  and  $\lambda_{min}$ , can be calculated as [27] (Figs. 5.18 and 5.19):

$$\lambda_{max} = \frac{(a + b) + \sqrt{(a - b)^2 + 4c^2}}{2} \quad (5.18)$$

$$\lambda_{min} = \frac{(a + b) - \sqrt{(a - b)^2 + 4c^2}}{2} \quad (5.19)$$

The *OCL* is defined in the range  $[0, 1]$  where a high value means good quality. The formula for calculating the *OCL* is therefore:

$$Ocl = 1 - \frac{\lambda_{min}}{\lambda_{max}} \quad (5.20)$$

### 5.4.2.2 Orientation Entropy

Considering the pixel-wise orientation,  $\theta(x, y)$ , as a random variable and using a sample rate,  $n$ , determining the amount of possible orientation values, a quantification scale  $\frac{\pi}{n}$  is used to discretise the orientation field. The discrete orientation field,  $\theta_n$ , with sample rate  $n$  can be computed as:

$$\forall x \in \{1..R\}, y \in \{1..C\} : \theta_n(x, y) = \frac{\pi}{n} \left( \frac{\theta(x, y)}{\frac{\pi}{n}} \mod n \right) \quad (5.21)$$

where  $R$  and  $C$  define the row count and column count of the orientation image, respectively. The entropy for each orientation is calculated according to the amount of different possibilities in the surrounding area. Blocks of size  $5 \times 5$  are used in the current implementation. The discrete entropy image,  $E_n$ , can be constructed as:

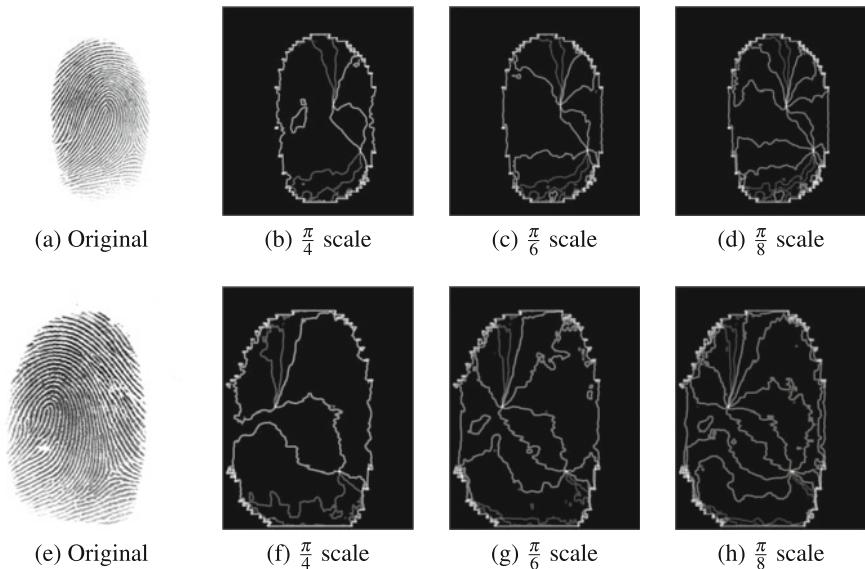
$$\forall x \in \{3..R-2\}, y \in \{3..C-2\} : E_n(x, y) = - \sum_{i=1}^n p_{x,y}(\frac{\pi}{i}) \log_{10} p_{x,y}(\frac{\pi}{i}) \quad (5.22)$$

where  $p_{x,y}(\frac{\pi}{i})$  is the probability of the discrete orientation  $\frac{\pi}{i}$  in the block belonging to  $(x, y)$ . Since blocks of size  $5 \times 5$  are used, pixels that don't have a complete border of  $2 \times 2$  surrounding pixels are not considered. Figures 5.20 and 5.21 show multiple scaled orientation entropy images of unaltered and altered fingerprint images, respectively.

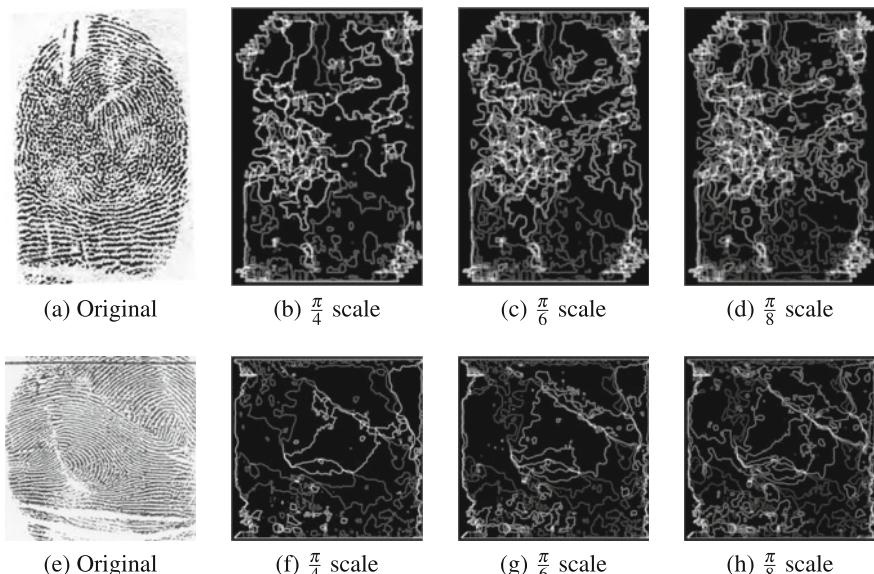
The characteristics of the entropy images are very distinct. The center points in the orientation entropy images which have a constant high entropy value are points with uncertain orientations. The difference between the entropy values in the illustrated unaltered and altered images can be summarized as: Unaltered fingerprint images have a constant high entropy value in the center of singular regions. On the contrary altered fingerprint images have many spurious positions with a constant high entropy value. Scarred regions and mutilated areas add uncertainties in the pixel-wise orientations which lead to the high entropy values. From Fig. 5.20 it becomes clear that it is possible to use the entropy to locate singular points. It can be concluded that altered images have a number of positions with a high orientation uncertainty. Based on both the uncertainty and entropy analysis it seems that altered images have a higher amount of positions in a fingerprint with characteristics similar to a singular point.

### 5.4.2.3 Poincaré Index

Methods for detecting singular points are commonly based on the Poincaré index which in this context analyses the change of direction around a given point as proposed by Kawagoe and Tojo [25]. The Poincaré index has shown high accuracy, but low robustness [36]. The method fails to correctly localize reference points in poor



**Fig. 5.20** Multiple scaled orientation entropy images of unaltered fingerprints. Images have sample rate  $n = 4, 6$ , and  $8$ . Constant high entropy values are in singular regions where they act as center points. The entropy varies in other positions of the fingerprint according to the scale. *Source a, e [6]*



**Fig. 5.21** Multiple scaled orientation entropy images of altered fingerprints. Images have sample rate 4, 6 and 8. Obliterated areas and scars generate high entropy readings. Even with multiple scaled entropy images, there are many central points with a constant high entropy value. *Source a [38], e [41]*

quality fingerprints with cracks and scars, dry skin or poor ridge and valley contrast [21]. Many false core and delta points can be produced when the orientation field is noisy, e.g. in low quality fingerprint areas. The proposed method takes advantage of this limitation. For correctly detecting singular points different heuristics are used to filter out false locations, e.g. iterative smoothing of the orientation field [5], using a modified gradient-based Poincaré method [3] and in combination with additional filters [4, 45]. No filtering process will be used in the current analysis which will therefore detect singular point candidates. In the current context, the Poincaré index for a given point  $P(x, y)$  can be defined as the cumulative change in the orientation field traveling clockwise around the neighboring points. The possible values of  $P(x, y)$  are 0,  $\pm\pi$  and  $\pm 2\pi$ :

$$P(x, y) = \begin{cases} 2\pi, & \text{if } (x, y) \text{ belongs to a whorl} \\ \pi, & \text{if } (x, y) \text{ belongs to a loop} \\ -\pi, & \text{if } (x, y) \text{ belongs to a delta} \\ -2\pi, & \text{if } (x, y) \text{ is in the center of a rare } \textit{diamond} \text{ shape} \\ 0, & \text{otherwise} \end{cases} \quad (5.23)$$

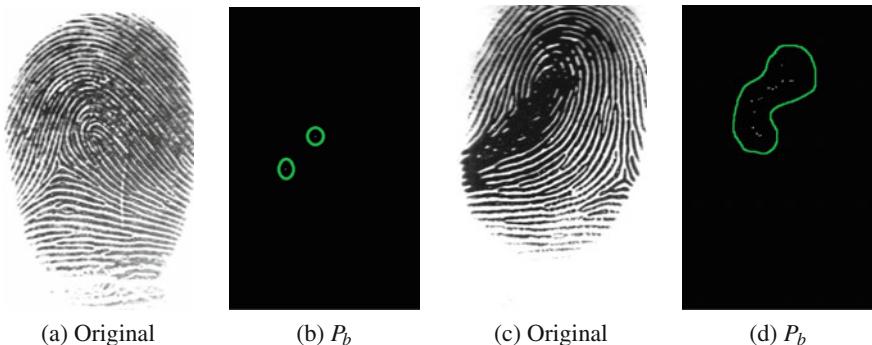
After calculating the Poincaré index for all pixels in the fingerprint image, the Poincaré matrix,  $P_b(x, y)$ , is binarised accordingly:

$$P_b(x, y) = \begin{cases} 1, & \text{if } P(x, y) \in \{\pm\pi, \pm 2\pi\} \\ 0, & \text{otherwise} \end{cases} \quad (5.24)$$

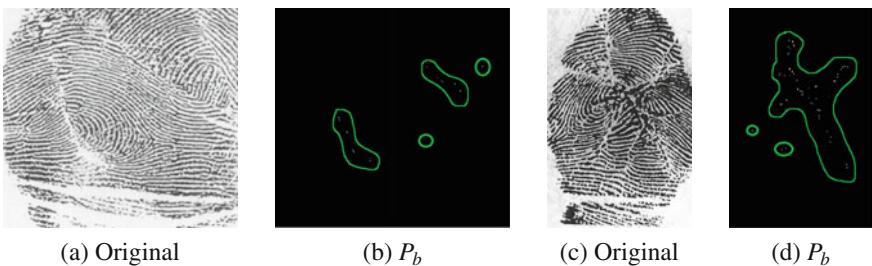
The resulting binarised Poincaré image will generally have a small cluster of one to four adjacent pixels with high values in the singular regions. However, noisy or low quality areas in unaltered images may lead to the detection of additional false singular points. The unnatural flow of friction ridges in altered fingerprints will generally result in a higher detection of false singular points. Figure 5.22 shows the resulting binarised Poincaré image,  $P_b$ , of two unaltered fingerprints. Figure 5.22d illustrates the problem with poor quality fingerprint images. Figure 5.23 shows  $P_b$  of altered images. An observation can be made on the spurious false singularities found in both altered and unaltered fingerprints that can further strengthen the analysis of singularities and help detect if a fingerprint has been altered or not. In the following section Gabor filters will be used to determine quality measures of the friction ridges with the intention to highlight *singular points* in good quality areas.

#### 5.4.2.4 Gabor Filters

The goal is to determine the clarity of friction ridges in a fingerprint so that found singularities can be classified based on the quality score of their position. Gabor



**Fig. 5.22** Singular point detection of unaltered fingerprints. **a** Fingerprint image [6], **b** its corresponding binarised Poincaré image with highlighted readings, **c** Poor quality fingerprint image [6], **d** its corresponding binarised Poincaré image. Low quality images will result in spurious false singularities



**Fig. 5.23** Singular point detection of altered fingerprints. **a** Distorted fingerprint [41], **b** its corresponding binarised Poincaré image with highlighted readings, **c** Heavily damaged fingertip [38], **d** its corresponding binarised Poincaré image

filters will be used to measure the quality of friction ridges. Gabor filters are band-pass filters that have both frequency and orientation properties; they can therefore be constructed to present friction ridge frequency and orientation.

A global quality measure based on Gabor filter responses was proposed by Olsen et al. [34]. This study uses an adaptation of this approach applying different filter bank sizes. The first step is to convolve the fingerprint image,  $I$ , with a two-dimensional Gaussian with  $\sigma = 1$ . The convolution is subtracted from the original image to give  $\bar{I}$  which is a high-pass filtered image. The quality of the ridge-valley structure of the friction ridges is found by convolving the fingerprint image,  $\bar{I}$ , with two-dimensional Gabor filters in  $n$  orientations ( $n = 8$ ). The orientations  $\theta$  are computed:

$$\theta = \pi \frac{k - 1}{n}, \quad k = 1, \dots, n \quad (5.25)$$

The general form of a complex two-dimensional Gabor filter in the spatial domain can be defined as:

$$h(x, y, \theta_k, f, \sigma_x, \sigma_y) = \exp\left(-\frac{1}{2}\left(\frac{x_{\theta_k}^2}{\sigma_x^2} + \frac{y_{\theta_k}^2}{\sigma_y^2}\right)\right) \exp(j2\pi f x_{\theta_k}), \quad k = 1, \dots, n \quad (5.26)$$

where

$n$  is the number of orientations used in the filter bank

$$x_{\theta_k} = x \sin \theta_k + y \cos \theta_k,$$

$$y_{\theta_k} = x \cos \theta_k - y \sin \theta_k,$$

$f$  is the frequency of the sinusoidal plane wave along the orientation  $\theta_k$ ,

$\sigma_x$  and  $\sigma_y$  are the parameters of the Gaussian window.

The filter frequency,  $f$ , is set to 0.1 as the average inter-ridge distance (corresponding to the wavelength) is approximately 10 pixels in a 500 dpi fingerprint image [21]. The standard deviation of the Gaussian envelope along the x and y axes are set as  $\sigma_x = \sigma_y = 6.0$ . The magnitude of the responses is convolved with a two-dimensional Gaussian with  $\sigma = 4$  to smoothen the magnitudes. Areas containing clear ridge-valley structures will have a strong response from one or more of the filter orientations, see Fig. 5.24 while low quality areas will have low responses. A pixel-wise standard deviation of the bank of Gabor responses is calculated,  $G_{std}$ .

The resulting Gabor quality matrix,  $G_{std}$ , is transformed to the interval  $[0, 1]$  by:

$$Q_G(x, y) = \begin{cases} G_{std}(x, y)/T, & \text{if } G_{std}(x, y) \leq T \\ 1, & \text{otherwise} \end{cases} \quad (5.27)$$

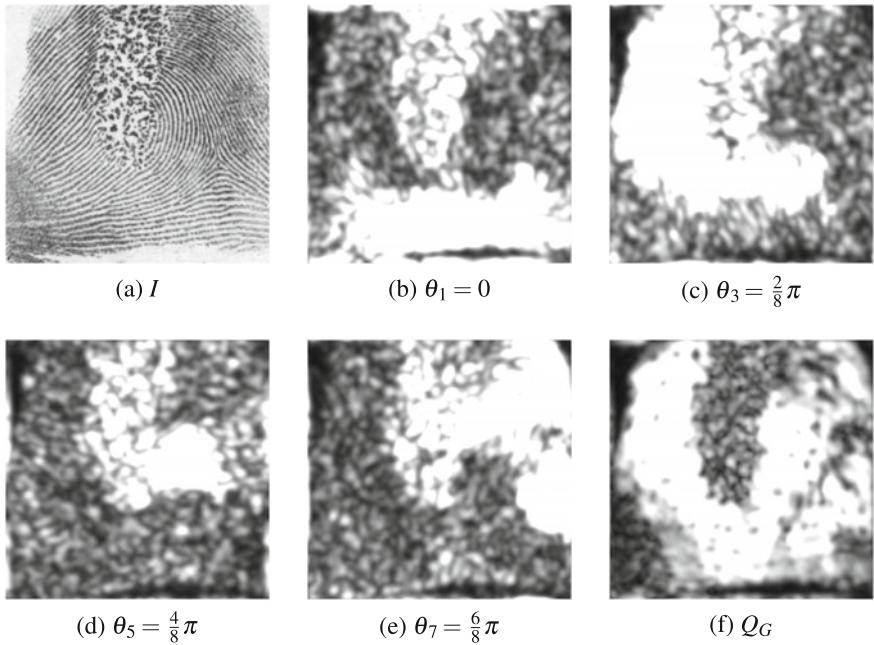
where  $T$  is a given threshold ( $T = 0.01$ ). The threshold is determined empirically.

#### 5.4.2.5 Density Map

The quality metrics matrix,  $Q_G$ , are factors that determine the quality of friction ridges in each pixel. The quality score is defined in the interval  $[0, 1]$  where 1 is the best quality. An initial density map,  $P_d$ , is created from combining the Poincaré matrix,  $P_b(x, y)$ , and the normalized Gabor quality matrix,  $Q_G$ , by multiplication:

$$P_d(x, y) = P_b(x, y) \cdot Q_G(x, y) \quad (5.28)$$

$P_d$  is a density map where singularities positioned within clear areas receive a higher value than singularities in poor quality areas. Similar to the previous minutia distribution analysis, the final density map,  $SP_d$ , is constructed by using the Parzen window method with uniform kernel function and smoothed using a low-pass

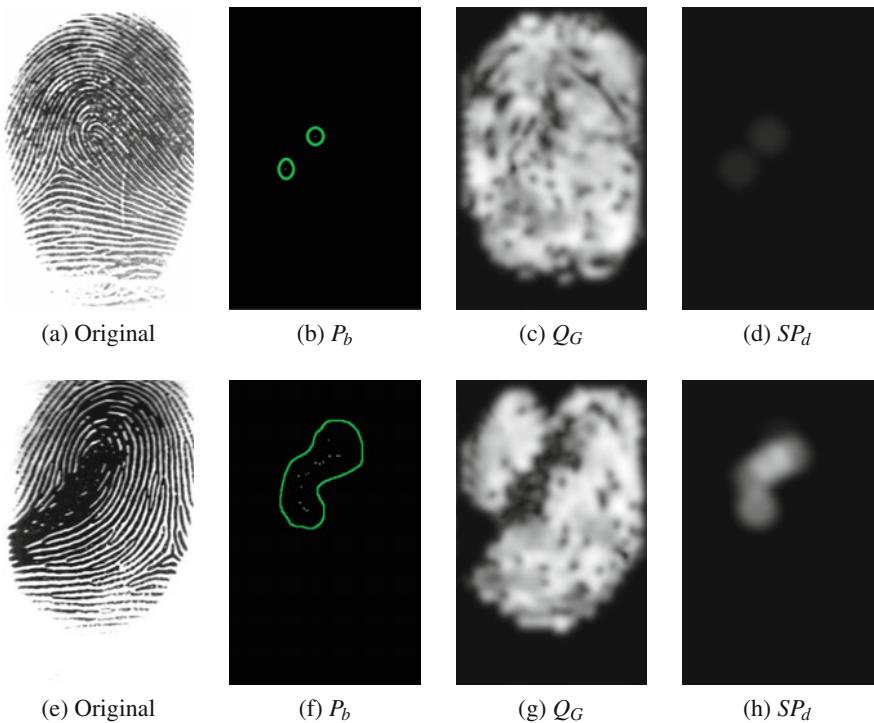


**Fig. 5.24** Gabor filter responses of an obliterated fingerprint. **a** Input image [44], **b–e** absolute values of Gabor filter responses in different orientations after Gaussian smoothening and **f** standard deviation of the Gabor responses with different orientations

Gaussian filter. Figures 5.25 and 5.26 show examples of the final density maps of unaltered and altered fingerprints, respectively. Notice that the altered fingerprints have a higher density of singularities than the unaltered in the final density maps.

### 5.4.3 Minutia Orientation Analysis

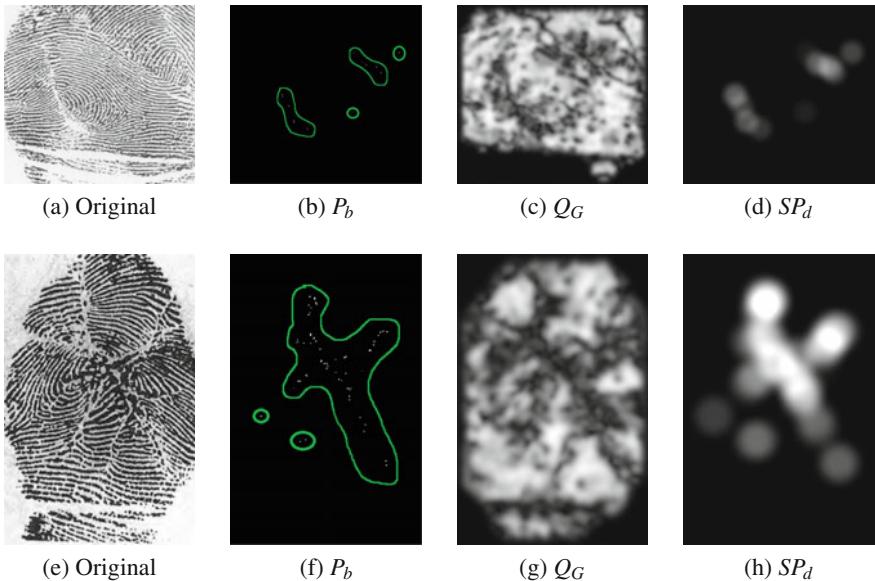
The Minutia Orientation Analysis (*MOA*) takes advantage of the excessive number of minutia that appears due to alteration. Instead of relying on the actual distribution of minutia as the *MDA* does, properties of minutia in local areas will be analyzed. Altered fingerprints have an unnatural orientation flow. Likewise, the *MOA* shows that minutia distribution of altered fingerprints often differs from that of natural fingerprints. Scars will produce broken ridges which will contribute to some of the additional minutiae that is extracted from altered fingerprints. Figure 5.27 shows an example of a z-cut alteration where a relatively large portion of minutiae is concentrated around the scarred areas. The idea is to register a metric for each minutia point, which describes the change in orientation in its surroundings. There are multiple options that can be considered. Two simple approaches based on analyzing orientations in a certain radius,  $r$ , from the minutia could be:



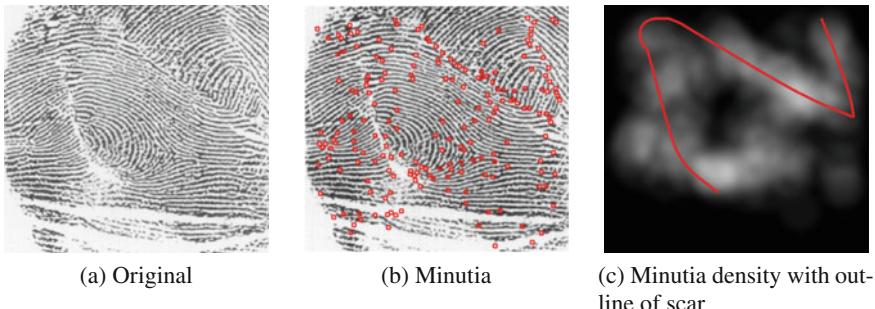
**Fig. 5.25** Singular point density analysis of unaltered fingerprints. **a** Input image, **b** its corresponding binarised Poincaré image with highlighted readings, **c** quality measurement of the friction ridges, **d** the final singular point density map. **e–g** show the same as **a–d** but with a lower quality image. Source **a, e** [6]

1. Using a rotating line. A circle of radius  $r$  is considered around each minutia point. Using the computed orientation field, the maximum difference of orientation between a series of orientation pairs on opposite sides of the circle around the minutia point can be found based on Fig. 5.28. Here, steps of  $\frac{\pi}{4}$  are used to find a set of 8 points around the minutia point. The minutia is assigned the maximum difference in orientation of the points on opposite sides of the minutia.
2. Based on minutia direction. The minutia extractor assigns a direction to every minutiae, normally in the range  $[0, 2\pi]$ , pointing towards the current friction ridge that it belongs to. Instead of analyzing multiple points around a circumference with radius  $r$  around the minutia, only the point in the given direction of the minutia and on the opposite side are measured like earlier. The minutia is assigned the orientation difference of these two points. See Fig. 5.29.

The problem with both of the above mentioned approaches is that minutiae close to high curvature regions will be assigned high values. The second approach is an improvement since it takes into account the actual direction of the minutia. A sim-

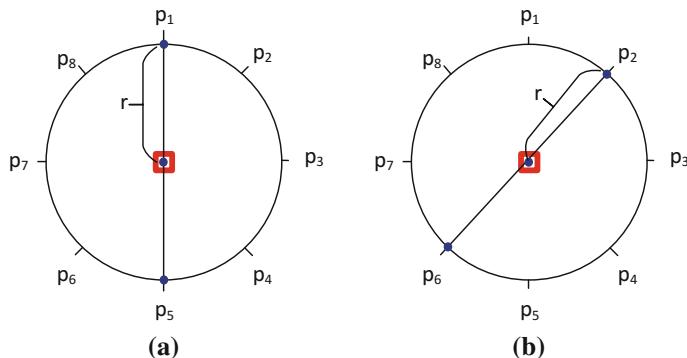


**Fig. 5.26** Singular point density analysis of altered fingerprints. **a** Input image, **b** its corresponding binarised Poincaré image with highlighted readings, **c** quality measurement of the friction ridges, **d** the final singular point density map. **e–g** show the same as **a–d** with a heavily damaged fingertip. Source **a** [41], **e** [38]

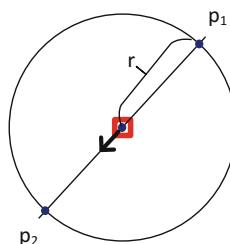


**Fig. 5.27** Scars cause higher concentration of minutiae. Source **a** [41]

ple method is proposed comparing neighboring minutiae that are closer than a given radius. To compare neighboring minutiae, it is desired that minutiae are found on both sides of scars in order to analyze changes. The problem comparing minutiae, is that minutiae extractors generally do a lot of analyses of candidate minutiae in order to remove false minutiae. Even though the density of minutiae is higher along scars, the minutiae extractor does a good job in limiting minutiae with conflicting orientations. While in principle every minutiae extractor could be used, we have employed the NIST MINDTCT.



**Fig. 5.28** Differences in orientations can be measured in a certain distance,  $r$ . A *line* is rotated clockwise in steps of  $\frac{\pi}{4}$ . **a** Orientation difference between points  $p_1$  and  $p_5$  are found, **b** the *line* is rotated and the difference between points  $p_2$  and  $p_6$  are found. The minutia is assigned the maximum orientation difference of the four measurements



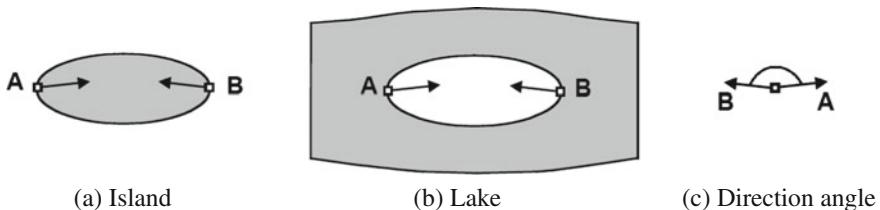
**Fig. 5.29** Differences in orientations can be measured based on points in a certain distance,  $r$ , from the minutia. Points are found by the actual orientation of the minutia. The minutia is assigned the orientation difference of points  $p_1$  and  $p_2$ .

### 5.4.3.1 Minutia Analyses

The NIST minutia feature extractor prepares output files in multiple formats and with additional information intended for miscellaneous fingerprint comparison algorithms.

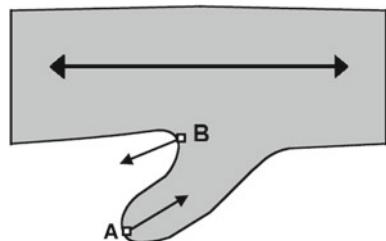
Based on the above dissection of the minutia extractor small modifications are made to the source in order to provide additional minutiae in strategic places. It is very important that the extractor still filters out minutiae; leaving out the whole filtering process would result in an excessive amount of minutiae. This would lead to large areas around singularities being identified as having unnatural changes in the flow of friction ridges based on the high curvatures. Modifications are the following:

**Islands and Lakes:** A pair of candidate minutiae will be detected at each end of islands and lakes. If two minutiae are found to be closer than a given threshold (16 pixels) with a minimum direction angle  $\geq 123.75^\circ$ , then the pair of minutiae is removed from the candidate list. An illustration of this is shown in Fig. 5.30.

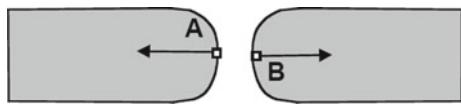


**Fig. 5.30** The minimum direction angle between a pair of minutiae points is used to determine if candidate minutiae are removed or not

**Fig. 5.31** Minutiae belonging to hooks have directions that differ from the friction ridge orientation



**Fig. 5.32** Discontinuities in ridges or valleys has a pair of minutiae points in opposite directions

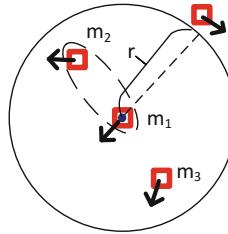


The modified version increases the threshold of the direction angle to  $157.5^\circ$ . This will give a larger amount of minutiae belonging to small islands and lakes.

**Hooks:** Scarred regions will produce hooks that protrude off the side of ridges. It would therefore seem obvious to keep candidate minutia belonging to hooks in the modified minutia extractor. The problem with hooks is that they have directions which do not correspond to the actual ridge flow in the corresponding area, see Fig. 5.31. Leaving minutiae belonging to hooks will increase the inconsistency of minutiae directions in unaltered regions of the fingerprint also. This is therefore unmodified in the current solution.

**Discontinuities:** Minutiae belonging to discontinuities (see Fig. 5.32) in ridges or valleys are removed in the standard feature extractor. Criteria are based on a combination of minutia distance, the difference of the directional angles between them and the direction of a line joining the two minutiae points.

Friction ridges on opposing side of scars can be considered as discontinuities. The modified feature extractor leaves out this check and does therefore not remove minutiae on the basis of discontinuities. Discontinuities in unaltered regions will not have a great impact on the analysis of minutiae orientations since they generally are consistent with the local friction ridge orientation. Alterations with a sufficient amount of friction ridges will reside in a region with an orientation—it will be unreliable—and will generally belong to the *ROI* of the fingerprint. Minutiae in these areas will typically have a lower quality score. Checks dealing with unreliable/invalid orienta-



**Fig. 5.33** Each minutia point finds the minutia with the largest orientation difference from its own within a given radius,  $r$ . Minutia point  $m_1$  is being processed. Only minutia within the given radius are considered. Minutia point  $m_2$  differs most from  $m_1$ ; the minutia which is being processed is assigned the orientation difference of these two points

tions are therefore unmodified in the feature extractor. The checks that test if ridges are to wide or too narrow do not greatly affect the minutiae points around scars and obliterated regions and will therefore be unmodified in the feature extractor. Only a couple of small modifications are done to the feature extractor. To summarize the above, only the process that removes islands and lakes has been modified while the process that checks for discontinuities has been removed.

#### 5.4.3.2 MOA Algorithm

The *MOA* is a relatively simple algorithm which analyses local orientation differences of minutiae. Each minutia point compares its orientation with every surrounding minutiae within a given radius, see Fig. 5.33. The largest orientation difference is registered and saved in a *orientation difference map*. A density map is then constructed from the orientation difference map.

Descriptions on how the orientation difference and orientation density maps are created is described below.

#### 5.4.4 Orientation Difference Map

A function is introduced  $\phi(\mathbf{u}, \mathbf{v}, r)$  which gives 1 if the distance between points  $\mathbf{u}$  and  $\mathbf{v}$  is less than a given radius (distance),  $r$ , and 0 otherwise:

$$\phi(\mathbf{u}, \mathbf{v}, r) = \begin{cases} 1 & \frac{|\mathbf{u}-\mathbf{v}|}{r} < 1, \\ 0 & \text{otherwise} \end{cases} \quad (5.29)$$

Minutia directions are in the range  $[0, 2\pi]$ . The directions are transformed to lie in the interval  $[0, \pi]$ , using modulo  $\pi$ , such as they now are orientations. The difference,  $d(\theta_i, \theta_j)$ , between two minutia orientations,  $\theta_i$  and  $\theta_j$ , is given as:

$$d(\theta_i, \theta_j) = \min(|\theta_i - \theta_j|, \pi - |\theta_i - \theta_j|) \quad (5.30)$$

Let  $S_m$  be the set of minutiae of the fingerprint containing the minutia's position and orientation, i.e.,

$$S_m = \{(\mathbf{x}, \theta) \mid \mathbf{x} = (x, y) \text{ is the position and } \theta \text{ is the orientation of minutia}\}. \quad (5.31)$$

A set,  $S_{diff}$ , containing the position of each minutia together with the largest orientation difference is given as:

$$S_{diff} = \{(\mathbf{x}, \theta_{max}) \mid (\mathbf{x}, \theta) \in S_m \wedge \theta_{max} = \text{Max}(\mathbf{x}, \theta, r, S_m)\}, \quad (5.32)$$

where  $r = 30$  and  $\text{Max}(\mathbf{x}, \theta, r, S_m)$  returns the largest orientation difference,  $a$ , between  $\theta$  and any minutia within the radius  $r$  by Eq. (5.33).

$$\forall b \in \{\phi(\mathbf{x}, \mathbf{x}_0, r) \cdot d(\theta, \theta_0) \mid (\mathbf{x}_0, \theta_0) \in S_m\} : (b \leq a) \quad (5.33)$$

The initial minutia orientation difference map,  $M'_{diff}(x, y)$ , is initiated by zeroes in the size of the fingerprint image. The minutia difference values,  $S_{diff}$ , are plotted into  $M'_{diff}$  such that the location of each minutia is assigned the corresponding minutia orientation difference.

The values of  $M'_{diff}(x, y)$  are transformed to lie in the interval  $[0, 1]$  by

$$M_{diff}(x, y) = \begin{cases} M'_{diff}(x, y)/T, & \text{if } M'_{diff}(x, y) \leq T, \\ 1, & \text{otherwise} \end{cases} \quad (5.34)$$

where  $T$  is a predetermined threshold ( $T$  is set to  $\pi/4$ ).

$M_{diff}$  is basically an image where each pixel  $M_{diff}(x, y)$  contains a normalized orientation difference of the minutia centered at  $(x, y)$ . If no minutia is in  $(x, y)$  then the value is 0.

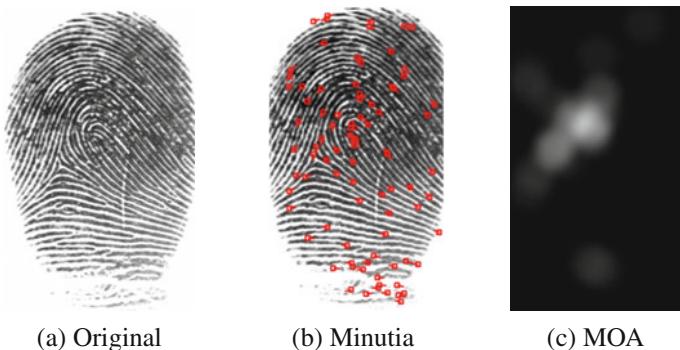
### 5.4.5 Orientation Density Map

The final orientation density map is computed as for the other density maps:

1. An initial density map,  $M'_{dens}$ , is constructed by

$$M'_{dens}(\mathbf{x}) = \sum_{\mathbf{x}_0 \in M_{diff}} K_r(\mathbf{x} - \mathbf{x}_0) \quad (5.35)$$

where  $K_r(\mathbf{x} - \mathbf{x}_0)$  is a kernel function centered at  $\mathbf{x}_0$  with a radius  $r$  (30 pixels).



**Fig. 5.34** MOA of an unaltered fingerprints. Minutia points generally share similar orientations to their neighbors. Singular regions will introduce variations; the MOA will therefore have higher amplitudes around singular regions. *Source a [6]*

2. Smoothening. The initial density map,  $M'_{dens}$ , is smoothed by a low-pass Gaussian filter of size  $30 \times 30$  with a standard deviation of 10 pixels.
3. Normalization. The values of  $M'_{dens}(x)$  are transformed to the interval  $[0, 1]$  by

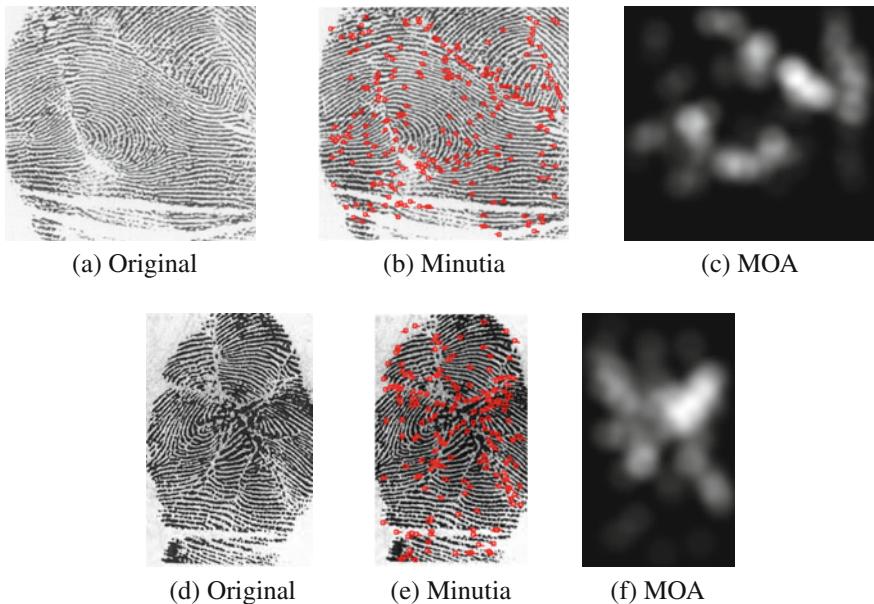
$$M_{dens}(x, y) = \begin{cases} M'_{dens}(x, y)/T, & \text{if } M'_{dens}(x, y) \leq T, \\ 1, & \text{otherwise} \end{cases} \quad (5.36)$$

where  $T$  is a predetermined threshold ( $T$  is set to 6.9).

Figures 5.34 and 5.35 show the resulting density map of the *MOA*. A natural fingerprint will generally only have small peaks around singular regions. The additional minutia around scars and obliterated areas with conflicting orientations in altered fingerprint images will produce areas with higher amplitudes such that the distribution of the density image will differ from an unaltered fingerprint image.

#### 5.4.5.1 Extraction

There are multitudes of different ways of analyzing data and extracting features to be used by an *SVM*. The feature extraction is equivalent to the previously presented state-of-the-art-method, since both of the proposed analysis algorithms are reliant on the distribution of special features across the fingerprint in a similar fashion to the predefined method. The final density maps from each analysis are images in the size of  $512 \times 480$  pixels with intensity values that are normalized to the range of  $[0, 1]$ . The feature extraction will construct a 189-dimensional vector from each analysis. This is done as follows:



**Fig. 5.35** MOA of two altered fingerprints. Regions around scars will add additional high amplitudes to the density map. *Source a [41], d [38]*

1. Columns of 16 pixel are removed from each side of the density map. This gives a density map of size  $480 \times 480$  pixels.
2. In order to separate and extract features from different sections of the image it is divided into  $3 \times 3$  cells. The size of each cell is thus  $160 \times 160$  pixels.
3. Histograms in 21 bins in the range  $[0,1]$  are computed for each of the nine cells.
4. The nine histograms are concatenated into a 189-dimensional feature vector.

The two feature vectors are fused by concatenation which results in a 378-dimensional feature vector. The feature vector is fed into a *SVM* for classification.

## 5.5 Evaluation and Results

In biometric research we commonly expect that proposed methods are evaluated on publicly accessible datasets, such that research results can be reproduced. For the research addressed in this work such a public dataset does not exist. Due to the nature of the target characteristic of interest it is neither possible to ask volunteers on large numbers to alter their fingerprints and to conduct a dedicated data collection. Seemingly, a possibility would be in using the forensic and immigration control databases. However, the individuals typically seek to avoid criminal prosecution or blacklisting by immigration control authorities, and the altered fingerprints, if identified, are

still subject to legal protection. This limitation forced the researchers working in this field to test their methods either using synthetically generated datasets of altered fingerprints or to benchmark them with the very few datasets that do contain altered fingerprint sample. The proposed method was tested on a dataset of altered fingerprints composed from the following sources:

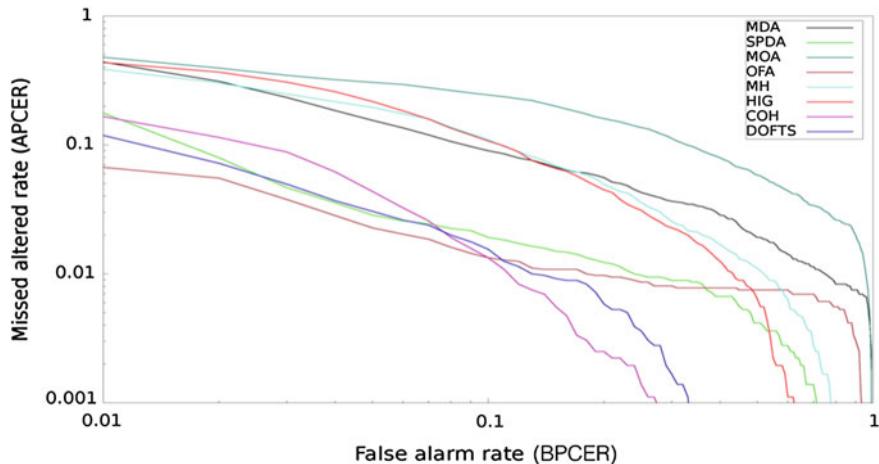
- Brno [2]. A collection of fingerprints containing a wide variety of dermatological diseases.
- GUC-100 [1]. An *in-house* database from Gjøvik University College (GUC) in Norway. A few images containing dermatological diseases have been used as altered.
- Samischenko [38]. The book *Atlas of the Unusual Papilla Patterns* by S.S. Samischenko contains a large collection of fingerprint images with unusual fingerprint patterns together with some natural ones. The database contains fingerprints from fingers altered by burns, acid burns, transplantation, miscellaneous injuries, and diseases.
- NIST SD14—fingerprints that have been identified as altered in the NIST Special Database 14

Moreover we added nonaltered images from public sources [1], FVC 2004 [6]. For the FVC 2004 Images from DB1 (set A) of the public fingerprint database collected for the Fingerprint Verification Competition 2004 (FVC 2004) have been used as unaltered images. The resulting dataset contains 116 altered and 180 unaltered fingerprint images normalized to the size  $512 \times 480$  pixels.

Both the resulting density maps from the Singular Point Density Analysis (SPDA) and Minutia Orientation Analysis (MOA) are classified using the approach used by Yoon et al. [43]. The density maps are cropped into a rectangular shape, normalized into the interval  $[0, 1]$  and divided into  $3 \times 3$  blocks, in total 9, blocks. For each block, a 21-bin histogram is computed that represents the distribution of the density values in the interval  $[0, 1]$ . For each of the density maps, all the histograms for all of the blocks are then concatenated into a single 189-dimensional feature vector that can be used for classification. For benchmarking the approach by Yoon et al. [43] was fully reimplemented. The results as presented by Ellingsgaard in [11] are listed in Table 5.3.

**Table 5.3** Initial evaluation results of Ellingsgaard [11]

Method	Analysis	Alteration Detection Rate (%)
Yoon et al. [43]	OFA	93.7
	MDA	77.5
Ellingsgaard [11]	SPDA	92.0
	MOA	81.6
Feature level fusion	<b>OFA/SPDA</b>	<b>94.6</b>



**Fig. 5.36** DET curves comparing the alteration detection performance. *Source* [14]

The results demonstrate that the proposed method based on SDPA and OA perform on the same accuracy level as the method published by Yoon et al. [43]. An improvement can be achieved by feature fusion of the Orientation Field Analysis (OFA) by Yoon et al. [43] and the proposed Singular Point Density Analysis (SPDA).

A more intensive evaluation also combined with very recently proposed algorithms can be found in Gottschlich et al. [14]. The accuracy of various fingerprint alteration detection methods are illustrated in Fig. 5.36.

Those results are presented in terms of the *Attack Presentation Classification Error Rate (APCER)* and *Bona Fide Presentation Classification Error Rate (BPCER)* metrics, which are derived from the international standardization project ISO/IEC 30107-3 [19]. The APCER metric is expressing the proportion of presentations of altered fingerprints incorrectly classified as bona fide presentation (a.k.a. normal presentations). The BPCER metric is the proportion of presentation of normal fingerprints incorrectly classified as being altered (i.e. attacks).

## 5.6 Conclusion

A novel method for detecting altered fingerprints has been developed that performs competitively with the state-of-the-art method by Yoon et al. [43] by achieving Alteration Detection Rate of **92 %** on the evaluation dataset. In addition, the classification performance can be further improved to **94.6 %** by combining the approach of Yoon et al. [43] and the proposed method. The future work would involve testing on a larger datasets of altered fingerprints from government sources.

**Acknowledgements** The authors would like to thank Anil Jain (Michigan State University), Christophe Champod (Universite de Lausanne), Martin Drahansky (Brno University of Technology, Faculty of Information Technology—STRaDe) and FN Olomouc that kindly provided access to the altered fingerprint data used in this work. Also thanks to Ctirad Sousedik and Martin Olsen for fruitful discussions. This work is carried out under the funding of the EU-FP7 INGRESS project (Grant No. SEC-2012-312792).

## References

1. GUC100 multisensor fingerprint database for in-house (Semipublic) performance test, January 2009. Gjøvik University College (GUC)
2. Fingerprint database. Alterations caused by diseases, March 2013. Faculty of Information Technology at Brno University of Technology and the research group STRaDe in collaboration with dermatologists from FN Olomouc
3. Bazen AM, Gerez SH (2001) Extraction of singular points from directional fields of fingerprints. In: The 7th annual CTIT workshop mobile communications in perspective workshop. Centre for Telematics and Information Technology, Enschede, pp 41–44
4. Bazen AM, Gerez SH (2002) Systematic methods for the computation of the directional fields and singular points of fingerprints. *IEEE Trans Pattern Anal Mach Intell* 24(7):905–919. 060.02
5. Bo J, Ping TH, Lan XM (2008) Fingerprint singular point detection algorithm by poincaré index. *WTOS* 7(12):1453–1462
6. Cappelli MM, Maio D, Maltoni D, Wayman JL, Jain AK (2004) FVC2004: Third fingerprint verification competition. In Proceedings of the first international conference on biometric authentication 1–7
7. Chikkerur SS, Cartwright AN, Govindaraju V (2005) Fingerprint image enhancement using STFT analysis. In IN PROC. ICAPR, pp 20–29
8. Commission, E. (2012) Visa information system—VIS. [http://ec.europa.eu/dgs/home-affairs/what-we-do/policies/borders-and-visas/visa-information-system/index\\_en.htm](http://ec.europa.eu/dgs/home-affairs/what-we-do/policies/borders-and-visas/visa-information-system/index_en.htm)
9. Cummins H (1935) Attempts to alter and obliterate finger-prints. *J Crim Law Criminol* 25:982–991
10. Doležel M, Drahanský M, Urbánek J (2013) Einfluss von hauterkrankungen auf den biometrischen erkennungsprozess. Datenschutz un Datensicherheit (DuD) Heft 6-2013, 358–362
11. Ellingsgaard J (2013) Fingerprint alteration detection. Master thesis, Technical University of Denmark. (June 2013)
12. Ellingsgaard J, Sousedik C, Busch C (2014) Detecting fingerprint alterations by orientation field and minutiae orientation analysis. In 2014 international workshop on biometrics and forensics (IWBF), (March 2014), pp 1–6
13. Feng J, Jain AK, Ross A (2009) Fingerprint alteration. Tech. Rep. MSU-CSE-09-30, Department of Computer Science, Michigan State University, East Lansing, Michigan. (December 2009)
14. Gottschlich C, Mikaelyan A, Olsen M, Bigun J, Busch C (2015) Improving fingerprint alteration detection. In Proceedings of 9th international symposium on image and signal processing and analysis (ISPA 2015)
15. Gu J, Zhou J (2003) A novel model for orientation field of fingerprints. In IEEE computer society conference on computer vision and pattern recognition, pp 493–498
16. Hoover JE, Collins FL (1943) The man without fingerprints. *Collier's Weekly* (January 1943), p 16
17. ISO/IEC JTC1 SC37 Biometrics (2009) ISO/IEC 29794-1:2009 Information Technology—Biometric Sample Quality—Part 1: Framework. International Organization for Standardization

18. ISO/IEC JTC1 SC37 Biometrics (2016) ISO/IEC 30107-1. Information Technology—Biometric presentation attack detection—Part 1: Framework. International Organization for Standardization
19. ISO/IEC JTC1 SC37 Biometrics (2016) ISO/IEC DIS 30107-3. Information Technology—Biometric presentation attack detection—Part 3: Testing and Reporting. International Organization for Standardization
20. Jain AK (1989) Fundamentals of digital image processing. Prentice-Hall Inc, Upper Saddle River, NJ
21. Jain AK, Prabhakar S, Hong L, Pankanti S (2000) Filterbank-based fingerprint matching. *IEEE Trans Image Process* 9:846–859
22. Jain AK, Yoon S (2012) Automatic detection of altered fingerprints. *IEEE Comput* 45(1):79–82
23. Jin C, Kim H, Elliott S (2007) Liveness detection of fingerprint based on band-selective fourier spectrum. In Proceedings of the 10th international conference on information security and cryptology, ICISC'07. Springer, Berlin, pp 168–179
24. Kamei T, Mizoguchi M (1995) Image filter design for fingerprint enhancement. In international symposium on computer vision, 1995. Proceedings, pp 109–114
25. Kawagoe M, Tojo A (1984) Fingerprint pattern classification. *Pattern Recognit* 17(3):295–303
26. Kim B-G, Park D-J (2002) Adaptive image normalisation based on block processing for enhancement of fingerprint image. *Electron Lett* 38(14):696–698
27. Lim E, Jiang X, Yau W-Y (2002) Fingerprint quality and validity analysis. In 2002 international conference on image processing (ICIP), vol 1, pp 469–472
28. Liu M, Jiang X, Kot AC (2005) Fingerprint reference-point detection. *EURASIP J. Appl Signal Process* 2005:498–509
29. MailOnline (2009) Calais migrants mutilate fingertips to hide true identity (July 2009). <http://www.dailymail.co.uk/news/article-1201126/Calais-migrants-mutilate-fingertips-hide-true-identity.html>
30. Maltoni D, Maio D, Jain AK, Prabhakar S (2009) Handbook of fingerprint recognition, 2nd edn. Springer Publishing Company, Incorporated
31. Merkle J, Ihmor H, Korte U, Niesing M, Schwaiger M (2010) Performance of the fuzzy vault for multiple fingerprints (extended version). [arXiv:1008.0807](https://arxiv.org/abs/1008.0807)
32. NIST (2012) Development of NFIQ 2.0—quality feature definitions. Tech. rep. (June 2012)
33. Olsen M, Smida V, Busch C (2015) Finger image quality assessment features—definitions and evaluation. *IET J Biom.* (December 2015)
34. Olsen MA, Xu H, Busch C (2012) Gabor filters as candidate quality measure for NFIQ 2.0. In 2012 5th IAPR international conference on biometrics (ICB), pp 158–163
35. Petrovici A, Lazar C (2010) Identifying fingerprint alteration using the reliability map of the orientation field. *The Annals of the University of Craiova. Series: Automation, Computers, Electronics and Mechatronics* 7(34), 45–52. (1)
36. Rajanna U, Erol A, Bebis G (2010) A comparative study on feature extraction for finger-print classification and performance improvements using rank-level fusion. *Pattern Anal Appl* 13(3):263–272
37. Ravishankar RA (1990) A taxonomy for texture description and identification. Springer, New York, NY
38. Samischenko S (2001) Atlas of the unusual papilla patterns/atlas neobychnykh papilliarnykh uzorov. Uriaprudentsiiia, Moscow
39. Watson C, Garris M, Tabassi C, Wilson RM (2012) NIST biometric image software. (December 2012). <http://www.nist.gov/itl/iad/ig/nbis.cfm>
40. Watson CI, Candela G, Grother P (1994) Comparison of fft fingerprint filtering methods for neural network classification. *NISTIR* 5493 (1994)
41. Wertheim K (1998) An extreme case of fingerprint mutilation. *J Forensic Identif* 48:466–477
42. Xie SJ, Yang JC, Yoon S, Park DS (2008) An optimal orientation certainty level approach for fingerprint quality estimation. In Second international symposium on intelligent information technology application, 2008. IITA'08. (2008), vol 3, pp 722–726

43. Yoon S, Feng J, Jain AK (2012) Altered fingerprints: analysis and detection. *IEEE Trans Pattern Anal Mach Intell* 34(3):451–464
44. Yoon S, Feng J, Jain AK, Ross A (2009) Automatic detection of altered fingerprints. ICPR. (August 2009, Presentation)
45. Zhou J, Chen F, Gu J (2009) A novel algorithm for detecting singular points from fingerprint images. *IEEE Trans Pattern Anal Mach Intell* 31(7):1239–1250

## **Part II**

# **Face and Video Analysis**

# **Chapter 6**

## **Face Sketch Recognition via Data-Driven Synthesis**

**Nannan Wang, Shengchuan Zhang, Chunlei Peng,  
Jie Li and Xinbo Gao**

**Abstract** In some real-world scenarios, there does not always exist a normal photo for face recognition or retrieval purpose, e.g. suspect searching for law enforcement. Under the circumstances, a sketch drawn by the artist is usually taken as the substitute for matching with the mug shot photos collected by the police office. However, due to the great discrepancy of the texture presentation between sketches and photos, common face recognition methods achieve limited performance on this task. In order to shrink this gap, sketches can be transformed to photos relying on some machine learning techniques and then synthesized photos are utilized to match with mug shot photos. Alternatively, photos can also be transformed to sketches and the probe sketch drawn by the artist is matched with the transformed sketches subsequently. Existing learning-based face sketch–photo synthesis methods are grouped into two major categories: data-driven methods (example-based methods) and model-based methods. This chapter would give a comprehensive analysis and comparison to advances on this topic.

---

N. Wang

State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, No. 2 Taibai South Road, Yanta District, Xi'an 710071, People's Republic of China  
e-mail: nwang@xidian.edu.cn

S. Zhang · C. Peng · J. Li · X. Gao (✉)

Lab of Video & Image Processing Systems, School of Electronic Engineering, Xidian University, No. 2 Taibai South Road, Yanta District, Xi'an 710071, People's Republic of China  
e-mail: xbgao@mail.xidian.edu.cn

S. Zhang

e-mail: zsc\_2007@163.com

C. Peng

e-mail: clp.xidian@gmail.com

## 6.1 Introduction

The intrinsic fluidity of face imaging and uncontrollable extrinsic imaging conditions (such as an intended target deliberately concealing his/her identity) means that suitable face images for processing and identifying a person cannot always be obtained [22]. For example, face images acquired by live surveillance cameras at a distance are usually in a low resolution. Though face super-resolution techniques could be utilized to enhance the resolution to some extent, there are still many limits to the problem [1]. An effective way to resolve this problem is to invite artists draw a sketch according to the surveillance video. Another case is that a photo of the culprit is not always available in suspect searching. As a substitute, face sketches are drawn by an artist based on the recollection of an eyewitness. Once a sketch is obtained in aforementioned scenarios, it can be employed to identify the suspect.

However, since face sketches and photos vary a lot in texture representation and imaging mechanism, traditional face recognition techniques perform poorly on this challenging task by matching a sketch with a mug shot database [17, 30]. In order to shrink the great gap between sketches and photos, they can be transformed to the same modality and then conventional face recognition techniques could be applied.

There have been many attempts to interactively or automatically synthesize facial sketches. For example, some commercial software designed for image processing is becoming widely popular, such as Adobe Photoshop,<sup>1</sup> Meitu,<sup>2</sup> and PaperArtist.<sup>3</sup> Figure 6.1 gives some generated sketches by these software and the sketch drawn by the artist. Li et al. [10] presented a simple filter-based framework for face photo-sketch synthesis. Although sketches generated by these analytic-based methods can achieve satisfying visual quality, they look more like photographs and are less stylistic. This would introduce great difficulty for face matching since these generated sketches are actually still in a heterogeneous modality with the sketch drawn by the artist. To really shrink the gap between sketches and photos, learning-based face sketch-photo synthesis methods described in this chapter have proven surprisingly successful in face identification from sketches [22]. Face sketch synthesis means transforming photos in the mug shot database to sketches and then the probe sketch can be matched with these synthesized sketches. Alternatively, face photo synthesis transform the probe sketch drawn by the artist to a photo. Indeed, these two strategies are invertible by just switching the roles of sketches and photos. Thus in following text, we would take face sketch synthesis as an example to introduce existing synthesis methods.

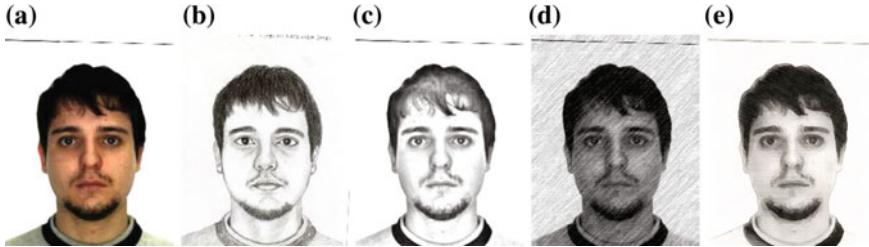
Learning-based methods generally synthesize a sketch relying on some sketch-photo examples through machine learning techniques. Existing learning-based methods incorporate two major categories: model-based methods and data-driven methods. Model-based methods first learn a model from training sketch-photo pairs which is then directly applied to generate sketches from input test photos.

---

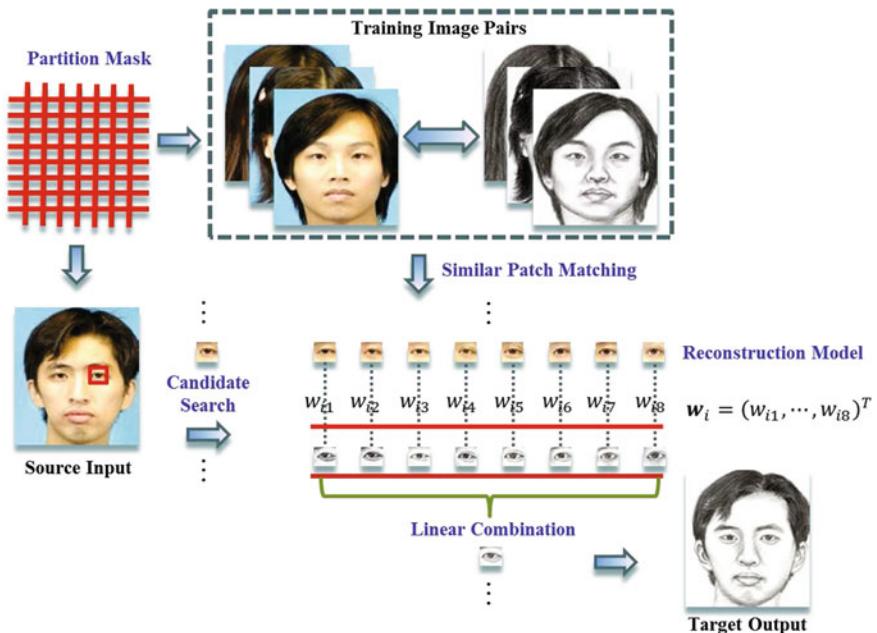
<sup>1</sup><http://www.adobe.com/products/photoshop.html>.

<sup>2</sup><http://xiuxiu.web.meitu.com/>.

<sup>3</sup><http://paperartist.net/>.



**Fig. 6.1** Sketch comparison generated from different sources: (a) input photo; (b) sketch drawn by the artist; (c) sketch generated by Meitu; (d) sketch generated by Photoshop; (e) sketch generated by Paper Artist



**Fig. 6.2** General procedures for data-driven methods

Gao et al. [7, 8, 25, 26, 31] employed embedded hidden Markov model (E-HMM) to learn the nonlinear relationship between sketches and their counterpart photo images. A coupled sketch–photo model is trained which can be simply seen as introducing a series of hidden variables into the Markov chain to control the generation process of a photo–sketch pair (See Fig. 6.2) [22]. The photo model (separated from the coupled model) explains the face photo generation process and the sketch model (separated from the coupled model) describes the generation process of corresponding sketch. Given an input test photo, the trained photo model is first utilized to interpret the test photo to obtain the hidden states. The trained sketch

model is then employed to generate the sketch corresponding to the input photo with the help of obtained hidden states. When E-HMM was first applied for face sketch synthesis [7, 31], the holistic face image is modeled as an E-HMM. Considering the fact that certain fine local features such as those associated with eyes, nose, and mouth, could not be learned, they afterwards proposed to divide each face image into some patches and each patch is taken as an E-HMM [8, 25, 26].

Chang et al. [2] proposed to explore multivariate output regression (ridge regression and relevance vector machine) to learn the mapping relation from photo patches to sketch patches. Each face image (both sketch and photo) is divided into some patches. A local regression function is learned from the training sketch–photo patch pairs corresponding to the same location. Inspired by the success application of sparse representation to image denoising [5] and image restoration [12, 27], sparse representation has been utilized to face sketch synthesis [3, 23]. Actually L1-norm constrained sparse representation relaxation is known as the Lasso, essentially a linear regression regularized with L1-norm on the coefficients [17]. Chang et al. [3] improved their previous multivariate regression method [2] by substituting the original L2-norm regularization with the L1-norm sparse constraint. Different from aforementioned sparse representation based methods, which assumed that the source input and the target output had the same sparse representation coefficients, Wang et al. [23] relaxed this assumption by supposing that they had their respective sparse representations. These two sparse representations are connected through a linear transformation. Then the objective function is composed of two sparse representation parts: one fidelity term between the sparse representation coefficients, and other the regularization term on the linear transformation matrix, under some scale constraints to each atom of dictionaries. They separated the objective function into three subproblems: sparse coding for training samples, dictionary updating and linear transformation matrix updating.

In the following text, we would review existing data-driven methods in Sect. 6.2. Our proposed two kinds of data-driven methods (sparse representation supported neighbor selection methods and graphical representation methods) are sequentially introduced in Sects. 6.3 and 6.4, respectively. Section 6.5 gives experimental results and some concluding remarks are summarized in Sect. 6.6.

## 6.2 Related Work

Data-driven methods synthesize a sketch by searching similar example sketches (or sketch patches) from the training dataset (see Fig. 6.2, the partition mask in this figure is utilized to divide an image into some even patches with some overlap). These selected sketches (or sketch patches) are usually linearly combined to generate the target sketch (or sketch patch) weighted by reconstruction coefficients obtained from reconstruction models. This category of methods generally assumes that a sketch and the corresponding photo share the same linear combination coefficients.

Principal component analysis (PCA) plays the role of the reconstruction model in the eigentransformation method proposed by Tang and Wang [16–18]. The eigen-transformation method works on the image level rather than the image patch level. The input test photo is first projected onto the whole photo training set to obtain the linear combination coefficients. The target sketch is then synthesized by linearly combining sketches in the training set with aforementioned coefficients. Here the whole sketch training set is taken as the similar sketch candidates to the target output sketch. Since these PCA-based methods performed on a holistic face image, this may result in the loss of some critical fine details and blurring effect.

In order to overcome limitations resulted from the holistic strategy in [16–18], Liu et al. [11] proposed a patch-based method assuming that sketch patches and their corresponding photo patches are sampled from two manifolds sharing a similar geometrical structure. The input photo patch is first reconstructed in a least square manner using its  $K$  nearest neighbors selected from the patches extracted from the photo training set. The reconstruction coefficients are computed in this process. The target sketch patch is then synthesized by linearly combining candidate sketch patches corresponding to those neighbor photo patches of the test photo patch, weighted by obtained reconstruction coefficients. After fusing all obtained target sketch patches, the final target sketch is generated.

Markov random field (MRF) is explored to model the interaction between sketch-photo pairs and neighborhoods through two compatibility functions [24]: data compatibility (constraints on the fidelity between the test photo patch and the target sketch patch) and spatial compatibility (constraints on the neighboring patches of the target sketch). Considering the strong structural property of face images, the uniform scale of MRF has limited ability to address the long-range dependency among local patches. Wang and Tang [24] further proposed a multiscale MRF model which constructs a hierarchical model of several MRFs corresponding to different scales (image patch size). Their method constructs the pairwise compatibility functions through the nearest neighbors searched from a training set across different scales. Under the maximum a posterior (MAP) rule, the best matched candidate sketch patch is taken as the target output sketch patch corresponding to the input photo patch. This can be interpreted as the weight coefficient corresponding to the selected candidate sketch patch is 1 and 0 s for other candidate sketch patches in Fig. 6.2.

Zhou et al. [32] claimed that above-mentioned MRF-based method [24] had two major drawbacks: cannot synthesize new sketch patches (i.e., each patch of final target output is from the training set) and NP-hard for the optimization problem in solving the MRF model. Then they proposed a weighted MRF method to model the relation between sketch and photo patches. By a linear combination of selected candidate sketch patches, their method could synthesize new sketch patches existing not in the training sketch set. In addition, the objective function is a convex optimization problem which has the unique optimal solution.

The core distinction between different data-driven methods lies in the fact that they either employ different patch matching strategies to select candidate sketch

patches or construct different reconstruction models to resolve reconstruction weights. Following we would introduce two methods for each of these two issues respectively.

## 6.3 Sparse Representation Supported Candidate Selection Methods

This section would introduce two sparse representation based face sketch synthesis methods. Different from aforementioned model-based face sketch synthesis methods [3, 23], the two proposed methods do not learn a coupled dictionary for face sketch synthesis. Instead, they utilize sparse representation as a tool for candidate sketch patch selection. We will introduce these two methods sequentially.

### 6.3.1 Sparse Feature Selection Based Face Sketch Synthesis

Aforementioned data-driven methods all search fixed number of candidate sketch (or sketch patches) to synthesize the target sketch (or sketch patch). It may result in some blurring effect or noise on the synthesized sketch. The underlying reason is that different parts of faces have different number of nearest neighbors in the training set. This can be easily understood by the following example: if the number of nearest neighbor  $K = 6$  and there are actually only 5 nearest neighbors most related to the test image patch. Then  $K$ -NN based method may still choose another image patch which is in fact a mismatched (noisy) patch. Inspired by the fact that sparse representation is capable of finding the fewest closely related dictionary atoms to reconstruct the input signal, we proposed a sparse feature selection (SFS) method for face sketch synthesis based on sparse representation [6].

Given a query photo  $\mathbf{P}_t$ , it is first divided into  $M$  even patches with some overlap  $\{\mathbf{t}_1, \dots, \mathbf{t}_i, \dots, \mathbf{t}_M\}$ . Each patch is ordered lexicographically as a column vector. Let  $\mathbf{D}_p$  denote a dictionary whose columns consist of patches sampled from the photo training set and  $\mathbf{D}_s$  indicate the dictionary with the sketch patches corresponding to photo patches in  $\mathbf{D}_p$  as its column vectors. Then the sparse representation of  $\mathbf{t}_i$  can be represented as follows:

$$\min_{\mathbf{c}_i} \|\mathbf{c}_i\|_1, \text{s.t. } \|\mathbf{D}_p \mathbf{c}_i - \mathbf{t}_i\|_2^2 \leq \varepsilon, \quad (6.1)$$

where  $\mathbf{c}_i = (c_{i1}, \dots, c_{il})$  indicates the coefficient vector and  $l$  is the number of columns in  $\mathbf{D}_p$ . We can solve the above optimization problem for  $\mathbf{c}_i$ . Afterwards the neighborhood of photo patch  $\mathbf{t}_i$  is achieved according to the following criteria:

$$\mathcal{N}(\mathbf{t}_i) = \{k | \delta(c_{ik}) \neq 0, 1 \leq k \leq l\}, \quad (6.2)$$

where  $\mathcal{N}(\mathbf{t}_i)$  denotes the neighborhood for  $\mathbf{t}_i$  and  $\delta(\cdot)$  is a neighbor selection function:

$$\delta(c_{ik}) = \begin{cases} c_{ik}, & |c_{ik}| > \sigma \\ 0, & \text{otherwise} \end{cases} \quad (6.3)$$

where  $\sigma$  is a small positive real number, which is set to 0.001 in our experiments. Once the neighborhood is determined, the weight for every neighbor in  $\mathcal{N}(\mathbf{t}_i)$  is calculated as follows:

$$c_{ik} = \frac{\delta(c_{ik})}{\sum_{k=1}^l \delta(c_{ik})}, \quad (6.4)$$

Finally, the corresponding target sketch patch can be synthesized as

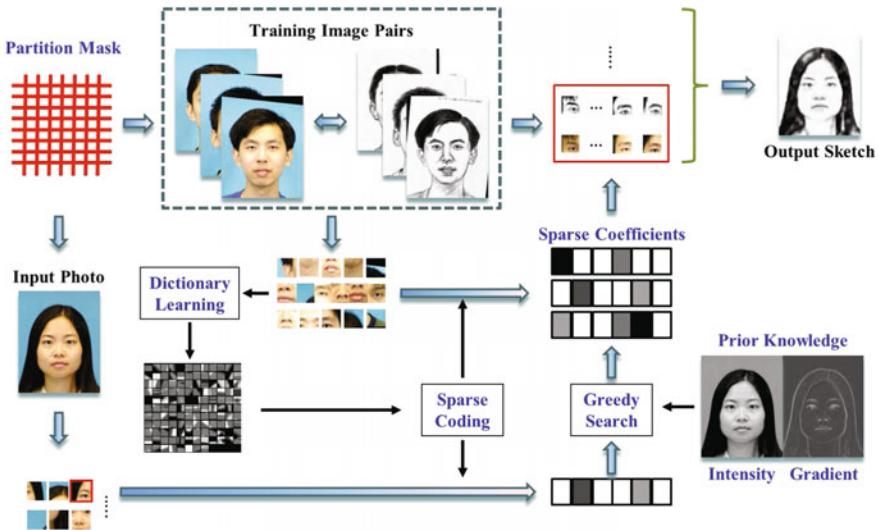
$$\mathbf{s}_i = \mathbf{D}_s \mathbf{c}_i. \quad (6.5)$$

For each photo patch in the query photo, we iterate the above steps and then fuse all these synthesized patches into a whole sketch by averaging the overlapping areas. In addition, in order to compensate the high frequency or detail information filtered with overlapping areas averaged, they employed support vector regression for further improving the quality of the synthesized sketch [20].

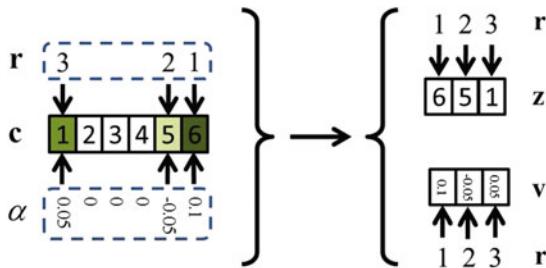
### 6.3.2 Sparse Representation Based Greedy Search for Face Sketch Synthesis

Most existing methods cannot handle some non-facial factors, such as hair styles, hairpins, glasses and backgrounds, if these factors are excluded in the training set. The reason is that these methods directly utilize image intensities as the query feature to search nearest neighbors and restrict the search range in local region to save computational time since images have been aligned by the positions of the eyes. We proposed a face sketch synthesis method via sparse representation based greedy search by considering a tradeoff between computational time and search range to overcome aforementioned drawbacks [28]. Their method adopted sparse coefficient values and dictionary atom selection orders as the query feature to find nearest neighbors which are robust to image backgrounds. In addition, they exploited the distribution of image patches extracted from the training set. After organizing these image patches in a tree structure, each query image patch was checked for its own leaf without costing much computational time. Since the search range was extended to the whole training set, their method can synthesize non-facial factors with facial elements (e.g., glasses can be substituted by face contours).

Suppose there are  $N$  photo-sketch pairs  $(\mathbf{p}_1, \mathbf{s}_1), \dots, (\mathbf{p}_i, \mathbf{s}_i), \dots, (\mathbf{p}_N, \mathbf{s}_N)$  in the training set (see Fig. 6.3), we divide them into overlapping photo-sketch patch pairs denoted as  $\{B_p, B_s\}$ ,  $B_p = \{\mathbf{p}_{11}, \dots, \mathbf{p}_{1M}, \dots, \mathbf{p}_{ij}, \dots, \mathbf{p}_{N1}, \dots, \mathbf{p}_{NM}\}$  is the sampled photo patches and  $B_s = \{\mathbf{s}_{11}, \dots, \mathbf{s}_{1M}, \dots, \mathbf{s}_{ij}, \dots, \mathbf{s}_{N1}, \dots, \mathbf{s}_{NM}\}$  is the corresponding



**Fig. 6.3** Diagram of face sketch synthesis via sparse representation based greedy search



**C** : Sparse Representation

**Z** : Index of Dictionary Atom being Selected

**r** : Dictionary Atom Selection Orders

**V** : Sparse Coefficient Values of Index

$\alpha$  : Sparse Coefficient Values

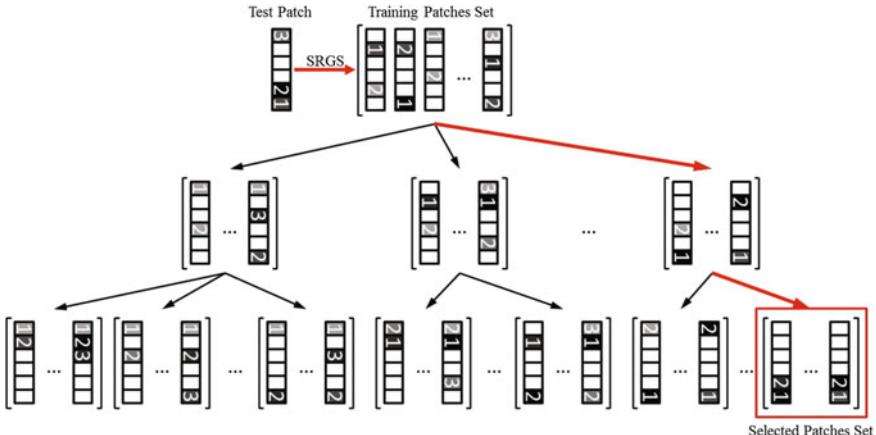
**Fig. 6.4** Illustration of dictionary atom selection orders  $\mathbf{r}$  and sparse coefficient values  $\alpha$ . Here the dictionary atom selection orders  $\mathbf{r}$  refers to the order of corresponding atom being selected. Given a sparse representation  $\mathbf{c}$ , we can rearrange  $\mathbf{r}$  and  $\alpha$  to obtain its index of dictionary atom being selected  $\mathbf{z}$  and sparse coefficient values of index  $\mathbf{v}$

sketch patches where  $M$  denotes the number of sampled patches in each image. Then a photo patch dictionary  $\mathbf{D}_p$  is learnt with the randomly selected photo patches from  $B_p$ . By sparse coding, the sparse representations  $C_p$  of  $B_p$  is obtained where  $C_p = \{\mathbf{c}_{11}, \dots, \mathbf{c}_{1M}, \dots, \mathbf{c}_{ij}, \dots, \mathbf{c}_{N1}, \dots, \mathbf{c}_{NM}\}$ . Each sparse representation  $\mathbf{c}_{ij}$  contains two variables: sparse coefficient values  $\alpha_{ij}$  and dictionary atom selection orders  $\mathbf{r}_{ij}$ . As shown in Fig. 6.4, these two variables can be further arranged to obtain index of

dictionary atom being selected  $\mathbf{z}_{ij}$  and sparse coefficient values of index  $\mathbf{v}_{ij}$ . Given an input photo  $\mathbf{x}$ , it is divided into overlapping patches  $\{\mathbf{x}_j\}_{j=1}^M$  in the same way as the training set. For each photo patch  $\mathbf{x}_j$ , its sparse representation  $\mathbf{c}_j$  including sparse coefficient values  $\alpha_j$ , dictionary atom selection orders  $\mathbf{r}_j$ , index of dictionary atom being selected  $\mathbf{z}_j$  and sparse coefficient values of index  $\mathbf{v}_j$  is first acquired by sparse coding with learnt photo patch dictionary  $\mathbf{D}_p$ . Then two criteria are applied to select nearest neighbors as follows.

$$\begin{aligned}\{\mathbf{c}\}_t &= \left\{ \mathbf{c} \mid \mathbf{z}_q(k) = \mathbf{z}_g(k) \right\} \\ \{\mathbf{c}\}_t &= \left\{ \mathbf{c} \mid \|\mathbf{v}_q(k) - \mathbf{v}_g(k)\| \leq \varepsilon \right\}\end{aligned}\quad (6.6)$$

where symbol  $q$  represents the test image patch (query patch) and symbol  $g$  denotes the training image patches (gallery patches).  $\varepsilon$  is the threshold.  $k$  is the index of dictionary atom selection order in sparse representations  $\mathbf{c}_q$  and  $\mathbf{c}_g$  where  $\mathbf{c}_q = \mathbf{c}_j$  and  $\mathbf{c}_g \in C_p$ .  $\{\mathbf{c}\}_t$  is the sparse representations subset after  $t$  iterations. Figure 6.5 illustrates the detail of data indexing structure. It can be seen that test patch searches through the data indexing structure until termination condition (all the nonzero coefficients in sparse representation of the test patch are traversed or the number of selected patches meets the requirement) is met. The sparse representations of final selected patches set are similar to that of test patch both in sparse coefficient values and dictionary atom selection orders directions. For each test photo patch, we can obtain its candidate sketch patches via above operations. The final sketch can be synthesized through the single-scale MRF-based model described in [24].



**Fig. 6.5** Data indexing structure. Brackets denote selected subsets. Numbers in the squares represent the dictionary atom selection order while different levels of square transparency represent different sparse coefficient values. Each arrow represents the neighbor searching process with the assistance of dictionary atom selection orders and sparse coefficient values

## 6.4 Graphical Representation Based Reconstruction Models

This section would introduce two graphical model based methods: the transductive method and the multiple representation based method.

### 6.4.1 Transductive Face Sketch Synthesis

Above face sketch–photo synthesis methods are under the inductive learning framework which may result in high losses for test samples. The underlying cause is that inductive learning only considers the empirical loss of training samples and neglects test samples. We presented a novel transductive face sketch–photo synthesis method that incorporates both the test samples and the training samples into the learning process [21]. In particular, it defines a probabilistic model to optimize both the reconstruction fidelity (whether the generated sketches are recognized as the original person in the test photo) of the input photo and the synthesis fidelity (whether the generated sketches are as close to the ones drawn by artists as possible) of the target output sketch, and efficiently optimizes this probabilistic model by alternating optimization.

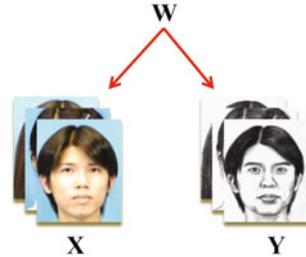
Let  $\mathbf{X}$  denote the data matrix containing both training and test photo patches,  $\mathbf{Y}$  represent the corresponding training sketch patches and target output sketch patches, and  $\mathbf{W}$  consist of the weight vectors. They are all arranged in a row form which means each row vector in  $\mathbf{X}$  denotes a photo patch and  $\mathbf{Y}$  and  $\mathbf{W}$  have a similar meaning:

$$\mathbf{X} = \begin{bmatrix} (\mathbf{p}_1^1)^T \\ \vdots \\ (\mathbf{p}_M^N)^T \\ \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_M^T \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{1\cdot} \\ \vdots \\ \mathbf{X}_{(NM)\cdot} \\ \mathbf{X}_{(NM+1)\cdot} \\ \vdots \\ \mathbf{X}_{((N+1)M)\cdot} \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} (\mathbf{s}_1^1)^T \\ \vdots \\ (\mathbf{s}_M^N)^T \\ \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_M^T \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_{1\cdot} \\ \vdots \\ \mathbf{Y}_{(NM)\cdot} \\ \mathbf{Y}_{(NM+1)\cdot} \\ \vdots \\ \mathbf{Y}_{((N+1)M)\cdot} \end{bmatrix}, \mathbf{W} = \begin{bmatrix} \mathbf{W}_{1\cdot} \\ \vdots \\ \mathbf{W}_{(NM)\cdot} \\ \mathbf{W}_{(NM+1)\cdot} \\ \vdots \\ \mathbf{W}_{((N+1)M)\cdot} \end{bmatrix}$$

where  $\mathbf{X}_{i\cdot}$ ,  $\mathbf{Y}_{i\cdot}$ , and  $\mathbf{W}_{i\cdot}$ ,  $i = 1, \dots, (N+1)M$  denote the  $i$ th row vector of  $\mathbf{X}$ ,  $\mathbf{Y}$ , and  $\mathbf{W}$  respectively (the same meaning as the following similar representations).

**Fig. 6.6** Illustration of the generative process of photo patches and sketch patches from common hidden parameters



We can model the generative process of photo patches and sketch patches by their common hidden parameters  $\mathbf{W}$  as shown in Fig. 6.6. Sketch–photo patch pairs are sampled from the joint distribution  $P(\mathbf{X}, \mathbf{Y}, \mathbf{W})$  governed by the weight matrix  $\mathbf{W}$ . We can then decompose the generative process as

$$\begin{aligned} P(\mathbf{Y}, \mathbf{X}, \mathbf{W}) \\ = P(\mathbf{Y}, \mathbf{X} | \mathbf{W})P(\mathbf{W}) \\ = P(\mathbf{Y} | \mathbf{X}, \mathbf{W})P(\mathbf{X} | \mathbf{W})P(\mathbf{W}) \end{aligned} \quad (6.7)$$

where  $\mathbf{W}$  controls the generation of sketches and photos,  $P(\mathbf{W})$  denotes the prior probability of the hidden parameters  $\mathbf{W}$ ,  $P(\mathbf{X} | \mathbf{W})$  is the conditional probability of generating photos  $\mathbf{X}$  given  $\mathbf{W}$ , and  $P(\mathbf{Y} | \mathbf{X}, \mathbf{W})$  indicates the conditional probability of generating sketches  $\mathbf{Y}$  given corresponding photos  $\mathbf{X}$  and parameters  $\mathbf{W}$ . The prior probability is usually treated as a regularization term on  $\mathbf{W}$ . In this paper, we explore the compatibility constraint on overlapping regions between neighboring sketch patches to model  $P(\mathbf{W})$

$$\begin{aligned} P(\mathbf{W}) &\propto \\ \prod_{(i,j) \in \mathcal{E}} \exp \Big\{ -\frac{\|\sum_{k \in \mathcal{N}(i)} \mathbf{W}_{ik} \mathbf{Y}_k^{(i,j)} - \sum_{l \in \mathcal{N}(j)} \mathbf{W}_{il} \mathbf{Y}_l^{(j,i)}\|^2}{2\sigma_r^2} \Big\} \\ \text{s.t. } \sum_{k \in \mathcal{V}} \mathbf{W}_{ik} = 1, \mathbf{W}_{ik} \geq 0, \forall i \in \mathcal{V} \end{aligned} \quad (6.8)$$

where  $\mathcal{N}(i)$  denotes the indices of neighbors of the node  $i$ ,  $\mathbf{W}_{ik}$  represents the element located on the  $i$ th row and the  $k$ th column of the matrix  $\mathbf{W}$ ,  $\mathbf{Y}_k^{(i,j)}$  means the pixel intensities of the overlapping region (determined by the neighboring relation of nodes  $i$  and  $j$ ) of the  $k$ th neighbor of the node  $i$ , and  $\mathbf{Y}_l^{(j,i)}$  means the pixel intensities of the overlapping region (determined by the neighboring relation of nodes  $i$  and  $j$ ) of the  $l$ th neighbor of the node  $j$ . Note that elements in the  $i$ th row  $i = 1, 2, \dots, (N+1)M$

row of the weight matrix  $\mathbf{W}$  are 0 except for elements located in the neighborhood of the  $i$ th node. This prior indicates the compatibility of the overlapping regions of neighboring sketch patches.

In face sketch synthesis,  $\mathbf{X}$  is observed, so  $P(\mathbf{X}|\mathbf{W})$  is the likelihood function. It can be modeled as a Gaussian distribution:

$$P(\mathbf{X}|\mathbf{W}) \propto \prod_{i \in \mathcal{V}} \exp \left\{ -\frac{\|\mathbf{X}_i - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{X}_j\|^2}{2\sigma_{dp}^2} \right\} \quad (6.9)$$

The probability  $P(\mathbf{X}|\mathbf{W})$  mainly considers the reconstruction fidelity between a photo patch  $\mathbf{X}_i$  and its nearest neighbors  $\mathbf{X}_j, j \in \mathcal{N}(i)$ .

From Fig. 6.6, we find that given  $\mathbf{W}$ ,  $\mathbf{Y}$  is conditionally independent of  $\mathbf{X}$ , i.e.,  $P(\mathbf{X}, \mathbf{Y}|\mathbf{W}) = P(\mathbf{X}|\mathbf{W})P(\mathbf{Y}|\mathbf{W})$  and then  $P(\mathbf{Y}|\mathbf{X}, \mathbf{W}) = P(\mathbf{Y}|\mathbf{W})$ . It is straightforward that the probability  $P(\mathbf{Y}|\mathbf{W})$  can be modeled as a product of a series of independent and identically distributed normal distributions:

$$\begin{aligned} P(\mathbf{Y}|\mathbf{X}, \mathbf{W}) &= P(\mathbf{Y}|\mathbf{W}) \\ &\propto \prod_{i \in \mathcal{V}} P\left(\mathbf{Y}_i - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{Y}_{(j)}\right) \\ &\propto \prod_{i \in \mathcal{V}} \exp \left\{ -\frac{\|\mathbf{Y}_i - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{Y}_j\|^2}{2\sigma_{ds}^2} \right\} \end{aligned} \quad (6.10)$$

$$\begin{aligned} P(\mathbf{X}, \mathbf{Y}, \mathbf{W}) &\propto \prod_{i \in \mathcal{V}} \left\{ \exp \left\{ -\frac{\|\mathbf{Y}_i - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{Y}_j\|^2}{2\sigma_{ds}^2} \right\} \exp \left\{ -\frac{\|\mathbf{X}_i - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{X}_j\|^2}{2\sigma_{dp}^2} \right\} \right\} \\ &\quad \prod_{(i,j) \in \mathcal{E}} \exp \left\{ -\frac{\|\sum_{k \in \mathcal{N}(i)} \mathbf{W}_{ik} \mathbf{Y}_k^{(i,j)} - \sum_{l \in \mathcal{N}(j)} \mathbf{W}_{il} \mathbf{Y}_l^{(j,i)}\|^2}{2\sigma_r^2} \right\} \\ &= \exp \left\{ -\sum_{i \in \mathcal{V}} \left\{ \frac{\|\mathbf{Y}_i - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{Y}_j\|^2}{2\sigma_{ds}^2} + \frac{\|\mathbf{X}_i - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{X}_j\|^2}{2\sigma_{dp}^2} \right\} \right. \\ &\quad \left. - \sum_{(i,j) \in \mathcal{V}} \frac{\|\sum_{k \in \mathcal{N}(i)} \mathbf{W}_{ik} \mathbf{Y}_k^{(i,j)} - \sum_{l \in \mathcal{N}(j)} \mathbf{W}_{il} \mathbf{Y}_l^{(j,i)}\|^2}{2\sigma_r^2} \right\} \end{aligned} \quad (6.11)$$

$$\begin{aligned} &\max_{\mathbf{W}, \mathbf{y}_1, \dots, \mathbf{y}_M} P(\mathbf{W}, \mathbf{y}_1, \dots, \mathbf{y}_M | \mathbf{x}_1, \dots, \mathbf{x}_M, \mathbf{X}_1, \dots, \mathbf{X}_{(NM)}, \mathbf{Y}_1, \dots, \mathbf{Y}_{(NM)}) \\ &\Leftrightarrow \max_{\mathbf{W}, \mathbf{y}_1, \dots, \mathbf{y}_M} \frac{P(\mathbf{Y}, \mathbf{X}, \mathbf{W})}{P(\mathbf{X}, \mathbf{Y}_1, \dots, \mathbf{Y}_{(NM)})} \\ &\Leftrightarrow \max_{\mathbf{W}, \mathbf{y}_1, \dots, \mathbf{y}_M} P(\mathbf{Y}, \mathbf{X}, \mathbf{W}) \end{aligned} \quad (6.12)$$

$$\begin{aligned}
& \min_{\mathbf{W}, \mathbf{y}_1, \dots, \mathbf{y}_M} \sum_{i \in \mathcal{V}} \left\{ \|\mathbf{Y}_i - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{Y}_j\|^2 + \alpha \|\mathbf{X}_i - \sum_{j \in \mathcal{N}(i)} \mathbf{W}_{ij} \mathbf{X}_j\|^2 \right\} \\
& + \beta \sum_{(i,j) \in \mathcal{V}} \left\| \sum_{k \in \mathcal{N}(i)} \mathbf{W}_{ik} \mathbf{Y}_k^{(i,j)} - \sum_{l \in \mathcal{N}(j)} \mathbf{W}_{il} \mathbf{Y}_l^{(j,i)} \right\|^2 \\
& s.t. \quad \sum_{k \in \mathcal{V}} \mathbf{W}_{ik} = 1, \mathbf{W}_{ik} \geq 0, \forall i \in \mathcal{V} \\
\Leftrightarrow & \min_{\mathbf{W}, \mathbf{y}_1, \dots, \mathbf{y}_M} \text{tr}(\mathbf{Y}^T \mathbf{M} \mathbf{Y}) + \alpha \text{tr}(\mathbf{X}^T \mathbf{M} \mathbf{X}) + \beta \sum_{(i,j) \in \mathcal{V}} \left\| \sum_{k \in \mathcal{N}(i)} \mathbf{W}_{ik} \mathbf{Y}_k^{(i,j)} - \sum_{l \in \mathcal{N}(j)} \mathbf{W}_{il} \mathbf{Y}_l^{(j,i)} \right\|^2 \\
& s.t. \quad \sum_{k \in \mathcal{V}} \mathbf{W}_{ik} = 1, \mathbf{W}_{ik} \geq 0, \forall i \in \mathcal{V}
\end{aligned} \tag{6.13}$$

We can then reformulate the joint probability as in Eq. (6.11) on the next page. In face sketch synthesis, we are given an input photo as the test image and a number of sketch–photo pairs as the training set, and the objective is to infer the sketch corresponding to the input. This can be formulated as a maximum a posterior probability estimation problem (6.12). According to (6.11), the above maximization problem (6.12) is equivalent to the minimization of the objective function (6.13), where  $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W})$ ,  $\alpha = \frac{\sigma_{ds}^2}{\sigma_{dp}^2}$ ,  $\beta = \frac{\sigma_{ds}^2}{\sigma_r^2}$ ,  $\mathbf{I}$  is the identity matrix with the same size as  $\mathbf{W}$ , the superscript  $T$  is the transpose operator and  $\text{tr}(\cdot)$  is the trace operator.

We apply the alternating minimization method to the above problem to obtain a local solution according to the following two steps:

(1) fixing  $\mathbf{W}$ , update  $\mathbf{y}_1, \dots, \mathbf{y}_M$  by solving (6.14)

$$\min_{\mathbf{y}_1, \dots, \mathbf{y}_M} \text{tr}(\mathbf{Y}^T \mathbf{M} \mathbf{Y}) \tag{6.14}$$

(2) then fixing  $\mathbf{y}_1, \dots, \mathbf{y}_M$  to be the above obtained value, and update  $\mathbf{W}$  by solving (6.15)

$$\begin{aligned}
& \min_{\mathbf{W}} \|\mathbf{U} - \mathbf{W}\mathbf{U}\|^2 \\
& + \beta \sum_{(i,j) \in \mathcal{V}} \left\| \sum_{k \in \mathcal{N}(i)} \mathbf{W}_{ik} \mathbf{Y}_k^{(i,j)} - \sum_{l \in \mathcal{N}(j)} \mathbf{W}_{il} \mathbf{Y}_l^{(j,i)} \right\|^2 \\
& s.t. \quad \sum_{k \in \mathcal{V}} \mathbf{W}_{ik} = 1, \mathbf{W}_{ik} \geq 0, \forall i \in \mathcal{V}
\end{aligned} \tag{6.15}$$

where  $\mathbf{U} = [\mathbf{X} \quad \sqrt{\alpha} \mathbf{Y}]$  and we name each row of  $\mathbf{U}$  a *pairwise patch* since it consists of two patches extracted from a sketch–photo pair. We alternately conduct (6.14) and (6.15) until convergence.

#### 6.4.2 Multiple Representation Based Face Sketch Synthesis

As we know, face images can be described using features from multiple aspects. However, most existing face sketch synthesis methods simply utilize image intensi-

ties to measure the similarity of two image patches which may lead to incorrect patch matching. In order to sufficiently dig out the information in image patches to correct patch matching, we presented a novel multiple representations based face sketch–photo synthesis method that adaptively combines multiple features to represent an image patch and further conduct more reasonable patch matching which is robust to light variation [15].

In order to synthesize the sketch patch  $\mathbf{y}_i$  corresponding to the input photo patch  $\mathbf{x}_i$ , where  $i = 1, 2, \dots, N$ , we find  $K$  candidate photo patches  $\{\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,K}\}$ , where  $\mathbf{x}_{i,k}$  represents the  $k$ th candidate photo patch for the  $i$ th input photo patch  $\mathbf{x}_i$  within the search region around the location of  $\mathbf{x}_i$ . The target sketch patch  $\mathbf{y}_i$  ( $i = 1, 2, \dots, N$ ) can then be obtained by the linear combination of the corresponding  $K$  candidate sketch patches  $\{\mathbf{y}_{i,1}, \mathbf{y}_{i,2}, \dots, \mathbf{y}_{i,K}\}$  weighted by the  $K$ -dimensional vector  $\mathbf{w}_i$ , where  $w_{i,k}$  represents the weight of the  $k$ th candidate sketch patch as follows:

$$\mathbf{y}_i = \sum_{k=1}^K w_{i,k} \mathbf{y}_{i,k} \quad (6.16)$$

where  $\sum_{k=1}^K w_{i,k} = 1$ .

To estimate the weights of the candidate sketch patches and generate a target sketch of high quality, we need to jointly model all the patches using the Markov network framework, similar to [21, 24, 32] and [29]. Since the target sketch patches are only dependent on the weights, as shown in Eq. (6.16), the joint probability of the input photo patches and the target sketch patches is equal to that of the input photo patches and the weights. It is defined as

$$\begin{aligned} & p(\mathbf{y}_1, \dots, \mathbf{y}_N, \mathbf{x}_1, \dots, \mathbf{x}_N) \\ &= p(\mathbf{w}_1, \dots, \mathbf{w}_N, \mathbf{x}_1, \dots, \mathbf{x}_N) \\ &= \prod_i \Phi(\mathbf{f}(\mathbf{x}_i), \mathbf{f}(\mathbf{w}_i)) \prod_{(i,j) \in \Xi} \Psi(\mathbf{w}_i, \mathbf{w}_j) \end{aligned} \quad (6.17)$$

where,  $(i, j) \in \Xi$  means that the  $i$ th image patch and the  $j$ th image patch are neighbors,  $\Phi(\mathbf{f}(\mathbf{x}_i), \mathbf{f}(\mathbf{w}_i))$  is the local evidence function, and  $\Psi(\mathbf{w}_i, \mathbf{w}_j)$  is the neighboring compatibility function. In the local evidence function,  $\mathbf{f}(\mathbf{x}_i) = [\mathbf{f}_1(\mathbf{x}_i), \mathbf{f}_2(\mathbf{x}_i), \dots, \mathbf{f}_L(\mathbf{x}_i)]$ , where  $\mathbf{f}_l(\mathbf{x}_i)$  means the  $l$ th representation of the photo patch  $\mathbf{x}_i$ , and  $\mathbf{f}(\mathbf{w}_i) = [\mathbf{f}_1(\mathbf{w}_i), \mathbf{f}_2(\mathbf{w}_i), \dots, \mathbf{f}_L(\mathbf{w}_i)]$  in which  $\mathbf{f}_l(\mathbf{w}_i) = \sum_{k=1}^K w_{i,k} \mathbf{f}_l(\mathbf{x}_{i,k})$ .

$\Phi(\mathbf{f}(\mathbf{x}_i), \mathbf{f}(\mathbf{w}_i))$  can be represented as

$$\begin{aligned} & \Phi(\mathbf{f}(\mathbf{x}_i), \mathbf{f}(\mathbf{w}_i)) \\ & \propto \exp\left\{-\sum_{l=1}^L \mu_{i,l} \|\mathbf{f}_l(\mathbf{x}_i) - \sum_{k=1}^K w_{i,k} \mathbf{f}_l(\mathbf{x}_{i,k})\|^2 / 2\delta_\Phi^2\right\} \end{aligned} \quad (6.18)$$

where  $\mu_i = [\mu_{i,1}, \mu_{i,2}, \dots, \mu_{i,L}]$ .  $\mu_{i,l}$  represents the weight of the distance of the  $l$ th representation between the  $i$ th photo patch  $\mathbf{f}_l(\mathbf{x}_i)$  and the combination of its candidates. The rationale behind the local evidence function is that if  $\sum_{k=1}^K w_{i,k} \mathbf{y}_{i,k}$  is a good estimation of  $\mathbf{y}_i$ ,  $\mathbf{f}_l(\sum_{k=1}^K w_{i,k} \mathbf{x}_{i,k})$  should be similar to  $\mathbf{f}_l(\mathbf{x}_i)$ . In order to exploit the relationship between the target image patch and its candidate image patches represented by multiple representations, we simply assume that  $\sum_{k=1}^K w_{i,k} \mathbf{f}_l(\mathbf{x}_{i,k})$  should also be similar to  $\mathbf{f}_l(\mathbf{x}_i)$  here, which is easier to be optimized, too.

The neighboring compatibility function  $\Psi(\mathbf{w}_i, \mathbf{w}_j)$  is defined as

$$\begin{aligned} & \Psi(\mathbf{w}_i, \mathbf{w}_j) \\ & \propto \exp\left\{-\left\|\sum_{k=1}^K w_{i,k} \mathbf{o}_{i,k}^j - \sum_{k=1}^K w_{j,k} \mathbf{o}_{j,k}^i\right\|^2 / 2\delta_\Psi^2\right\} \end{aligned} \quad (6.19)$$

where  $\mathbf{o}_{i,k}^j$  represents the overlapping area of the candidate sketch patch  $\mathbf{y}_{i,k}$  with the  $j$ th patch. This term is utilized to guarantee that neighboring patches have compatible overlaps. In Eqs. (6.18) and (6.19),  $\delta_\phi$  and  $\delta_\psi$  are two parameters balancing the local evidence function and the neighboring compatibility function separately.

To avoid the weight of multiple representations overfitting to one representation [9], a regularization term  $\exp\{-\lambda_i \|\mu_i\|^2\}$  is added to the joint probability in Eq. (6.17):

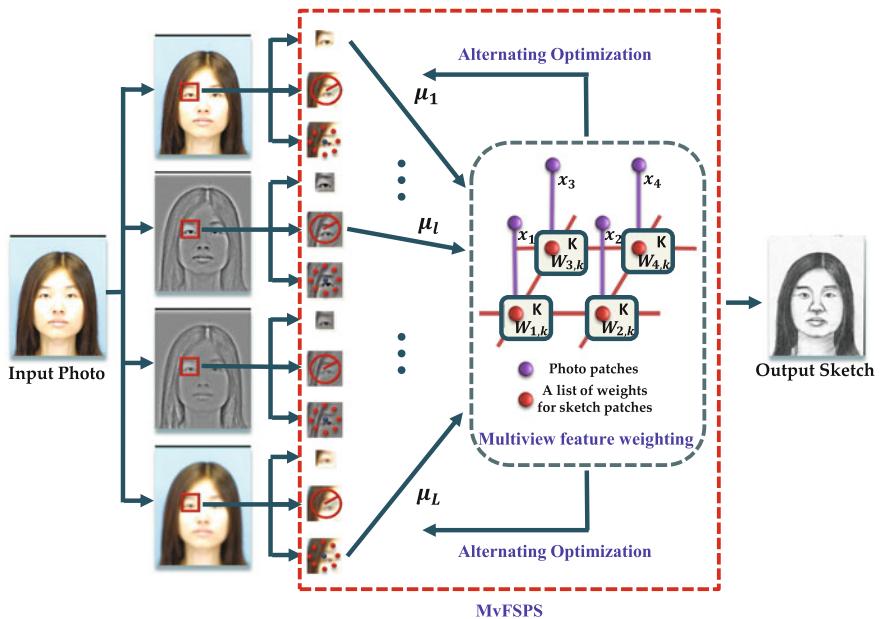
$$\begin{aligned} & p(\mathbf{y}_1, \dots, \mathbf{y}_N, \mathbf{x}_1, \dots, \mathbf{x}_N) \\ & = p(\mathbf{w}_1, \dots, \mathbf{w}_N, \mathbf{x}_1, \dots, \mathbf{x}_N) \\ & = \prod_{(i,j) \in \Xi} \Psi(\mathbf{w}_i, \mathbf{w}_j) \prod_i \Phi(\mathbf{f}(\mathbf{w}_i), \mathbf{f}(\mathbf{x}_i)) \prod_i \exp\{-\lambda_i \|\mu_i\|^2\} \end{aligned} \quad (6.20)$$

where  $\mu_i = [\mu_{i,1}, \mu_{i,2}, \dots, \mu_{i,L}]$  and  $\lambda_i$  balances the regularization term with the other two terms. To obtain the optimal weights for sketch synthesis, we need to maximize the joint probability in the above equation. The optimization problem could be solved in an alternative manner similar as that described in the aforementioned transductive method. The proposed approach is outlined in Fig. 6.7.

## 6.5 Experimental Results

We conduct experiments on the CUHK face sketch database (CUFS) [24]. The CUHK face sketch database consists of 606 sketches corresponding to 606 faces collected from the CUHK student database (including 188 photos), the AR database (including 123 photos) [13], and the XM2VTS database (including 295 photos) [14]. Examples of CUFS database are shown in Fig. 6.8.

For the synthesis purpose, in the CUHK student database, we choose 88 sketch-photo pairs for training and the rest 100 pairs for testing. In the AR database, a leave-one-out likelihood strategy is adopted in which we leave 23 out. In the XM2VTS database, 100 sketch-photo pairs are collected as the training set and the rest as the



**Fig. 6.7** The framework of the proposed multiple representations based face sketch synthesis



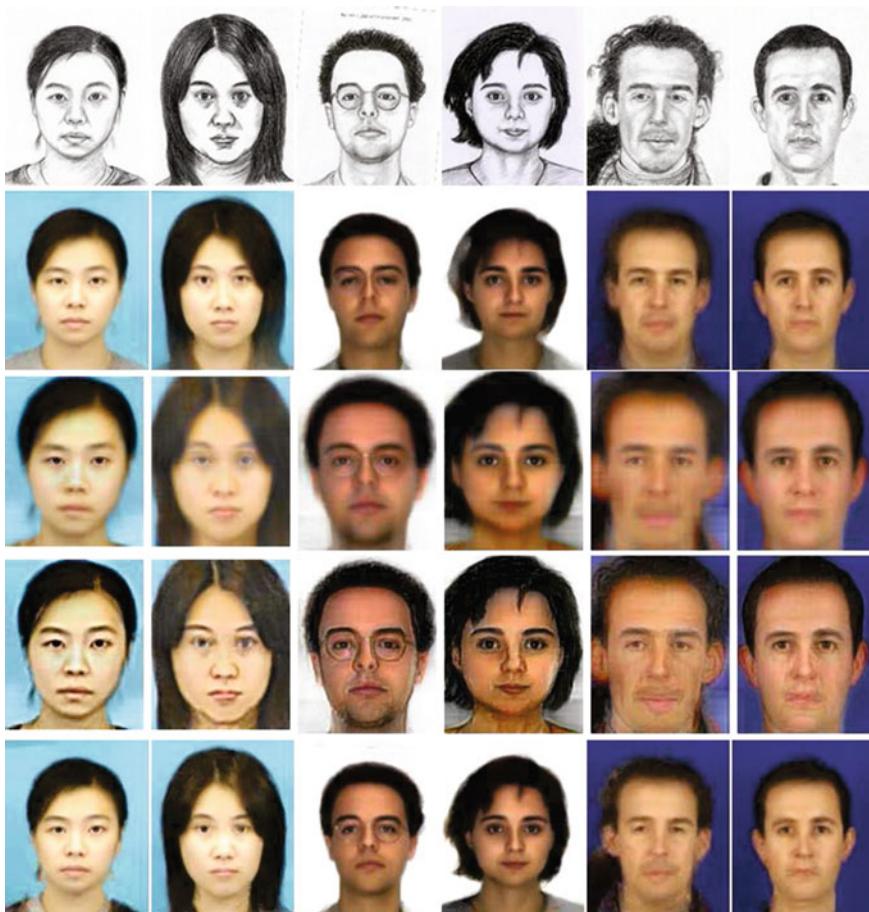
**Fig. 6.8** Examples used in experiments: the two leftmost columns come from the CUHK student database, the second two columns come from the AR database, and the two rightmost columns come from XM2VTS database



**Fig. 6.9** Comparison of the Meitu software, the approaches [6, 15, 21, 28, 32]. Top row: input photos. Second row: results of Meitu software. Third row: results of method [15]. Fourth row: results of method [6]. Fifth row: results of method [20]. Sixth row: results of method [28]. Last row: results of method [21]

test set. As a result, we totally possess 418 synthesized images (photo or sketch) which include 100 from the CUHK student database, 123 from the AR database and 195 from the XM2VTS database. Some synthesized sketches and photos are shown in Figs. 6.9 and 6.10.

We performed face sketch recognition on the three databases together in the following two ways: (a) first transform all face photos in the gallery to sketches using our proposed face sketch synthesis method, and then a query sketch is identified from the synthesized gallery sketches; (b) first transform the query sketch to a photo exploiting the above synthesis method by switching the role of sketches and photos, and subsequently, match the synthesized photo to the gallery photo to identify the



**Fig. 6.10** Comparison of the approaches [6, 15, 21, 32]. Top row: input sketches. Second row: results of method [15]. Third row: results of method [6]. Fourth row: results of method [20]. Last row: results of method [21]

**Table 6.1** Synthesized sketch based rank one recognition accuracy using different synthesis methods and face recognition approaches (%). The number in the bracket is the dimension at which the corresponding recognition method achieves the best recognition accuracy

	Meitu	MrFSS	SFS	SFS-SVR	SRGS	TFSPS
Eigenface [19]	40.24 (173)	74.09 (155)	68.90 (165)	70.43 (178)	46.65 (156)	69.21(123)
NLDA [4]	71.65 (88)	92.38 (86)	86.89 (89)	82.93 (73)	73.78 (86)	83.84 (74)

**Table 6.2** Synthesized photo based rank one recognition accuracy using different synthesis methods and face recognition approaches (%). The number in the bracket is the dimension at which the corresponding recognition method achieves the best recognition accuracy

	MrFSS	SFS	SFS-SVR	TFSPS
Eigenface [19]	41.16 (179)	52.44 (159)	66.46 (164)	56.40 (150)
NLDA [4]	69.82 (87)	64.63 (83)	82.01 (89)	81.10 (84)

person. For strategy (a), we will obtain 418 pairs of pseudo-sketch and corresponding original sketch. These 418 pairs are separated into tow subsets. Subset 1 containing 90 pairs (30 pairs come from above three databases respectively) is used for training the classifiers. Subset 2 consists of the remaining 328 pairs and is taken as the test set. For strategy (b), we will obtain 418 pairs of pseudo-photo and corresponding original photo instead and then perform face sketch recognition similar to strategy (a). We apply two face recognition methods to perform face recognition: Eigenface [19] and Null-space LDA (NLDA) [4]. Tables 6.1 and 6.2 compare the recognition accuracies of different synthesis methods Meitu, [6, 15, 20, 21, 28].

## 6.6 Conclusion

This chapter mainly introduces some data-driven face sketch synthesis methods for sketch-based face recognition. Theoretically, model-based methods can possess a fast generation process because it exploits the learnt mapping function to evaluate the target sketch which inevitably looks less stylistic than the one generated by data-driven methods. Data-driven methods can preserve sketch style well since it directly applies the training samples to synthesize the target sketch. However, the computational time in synthesis- based generation model grows linearly with the amounts of training samples. Among data-driven approaches, graphical representation based methods could achieve more satisfying visual perception than other methods.

**Acknowledgements** This work was supported in part by the National Natural Science Foundation of China (under Grant 61501339 and 61671339).

## References

1. Baker S, Kanade T (2002) Limits on super-resolution and how to break them. *IEEE Trans Pattern Anal Mach Intell* 24(9):1167–1183
2. Chang L, Zhou M, Deng X, Han Y (2011) Face sketch synthesis via multivariate output regression. In: Proceedings of international conference on human-computer interaction, pp 555–561
3. Chang L, Zhou M, Han Y, Deng X (2010) Face sketch synthesis via sparse representation. In: Proceedings of international conference on pattern recognition, pp 2146–2149
4. Chen L, Liao H, Ko M, Lin J, Yu G (2000) A new lda-based face recognition system which can solve the small sample size problem. *Pattern Recognit* 33(10):1713–1726
5. Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans Image Process* 15(12):3736–3745
6. Gao X, Wang N, Tao D, Li X (2012) Face sketch-photo synthesis and retrieval using sparse representation. *IEEE Trans Circuits Syst Video Technol* 22(8):1213–1226
7. Gao X, Zhong J, Li J, Tian C (2008) Face sketch synthesis algorithm using e-hmm and selective ensemble. *IEEE Trans Circuits Syst Video Technol* 18(4):487–496
8. Gao X, Zhong J, Tao D, Li X (2008) Local face sketch synthesis learning. *Neurocomputing* 71(10–12):1921–1930
9. Geng B, Tao D, Xu C, Yang L, Hua X (2012) Ensemble manifold regularization. *IEEE Trans Pattern Anal Mach Intell* 34(6):1227–1233
10. Li X, Cao X (2012) A simple framework for face photo-sketch synthesis. *Math Probl Eng*:1–19
11. Liu Q, Tang X, Jin H, Lu H, Ma S (2005) A nonlinear approach for face sketch synthesis and recognition. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1005–1010
12. Mairal J, Sapiro G, Elad M (2008) Learning multiscale sparse representations for image and video restoration. *SIAM Multiscale Model Simul* 17:214–241
13. Martinez A, Benavente R (1998) The ar face database. Technical report, CVC Technical Report #24
14. Messer K, Matas J, Kittler J, Luettin J, Maitre G (1999) Xm2vtsdb: the extended m2vts database. In: Proceedings of international conference on audio- and video-based biometric person authentication, pp 72–77
15. Peng C, Gao X, Wang N, Tao D, Li X, Li J (2015) Multiple representations based face sketch-photo synthesis. *IEEE Trans Neural Netw Learn Syst*:1–13
16. Tang X, Wang X (2002) Face photo recognition using sketches. In: Proceedings of IEEE international conference on image processing, pp 257–260
17. Tang X, Wang X (2003) Face sketch synthesis and recognition. In: Proceedings of IEEE international conference on computer vision, pp 687–694
18. Tang X, Wang X (2004) Face sketch recognition. *IEEE Trans Circuits Syst Video Technol* 14(1):1–7
19. Turk M, Pentland A (1991) Face recognition using eigenfaces. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 586–591
20. Wang N, Li J, Tao D, Li X, Gao X (2013) Heterogeneous image transformation. *Pattern Recognit Lett* 34(1):77–84
21. Wang N, Tao D, Gao X, Li X, Li J (2013) Transductive face sketch-photo synthesis. *IEEE Trans Neural Netw Learn Syst* 24(9):1364–1376
22. Wang N, Tao D, Gao X, Li X, Li J (2014) A comprehensive survey to face hallucination. *Int J Comput Vision* 106(1):9–30
23. Wang S, Zhang L, Liang Y, Pan Q (2012) Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 2216–2223
24. Wang X, Tang X (2009) Face photo-sketch synthesis and recognition. *IEEE Trans Pattern Anal Mach Intell* 31(11):1955–1967
25. Xiao B, Gao X, Tao D, Li X (2009) A new approach for face recognition by sketches in photos. *Signal Process* 89(8):1576–1588

26. Xiao B, Gao X, Tao D, Yuan Y, Li J (2010) Photo-sketch synthesis and recognition based on subspace learning. *Neurocomputing* 73(4–6):840–852
27. Yang J, Wright J, Huang T, Ma, Y.: Image super-resolution as sparse representation of raw image patches. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1–8
28. Zhang S, Gao X, Wang N, Li J, Zhang M (2015) Face sketch synthesis via sparse representation-based greedy search. *IEEE Trans Image Process* 24(8):2466–2477
29. Zhang W, Wang X, Tang X (2010) Lighting and pose robust face sketch synthesis. In: Proceedings of European conference on computer vision, pp 420–423
30. Zhao W, Chellappa R, Phillips P, Rosenfeld A (2003) Face recognition: a literature survey. *ACM Comput Surv* 35(4):399–458
31. Zhong J, Gao X, Tian C (2007) Face sketch synthesis using e-hmm and selective ensemble. In: Proceedings of IEEE international conference on acoustics, speech, and signal processing, pp 485–488
32. Zhou H, Kuang Z, Wong K (2012) Markov weight fields for face sketch synthesis. In: Proceedings of IEEE conference on computer vision and pattern recognition, pp 1091–1097

# Chapter 7

## Recent Developments in Video-Based Face Recognition

Jingxiao Zheng, Vishal M. Patel and Rama Chellappa

**Abstract** Face recognition with its wide range of commercial and law enforcement applications has been one of the most active areas of research in the field of computer vision and pattern recognition. Personal identification systems based on faces have the advantage that facial images can be obtained from a distance without requiring cooperation of the subject, as compared to other biometrics such as fingerprint, iris, etc. Face recognition is concerned with identifying or verifying one or more persons from still images or video sequences using a stored database of faces. Depending on the particular application, there can be different scenarios, ranging from controlled still images to uncontrolled videos. Since face recognition is essentially the problem of recognizing a 3D object from its 2D image or a video sequence, it has to deal with significant appearance changes due to illumination and pose variations. Current algorithms perform well in controlled scenarios, but their performance is far from satisfactory in uncontrolled scenarios. Most of the current research in this area is focused toward recognizing faces in uncontrolled scenarios. This chapter presents an overview of recent video-based face recognition methods. In particular, recent sparse coding-based, manifold-based, probabilistic, geometric model-based, and dynamic model-based methods are reviewed.

---

J. Zheng · R. Chellappa (✉)  
Center for Automation Research, University of Maryland,  
College Park, MD 20742, USA  
e-mail: rama@umiacs.umd.edu

J. Zheng  
e-mail: jxzheng@umiacs.umd.edu

V.M. Patel  
Rutgers, The State University of New Jersey,  
94 Brett Road, Piscataway, NJ 08854, USA  
e-mail: vishal.m.patel@rutgers.edu

## 7.1 Introduction

Video-based face recognition has received a significant amount of attention in recent years. This is mainly due to the fact that large amounts of video data are becoming available everyday. Millions of cameras have been installed in buildings, streets, and airports around the world, and people are using billions of handheld devices that are capable of capturing videos. As a result, 350 million photos are uploaded to Facebook every day and 100 h of video are uploaded to YouTube each minute.

For video-based face recognition problem, the identification and verification tasks are all based on videos rather than still images compared to the classical image-based face recognition problem. Approaches for video-based face recognition need to identify a person in a video, given some possible candidates, or to decide whether the two people in two different videos are the same person.

In most of the video-based face recognition methods, given video data, tracking algorithms like [38] are first used to detect faces in the video frames. Then fiducial extraction methods like [47] are applied to align the detected faces. After the alignment, traditional feature extraction techniques such as SIFT [30], Hog [14], LBP [31] or the very popular DCNN features [26, 35, 36] are used to extract features for matching.

In video-based face recognition, a key challenge is in exploiting the extra information available in a video, e.g., face, body, and motion identity cues. In addition, different video sequences of the same subject may contain variations in resolution, illumination, pose, and facial expressions. These variations contribute to the challenges in designing an effective video-based face recognition algorithm. Whether the temporal information is considered or not, most video-based face recognition can be divided into sequence-based methods or set-based methods. Sequence-based face recognition methods consider the video as a sequence of images and make use of the temporal information for recognition. On the other hand, set-based face recognition methods only consider the video as a set of images and ignore their order.

Besides using temporal information, video-based face recognition can also be sorted by the techniques used to model the video. These include sparse coding-based methods, manifold-based methods, probabilistic methods, geometrical model-based methods, and dynamical model-based methods. In this chapter, we give an overview of some of these modeling approaches.

## 7.2 Sparse Coding-Based Methods

For sparse coding-based methods, faces (or features extracted from faces) in videos are modeled as dictionaries, which are overcomplete atoms learned from the training data with sparsity constraints.

Given  $L$  video frames with faces of dimension  $M$  concatenated in a matrix  $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_L] \in \mathbb{R}^{M \times L}$ , the problem of learning a dictionary, which minimizes

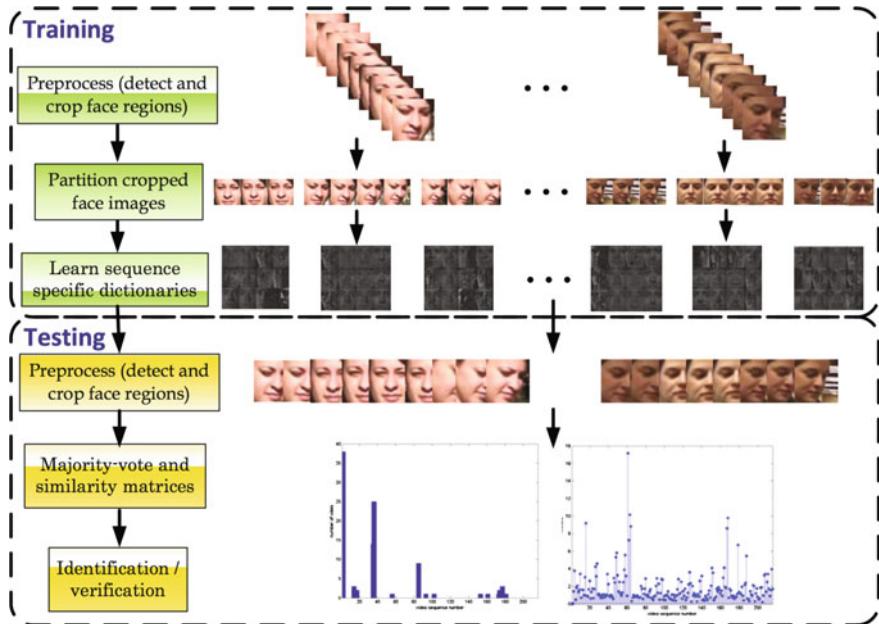


Fig. 7.1 Dictionary-based face recognition from video [12]

the representation error with a sparseness constraint is equivalent to solving the following optimization problem

$$\min_{\mathbf{D}, \mathbf{X}} \|\mathbf{Y} - \mathbf{DX}\|_F^2 \text{ s.t. } \|\mathbf{x}_i\|_0 \leq T, \quad \mathbf{d}_i^T \mathbf{d}_i = 1 \quad \forall i, \quad (7.1)$$

where  $\|\cdot\|_F$  is the Frobenius norm,  $\|\mathbf{x}\|_0$  is the  $\ell_0$  norm of  $\mathbf{x}$  which counts the number of nonzero elements in  $\mathbf{x}$ ,  $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_S] \in \mathbb{R}^{M \times S}$  is the dictionary,  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in \mathbb{R}^{S \times N}$  is the corresponding collection of sparse coefficients,  $S$  is the number of atoms in the dictionary, and  $T$  is a sparsity parameter. Because of the sparsity constraint, the learned dictionaries are robust to different kinds of variations in video sequences.

[12] proposed a generative dictionary learning method for video-based face recognition. The main idea of the method is to partition the video frames into clusters with different poses and illuminations and learn a set of sub-dictionaries for each cluster. Then the concatenation of the sub-dictionaries removes the temporal redundancy in the videos and can handle large variations on poses and illumination variations. An overview of this method is shown in Fig. 7.1.

For each frame in a video sequence, the face regions are first detected and cropped. Then all the cropped face images are partitioned into  $K$  different partitions by a  $K$ -means clustering type of algorithm. For each partition, a dictionary is learned with the minimum representation error under a sparseness constraint using (7.1).

Thus, there will be  $K$  sub-dictionaries built to represent a video sequence. Then the video sequence-specific dictionary is constructed by concatenating these partition-level sub-dictionaries as  $\mathbf{D}_p = [\mathbf{D}_p^1, \mathbf{D}_p^2, \dots, \mathbf{D}_p^K]$ . Due to changes in pose and lighting in a video sequence, the number of face images in a partition will vary. Those partitions with very few images will be augmented by synthesized face images. This is done by creating horizontally, vertically, or diagonally position shifted face images, or by in-plane rotated face images.

For identification task, testing videos are partitioned into  $K$  partitions as well. Given a testing frame  $\mathbf{q}_{l,k}$  from the  $k$ th partition, the frame-level decision  $\hat{p}_{l,k}$  is the sequence  $p$  with the minimum residual error from its projection onto the subspace spanned by  $\mathbf{D}_p$  as

$$\hat{p}_{l,k} = \operatorname{argmin}_p \|\mathbf{q}_{l,k} - \mathbf{D}_p \mathbf{D}_p^\dagger \mathbf{q}_{l,k}\|_2 \quad (7.2)$$

The sequence-level decision  $\hat{p}$  is then the weighted sum of votes from  $K$  partitions as

$$\hat{p} = \operatorname{argmax}_i \sum_{k=1}^K w_k \sum_l \mathbf{1}\{\hat{p}_{l,k} = i\} \quad (7.3)$$

For verification task, given a query video sequence  $m$  and gallery video sequence  $p$  (with learned dictionary  $\mathbf{D}_p$ ), the similarity score is

$$\mathbf{R}^{m,p} = \min_k \min_l \|\mathbf{q}_{l,k}^m - \mathbf{D}_p \mathbf{D}_p^\dagger \mathbf{q}_{l,k}^m\|_2. \quad (7.4)$$

which is the minimum residual among all  $l$  and all  $k$ , between the frames from query video sequence  $m$  and gallery dictionary  $\mathbf{D}_p$ .

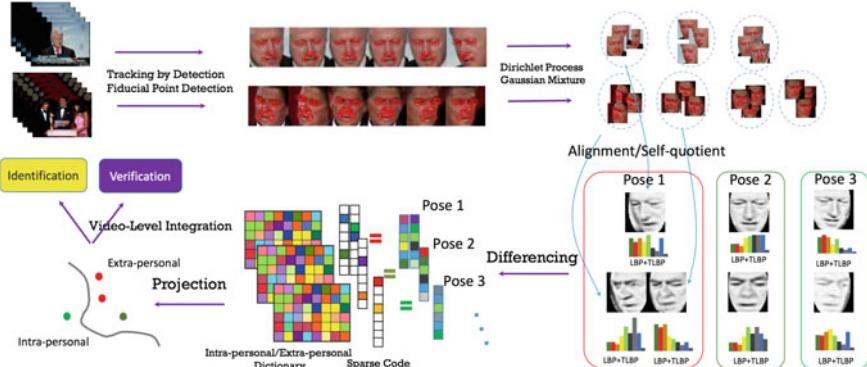
[11] further introduced the joint sparsity constraints into their dictionary learning algorithm. Given video frames sets  $\{\mathbf{Y}^k\}$ , instead of learning dictionaries from each frame partition independently as

$$\min_{\mathbf{D}^k, \mathbf{X}^k} \|\mathbf{Y}^k - \mathbf{D}^k \mathbf{X}^k\|_F^2 \text{ s.t. } \|\mathbf{x}_i^k\|_0 \leq T, \quad \mathbf{d}_i^{kT} \mathbf{d}_i^k = 1 \quad \forall i \quad (7.5)$$

based on the joint sparse constraints, the dictionaries are learned jointly as

$$\min_{\{\mathbf{D}^k\}, \mathbf{X}} \frac{1}{2} \sum_{k=1}^K \|\mathbf{Y}^k - \mathbf{D}^k \mathbf{X}^k\|_F^2 + \lambda \|\mathbf{X}\|_{1,2} \text{ s.t. } \mathbf{d}_i^{kT} \mathbf{d}_i^k = 1 \quad \forall i \quad (7.6)$$

where  $\|\mathbf{X}\|_{1,2} = \sum_{i=1}^d \|\mathbf{x}_i\|_2$  is the sparse constraint on  $\mathbf{X} = [\mathbf{X}^1, \dots, \mathbf{X}^K]$ . It enforces the sparse pattern for each column of  $\mathbf{X}$  to be similar, which makes the learned dictionaries more robust to noise and occlusion. [11] also introduced a kernel version of their algorithm to deal with those non-linearly separable cases and improve the performance.



**Fig. 7.2** Video-based face recognition using the intra/extrapersonal difference dictionary [16]

Du and Chellappa [16] proposed a video-based face recognition method based on intra/extrapersonal difference dictionary. Since pose variations often cause within-class variance to exceed between-class variance in face recognition, instead of learning dictionaries from the face features directly, pose-specific dictionaries are learned from those intra/extrapersonal difference features. Also, instead of learning generative dictionaries by merely minimizing the reconstruction error, it jointly learns dictionaries and discriminative projection matrices, which improves performance. The overall algorithm is shown in Fig. 7.2.

In their algorithm, given a video  $\mathbf{V}$ , faces are first detected and cropped from the videos by using a tracking algorithm. Fiducial points are then detected by a structural SVM approach. These cropped faces are aligned and clustered by the  $K$ -means algorithm into  $K$  clusters according to their poses. Then a given video can be characterized by its  $K$  cluster centers  $\{\mathbf{v}_k, k = 1, 2, \dots, K\}$  considered as representative images.

For the training videos, the intrapersonal difference features  $\{\mathbf{x}_{In} = \mathbf{v}_i^m - \mathbf{v}_j^n, ID(\mathbf{V}_i) = ID(\mathbf{V}_j)\}$  and the extrapersonal ones  $\{\mathbf{x}_{Ex} = \mathbf{v}_i^m - \mathbf{v}_j^n, ID(\mathbf{V}_i) \neq ID(\mathbf{V}_j)\}$  are employed to learn the dictionary  $\mathbf{D}$  and the projection matrix  $\mathbf{W}$  simultaneously for each pair of poses by solving the following Label-Consistent K-SVD problem (LC-K-SVD):

$$\min_{\mathbf{D}, \mathbf{A}} \|\mathbf{X} - \mathbf{DA}\|_2^2 + \mu \|\mathbf{Q} - \mathbf{BA}\|_2^2 + \sigma \|\mathbf{F} - \mathbf{WA}\|_2^2 + \lambda \sum_i \|\boldsymbol{\alpha}_i\|_1, \quad (7.7)$$

where  $\mathbf{X} = [\mathbf{X}_{In} \ \mathbf{X}_{Ex}]$  is the concatenation of intrapersonal and extrapersonal features. The columns of  $\mathbf{F} \in \mathbb{R}^{2 \times N}$  are the corresponding labels (same or different), represented using the 1-of- $K$  coding scheme. It enforces  $\mathbf{W}$  to encode discriminative information from the sparse codes.  $\mathbf{B} \in \mathbb{R}^{K \times d}$  is a linear transformation that encourages the samples from the same class to be reconstructed using the entries in the sub-dictionary of that class.  $\mathbf{Q} \in \mathbb{R}^{K \times N}$  has a block diagonal form: The  $c$ -th block contains entry  $\mathbf{Q}_{ij}$ ,  $i \in v_c, j \in h_c$ , where  $v_c$  are the indices of atoms from class  $c$

(i.e., intrapersonal or extrapersonal) and  $h_c$  are the indices of training instances from class  $c$ . All the nonzero entries in  $\mathbf{Q}$  are assigned with unit value. This problem can be converted to a typical K-SVD [3] objective function and solved using the same procedure.

At the testing stage, for every probe-gallery video pair  $\{\mathbf{V}_p, \mathbf{V}_g\}$ , feature difference vectors  $\{\mathbf{x}_{p,g}^{m,n} = \mathbf{v}_p^m - \mathbf{v}_g^n\}$  from each pair of poses are calculated. The sparse representation of  $\mathbf{x}_{p,g}^{m,n}$  is obtained by solving  $\boldsymbol{\alpha}_{p,g}^{m,n} = \operatorname{argmin}_{\boldsymbol{\alpha}} \sum_{i=1}^N \frac{1}{2} \|\mathbf{x}_{p,g}^{m,n} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1$  using the learned dictionary  $\mathbf{D}$  in the training stage. The similarity score for this video pair is then calculated as

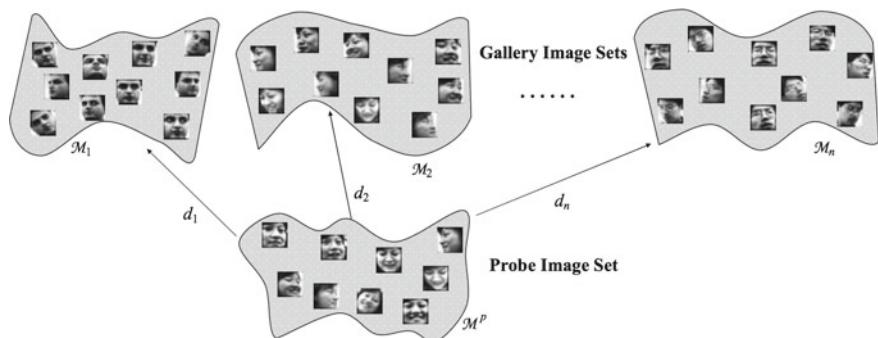
$$s(p, g) = \sum_{m=1}^M \sum_{n=1}^N \mathbf{1}(\mathbf{t}_1 \mathbf{W} \boldsymbol{\alpha}_{p,g}^{m,n} > \mathbf{t}_0 \mathbf{W} \boldsymbol{\alpha}_{p,g}^{m,n}) / MN \quad (7.8)$$

where  $\mathbf{t}_0 = [0, 1]^T$  and  $\mathbf{t}_1 = [1, 0]^T$  are the 1-of- $K$  coding label for intrapersonal and extrapersonal class, respectively. For video-based recognition, the decision is made by  $ID(\mathbf{V}_p) = \operatorname{argmax}_g s(p, g)$ .

Some of the other sparse dictionary learning-based methods for video-based face recognition include [18, 32].

### 7.3 Manifold-Based Methods

In manifold-based methods, videos are usually modeled as image sets. These image sets are considered as the approximation of manifolds and the problem actually turns into looking for a discriminant distance metric between manifolds. The basic idea is shown in Fig. 7.3.



**Fig. 7.3** Manifold-based face recognition [41]

In [41], the image set classification problem is based on the computation of manifold–manifold distance. It models the image sets cropped from videos as manifolds which consist of component linear subspaces. Then the manifold-to-manifold distance can be considered as the similarity between two videos.

Given face image set  $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$  from a video, it is partitioned into a collection of disjoint Maximal Linear Patches  $\{C_i\}$ . Each video is considered as a manifold consisting of these local linear patches which can be obtained by Algorithm 1.

---

**Algorithm 1** Local model construction.

---

1. Initialize that  $i = 1, C_i = \emptyset, X_T = \emptyset, X_R = X$ .
  2. While( $X_R \neq \emptyset$ )
    - 2.1 Randomly select a seed point from  $X_R$  as  $\mathbf{x}_1^{(i)}$ , update  $C_i = \{\mathbf{x}_1^{(i)}\}, X_R = X_R - \{\mathbf{x}_1^{(i)}\}$ .
    - 2.2 For ( $\forall \mathbf{x}_m^{(i)} \in C_i$ )
 

Identify each of its  $k$ -NNs  $\mathbf{x}_c$  as *candidate*. If  $\mathbf{x}_c$  satisfies simultaneously  $\mathbf{x}_c \in X_R$  and

$$D_G(\mathbf{x}_c, \mathbf{x}_n^{(i)})/D_E(\mathbf{x}_c, \mathbf{x}_n^{(i)}) < \theta, \forall \mathbf{x}_n^{(i)} \in C_i \quad (7.9)$$

then update  $C_i = C_i \cup \{\mathbf{x}_c\}, X_R = X_R - \{\mathbf{x}_c\}$ , until no candidate point can be added into  $C_i$ .
    - 2.3  $X_T = \bigcup_{j=1}^i C_j, X_R = X - X_T, i = i + 1, C_i = \emptyset$ .
- 

Here,  $D_E(\cdot)$  denotes the Euclidean distance and  $D_G(\cdot)$  denotes the geodesic distance. Their ratio reflects the linear deviation of the local linear subspaces. Threshold  $\theta$  controls the degree of linear deviation. Thus larger  $\theta$  implies fewer local structures but large linear deviation in each structure. After obtaining the local linear subspaces for each video, the manifold-manifold distance between two video manifolds  $M_1$  and  $M_2$  can be computed as

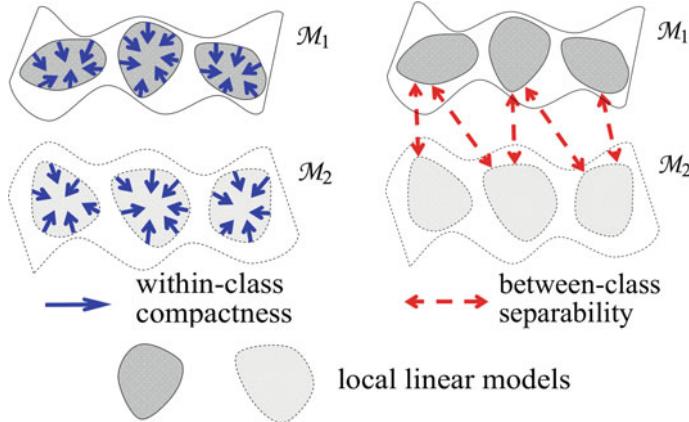
$$d(M_1, M_2) = \min_{C_i \in M_1} \min_{C_j \in M_2} d(C_i, C_j) \quad (7.10)$$

which is the distance between the closest local subspace pair.

Suppose  $\mathbf{e}_i, \mathbf{e}_j$  and  $\mathbf{P}_i \in \mathbb{R}^{D \times d_1}, \mathbf{P}_j \in \mathbb{R}^{D \times d_2}$  are the exemplars (means) and orthonormal bases of two subspaces  $C_i$  and  $C_j$ .  $r = \min(d_i, d_j)$ . The SVD of  $\mathbf{P}_1^T \mathbf{P}_2$  is  $\mathbf{P}_1^T \mathbf{P}_2 = \mathbf{Q}_{12} \mathbf{\Lambda} \mathbf{Q}_{21}^T$  and  $\mathbf{\Lambda} = \text{diag}(\sigma_1, \dots, \sigma_r)$ . The distance between two local subspace is defined as

$$d(C_i, C_j) = (1 - \alpha)d_E(C_i, C_j) + \alpha d_V(C_i, C_j). \quad (7.11)$$

Here,  $d_E(C_i, C_j) = \|\mathbf{e}_i\| \|\mathbf{e}_j\| / \mathbf{e}_i^T \mathbf{e}_j$  is called the *exemplar distance measure*, which measures how similar the two sets are.  $d_V(C_i, C_j) = r / \sum_{k=1}^r \sigma_k$  is called the *variation distance measure* which measures how close the common variation modes of the two sets. By fusing these distance measures, the overall manifold–manifold



**Fig. 7.4** Manifold discriminant analysis [39]

distance captures the difference of both average appearance and variation information between two sets.

Finally, for verification task, the similarity score between any gallery and probe video pair is the manifold–manifold distance between their corresponding manifolds. For identification task, decision is made by finding the video with the minimum manifold–manifold distance.

A manifold-based discriminative learning method called Manifold Discriminant Analysis for image set classification was proposed in [39]. It learns an embedding space where manifolds with different class labels are better separated and local data compactness within each manifold is enhanced. An overview of this method is shown in Fig. 7.4.

Like [41], given image sets considered as manifolds, local linear models are first extracted as  $M_i = \{C_{i,k}\}$ . The learning method is formulated as:

1. Two graphs are constructed, which are intrinsic graph  $G$  and penalty graph  $G'$ . In both graphs, nodes are all the images in the training set  $\mathbf{X} = \{\mathbf{x}_m\}$ . In  $G$ , nodes  $\mathbf{x}_m$  and  $\mathbf{x}_n$  are connected if  $\mathbf{x}_m \in C_{i,k}, \mathbf{x}_n \in C_{j,l}, i = j$  and  $k = l$ , which means only the nodes come from the same local linear model are connected. In  $G'$ , nodes  $\mathbf{x}_m$  and  $\mathbf{x}_n$  are connected if their class labels are different and  $C_{i,k}$  is among the  $k'$ -nearest between-class neighbors of  $C_{j,l}$ .
2. The weight matrix  $\mathbf{W} = \{w_{mn}\}$  for  $G$  is computed as

$$w_{mn} = \begin{cases} 1 & \text{if } \mathbf{x}_m \text{ and } \mathbf{x}_n \text{ are connected} \\ 0 & \text{otherwise} \end{cases} \quad (7.12)$$

$\mathbf{W}'$  for  $G'$  is computed in the same way.  $\mathbf{D}$  and  $\mathbf{D}'$  are diagonal matrices with diagonal elements  $d_{mm} = \sum_n w_{mn}$  and  $d'_{mm} = \sum_n w'_{mn}$ .  $\mathbf{L}_w = \mathbf{D} - \mathbf{W}$  and  $\mathbf{L}_b = \mathbf{D}' - \mathbf{W}'$  are their Laplacian matrices, respectively.

3. A linear embedding  $\mathbf{z} = \mathbf{V}^T \mathbf{x}$  based on linear projection is learned, where  $\mathbf{V} \in \mathbb{R}^{d \times l}$  with  $l \ll d$ . For each column of  $\mathbf{V}$ , learning is fulfilled by maximizing the between-class scatter  $S_b = \sum_{m,n} \|\mathbf{v}^T \mathbf{x}_m - \mathbf{v}^T \mathbf{x}_n\|^2 w'_{m,n} = 2\mathbf{v}^T \mathbf{XL}_b \mathbf{X}^T \mathbf{v}$  and minimizing the within-class scatter  $S_w = \sum_{m,n} \|\mathbf{v}^T \mathbf{x}_m - \mathbf{v}^T \mathbf{x}_n\|^2 w_{m,n} = 2\mathbf{v}^T \mathbf{XL}_w \mathbf{X}^T \mathbf{v}$ . This is equivalent to solving the optimization problem:

$$\underset{\mathbf{v}}{\text{maximize}} = \frac{S_b}{S_w} = \frac{\mathbf{v}^T \mathbf{XL}_b \mathbf{X}^T \mathbf{v}}{\mathbf{v}^T \mathbf{XL}_w \mathbf{X}^T \mathbf{v}} \quad (7.13)$$

The columns of the optimal  $\mathbf{V}$  are the generalized eigenvectors corresponding to the  $l$  largest eigenvalues in

$$\mathbf{XL}_b \mathbf{X}^T \mathbf{v} = \lambda \mathbf{XL}_w \mathbf{X}^T \mathbf{v}. \quad (7.14)$$

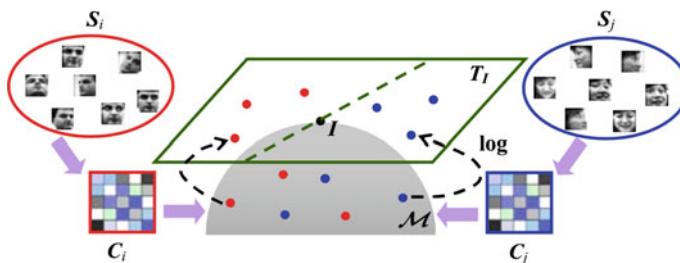
Finally, for verification task, given two manifolds  $M_k$  and  $M_l$ , their distance is calculated as  $d(M_k, M_l) = \min_{i,j} d(C_{i,k}, C_{j,l})$ , which is same as the manifold-to-manifold distance proposed in [41].  $d(C_{i,k}, C_{j,l}) = \|\mathbf{e}_{i,k} - \mathbf{e}_{j,l}\|$  is the *empirical distance* between each pair of local linear models, where  $\mathbf{e}_{i,k} = \frac{1}{N_{i,k}} \sum_{n=1}^{N_{i,k}} \mathbf{V}^T \mathbf{x}_{i,k}^n$  is the sample mean of  $C_{i,k}$  and  $\mathbf{e}_{j,l} = \frac{1}{N_{j,l}} \sum_{n=1}^{N_{j,l}} \mathbf{V}^T \mathbf{x}_{j,l}^n$  is the sample mean of  $C_{j,l}$ , both in the learned embedding space.

Similarly, Wang et al. [40] proposed a discriminative learning approach for image set classification by modeling the image set using its covariance matrix. The conceptual illustration of this method is shown in Fig. 7.5.

Given face images from a video,  $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n]$ , the samples covariance of this image set is

$$\mathbf{C} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{s}_i - \bar{\mathbf{s}})(\mathbf{s}_i - \bar{\mathbf{s}})^T \quad (7.15)$$

where  $\bar{\mathbf{s}}$  is the sample mean. The video is thus characterized by its covariance matrix  $\mathbf{C}$ . Since  $\mathbf{C}$  is an SPD matrix, it lies on a Riemannian manifold. It is not easy to train a classifier on the manifold because most of the classic classifiers are



**Fig. 7.5** Covariance discriminative learning [40]

designed for Euclidean metrics. In the paper, a distance metric, Log-Euclidean distance (LED) is introduced as  $d_{LED}(\mathbf{C}_1, \mathbf{C}_2) = \|\log(\mathbf{C}_1) - \log(\mathbf{C}_2)\|_F$  where  $\log(\cdot)$  here is the ordinary matrix logarithm operator. If  $\mathbf{C} = \mathbf{U}\Sigma\mathbf{U}^T$  is an SPD matrix,  $\log(\mathbf{C}) = \mathbf{U}\log(\Sigma)\mathbf{U}^T$ .

Given training videos  $\{\mathbf{S}_i^{tr}\}$  from  $C$  different classes, first the covariance matrices  $\{\mathbf{C}_i^{tr}\}$  are calculated. Then, two different learning methods are used:

### 1. Kernel LDA

The KLDA optimization problem is:

$$\boldsymbol{\alpha}_{opt} = \underset{\boldsymbol{\alpha}}{\operatorname{argmax}} \frac{\boldsymbol{\alpha}^T \mathbf{K} \mathbf{W} \mathbf{K} \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \mathbf{K} \mathbf{K} \boldsymbol{\alpha}}, \quad (7.16)$$

where  $\mathbf{K}_{ij} = k(\mathbf{S}_i^{tr}, \mathbf{S}_j^{tr}) = d_{LED}(\mathbf{C}_i^{tr}, \mathbf{C}_j^{tr})$ . And  $\mathbf{W}$  is defined as:

$$\mathbf{W}_{ij} = \begin{cases} 1/n_k & \text{if } \mathbf{S}_i^{tr}, \mathbf{S}_j^{tr} \text{ are both in the } k\text{th class} \\ 0 & \text{otherwise} \end{cases} \quad (7.17)$$

and  $n_k$  is the number of videos in the  $k$ th class.

The solution to (7.16) is the eigenvector corresponding to the largest eigenvalue of the problem  $\mathbf{K} \mathbf{W} \mathbf{K} \boldsymbol{\alpha} = \lambda \mathbf{K} \mathbf{K} \boldsymbol{\alpha}$ . Then given a testing video  $\mathbf{S}^{te}$  with its covariance matrix  $\mathbf{S}^{te}$ , its projection in the  $C - 1$  dimensional discriminant subspace is:

$$\mathbf{z}^{te} = \mathbf{A}^T \mathbf{K}^{te} \quad (7.18)$$

where  $\mathbf{A} = [\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_{C-1}]$  is the collection of  $C - 1$  largest eigenvectors and  $\mathbf{K}^{te} = [k(\mathbf{S}_1^{tr}, \mathbf{S}^{te}), k(\mathbf{S}_2^{tr}, \mathbf{S}^{te}), \dots]^T$ . Nearest Neighbor classification in the discriminant subspace based on Euclidean distance is then performed.

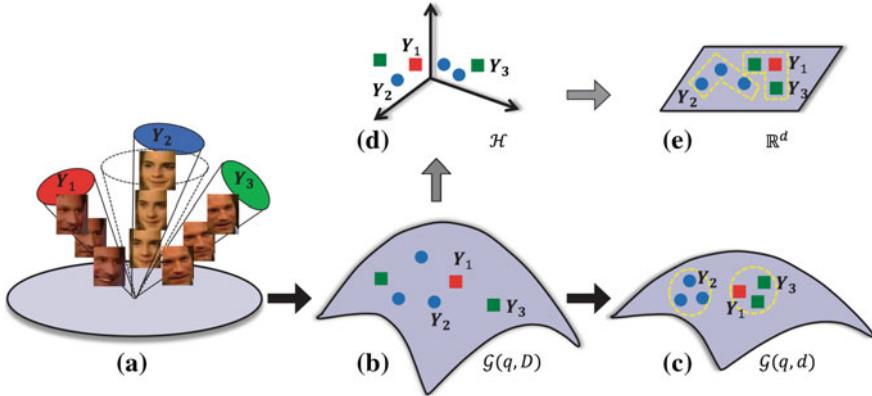
### 2. Kernel PLS

Different from KLDA, KPLS directly learns a regression model between training observations  $\{\mathbf{S}_i^{tr}\}$  and their 1-of- $K$  coding labels  $\mathbf{Y}^{tr}$  (refer to [34] for more details). Then given testing video  $\mathbf{S}^{te}$ , its KPLS prediction is given by

$$\mathbf{y}^{te} = \mathbf{K}^{teT} \mathbf{U} (\mathbf{T}^T \mathbf{K} \mathbf{U})^{-1} \mathbf{T}^T \mathbf{Y}^{tr} \quad (7.19)$$

where  $\mathbf{U}$  and  $\mathbf{T}$  are regression parameters learned by KPLS,  $\mathbf{K}$  and  $\mathbf{K}^{te}$  are the same as in KLDA. The entry index with the largest response in  $\mathbf{y}^{te}$  determines the label of the video.

Furthermore, [22] introduced a method that learns the projection metric directly on the Grassmann manifold rather than in Hilbert space. It performs a geometry-aware dimensionality reduction from the original Grassmann manifold to a lower dimensional, more discriminative Grassmann manifold. The method is demonstrated in Fig. 7.6.



**Fig. 7.6** Projection metric learning on Grassmann manifold [22]

Given face frames \$\{\mathbf{X}\_i\}\$ from videos where \$\mathbf{X}\_i \in \mathbb{R}^{D \times n\_i}\$ describes a data matrix of the \$n\_i\$ frames from the \$i\$th video, \$\mathbf{X}\_i\$ is represented by a \$q\$-dimensional linear subspace spanned by an orthonormal basis matrix \$\mathbf{Y}\_i \in \mathbb{R}^{D \times q}\$. This is calculated by \$\mathbf{X}\_i \mathbf{X}\_i^T \simeq \mathbf{Y}\_i \Lambda\_i \mathbf{Y}\_i^T\$, \$\Lambda\_i\$ and \$\mathbf{Y}\_i\$ correspond to the \$q\$ largest eigenvalues and eigenvectors, respectively.

The linear subspace span(\$\mathbf{Y}\_i\$) lies on a Grassmann manifold \$\mathcal{G}(q, D)\$. It can be represented by the projection mapping \$\Phi(\mathbf{Y}\_i) = \mathbf{Y}\_i \mathbf{Y}\_i^T\$ since there is a one-to-one mapping between each projection matrix and the point on the Grassmann manifold. The projection distance metric between \$\mathbf{Y}\_i \mathbf{Y}\_i^T\$ and \$\mathbf{Y}\_j \mathbf{Y}\_j^T\$ is defined as

$$d_p(\mathbf{Y}_i \mathbf{Y}_i^T, \mathbf{Y}_j \mathbf{Y}_j^T) = 2^{-1/2} \|\mathbf{Y}_i \mathbf{Y}_i^T - \mathbf{Y}_j \mathbf{Y}_j^T\|_F. \quad (7.20)$$

The method learns a mapping \$f : \mathcal{G}(q, D) \rightarrow \mathcal{G}(q, d)\$ which is defined as

$$f(\mathbf{Y}_i \mathbf{Y}_i^T) = \mathbf{W}^T \mathbf{Y}_i \mathbf{Y}_i^T \mathbf{W} = (\mathbf{W}^T \mathbf{Y}_i)(\mathbf{W}^T \mathbf{Y}_i)^T \quad (7.21)$$

where \$\mathbf{W} \in \mathbb{R}^{D \times d}\$ is the column full rank transformation matrix. Here, \$\mathbf{W}^T \mathbf{Y}\_i\$ is not a orthonormal basis in general, which doesn't lie on a Grassmann manifold. Thus, \$\mathbf{W}^T \mathbf{Y}\_i\$ is replaced by \$\mathbf{W}^T \mathbf{Y}'\_i\$, which is an orthonormal basis of \$\mathbf{W}^T \mathbf{Y}\_i\$.

After transformation, the projection distance between \$\mathbf{W}^T \mathbf{Y}'\_i \mathbf{Y}'\_i^T \mathbf{W}\$ and \$\mathbf{W}^T \mathbf{Y}'\_j \mathbf{Y}'\_j^T \mathbf{W}\$ is

$$d_p^2(\mathbf{W}^T \mathbf{Y}'_i \mathbf{Y}'_i^T \mathbf{W}, \mathbf{W}^T \mathbf{Y}'_j \mathbf{Y}'_j^T \mathbf{W}) = \frac{1}{2} \|\mathbf{W}^T \mathbf{Y}'_i \mathbf{Y}'_i^T \mathbf{W} - \mathbf{W}^T \mathbf{Y}'_j \mathbf{Y}'_j^T \mathbf{W}\|_F^2 = \frac{1}{2} \text{tr}(\mathbf{P} \mathbf{A}_{ij} \mathbf{A}_{ij}^T \mathbf{P}), \quad (7.22)$$

where \$\mathbf{A}\_{ij} = \mathbf{Y}'\_i \mathbf{Y}'\_i^T - \mathbf{Y}'\_j \mathbf{Y}'\_j^T\$ and \$\mathbf{P} = \mathbf{W} \mathbf{W}^T\$, which is a rank-\$d\$ \$D \times D\$ PSD matrix.

The method learns \$\mathbf{P}\$ by minimizing the projection distances of any within-class subspace pairs and maximizing the projection distances of between-class subspace pairs. The corresponding objective function \$J(\mathbf{P})\$ is defined as

$$\mathbf{P}^* = \operatorname{argmin}_{\mathbf{P}} J(\mathbf{P}) = \operatorname{argmin}_{\mathbf{P}} (J_w(\mathbf{P}) - \alpha J_b(\mathbf{P})), \quad (7.23)$$

where  $\alpha$  is the trade-off parameter between within-class scatter  $J_w(\mathbf{P})$  and between-class scatter  $J_b(\mathbf{P})$ ,

$$\begin{aligned} J_w(\mathbf{P}) &= \frac{1}{N_w} \sum_{i=1}^m \sum_{j: C_i = C_j} \operatorname{tr}(\mathbf{P} \mathbf{A}_{ij} \mathbf{A}_{ij}^T \mathbf{P}) = \operatorname{tr}(\mathbf{P} \mathbf{S}_w \mathbf{P}) \\ J_b(\mathbf{P}) &= \frac{1}{N_b} \sum_{i=1}^m \sum_{j: C_i \neq C_j} \operatorname{tr}(\mathbf{P} \mathbf{A}_{ij} \mathbf{A}_{ij}^T \mathbf{P}) \operatorname{tr}(\mathbf{P} \mathbf{S}_b \mathbf{P}). \end{aligned} \quad (7.24)$$

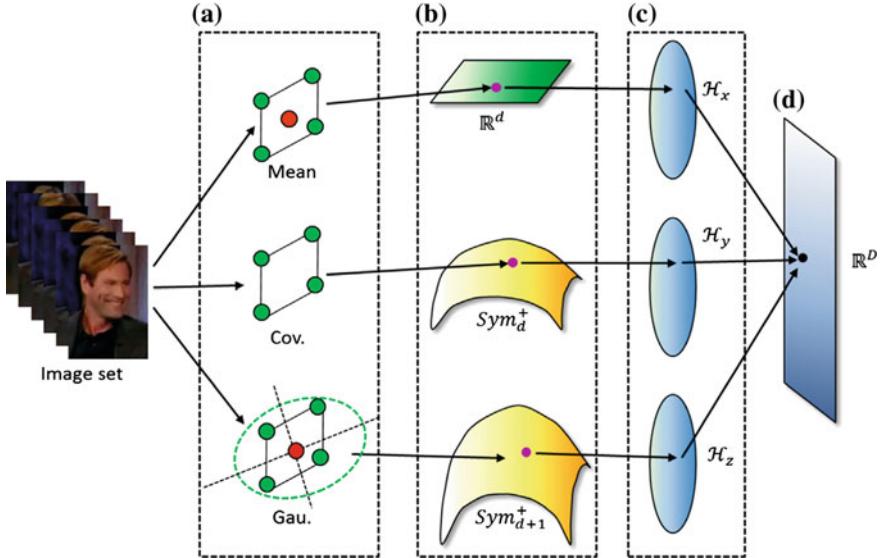
Since  $\mathbf{WY}'_i$  need to be orthogonal all the time, an algorithm is proposed to optimize  $\mathbf{P}$  and solving  $\mathbf{Y}'_i$  iteratively. In each iteration,  $\mathbf{W}^T \mathbf{Y}'_i$  is first decomposed into  $\mathbf{W}^T \mathbf{Y}'_i = \mathbf{Q}_i \mathbf{R}_i$  by QR-decomposition.  $\mathbf{Y}'_i$  is normalized by  $\mathbf{Y}'_i = \mathbf{Y}_i \mathbf{R}_i^{-1}$ . Then  $\mathbf{P}$  is solved using Riemannian Conjugate Gradient algorithm [1]. Finally, for verification task, given two videos, their projection distance in the low-dimensional space can be calculated using (7.22).

In [21] a hybrid metric learning method for image set-based face recognition was proposed, which is essentially an extension of [20]. The image sets are modeled simultaneously by mean, covariance matrix and Gaussian distribution and fused together for robustness. Another highlight of this paper is that the metrics are learned based on deep learning features. Combining set-based face recognition algorithm and the power of deep learning, the proposed method achieved state-of-the-art results in many challenging datasets. The conceptual illustration of the method is shown in Fig. 7.7.

Given an image set  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ ,  $\mathbf{x}_i \in \mathbb{R}^d$ , the DCNN features  $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$  are first extracted. Here, according to [9], the DCNN network is trained on 256 by 256 pixel face images. The face images are normalized using detected eye positions. The network has 17 layers, including 14 convolution layers, 2 fully connected layers, and 1 soft-max layer. The training of the DCNN network consists of pretraining and fine-tuning. The pretraining is conducted on “Celebrities on the Web” (CFW) dataset [44]. The fine-tuning is carried using the training part of the given dataset. Finally, the output of the second fully connected layer of the trained DCNN network is used as the deep feature. All the network training and feature extraction are accomplished by the Caffe deep learning framework [24].

After the deep features are obtained, the first statistic, the sample mean is defined by  $\mathbf{m} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i$ , which lies in Euclidean space  $\mathbb{R}^d$ . The second statistic, the covariance matrix, is defined by  $\mathbf{C} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{y}_i - \mathbf{m})(\mathbf{y}_i - \mathbf{m})^T$ , which lies on Riemannian manifold  $Sym_+^d$ . The third statistic, the Gaussian Mixture Model, is learned by Expectation Maximization algorithm. It can be written as

$$G = \sum_{i=1}^M w_i \mathcal{N}(\mathbf{y} | \tilde{\mathbf{m}}_i, \tilde{\mathbf{C}}_i), \quad (7.25)$$



**Fig. 7.7** Hybrid Euclidean-and-Riemannian metric learning [21]

where  $\tilde{\mathbf{m}}_i$  and  $\tilde{\mathbf{C}}_i$  are the mean and covariance matrix for the  $i$ th Gaussian component. According to the information geometry theory in [29], it can be embedded into  $\text{Sym}^{d+1}_+$  and represented by a  $(d+1) \times (d+1)$ -dimensional SPD matrix  $\mathbf{P}$  as

$$\mathcal{N}(\tilde{\mathbf{m}}_i, \tilde{\mathbf{C}}_i) \sim \mathbf{P} = |\mathbf{Q}|^{-2/(d+1)} \begin{bmatrix} \mathbf{Q}\mathbf{Q}^T + \tilde{\mathbf{m}}_i\tilde{\mathbf{m}}_i^T & \tilde{\mathbf{m}}_i \\ \tilde{\mathbf{m}}_i^T & 1 \end{bmatrix}, \quad (7.26)$$

where  $\tilde{\mathbf{C}} = \mathbf{Q}\mathbf{Q}^T$  and  $|\mathbf{Q}| > 0$ . For mean vectors, the linear kernel is directly used, which is

$$K_m(\mathbf{m}_i, \mathbf{m}_j) = \mathbf{m}_i^T \mathbf{m}_j. \quad (7.27)$$

For covariance matrices, the Log-Euclidean Distance is used, which is  $d(\mathbf{C}_i, \mathbf{C}_j) = \|\log(\mathbf{C}_i) - \log(\mathbf{C}_j)\|_F$ . It leads to the kernel

$$K_C(\mathbf{C}_i, \mathbf{C}_j) = \text{tr}(\log(\mathbf{C}_i) \log(\mathbf{C}_j)). \quad (7.28)$$

For GMMs, the LED metric is used as well. The kernel function is

$$K_G(\mathbf{G}_i, \mathbf{G}_j) = \sum_{a=1}^{M_a} \sum_{b=1}^{M_b} w_a w_b \text{tr}(\log(\mathbf{P}_i^a) \log(\mathbf{P}_j^b)), \quad (7.29)$$

where  $\mathbf{P}_i^a$  is the  $a$ th Gaussian component of the  $i$ th GMM.

Given training sets  $\mathbf{X}_i$  and  $\mathbf{X}_j$ , let  $\Phi_i^r$  and  $\Phi_j^r$  be the high dimensional features in RKHS of the  $r$ th statistic feature. The distance metric is defined as

$$d_{A_r}(\Phi_i^r, \Phi_j^r) = \text{tr}(\mathbf{A}_r(\Phi_i^r - \Phi_j^r)(\Phi_i^r - \Phi_j^r)^T), \quad (7.30)$$

where  $\mathbf{A}_r$  is the learned Mahalanobis matrix for the  $r$ th statistic in the high dimensional RKHS ( $r = 1, \dots, 3$  here). Using the Information-Theoretic Metric Learning method proposed in [15], the objective function for learning  $\{\mathbf{A}_r\}$  is formulated as

$$\begin{aligned} & \min_{\mathbf{A}_1 \geq 0, \dots, \mathbf{A}_R \geq 0, \xi} \frac{1}{R} \sum_{r=1}^R D_{\ell d}(\mathbf{A}_r, \mathbf{A}_0) + \gamma D_{\ell d}(\text{diag}(\xi), \text{diag}(\xi_0)), \\ & \text{s.t. } \frac{\delta_{ij}}{R} \sum_{r=1}^R d_{A_r}(\Phi_i^r, \Phi_j^r) \leq \xi_{ij}, \quad \forall i, j \end{aligned} \quad (7.31)$$

where  $D_{\ell d}(\mathbf{A}_r, \mathbf{A}_0) = \text{tr}(\mathbf{A}_r \mathbf{A}_0^{-1}) - \log \det(\mathbf{A}_r \mathbf{A}_0^{-1}) - d$ ,  $d$  is the dimensionality of the data.  $\xi$  is a vector of slack variables and is initialized to  $\xi_0$ , where  $\xi_{0ij} = \delta_{ij}\rho - \zeta\tau$ ,  $\rho$  is the threshold for distance comparison,  $\tau$  is the margin, and  $\zeta$  is the tuning scale

of the margin.  $\delta_{ij} = \begin{cases} 1 & \text{if } \mathbf{X}_i \text{ and } \mathbf{X}_j \text{ come from the same class} \\ -1 & \text{otherwise} \end{cases}$

Learning  $\mathbf{A}_r$  is equivalent to learning  $\mathbf{W}_r$  such that  $\mathbf{A}_r = \mathbf{W}_r \mathbf{W}_r^T$ . By applying the kernel trick, explicit computation of  $\Phi^r$  can be avoided. Assume that every column of  $\mathbf{W}_r$  is a linear combination of all the training samples in RKHS,  $\mathbf{w}_k^r$  can be expressed by  $\mathbf{w}_k^r = \sum_{j=1}^N \mathbf{u}_j^k \Phi_j^r$ ,  $\mathbf{u}^k$  are the expansion coefficients here. Let  $\mathbf{U}_r = [\mathbf{u}^1, \dots, \mathbf{u}^N]$ ,  $\mathbf{W}_r = \Phi^r \mathbf{U}_r$ , instead of learning  $\mathbf{W}_r$  directly,  $\mathbf{U}_r$  can be learned. Then the objective function can be rewritten as

$$\begin{aligned} & \min_{\mathbf{B}_1 \geq 0, \dots, \mathbf{B}_R \geq 0, \xi} \frac{1}{R} \sum_{r=1}^R D_{\ell d}(\mathbf{B}_r, \mathbf{B}_0) + \gamma D_{\ell d}(\text{diag}(\xi), \text{diag}(\xi_0)), \\ & \text{s.t. } \frac{\delta_{ij}}{R} \sum_{r=1}^R d_{B_r}(\mathbf{K}_{i,i}^r, \mathbf{K}_{j,j}^r) \leq \xi_{ij}, \quad \forall i, j, \end{aligned} \quad (7.32)$$

where  $\mathbf{B}_r = \mathbf{U}_r \mathbf{U}_r^T$  is the new Mahalanobis matrix.  $d_{B_r}(\mathbf{K}_{i,i}^r, \mathbf{K}_{j,j}^r) = \text{tr}(\mathbf{B}_r(\mathbf{K}_{i,i}^r - \mathbf{K}_{j,j}^r)(\mathbf{K}_{i,i}^r - \mathbf{K}_{j,j}^r)^T)$ .  $\mathbf{K}_{i,i}^r$  is the  $i$ th column of  $\mathbf{K}^r$ . The proposed method adopted the cyclic Bregman projection method [10] to solve this problem.

After  $\{\mathbf{B}_r\}_{r=1}^3$  are learned for all statistics, for verification task, given two image sets  $\mathbf{X}_i$  and  $\mathbf{X}_j$ , their corresponding DCNN features are first calculated. Means, covariance matrices and GMMs are then computed. Then the kernels between these testing samples and the training samples are computed as  $\mathbf{k}_i^r$  and  $\mathbf{k}_j^r$ . Finally, their distance is calculated by

$$d(\mathbf{X}_1, \mathbf{X}_2) = \sum_{r=1}^3 d_{\mathbf{B}_r}(\mathbf{k}_i^r, \mathbf{k}_j^r) = \sum_{r=1}^3 \text{tr}(\mathbf{B}_r(\mathbf{k}_i^r - \mathbf{k}_j^r)(\mathbf{k}_i^r - \mathbf{k}_j^r)^T). \quad (7.33)$$

Besides the methods mentioned above, Wang et al. [42] proposed a face recognition method for image sets using Gaussian Mixture Model which lies on specific Riemannian manifold. Huang et al. [23] provided a image set-based metric learning method using Log-Euclidean metric on SPD manifold. Arandjelovic and Cipolla [6] built a pose-wise linear illumination manifold model for video-based face recognition. Arandjelovic and Cipolla [5] modeled the video faces by shape-illumination manifolds which are robust to different variations. Kim et al. [25] introduced canonical correlations between two subspaces for image set recognition. Huang et al. [19] proposed the Euclidean-to-Riemannian Metric for Point-to-Set Classification on Riemannian manifold.

## 7.4 Probabilistic Methods

Probabilistic methods provide flexibility so that the similarity scores can either be modeled as “distance” or as “likelihood”.

In [27], a video-based face recognition algorithm based on probabilistic appearance manifolds is introduced. The image set of a given object can be treated as a low-dimensional appearance manifold  $M_k$  in the image space. Given a testing image  $I$ , identity  $k^*$  is determined by finding the manifold  $M_k$  with minimal distance to  $I$ , which is

$$k^* = \underset{k}{\operatorname{argmin}} d_H(I, M_k), \quad (7.34)$$

where  $d_H$  denotes the  $L^2$ -Hausdorff distance between the image  $I$  and  $M_k$ . Probabilistically, let

$$P(k|I) = \frac{1}{\Lambda} \exp(-\frac{1}{2\sigma^2} d_H^2(I, M_k)), \quad (7.35)$$

where  $\Lambda$  is a normalization term. Thus, (7.34) turns into

$$k^* = \underset{k}{\operatorname{argmax}} P(k|I). \quad (7.36)$$

Since  $M_k$  is usually not known and can only be estimated by samples,  $d_H(I, M_k)$  cannot be calculated directly. Let  $p_{M_k}(x|I)$  be the probability that  $x$  is the point on  $M_k$  at minimal  $L^2$  distance to  $I$ . Also, since the appearance manifold is complex and non-linear, it is decomposed into a collection of  $m$  simpler disjoint manifolds as  $M_k = C^{k1} \cup \dots \cup C^{km}$  where  $C^{ki}$  is called a pose manifold. Each pose manifold is further

approximated by an affine plane through PCA.  $P(C^{ki}|I)$  denotes the probability that  $C^{ki}$  contains point  $x^*$  with minimal distance to  $I$ . Then

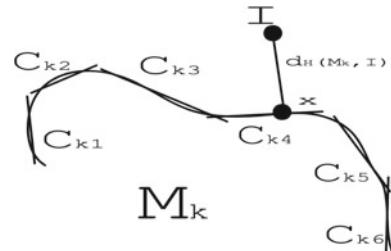
$$\begin{aligned} d_H(I, M_k) &= \int_{M_k} d(x, I) p_{M_k}(x|I) dx = \sum_{i=1}^m P(C^{ki}|I) \int_{C^{ki}} d_H(x, I) p_{C^{ki}}(x|I) dx \\ &= \sum_{i=1}^m P(C^{ki}|I) d_H(I, C^{ki}), \end{aligned} \quad (7.37)$$

which is the average expected distance between  $I$  and each pose manifold  $C^{ki}$ . This is shown in Fig. 7.8.

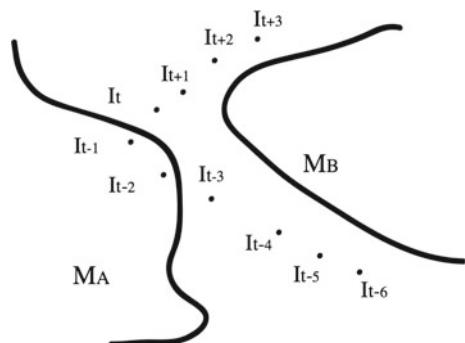
For video-based face recognition, the temporal coherence between consecutive image frames can be exploited. As the example shown in Fig. 7.9,  $\{I_t\}$  probably originate from  $M_B$  by looking at the whole sequence. But because of the appearance variations, some of the frames are closer to  $M_A$ . By considering the temporal coherence, the image-to-manifold can be estimated more robustly.

Given previous frames  $I_{0:t-1}$  at time  $t$ , assume  $I_t$  and  $I_{0:t-1}$  are independent given  $C_t^{ki}$ ,  $C_t^{ki}$  and  $I_{0:t-1}$  are independent given  $C_{t-1}^{ki}$ ,  $P(C_t^{ki}|I_t, I_{0:t-1})$  can be calculated as

**Fig. 7.8**  $d_H(I, M_k)$  [27]



**Fig. 7.9** Exploit temporal coherence [27]



$$\begin{aligned}
P(C_t^{ki}|I_t, I_{0:t-1}) &= \alpha P(I_t|C_t^{ki}, I_{0:t-1})P(C_t^{ki}|I_{0:t-1}) \\
&= \alpha P(I_t|C_t^{ki}) \sum_{j=1}^m P(C_t^{ki}|C_{t-1}^{kj}, I_{0:t-1})P(C_{t-1}^{kj}|I_{0:t-1}) \\
&= \alpha P(I_t|C_t^{ki}) \sum_{j=1}^m P(C_t^{ki}|C_{t-1}^{kj})P(C_{t-1}^{kj}|I_{t-1}, I_{0:t-2}),
\end{aligned} \tag{7.38}$$

where  $\alpha$  is a normalization constant.  $P(C_t^{ki}|C_{t-1}^{kj})$  is the probability of  $x_t^* \in C^{ki}$  given  $x_{t-1}^* \in C^{kj}$ . Because of the temporal coherency between consecutive frames,  $x_{t-1}^*$  and  $x_t^*$  should have small geodesic distance on  $M_k$ .  $P(C_t^{ki}|C_{t-1}^{kj})$  is thus related to their geodesic distance. Equation (7.38) can be computed recursively if  $P(I_t|C_t^{ki})$  and  $P(C_t^{ki}|C_{t-1}^{kj}) \forall i, j, t$  are known.

Given training image sets  $\{S_k\}$  from videos,  $K$ -means algorithm is used to partition these sets into  $m$  disjoint subsets  $\{S_{k1}, \dots, S_{km}\}$ . For each  $S_{ki}$ , a linear approximation  $L_{ki}$  of local manifold  $C^{ki}$  is obtained by PCA.  $P(C^{ki}|C^{kj})$  is then calculated by

$$P(C^{ki}|C^{kj}) = \frac{1}{\Lambda_{kj}} \sum_{t=2}^l \delta(I_{t-1} \in S_{ki}) \delta(I_t \in S_{kj}), \tag{7.39}$$

which is counting the actual transitions in the corresponding training set.  $\Lambda_{kj}$  is a normalization constant.  $P(I|C^{ki})$  is calculated by

$$P(I|C^{ki}) = \frac{1}{\Lambda_{ki}} \exp\left(-\frac{1}{2\sigma^2} d_H(I, L_{ki})\right), \tag{7.40}$$

where  $L_{ki}$  is the low-dimensional linear approximation of manifold  $C^{ki}$ .  $\Lambda_{ki}$  is a normalization constant.  $d_H(I, L_{ki}) = d_H(I, C^{ki})$  is the distance between  $I$  and  $C^{ki}$ . Finally, for identification task, given an image  $I_t$  from a testing video sequence  $\{I_t\}$ ,  $P(C_t^{ki}|I_t, I_{0:t-1}), \forall k, i$  are calculated recursively by (7.38).  $d_H(I, M_k)$  is then obtained by (7.37). The decision is made by (7.36).

A probability distribution-based method for video-based face recognition was proposed in [7]. The Kullback–Leibler divergence is used as the distance measure between the distributions of videos. Given image sets collected from videos, Gaussian mixture models  $\hat{p}$  are learned for each image set. This is done using the Expectation Maximization algorithm. EM is initialized by  $K$ -means clustering and constrained to diagonal covariance matrices. The number of components is selected according to the minimal description length criterion [8]. Then for each training and testing video pair  $(V^{te}, V^{tr})$ , the KL divergence between the learned distributions  $\hat{p}^{te}$  and  $\hat{p}^{tr}$  is used as the distance measure, which is

$$d(V^{te}, V^{tr}) = D_{KL}(\hat{p}^{te} || \hat{p}^{tr}) = \int \hat{p}^{te}(\mathbf{x}) \log \left( \frac{\hat{p}^{te}(\mathbf{x})}{\hat{p}^{tr}(\mathbf{x})} \right) d\mathbf{x}. \tag{7.41}$$

The KL divergence  $D_{KL}(p||q)$  quantifies how well the distribution  $p$  describes samples from  $q$ . It is nonnegative and equal to zero if  $p \equiv q$ . Since the calculation of the KL divergence involves integration, there is no closed form when  $\hat{p}^{te}$  and  $\hat{p}^{tr}$  are GMMs. However, according to the law of large numbers, the KL divergence can still be approximated by sampling using Monte-Carlo simulation:

$$D_{KL}(\hat{p}^{te} || \hat{p}^{tr}) \approx \frac{1}{N} \sum_{k=1}^N \log \left( \frac{\hat{p}^{te}(\mathbf{x}_k)}{\hat{p}^{tr}(\mathbf{x}_k)} \right), \quad (7.42)$$

where  $\mathbf{x}_k$  are samples drawn from distribution  $\hat{p}^{te}$ . Then for identification task, the similarity between every training and testing video pair is computed using (7.42).

Liu and Chen [28] proposed a Hidden Markov Models based method to perform video-based face recognition. When training, the statistics and the temporal information of training videos are learned by HMMs. During the recognition phase, the temporal characteristics of the testing videos are analyzed by the HMM corresponding to each subject. The decision is made by finding the highest likelihood scores provided by the HMMs.

A continuous HMM model is defined as the triplet  $\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$ .  $\mathbf{A} = \{a_{ij}\}$  is the transition probability matrix, where  $a_{ij} = P(q_t = S_j | q_{t-1} = S_i)$ ,  $1 \leq i, j \leq N$ .  $\mathbf{B} = \{b_i(\mathbf{o})\}$  is the observation probability density function, where  $b_i(\mathbf{o}) = \sum_{k=1}^M c_{ik} \mathcal{N}(\mathbf{o}; \boldsymbol{\mu}_{ik}, \boldsymbol{\Sigma}_{ik})$ .  $\boldsymbol{\pi} = \{\pi_i\}$  is the initial state distribution, where  $\pi_i = P(q_1 = S_i)$ ,  $1 \leq i \leq N$ . Here  $\mathbf{S} = \{S_1, S_2, \dots, S_N\}$  is the set of states in the model.  $\mathbf{O} = \{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_T\}$  are the observations and  $\mathbf{Q} = \{q_1, q_2, \dots, q_T\}$  are the corresponding hidden state variables. Given state  $S_i$ ,  $b_i(\mathbf{o})$  is a Gaussian Mixture Model with  $M$  Gaussians.  $c_{ik}$ ,  $\boldsymbol{\mu}_{ik}$  and  $\boldsymbol{\Sigma}_{ik}$  are the mixture coefficient, mean and covariance for the  $k$ th Gaussian, respectively.

Given training videos, the face images are first projected to a low-dimensional space using PCA. Then each video is modeled as an HMM with these low-dimensional features as observations  $\mathbf{O}$ , which is shown in Fig. 7.10.

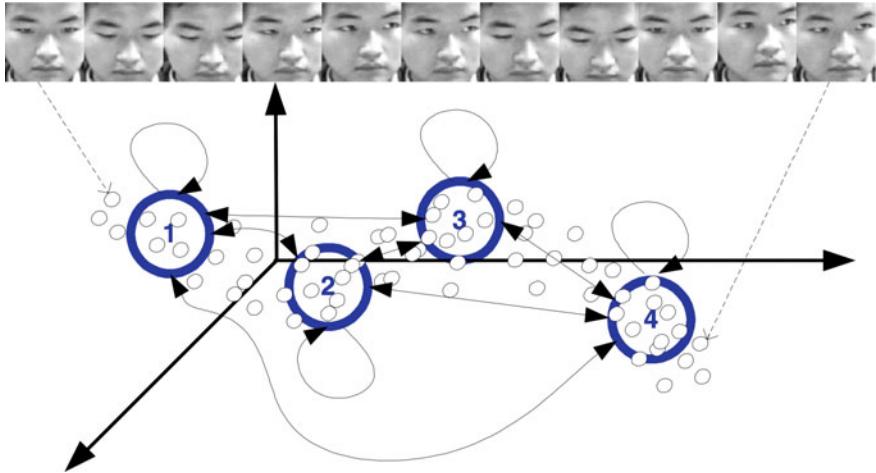
The estimation for HMM parameter  $\lambda = (\mathbf{A}, \mathbf{B}, \boldsymbol{\pi})$  is as follows:

1.  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\boldsymbol{\pi}$  are initialized (observations are clustered into  $M$  Gaussians.  $c_{ik}^0$ ,  $\boldsymbol{\mu}_{ik}^0$  and  $\boldsymbol{\Sigma}_{ik}^0$  are estimated for each Gaussian).  $n = 0$ .
2. Do.

- 2.1. Reestimate  $\lambda$  using the expectation maximization algorithm, in order to maximize the likelihood  $p(\mathbf{O}|\lambda)$ . The reestimation is defined as

$$\pi_i^{n+1} = \frac{P(\mathbf{O}, q_1 = i | \lambda^n)}{p(\mathbf{O} | \lambda^n)} \quad (7.43)$$

$$a_{ij}^{n+1} = \frac{\sum_{t=1}^T p(\mathbf{O}, q_{t-1} = i, q_t = j | \lambda^n)}{\sum_{t=1}^T p(\mathbf{O}, q_{t-1} = i | \lambda^n)} \quad (7.44)$$



**Fig. 7.10** Temporal HMM for modeling face sequences [28]

$$c_{ik}^{n+1} = \frac{\sum_{t=1}^T P(q_t = i, m_{q,t} = k | \mathbf{O}, \lambda^n)}{\sum_{t=1}^T \sum_{k=1}^M P(q_t = i, m_{q,t} = k | \mathbf{O}, \lambda^n)} \quad (7.45)$$

$$\mu_{ik}^{n+1} = \frac{\sum_{t=1}^T \mathbf{o}_t P(q_t = i, m_{q,t} = k | \mathbf{O}, \lambda^n)}{\sum_{t=1}^T P(q_t = i, m_{q,t} = k | \mathbf{O}, \lambda^n)} \quad (7.46)$$

$$\Sigma_{ik}^{n+1} = (1 - \alpha) \mathbf{C} + \alpha \frac{\sum_{t=1}^T (\mathbf{o}_t - \mu_{ik}^{n+1})(\mathbf{o}_t - \mu_{ik}^{n+1})^T P(q_t = i, m_{q,t} = k | \mathbf{O}, \lambda^n)}{P(q_t = i, m_{q,t} = k | \mathbf{O}, \lambda^n)} \quad (7.47)$$

where  $m_{q,t}$  indicates the mixture component of state  $q_t$  and time  $t$ .  $\mathbf{C}$  is a general model for the variance of all videos.  $\alpha$  is a weighting factor, which prevents the estimated  $\Sigma$  to be singular.

## 2.2 $n = n + 1$

### 3. Until $p(\mathbf{O} | \lambda)$ converges.

For identification task, after the HMM models  $\{\lambda_c^{tr}\}$  are estimated for training videos, given a testing video, the face images are projected onto the same low-dimensional space as the training samples and obtain the testing observation  $\mathbf{O}^{te}$ . Then the likelihood score  $p(\mathbf{O}^{te} | \lambda_c^{tr})$  of the observation given the training testing HMM models are computed. The identification decision is made by  $p = \text{argmax}_c p(\mathbf{O}^{te} | \lambda_c^{tr})$ , which finds the highest likelihood score.

In addition to the methods discussed above, Zhou et al. [45] introduced an appearance-adaptive model-based on particle filter to realize robust visual tracking and recognition. Zhou et al. [46] proposed a time series based method for video-based face recognition. Arandjelovic and Cipolla [4] provided another method based on kernelized distribution-to-distribution distance. Wang et al. [43] introduced a probabilistic nearest neighbor search method for image set classification.

## 7.5 Geometrical Model-Based Methods

Geometrical model-based methods construct certain geometrical models for faces in the videos. Then the texture map of the faces are projected on to these models and features are extracted. The recognition will based on these features. The models can vary from the simple spherical head models to the human-specific 3D head models. Geometrical model based methods are more robust to illumination and pose variations because they exploit the geometrical structures from the faces.

Sankaranarayanan and Chellappa [17] proposed a novel feature for robust video-based face recognition in camera networks. It is developed using the spherical harmonic representation of the face texture mapped onto a spherical head model. Spherical harmonics are a set of orthonormal basis functions defined over the unit sphere, and can be used to linearly expand any square-integrable function on  $\mathbb{S}^2$  as

$$f(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-l}^l f_{lm} Y_{lm}(\theta, \phi), \quad (7.48)$$

where  $Y_{lm}(\cdot, \cdot)$  defines the SH basis function of degree  $l \geq 0$  and order  $m \in (-l, -l + 1, \dots, l - 1, l)$ .  $f_{lm}$  is the coefficient associated with the basis function  $Y_{lm}$  for the function  $f$ . The spherical coordinate system is used here.  $\theta \in (0, \pi)$  and  $\phi \in (0, 2\pi)$  are the zenith and azimuth angles, respectively. There are  $2l + 1$  basis functions for a given order  $l$ . The SH basis function for degree  $l$  and order  $m$  has the following form (shown in Fig. 7.11):

$$Y_{lm}(\theta, \phi) = K_{lm} P_l^m(\cos \theta) e^{im\phi}, \quad (7.49)$$

where  $K_{lm}$  denotes a normalization constant such that

$$\int_0^\pi \int_0^{2\pi} Y_{lm} Y_{lm}^* d\phi d\theta = 1. \quad (7.50)$$

Here,  $P_{lm}(x)$  are the associated Legendre functions. As with Fourier expansion, the SH expansion coefficients  $f_l^m$  can be computed as

$$f_l^m = \int_\theta \int_\phi f(\theta, \phi) Y_l^m(\theta, \phi) d\theta d\phi. \quad (7.51)$$

Given two multiview videos, the head centers in these videos are first obtained using a multiview tracking algorithm proposed in the paper. Then a spherical head model for each head is build. The SH spectrum features are extracted from the texture map projected on the models from all views. The projection of the texture map is shown in Fig. 7.12.

These features are projected into a reproducing kernel Hilbert space (RKHS), which is performed via an Radial Basis Function (RBF) kernel. The limiting

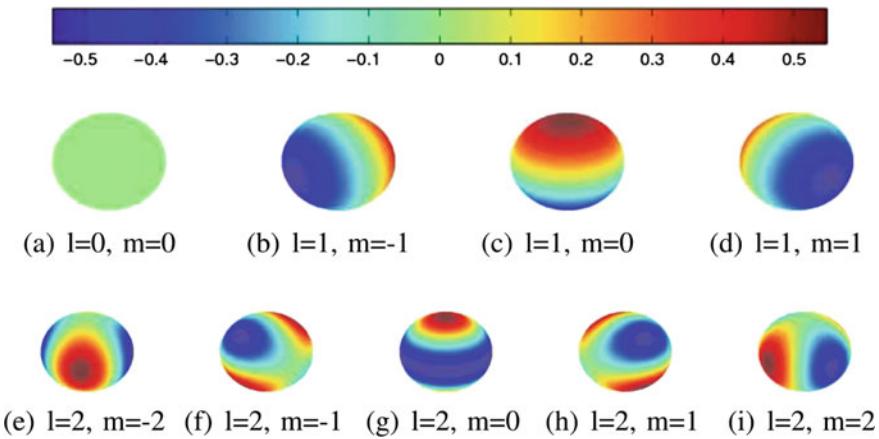


Fig. 7.11 Visualization of the first three degree of Spherical Harmonics [17]

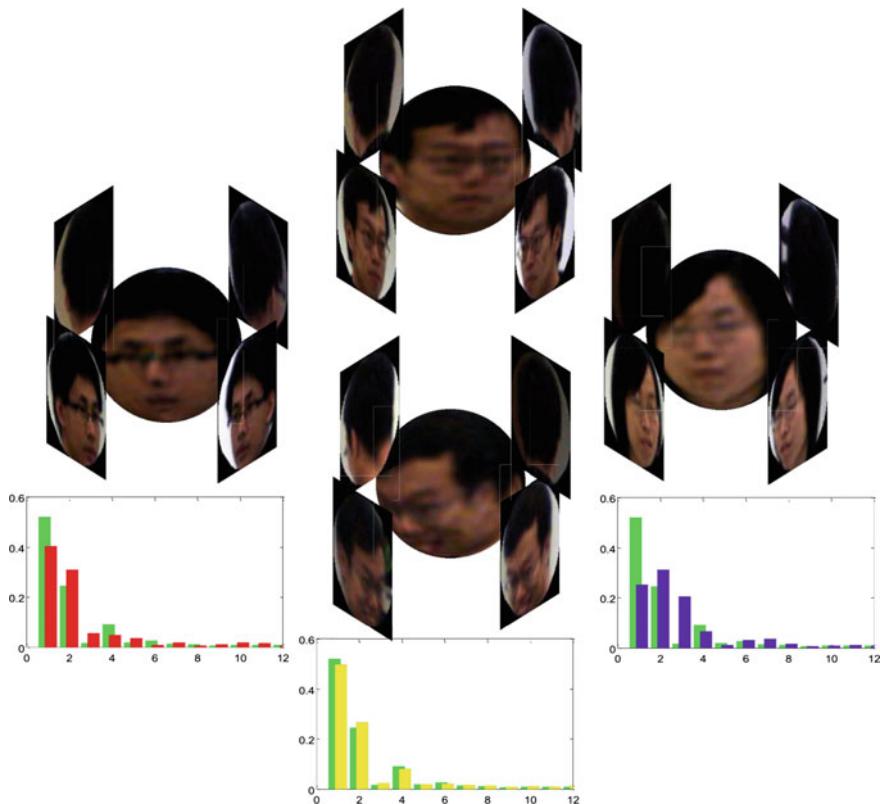


Fig. 7.12 Texture map projection [17]

Bhattacharyya distance between these probability distributions in RKHS (assume to be Gaussian) is considered as the distance measure. The limiting Bhattacharyya distance in this case is

$$D = \frac{1}{8}(\alpha_{11} + \alpha_{22} - 2\alpha_{12}), \quad (7.52)$$

where

$$\alpha_{ij} = \boldsymbol{\mu}_i^T \left( \frac{1}{2} \mathbf{C}_i + \frac{1}{2} \mathbf{C}_j \right)^{-1} \boldsymbol{\mu}_j. \quad (7.53)$$

$\boldsymbol{\mu}_i$  and  $\mathbf{C}_i$  are the means and covariance matrices in RKHS which cannot be directly calculated. Denote the Gram matrix as  $\mathbf{K}_{ij}$ , where  $i, j \in \{1, 2\}$  are the indices of videos.  $\mathbf{K}_{11}$  and  $\mathbf{K}_{22}$  are centered by

$$\mathbf{K}'_{ii} = \mathbf{J}_i^T \mathbf{K}_{ii} \mathbf{J}_i, \mathbf{J}_i = N_i^{-\frac{1}{2}} (\mathbf{I}_N - \mathbf{s} \mathbf{1}^T), \quad (7.54)$$

where  $\mathbf{s} = N_i^{-1} \mathbf{1}$ ,  $\mathbf{1}$  is a  $N_i \times 1$  vector of 1s and  $N_i$  is the number of features from video  $i$ . Then  $\alpha_{ij}$  is calculated by

$$\alpha_{ij} = \mathbf{s}_i^T \mathbf{K}'_{ij} \mathbf{s}_j - \mathbf{s}_i^T [\mathbf{K}_{i1} \ \mathbf{K}_{i2}] \mathbf{B} \begin{bmatrix} \mathbf{K}_{j1} \\ \mathbf{K}_{j2} \end{bmatrix} \mathbf{s}_j, \quad (7.55)$$

where

$$\mathbf{B} = \mathbf{P} \mathbf{L}^{-1} \mathbf{P}, \mathbf{L} = \mathbf{P}^T \begin{bmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{bmatrix} \mathbf{P} \quad (7.56)$$

and

$$\mathbf{P} = \begin{bmatrix} \sqrt{\frac{1}{2}} \mathbf{J}_1 \mathbf{V}_1 & 0 \\ 0 & \sqrt{\frac{1}{2}} \mathbf{J}_2 \mathbf{V}_2 \end{bmatrix} \quad (7.57)$$

$\mathbf{V}_i$  is the matrix which stores the first  $r$  eigenvectors of  $\mathbf{K}'_{ii}$  (i.e., corresponding to the  $r$  largest eigenvalues). For identification and verification tasks, the similarity between the two set of features is measured by the computed limiting Bhattacharyya distance between them.

Park and Jain [33] also provided a video-based face recognition method which reconstructs 3D face models from the videos and recognizes faces at frontal view.

## 7.6 Dynamical Model-Based Methods

Dynamical model-based methods are sequence-based methods. They consider videos as dynamical systems with video frames as the observation of these systems. The advantage of these methods is that the extra temporal information is exploited.

Dynamical models are often used to represent motions or activities, but there are some publications that use dynamical models for face recognition.

In [2], the video-to-video face recognition problem is transferred into a dynamical system identification and classification problem. Videos are modeled by dynamical systems. Here, the ARMA model is used for the dynamical system. The ARMA model is defined as

$$\mathbf{x}(t+1) = \mathbf{Ax}(t) + \mathbf{v}(t) \quad (7.58)$$

$$\mathbf{y}(t) = \mathbf{Cx}(t) + \mathbf{w}(t), \quad (7.59)$$

where  $\mathbf{x}(t)$  is the state vector,  $\mathbf{y}(t)$  is the observation.  $\mathbf{A}$  and  $\mathbf{C}$  are transition matrix and observation matrix, respectively. The system is driven by the IID process  $\mathbf{v}(t)$ .  $\mathbf{w}(t)$  is the observation noise.

Suppose  $\mathbf{v}(t) \sim \mathcal{N}(0, \mathbf{Q})$  and  $\mathbf{w}(t) \sim \mathcal{N}(0, \mathbf{R})$ , given a video sequence  $\mathbf{Y}^\tau = [\mathbf{y}(1), \dots, \mathbf{y}(\tau)]$ , (7.59) can be rewritten as

$$\mathbf{Y}^\tau = \mathbf{CX}^\tau + \mathbf{W}^\tau, \quad (7.60)$$

where  $\mathbf{X}$  and  $\mathbf{W}$  are similarly defined. Then the model parameters can be estimated by

$$\hat{\mathbf{C}}(\tau) = \mathbf{U} \quad (7.61)$$

$$\hat{\mathbf{X}}(\tau) = \mathbf{\Sigma V}^T \quad (7.62)$$

$$\hat{\mathbf{A}}(\tau) = \mathbf{\Sigma V}^T \mathbf{D}_1 \mathbf{V} (\mathbf{V}^T \mathbf{D}_2 \mathbf{V})^{-1} \mathbf{\Sigma}^{-1} \quad (7.63)$$

$$\hat{\mathbf{Q}}(\tau) = \frac{1}{\tau} \sum_{t=1}^{\tau} \hat{\mathbf{v}}(t) \hat{\mathbf{v}}^T(t), \quad (7.64)$$

where  $\mathbf{Y}^\tau = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T$  is the SVD of  $\mathbf{Y}^\tau$ .  $\mathbf{D}_1 = \begin{bmatrix} 0 & 0 \\ \mathbf{I}_{\tau-1} & 0 \end{bmatrix}$  and  $\mathbf{D}_2 = \begin{bmatrix} \mathbf{I}_{\tau-1} & 0 \\ 0 & 0 \end{bmatrix}$ .  $\hat{\mathbf{v}}(t) = \hat{\mathbf{x}}(t+1) - \hat{\mathbf{A}}(\tau) \hat{\mathbf{x}}(t)$ .

Given video pairs  $\mathbf{V}_1$  and  $\mathbf{V}_2$ , their model parameters  $M_1$  and  $M_2$  are first estimated, respectively. Then the distance between two ARMA models is calculated by

$$d_M(M_1, M_2)^2 = \ln \prod_{i=1}^n \frac{1}{\cos^2 \theta_i} \quad (7.65)$$

$$d_g(M_1, M_2) = \sin \theta_{max} \quad (7.66)$$

$$d_f(M_1, M_2)^2 = 2 \sum_{i=1}^n \sin^2 \theta_i, \quad (7.67)$$

where  $d_M(M_1, M_2)$  is the Martin distance,  $d_g(M_1, M_2)$  is the gap distance and  $d_f(M_1, M_2)$  is the distance based on Frobenius norm.  $\theta_i$ 's are the subspace angles between  $M_1$

and  $M_2$  (see [13] for more details). Different distances can be chosen for different scenarios or fused together to improve the performance.

Turaga [37] also considered videos as ARMA models and treated each video as a point on the Grassmann manifold for recognition.

## 7.7 Conclusion and Future Directions

As we saw in this chapter, most of the modeling approaches for video-based face recognition focus on how to define the similarity scores (or the “distances”) between videos. Sparse coding-based methods model videos as dictionaries and use reconstruction error as the similarity score. Manifold-based methods use special kernels between manifolds as the similarity. Probabilistic methods are more flexible. The similarity scores can be the KL divergence between distributions, or the expected distance under some certain distributions. Dynamical model-based methods consider videos as dynamical systems. The similarity scores are the distance between two systems on a certain manifold. Geometrical model-based methods are slightly different from the others since their main objective is to construct geometrical models from videos and project texture maps onto them.

Since deep learning is becoming increasingly important recently, one of the future directions for video-based face recognition will be the classic methods combined with deep learning-based methods. The special statistical and geometrical properties of deep features will lead to new modeling techniques. Another possible direction would be to build 3D DCNN networks, where the convolutions are applied through the time-axis as well, in order to capture the temporal information between consecutive frames. Also, thanks to the fast developments in deep learning-based detection and landmark extraction techniques, face detection and alignment are becoming more and more precise, which can provide geometrical model-based methods with improved performance.

**Acknowledgements** This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2014-14071600012. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

## References

1. Absil PA, Mahony R, Sepulchre R (2007) Optimization algorithms on matrix manifolds. Princeton University Press, Princeton, NJ, USA
2. Aggarwal G, Chowdhury A, Chellappa R (2004) A system identification approach for video-based face recognition. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, vol 4, pp 175–178
3. Aharon M, Elad M, Bruckstein A (2006) K-svd: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process* 54(11):4311–4322
4. Arandjelovic O, Cipolla R (2004) Face recognition from face motion manifolds using robust kernel resistor-average distance. In: Conference on Computer Vision and Pattern Recognition Workshop, 2004. CVPRW '04, p 88
5. Arandjelovic O, Cipolla R (2006) Face Recognition from Video Using the Generic Shape-Illumination Manifold. In: Proceedings of Computer Vision—ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, 7–13 May 2006, Part IV. Springer, Berlin Heidelberg, pp 27–40
6. Arandjelovic O, Cipolla R (2009) A pose-wise linear illumination manifold model for face recognition using video. *Comput Vis Image Underst* 113(1):113–125
7. Arandjelovic O, Shakhnarovich G, Fisher J, Cipolla R, Darrell T (2005) Face recognition with image sets using manifold density divergence. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, CVPR 2005. vol 1 (2005), pp 581–588
8. Barron A, Rissanen J, Yu B (1998) The minimum description length principle in coding and modeling. *IEEE Trans Inf Theory* 44(6):2743–2760
9. Beveridge J, Zhang H, Draper B, Flynn P, Feng Z, Huber P, Kittler J, Huang Z, Li S, Li Y, Kan M, Wang R, Shan S, Chen X, Li H, Hua G, Struc V, Krizaj J, Ding C, Tao D, Phillips P (2015) Report on the fg 2015 video person recognition evaluation. In: 2015 11th IEEE international conference and workshops on Automatic Face and Gesture Recognition (FG), vol 1, pp 1–8
10. Bregman L (1967) The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput Math Math Phys* 7(3):200–217
11. Chen YC, Patel V, Shekhar S, Chellappa R, Phillips P (2013) Video-based face recognition via joint sparse representation. In: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp 1–8
12. Chen YC, Patel VM, Phillips PJ, Chellappa R (2012) Dictionary-Based Face Recognition from Video. In: Computer Vision—ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, 7–13 Oct 2012, Proceedings. Springer, Berlin, pp 766–779
13. Cock KD, Moor BD (2002) Subspace angles between arma models. *Syst Control Lett* 46(4):265–270
14. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), CVPR '05, vol 1. IEEE Computer Society, Washington, DC, USA, pp 886–893
15. Davis JV, Kulis B, Jain P, Sra S, Dhillon IS (2007) Information-theoretic metric learning. Proceedings of the 24th International Conference on Machine Learning., ICML '07ACM, New York, NY, USA, pp 209–216
16. Du M, Chellappa R (2014) Video-based face recognition using the intra-personal/extrAPERSONAL difference dictionary. In: Proceedings of the British Machine Vision Conference. BMVA Press
17. Du M, Sankaranarayanan A, Chellappa R (2014) Robust face recognition from multi-view videos. *IEEE Trans Image Process* 23(3):1105–1117
18. Hu Y, Mian A, Owens R (2011) Sparse approximated nearest points for image set classification. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 121–128

19. Huang Z, Wang R, Shan S, Chen X (2014) Learning euclidean-to-riemannian metric for point-to-set classification. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 1677–1684 (2014)
20. Huang Z, Wang R, Shan S, Chen X (2015) Hybrid Euclidean-and-Riemannian Metric Learning for Image Set Classification. In: Computer Vision—ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, 1–5 Nov 2014, Revised Selected Papers, Part III. Springer International Publishing, Cham, pp 562–577
21. Huang Z, Wang R, Shan S, Chen X (2015) Face recognition on large-scale video in the wild with hybrid euclidean-and-riemannian metric learning. *Patt Recogn* 48(10):3113–3124
22. Huang Z, Wang R, Shan S, Chen X (2015) Projection metric learning on grassmann manifold with application to video based face recognition. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 140–149
23. Huang Z, Wang R, Shan S, Li X, Chen X (2015) Log-euclidean metric learning on symmetric positive definite manifold with application to image set classification. In: Blei D, Bach E (eds.) Proceedings of the 32nd International Conference on Machine Learning (ICML-15), JMLR Workshop and Conference Proceedings, pp 720–729
24. Jia Y, Shelhamer E, Donahue J, Karayev S, Long, J, Girshick, R, Guadarrama S, Darrell T (2014) Caffe: convolutional architecture for fast feature embedding. arXiv preprint [arXiv:1408.5093](https://arxiv.org/abs/1408.5093)
25. Kim TK, Kittler J, Cipolla R (2007) Discriminative learning and recognition of image set classes using canonical correlations. *IEEE Trans Pattern Anal Mach Intell* 29(6):1005–1018
26. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Pereira E, Burges C, Bottou L, Weinberger K (eds) Advances in neural information processing systems, vol 25. Curran Associates, Inc., pp 1097–1105
27. Lee KC, Ho J, Yang MH, Kriegman D (2003) Video-based face recognition using probabilistic appearance manifolds. Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR’03. IEEE Computer Society, Washington, DC, USA, pp 313–320
28. Liu X, Chen T (2003) Video-based face recognition using adaptive hidden markov models. Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR’03. IEEE Computer Society, Washington, DC, USA, pp 340–345
29. Lovri M, Min-Oo M, Ruh EA (2000) Multivariate normal distributions parametrized as a riemannian symmetric space. *J Multivariate Anal* 74(1):36–48
30. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vision* 60(2):91–110
31. Ojala T, Pietikainen M, Maenpaa T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 24(7):971–987
32. Ortiz E, Wright A, Shah M (2013) Face recognition in movie trailers via mean sequence sparse representation-based classification. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 3531–3538
33. Park U, Jain AK (2007) 3D Model-based face recognition in video. In: Advances in Biometrics: International Conference, ICB 2007, Seoul, Korea, 27–29 Aug 2007. Proceedings. Springer, Berlin, pp 1085–1094
34. Rosipal R, Kramer N (2006) Overview and recent advances in partial least squares. In: Proceedings of the 2005 International Conference on Subspace, Latent Structure and Feature Selection, SLSFS’05. Springer, Berlin, pp 34–51
35. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. CoRR [abs/1409.1556](https://arxiv.org/abs/1409.1556)
36. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
37. Turaga P, Veeraraghavan A, Srivastava A, Chellappa R (2011) Statistical computations on grassmann and stiefel manifolds for image and video-based recognition. *IEEE Trans Pattern Anal Mach Intell* 33(11):2273–2286

38. Viola P, Jones MJ (2004) Robust real-time face detection. *Int J Comput Vision* 57(2):137–154
39. Wang R, Chen X (2009) Manifold discriminant analysis. In: IEEE Conference on Computer Vision and Pattern Recognition, 2009, CVPR 2009. pp 429–436
40. Wang R, Guo H, Davis L, Dai Q (2012) Covariance discriminative learning: a natural and efficient approach to image set classification. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2496–2503
41. Wang R, Shan S, Chen X, Gao W (2008) Manifold-manifold distance with application to face recognition based on image set. In: IEEE Conference on Computer Vision and Pattern Recognition, 2008, CVPR 2008, pp 1–8
42. Wang W, Wang R, Huang Z, Shan S, Chen X (2015) Discriminant analysis on Riemannian manifold of Gaussian distributions for face recognition with image sets. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2048–2057
43. Wang W, Wang R, Shan S, Chen X (2015) Probabilistic nearest neighbor search for robust classification of face image sets. In: 2015 11th IEEE international conference and workshops on automatic Face and Gesture Recognition (FG) (Vol. 1, pp. 1–7)
44. Zhang X, Zhang L, Wang XJ, Shum HY (2012) Finding celebrities in billions of web images. *IEEE Trans Multimedia* 14(4):995–1007
45. Zhou S, Chellappa R, Moghaddam B (2004) Visual tracking and recognition using appearance-adaptive models in particle filters. *IEEE Trans Image Process* 13(11):1491–1506
46. Zhou S, Krueger V, Chellappa R (2003) Probabilistic recognition of human faces from video. *Comput Vis Image Understand* 91(12):214–245. Special Issue on Face Recognition
47. Zhu X, Ramanan D (2012) Face detection, pose estimation, and landmark localization in the wild. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 2879–2886

# **Chapter 8**

# **Face Recognition Technologies for Evidential Evaluation of Video Traces**

**Xingjie Wei and Chang-Tsun Li**

**Abstract** Human recognition from video traces is an important task in forensic investigations and evidence evaluations. Compared with other biometric traits, face is one of the most popularly used modalities for human recognition due to the fact that its collection is non-intrusive and requires less cooperation from the subjects. Moreover, face images taken at a long distance can still provide reasonable resolution, while most biometric modalities, such as iris and fingerprint, do not have this merit. In this chapter, we discuss automatic face recognition technologies for evidential evaluations of video traces. We first introduce the general concepts in both forensic and automatic face recognition, then analyse the difficulties in face recognition from videos. We summarise and categorise the approaches for handling different uncontrollable factors in difficult recognition conditions. Finally we discuss some challenges and trends in face recognition research in both forensics and biometrics. Given its merits tested in many deployed systems and great potential in other emerging applications, considerable research and development efforts are expected to be devoted in face recognition in the near future.

## **8.1 Introduction**

The UK currently has the most widely deployed CCTV coverage in the world. In 2013, the British Security Industry Authority (BSIA) estimated that there are up to 5.9 million CCTV in the UK equating to 1 camera for every 11 people. With increasing emphasis on national and global security, there is a growing and acute need for human recognition (e.g. identifying or searching victims/witnesses/suspects) from videos.

---

X. Wei (✉)

School of Computing Science, Newcastle University, Newcastle upon Tyne NE1 7RU, UK  
e-mail: xingjie.wei@ncl.ac.uk

C.-T. Li

Department of Computer Science, University of Warwick, Coventry CV4 7AL, UK  
e-mail: C.T.Li@warwick.ac.uk

Biometrics is the science of identifying an individual based on the physiological and behavioural characteristics. The physiological characteristics are related to the shape of the body including face, iris, retina, fingerprint, palmprint, palm vein, hand geometry, DNA, earlobe, etc. The behavioural characteristics are related to the pattern of behaviour of a person such as gait, signature, keystroke dynamics, voice, etc. Among these biometric traits, face is the most commonly seen and used for human recognition due to the fact that its collection is non-intrusive and requires less cooperation from the subjects. Everyone has a face and it is widely accepted as a means of recognition.

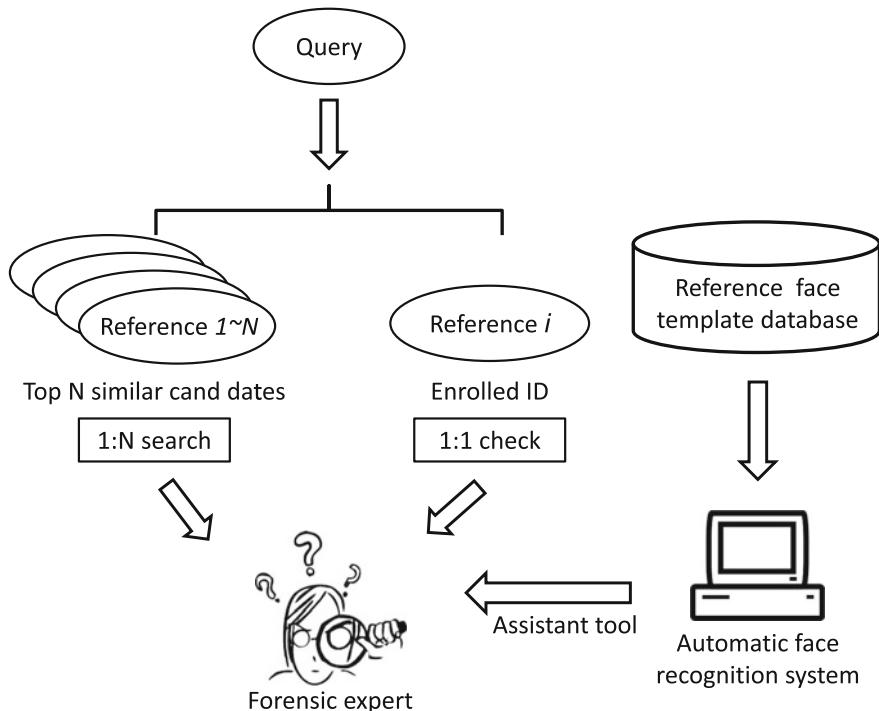
Forensic science is the scientific knowledge and technical methods in gathering and examining traces which might be useful for the investigations of crime. The biometric traces are of great interest for both biometrics and forensics research [40]. A large number of biometric applications such as fingerprint identification systems and computerised DNA databases, were implemented to serve forensic purposes. Among them, the face recognition system is becoming increasingly important due to the abundance of image and video data provided by surveillance cameras, mobile devices, Internet social networks, etc. The videos collected during an investigation needs to be evaluated by investigators throughout the forensics process. The evaluation is to explore whether additional line of questions can be identified and make sure that the current investigated actions have been completed.

There are broadly two types of evaluation during a forensics process<sup>1</sup>: *investigative evaluation* and *evidential evaluation*. The investigative evaluation concerns (1) *what is known* and (2) *what is unknown*, (3) *what are the consistencies* and (4) *conflicts* in the cases. On the other hand, the evidential evaluation focuses on the *relevance*, *reliability* and *admissibility* of the materials considering issues such as *the overall strength of the case* and *whether sufficient evidence exists against the offender to proceed to charge*.

In a typical forensic face recognition scenario, a forensic expert is given face images of a suspect and a person in question, who may or may not be the suspect. The forensic expert gives a value which represents the degree to which these images appear to be of the same person. When a large amount of images and videos have been gathered, identifying the possible suspects manually is extremely time consuming. Face recognition technologies can be used as a tool in the forensic scenarios for helping with queries against a large enrolled database. The query can be a *one-to-many* search, or a *one-to-one* check, as shown in Fig. 8.1. The West Yorkshire Police in UK has tested a face recognition system for matching CCTV images against a database of mugshots and they have reported that although the technology is not fully mature, it has proven to be a useful investigation tool [46]. We will introduce the face recognition technologies for forensics evaluation of video traces in the rest of this chapter.

---

<sup>1</sup>National Police Library, <http://www.college.police.uk/>.



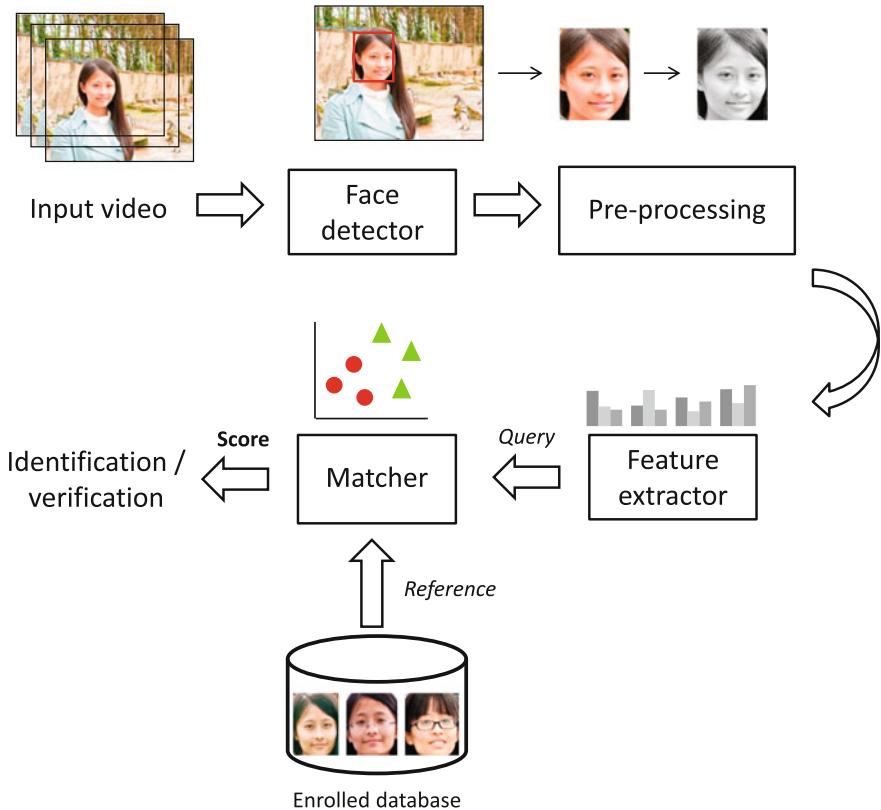
**Fig. 8.1** The process of the forensic face recognition augmented with automatic face recognition system on queries against a large enrolled database. The query can be a one-to-many search or a one-to-one check

## 8.2 Automatic Face Recognition

A general automatic face recognition system usually consists of the following modules: a face detector, a feature extractor and a matcher, as shown in Fig. 8.2. The face detector first determines which image area contains a face. With the detected face pre-processed, facial features are extracted by the feature extractor and are fed to the matcher as input for comparison against the features of the enrolled images. A similarity score between the query face and each enrolled face is generated by the matcher and used for recognition in either identification mode or verification mode.

### 8.2.1 Face Detection

The face detector locates and tracks the face area from the background of an image or video frames. Facial components such as eyes, nose and mouth are located based on the facial landmarks. As shown in Fig. 8.2, this *first-detection/tracking-then-*



**Fig. 8.2** Basic framework of the automatic face recognition. It usually consists of a face detector, a feature extractor and a matcher. Recognition typically works in identification mode or verification mode

*recognition* framework is applied in most image-based face recognition systems. For face recognition from videosFace recognition from videos, approaches based on this scheme first selects key frames (with good image quality, ideal face size, pose, etc.) and then perform detection/tracking and recognition sequentially. On the other hand, in order to deal with the uncertainties in tracking as well as in recognition, some methods [32, 57] perform simultaneous tracking and recognition using the temporal information in videos. Such *tracking-and-recognition* provides more flexibility for face recognition from videos. It is applicable to the *image-to-video* and *video-to-video* recognition scenarios.

### 8.2.2 Feature Extraction

Usually preprocessing such as face alignment and normalisation by the facial landmarks are performed before feature extraction. The face area is cropped from an image and normalised according to its geometrical properties (e.g. size and pose) using geometrical transforms or morphing. Usually a detected face area is further normalised with respect to its photometrical properties (e.g. illumination). After that, the feature extractor extracts discriminative information from the face image for distinguishing different individuals. The features can be holistic representations of the face [7, 47], or local patterns around certain salient points of the face [3, 38], depending on the methodology of the matching algorithm. Formulating an effective feature set is a nontrivial task. An effective feature set ideally should be able to characterise the discriminating features in a compact manner. Redundant features not only add little value, but also reduce the power of other useful features. An overly expressive feature set may also lead to the so-called *Curse of Dimensionality*, which requires dimension reduction in order to facilitate efficient matching.

### 8.2.3 Matching

The matcher compares two faces according to the extracted features producing a similarity score for the ensuing face recognition to be based on. There are two modes in face recognition: *identification* and *verification*, corresponding to the forensic tasks in Fig. 8.1. In the identification mode, the face recognition system matches the features of the query face to the features of each enrolled reference face template in the database. The templates and query sample can be still images or videos. An ordered list of the *top n* most similar candidates are returned as the possible identities of the query according to the similarity scores. The performance of the system in the identification mode is measured in terms of *rank-n* identification rate which is the rate at which the true association is included in the top *n* matches. The identification rate usually refers to *rank-1* identification rate where the system returns a single match (the best match), as the most probable association with the query face. This rate is also called the *recognition rate*.

On the other hand, verification is the task where the recognition system attempts to confirm an individual's claimed identity by comparing the query sample to the individual's previously enrolled reference templates. Verification is based on a decision threshold which is set by computing the similarity scores of all samples pairs in the database. The threshold is chosen to separate the genuine (i.e. match) similarity scores distribution from the impostor (i.e. non-match) similarity scores distribution and gives the best performance based on one of the following metrics. Here an impostor is a person who submits a sample attempting to claim the identity of another person.

- *False acceptance rate* (FAR) is the rate at which the comparison between two different individuals' samples is erroneously accepted by the system as the true match. In other words, FAR is the percentage of the impostor scores which are higher than the decision threshold.
- *False rejection rate* (FRR) is the percentage of times when an individual is not matched to his/her own existing reference templates. In other words, FRR is the percentage of the genuine scores which are lower than the decision threshold.
- *Equal error rate* (EER) is the rate at which both acceptance and rejection errors are equal (i.e.  $\text{FAR} = \text{FRR}$ ). Generally, the lower the EER value, the higher the accuracy of the system.

The first fully automatic face recognition system was presented by Kanade [28] in 1973, which marked a milestone at that time. Over the past four decades, there has been significant progress in automatic face recognition. However, albeit the fact that some unconstrained factors have been considered in the still-image based face recognition, recognising face from videos remains a relatively less explored area due to various factors, such as low resolution, occlusions, lack of subject's cooperation, constantly changing imaging conditions, etc. In the next section we will briefly introduce the challenges in face recognition from videos in the real-world environment.

### 8.3 Face Recognition from Videos Traces

During the *London Riots* in 2011, the London Metropolitan Police used automatic face recognition software attempting to find the suspects from video traces, but in many cases it failed [16]. The poor quality of most CCTV footage makes it very difficult to trust standard automatic facial recognition techniques. Changes in illumination, image quality, background and orientation can easily fool the face recognition systems.

In 2013, Klontz and Jain [30] conducted a case study that used the photographs from the CCTV footage of the two suspects in the *Boston Marathon bombings* to match against a background set of mugshots. The suspects' images released by the FBI were captured in the uncontrolled environment and their faces were partially occluded by sunglasses and hats. The study showed that current commercial automatic face recognition systems have the notable potential to assist law enforcement. But the matching accuracy was not high enough, suggesting that further technical progress needs to be made in order to be able to recognise faces from CCTV footage taken from unconstrained environments.

Compared with still images, videos contain more information: e.g. multiple views of the face or different illumination conditions and temporal information. However, face recognition from CCTV footages is an even more challenging problem than recognising faces from still images taken from a short distance. First of all, there are two main difficulties for face tracking [59]: (1) real-world videos (e.g. surveillance

**Table 8.1** Unconstrained variations in face recognition from video traces

Caused by users	Pose
	Facial expression
	Occlusion
Caused by environment	Illumination
	Low resolution
	Blur (motion blur and focus blur)

videos) are usually of low resolution, which contains less useful information of faces, and (2) illumination changes have a strong impacts on the accuracy of tracking.

In addition, face recognition from videos in real scenarios usually involves the remote face recognition (i.e. *face recognition at a distance* (FRAD)), which is also one of the most challenging problems in face recognition. *Remote face* usually refers to the face data captured at 10 m or further away from the cameras. FRAD is often related to the security and defense applications. Compared with the traditional face recognition (e.g. near-distance face recognition), FRAD is faced with two difficulties: (1) less or even no user-cooperation is available, and (2) the environment condition is more difficult to control. The FRAD algorithms need to consider more serious unconstrained variations in face data in the real-world scenarios. These variations are mainly due to two main factors: the user and the environment, as shown in Table 8.1.

The approaches of recognising face from videos can be divided into three categories [6, 9]: (1) *set-based approaches*, *sequence-based approaches* and *dictionary-based approaches*. The set-based approaches [49] regard a video as a set of unordered images and the recognition robustness is achieved by exploiting the multiple viewing conditions contained in the image set. Traditional still-image based methods can be extended to videos image sets. Recognition is performed by fusing the matching results from multiple frame-pairs at different levels (i.e. feature level, score level and decision level), or by measuring the similarity between the query and reference subspaces/manifolds which are learned from image sets. On the other hand, sequence-based approaches explicitly utilise the temporal information between video frames during recognition. This category of approaches exploits the appearance as well as the motion information of faces to obtain better recognition accuracy. To utilise the temporal information, time series state models such as sequential importance sampling (SIS) [56], and Hidden Markov Models [37] are used to model the video sequences. In recent years, matching face images based on sparse representation via a dictionary becomes popular [13]. *Dictionary* here is a large collection of face images where every individual has multiple images containing a wide range of variations such as poses, expressions and illumination changes. Face frames from a given query video are projected onto the subspace spanned by the atoms in the dictionary and the representation residuals are computed for matching.

## 8.4 Handling Uncontrollable Factors Present in Videos

Videos provide abundant information in the form of multiple frames and temporal information compared to still images. These information can be exploited for improving the performance of face recognition and provide robustness to large variations in facial poses, expressions and illumination conditions, occlusions and low image quality factors. In the following sub-sections, we introduce face recognition approaches according to the uncontrollable factors they are dealing with. Although some approaches are developed for still images, they do shed light on the new development for video-based systems.

### 8.4.1 Approaches for Handling Pose Variations

Forensics experts, or even ordinary people are able to recognise a person by the face from different views. However, due to the complex 3D structures of faces, the view generalisation is nontrivial and often ill-posed task for automatic face recognition systems. The main challenge of pose problem is that appearance variations caused by variable poses of the same person (i.e. intra-class variations) can be larger than those caused by identity differences (i.e. inter-class variations). The similarity between two faces from different persons in the same pose can be larger than that of the same person in different poses.

For pose-invariant face recognition, a natural solution is to collect multi-view images/videos which increase the chances of the face being captured in a favourable frontal pose. This is straight-forward and the frontal face recognition algorithms can be directly extended to solve the non-frontal pose problem. Usually pose estimation, which is the process of inferring the orientation of a head, is needed when processing the multi-view data. Then view selection is performed to select images/frames with ideal view conditions based on the pose information. Li et al. [35] estimate the head pose using the SVM regression. The pose information is used to choose the appropriate face detector for multi-view face detection which provides improved performance in terms of accuracy and speed. Only the frontal faces are retained for recognition.

Another direction is to model the pose variations. One of the popular methods is the face manifold. An image of a given size can be viewed as a high-dimensional vector in a Euclidean image space where the dimensionality equals to the number of pixels. The surface and texture of a face is mostly smooth confining it to an embedded face manifold. The basic idea of face manifold-based methods is to cluster the images of similar pose and train a linear subspace to represent each pose cluster. Face manifold is approximated by the linear subspace model. In this way, two face manifolds of similar poses can be compared under a small variation in pose parameters. Arandjelović and Cipolla [5] propose a method, which first decomposes an appearance manifold to Gaussian pose clusters then fuses the fixed-pose comparison

results using a neural network for recognition. Wang et al. [49] define the manifold–manifold distance as the distance between the gallery manifold learned from the training gallery image set and the probe manifold learned from the probe image set.

The above methods deal with the pose variations by performing explicit pose estimation or model registration from the multi-view image set or videos. In a surveillance environment, these processes are still very challenging due to the fact that the image quality is poor and the calibration of cameras is not accurate. Du et al. propose a pose-insensitive feature [14] which does not require explicitly estimate the pose of the face. The proposed feature is developed using the spherical harmonic representation of the face texture-mapped onto a sphere. The texture map itself is generated by back-projecting the multi-view video data. One limitation of the method is that the pose-insensitive feature relies on the assumption that the spherical function remains unchanged other than a rotation. The assumption works in normal illumination conditions but does not hold in extreme lightings.

Another limitation for pure 2D image-based methods as described above is that they assume the pose transformation is continuous within the 2D space. On the other hand, approaches with assistance of 3D models [11, 27] which are estimated from videos achieve better performance when addressing pose variations. Compared with 2D image-based methods, 3D model-based methods usually incur a higher computational cost.

#### 8.4.2 Approaches for Handling Occlusion Variations

In the real-world environments, faces are easily occluded by facial accessories (e.g. sunglasses, scarf, hat, veil) or objects in front of the face (e.g. hand, food, mobile phone). These occlusions can largely change the appearance of the face, which makes the face detection more difficult. In forensic investigations, although occluded faces can be manually detected and cropped, recognising partially occluded faces images is still challenging for automatic systems.

One intuitive idea is to first determine whether a face image is occluded or not [44], and then reject the occluded images in applications. This rejection mechanism is not always suitable for face recognition especially in forensic scenarios where no alternative image can be obtained due to the lack of subject cooperation. Some approaches first segment the occluded regions from face images and then perform recognition based on the remaining parts [25, 26, 41]. These models require a skin colour-based *occlusion mask* to detect the occluded areas in faces. The occlusion detectors are usually trained on specific types of occlusions (i.e. the training is data-dependent) and hence generalise poorly to various types of occlusions in real-world environments.

So performing recognition with the presence of occlusions is very challenging. There are three main categories of approaches for face recognition without detecting occlusion in advance: (1) *reconstruction-based approaches*, (2) *local matching-based approaches* and (3) *occlusion-insensitive feature-based approaches*.

The first category, reconstruction-based approaches treats occluded face recognition as a reconstruction problem [51, 53, 55]. A clean image is reconstructed from an occluded query image by a linear combination of reference images. Then the occluded image is assigned to the class with the minimal reconstruction error. A common drawback of reconstruction-based methods is that they usually require a large number of samples per subject to represent a query image. Most of them assume that the reference/training images are captured in well controlled conditions. However, this assumption does not usually hold in real-world scenarios. Another drawback is this category of approaches usually incur a high computational cost [53].

The second category is the local matching-based approaches [39, 45, 50, 52]. Facial features are extracted from local areas of a face, for example, overlapping or non-overlapping patches of an image, so the affected and unaffected parts of the face can be analysed in isolation. In order to minimise matching errors due to occluded parts, different strategies such as local space learning [39, 45], multi-task sparse representation learning [36] or voting [50] are performed. The common intuition behind the local matching-based approaches is based on that the facial appearance changes caused by occlusions are local in nature. Only part of a face is distorted by occlusions while others are less affected and reliable for recognition. So compared with the reconstruction-based methods, local matching-based methods are less likely to be able to handle the situation in which more than half of the face is occluded.

In addition to the above approaches, which focus on improving the robustness to occlusions during the matching stage, researchers also pay attention to image representation. The third category is occlusion-insensitive feature-based approaches [48, 58] which attempts to extract occlusion-insensitive features from face images. Tzimiropoulos et al. [48] pointed out that PCA learning in the gradient orientation domain with a cosine-based distance measure helps reduce the effects due to occlusions in face images. The distribution of image gradient orientation (IGO) differences and the cosine kernel provide a powerful way to measure image correlation/similarity when image data are corrupted by occlusions. The face representations learned from the image gradient orientations are relatively invariant to the occlusion effects. Inspired by their work, Zhu et al. [58] further proposed a Gabor phase difference representation for occluded face recognition. They find that the Gabor phase (GP) difference is more stable and robust than gradient orientation to occlusions.

### ***8.4.3 Approaches for Handling Illumination Variations***

Due to the 3D structure and various surface reflectance of faces, light sources can cast shading and shadows, creating nonuniform illumination on faces, which accentuates or diminishes certain facial features. The differences induced by this impact in the facial appearance can be greater than that between individuals.

There are two categories of approaches for addressing the illumination problem—*active* and *passive* ways [60]. The active approaches attempt to obtain face images which are invariant to the illumination changes. Usually specific devices such as 3D scanners, thermal cameras, infrared cameras, etc. other than the visible light cameras are required. A good survey on 3D face recognition can be found in [1]. Thermal images and near-infrared images are more insensitive to large illumination changes as compared to visible light images. Introductions to illumination invariant face recognition using thermal images and near-infrared images are presented in [17, 34], respectively.

On the other hand, the passive approaches attempt to directly deal with images which have already been affected by illuminations. There are usually three classes of approaches. (1) Illumination normalisation [12, 15], which seeks to suppress the illumination variations either by image transformations or by synthesising an unaffected image from affected ones, (2) illumination invariant representation [3, 18], which attempts to extract features invariant to illumination changes, and (3) illumination variation modelling. Illumination core [8, 20] and sparse representation [53] are based on the theoretical principle that the set of images of a convex Lambertian object [31] obtained in a wide variety of illumination conditions can be approximated by a low-dimensional linear subspace in which the recognition can be performed. These methods require well-registered face images and the availability of face images with different illumination conditions. Such requirements limit their applicability in practice, especially for videos captured in unconstrained environments. Some models for pose variations are also applicable to solving of the illumination problem. Using the manifold model introduced in Sect. 8.4.1, the illumination variation for each of the pose clusters can be modelled using a linear, pose-specific illumination subspace. Given a reference template and a novel cluster with the same pose, the illumination changes can be normalised by adding a vector from the pose illumination subspace to the frame so that its distance from the reference cluster is minimised [5].

#### 8.4.4 *Approaches for Handling Low Image Quality Variations*

In security-related face recognition applications such as surveillance, usually the face images/videos captured are degraded by low resolution and blur effects. When sufficient videos are available, one simple idea is to select a set of frames which yield the best recognition accuracy by a classifier. This can help to remove or give lower weight to the poor quality frames during recognition.

For the low resolution problem, another intuitive solution is the super-resolution (SR) based method [4]. SR is a technique for synthesising high-resolution images from low resolution images for visual enhancement. After applying SR, a higher resolution image can be obtained and then used for recognition, for example, matching a face in a low resolution surveillance footage against a set of higher quality gallery

sequences enrolled in a database. One common drawback of SR-based face recognition approaches is that SR does not directly contribute to recognition. The identity information may be contaminated by some artefacts attributed to the SR process.

Another category of approaches do not apply the SR preprocessing to low resolution images. Li et al. [33] proposed a method to learn coupled mappings (CMs), which minimises the difference between the low resolution image and its high-resolution counterpart. Then the low resolution image is projected onto a unified feature space where higher recognition performance can be achieved. Biswas et al. [10] proposed a method using Multidimensional Scaling (MDS) to transform the low resolution gallery and probe images into an Euclidean space such that the distances between them approximates the best distances. Shekhar et al. [43] propose a generative approach to low resolution image-based on learning class specific dictionaries, which is also robust to illumination variations.

There are two types of effect attributed to the blur problem: *focus blur* and *motion blur*. A focus is the point where lights originating from a point on the object converge. When the light from object points is not well converged, an out-of-focus image with the blur effect will be generated by the sensor (e.g. camera). The work in [24] analysed the impact of out-of-focus blur on face recognition performance. On the other hand, motion blur is due to the rapid object movement or camera shaking. Blurring affects the appearance of faces in images, causing two main problems [42]: (1) the appearance of face images from the same individual changes significantly due to blur, and (2) different individuals tend to appear more similar when blurred due to the loss of discriminative features. There are two main categories of approaches to improve the quality of the blurred face images: (1) blurred image modelling through subspace analysis [42] or sparse representation [54], and (2) blur-tolerant descriptors which attempt to extract blur insensitive features such as Local Phase Quantisation (LPQ) [2, 23] to represent the face images.

## 8.5 Future Trends

As introduced in the last section, the appearance variations caused by the unconstrained conditions are still challenging for face recognition from images and videos. This section will discuss several specific face recognition problems, which are the new trends of research in both biometrics and forensics communities.

### 8.5.1 Combining with Other Biometric Traits

When faces are heavily occluded or degraded due to extreme conditions, face recognition technologies become ineffective. In unconstrained environments, the face is not the only trait used by humans to recognise each other. It is natural to combine face and other biometric data to improve the recognition performance. These data is

either from other modalities such as voice, gait [21] or soft biometric features such as height, gender, hair, skin and clothing colour. The advantages of using such features are: (1) they can be captured without constraint in uncontrollable environments, and (2) they can be captured along with the face using the same sensor such as a CCTV camera. How to represent the features from different modalities and how to fuse these features and matching scores will be the important issues for future investigations.

### 8.5.2 Contending with the Face Ageing Issue

Facial ageing is a complex process that affects both the shape and texture (e.g. skin tone or wrinkles) of a face. The typical scenario of face recognition across ageing is to detect if a particular person is present in a previously recorded database. Applications include missing children identification, suspect watch-list check, etc. For still-image based recognition, ageing effect has been studied in two directions: (1) developing *age estimation techniques* to classify face images-based on age [19, 22] and (2) developing *ageing robust systems* to perform recognition. However, the ageing issue are seldom considered in video-based face recognition algorithms. One most challenging aspect of face recognition involving the ageing issue is that it must address all other ‘historical’ unconstrained variations as well. Pose, expression, illumination changes and occlusions can occur when images are taken years apart.

Compared to still images, videos contain the temporal information which is of great value for face recognition. It is interesting to investigate into the ways of utilising the temporal information effectively to deal with the ageing issue.

### 8.5.3 Different Imaging Modalities

Face recognition across different imaging modalities, also called *heterogeneous face recognition* [29], is another interesting area for further explorations. It involves matching two face images from different imaging modalities, which is of great practical value in forensic scenarios. The images of suspects may come from various sources, e.g. still images captured from CCTV, footages taken by the police helicopters or images snapped by members of the public. In addition, in some extreme situations, only a particular modality of a face image is available. For example, in night-time environments, infrared imaging may be the only modality for acquiring a useful face image of a suspect. But the mug-shots held by the police are visible band images. Another example is the sketch-photograph matching. When no photograph of a suspect is available, a forensic sketch is often generated according to the description of an eye-witness. Matching sketches against face photographs is very important for forensic investigation. On the other hand, 2D-3D face matching is expected to attract intensive research efforts in the near future since face can be represented by heterogeneous features in the 3D and 2D modalities in the real-world cases.

### 8.5.4 Other Issues in Forensic Tasks

The face recognition technologies discussed in previous sections mainly focus on how to improve the recognition accuracy from face images and videos. This is essential in forensic investigation and case linking. Besides that, there are other requirements in other forensic tasks [40]. For example, in forensic identification, such as identifying missing people, besides recognition accuracy, the other challenges lie in the development and management of reference databases. How to increase the integrity, quality and interoperability of the template data with the help of face image processing or analysis technologies is an important issue. For forensic evidence evaluation, the challenges are not only about the development of automatic methods, but also the integration of expert-based and automatic methods into hybrid methods.

## 8.6 Summary

The face is one of the most popular biometric traits used in the daily life for human recognition. The widespread use of CCTV cameras for surveillance and security applications have stirred extensive research interests in video-based face recognition. Face recognition can play an important role in identifying perpetrators of crime activities as well as missing peoples. Automatic face recognition technology is becoming an indispensable tool for modern forensic investigations.

In this chapter we have introduced the advanced face recognition technologies. The past decades have seen significant progress in automatic face recognition. But the performance of the face recognition from videosFace recognition from videos taken in unconstrained environments is still unsatisfactory. Uncontrollable illumination, pose changes, low image quality and occlusions pose acute challenges to face recognition techniques. Therefore, intensive research efforts to contend with these interweaving factors are required in the years to come.

## References

1. Abate AF, Nappi M, Riccio D, Sabatino G (2007) 2D and 3D face recognition: a survey. *Pattern Recognit Lett* 28(14):1885–1906. *Image: Information and Control*
2. Ahonen T, Rahtu E, Ojansivu V, Heikkila J (2008) Recognition of blurred faces using local phase quantization. In: International conference on pattern recognition (ICPR), pp 1–4
3. Ahonen T, Hadid A, Pietikainen M (2006) Face description with local binary patterns: application to face recognition. *IEEE Trans Pattern Anal Mach Intell* 28(12):2037–2041
4. Arandjelović O, Cipolla R (2007) A manifold approach to face recognition from low quality video across illumination and pose using implicit super-resolution. In: IEEE international conference computer vision (ICCV), pp 1–8
5. Arandjelović O, Cipolla R (2009) A pose-wise linear illumination manifold model for face recognition using video. *Comput Vis Image Underst* 113(1):113–125

6. Barr JR, Bowyer KW, Flynn PJ, Biswas S (2012) Face recognition from video: a review. *Int J Pattern Recognit Artif Intell* 26(5)
7. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans Pattern Anal Mach Intell* 19(7):711–720
8. Belhumeur PN, Kriegman D (1996) What is the set of images of an object under all possible lighting conditions? In: IEEE conference computer vision and pattern recognition (CVPR), pp 270–277
9. Bhatt HS, Singh R, Vatsa M (2014) On recognizing faces in videos using clustering-based re-ranking and fusion. *IEEE Trans Inf Forensics Secur* 9(7):1056–1068
10. Biswas S, Bowyer KW, Flynn PJ (2010) Multidimensional scaling for matching low-resolution facial images. In: IEEE international conference on biometrics: theory, applications, and systems (BTAS), pp 1–6
11. Castillo CD, Jacobs DW (2009) Using stereo matching with general epipolar geometry for 2d face recognition across pose. *IEEE Trans Pattern Anal Mach Intell* 31(12):2298–2304
12. Chen T, Yin W, Zhou XS, Comaniciu D, Huang TS (2006) Total variation models for variable lighting face recognition. *IEEE Trans Pattern Anal Mach Intell* 28(9):1519–1524
13. Chen YC, Patel VM, Phillips PJ, Chellappa R (2012) Dictionary-based face recognition from video. In: Fitzgibbon A, Lazebnik S, Perona P, Sato Y, Schmid C (eds) European conference computer vision (ECCV), vol 7577. Lecture notes in computer science. Springer, Berlin, pp 766–779
14. Du M, Sankaranarayanan AC, Chellappa R (2014) Robust face recognition from multi-view videos. *IEEE Trans Image Process* 23(3):1105–1117
15. Du S, Ward R (2005) Wavelet-based illumination normalization for face recognition. In: IEEE international conference on image processing (ICIP), vol 2, pp II-954–7
16. Firth N (2011) Face recognition technology fails to find UK rioters. *New Sci*
17. Gabriel H, del Solar JR, Verschae R, Correa M (2012) A comparative study of thermal face recognition methods in unconstrained environments. *Pattern Recognit* 45(7):2445–2459
18. Gao Y, Leung MKH (2002) Face recognition using line edge map. *IEEE Trans Pattern Anal Mach Intell* 24(6):764–779
19. Geng X, Zhou Z-H, Smith-Miles K (2007) Automatic age estimation based on facial aging patterns. *IEEE Trans Pattern Anal Mach Intell* 29(12):2234–2240
20. Georghiades AS, Belhumeur PN, Kriegman D (2001) From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Trans Pattern Anal Mach Intell* 23(6):643–660
21. Guan Y, Wei X, Li C-T, Marcialis GL, Roli F, Tistarelli M (2013) Combining gait and face for tackling the elapsed time challenges. In: IEEE international conference on biometrics: theory, applications, and systems (BTAS), pp 1–8
22. Guodong G, Fu Y, Dyer CR, Huang TS (2008) Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Tran Image Process* 17(7):1178–1188
23. Hadid A, Nishiyama M, Sato Y (2010) Recognition of blurred faces via facial deblurring combined with blur-tolerant descriptors. In: 2010 20th international conference on pattern recognition (ICPR), pp 1160–1163
24. Hua F, Johnson P, Sazonova N, Lopez-Meyer P, Schuckers S (2012) Impact of out-of-focus blur on face recognition performance based on modular transfer function. In: IAPR international conference biometrics (ICB), pp 85–90
25. Jia H, Martínez AM (2008) Face recognition with occlusions in the training and testing sets. In: IEEE international conference automatic face and gesture recognition (FG), pp 1–6
26. Jia H, Martínez AM (2009) Support vector machines in face recognition with occlusions. In: IEEE conference computer vision and pattern recognition (CVPR), pp 136–141
27. Jiang D, Hu Y, Yan S, Zhang L, Zhang H, Gao W (2005) Efficient 3d reconstruction for face recognition. *Pattern Recognit* 38(6):787–798. *Image Understanding for Photographs*
28. Kanade T (1973) Picture processing system by computer complex and recognition of human faces. In: Doctoral dissertation, Kyoto University

29. Klare B, Jain AK (2010) Heterogeneous face recognition: matching NIR to visible light images. In: International conference on pattern recognition (ICPR), pp 1513–1516
30. Klontz JC, Jain AK (2013) A case study on unconstrained facial recognition using the Boston marathon bombings suspects. Technical Report MSU-CSE-13-4
31. Lambert J (1760) Photometria sive de mensura et gradibus luminis. Colorum et Umbrae, Eberhard Klett
32. Li B, Chellappa R (2002) A generic approach to simultaneous tracking and verification in video. *IEEE Trans Image Process* 11(5):530–544
33. Li B, Chang H, Shan S, Chen X (2010) Low-resolution face recognition via coupled locality preserving mappings. *IEEE Signal Process Lett* 17(1):20–23
34. Li SZ, Chu R, Liao S, Zhang L (2007) Illumination invariant face recognition using near-infrared images. *IEEE Trans Pattern Anal Mach Intell* 29(4):627–639
35. Li Y, Gong S, Sherrah J, Liddell H (2004) Support vector machine based multi-view face detection and recognition. *Image Vis Comput* 22(5):413–427
36. Liao S, Jain AK, Li SZ (2013) Partial face recognition: alignment-free approach. *IEEE Trans Pattern Anal Mach Intell* 35(5):1193–1205
37. Liu X, Chen I (2003) Video-based face recognition using adaptive hidden markov models. In: IEEE conference computer vision and pattern recognition (CVPR), vol 1, pp I–340–I–345
38. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *Int J Comput Vis* 60(2):91–110
39. Martínez AM (2002) Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Trans Pattern Anal Mach Intell* 24(6):748–763
40. Meuwly D, Veldhuis R (2012) Forensic biometrics: from two communities to one discipline. In: International conference of the biometrics special interest group BIOSIG, Darmstadt, Germany, pp 1–12
41. Min R, Hadid A, Dugelay J-C (2011) Improving the recognition of faces occluded by facial accessories. In: IEEE international conference automatic face and gesture recognition (FG), pp 442–447
42. Nishiyama M, Hadid A, Takeshima H, Shotton J, Kozakaya T, Yamaguchi O (2011) Facial deblur inference using subspace analysis for recognition of blurred faces. *IEEE Trans Pattern Anal Mach Intell* 33(4):838–845
43. Shekhar S, Patel VM, Chellappa R (2011) Synthesis-based recognition of low resolution faces. In: IEEE international joint conference on biometrics (IJCB), pp 1–6
44. Storer M, Urschler M, Bischof H (2010) Occlusion detection for ICAO compliant facial photographs. In: IEEE conference computer vision and pattern recognition workshops (CVPRW), pp 122–129
45. Tan X, Chen S, Zhou Z-H, Liu J (2009) Face recognition under occlusions and variant expressions with partial similarity. *IEEE Trans Inf Forensics Secur* 4(2):217–230
46. Travis A (2008) Police trying out national database with 750,000 mugshots. MPs told. *The guardian*
47. Turk MA, Pentland AP (1991) Face recognition using eigenfaces. In: IEEE conference computer vision and pattern recognition (CVPR), pp 586–591
48. Tzimiropoulos G, Zafeiriou S, Pantic M (2012) Subspace learning from image gradient orientations. *IEEE Trans Pattern Anal Mach Intell* 34(12):2454–2466
49. Wang R, Shan S, Chen X, Gao W (2008) Manifold-manifold distance with application to face recognition based on image set. In: IEEE conference computer vision and pattern recognition (CVPR), pp 1–8
50. Wei X, Li C-T (2013) Fixation and saccade based face recognition from single image per person with various occlusions and expressions. In: IEEE conference computer vision and pattern recognition workshops (CVPRW), pp 70–75
51. Wei X, Li C-T, Hu Y (2012) Robust face recognition under varying illumination and occlusion considering structured sparsity. In: International conference digital image computing techniques and applications (DICTA), pp 1–7

52. Wei X, Li C-T, Lei Z, Yi D, Li SZ (2014) Dynamic image-to-class warping for occluded face recognition. *IEEE Trans Inf Forensics Secur* 9(12):2035–2050
53. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009) Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell* 31(2):210–227
54. Zhang H, Yang J, Zhang Y, Nasrabadi NM, Huang TS (2011) Close the loop: joint blind image restoration and recognition with sparse representation prior. In: IEEE international conference computer vision (ICCV), pp 770–777
55. Zhang L, Yang M, Feng X (2011) Sparse representation or collaborative representation: which helps face recognition? In: IEEE international conference computer vision (ICCV), pp 471–478
56. Zhou S, Krueger V, Chellappa R (2003) Probabilistic recognition of human faces from video. *Comput Vis Image Underst* 91(1–2):214–245. Special issue on face recognition
57. Zhou SK, Chellappa R, Moghaddam B (2004) Visual tracking and recognition using appearance-adaptive models in particle filters. *IEEE Trans Image Process* 13(11):1491–1506
58. Zhu J, Cao D, Liu S, Lei Z, Li SZ (2012) Discriminant analysis with Gabor phase for robust face recognition. In: IAPR international conference biometrics (ICB), pp 13–18
59. Zou WW, Yuen PC, Chellappa R (2013) Low-resolution face tracker robust to illumination variations. *IEEE Trans Image Process* 22(5):1726–1739
60. Zou X, Kittler J, Messer K (2007) Illumination invariant face recognition: a survey. In: IEEE international conference on biometrics: theory, applications, and systems (BTAS), pp 1–8

# **Chapter 9**

## **Human Factors in Forensic Face Identification**

**David White, Kristin Norell, P. Jonathon Phillips and Alice J. O'Toole**

**Abstract** Facial identification by forensic examiners is a core component of criminal investigations and convictions. These identifications are often done in challenging circumstances that require experts to match identity across images and videos taken at various distances, under different illumination conditions, and across a wide range of poses. Until recently, laboratory studies of human face identification have concentrated, almost exclusively, on face identification by untrained (naïve) observers, with only a handful of studies focusing directly on the accuracy of expert forensic facial examiners. Over the last two decades, DNA-based exonerations of convicted criminals in the United States have revealed weaknesses in the forensic identification process due to *human factors*. In this paper, we review and analyze the factors known to impact facial identification accuracy for both naïve participants and trained experts. Combined, these studies point to a set of challenges that impact accuracy for both groups of participants. They also lead to an understanding of the specific conditions under which forensic facial examiners can surpass naïve observers at the task of face identification. Finally, we consider the role that computer-based face recognition systems can play in the future of forensic facial identification. These systems have made remarkable strides in recent years, raising new questions about how human and machine strengths at face identification can be combined to achieve optimum accuracy.

---

In (Eds. M. Tistarelli, C. Champod) *Biometrics in Forensic Sciences*.

---

D. White  
University of New South Wales, Sydney, Australia

K. Norell  
Swedish National Forensic Centre, Linköping, Sweden

P.J. Phillips  
National Institute of Standards and Technology, Gaithersburg, USA

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

## 9.1 Introduction

Identification by forensic facial analysis experts contributes evidence in criminal cases in most countries that have functioning governments and court systems. Societies and individuals have a vested interest in assuring the accuracy and credibility of these identification judgments. Accurate judgments contribute to communities that are safe from dangerous criminals and also uphold the rights of innocent people from being falsely accused and convicted of crimes. What do we know *scientifically* about the accuracy of these forensic facial identification judgments? In a perfect world, forensic examiners would receive feedback on the correctness of every judgment they make in a criminal or court case. This is not possible when corroborating evidence in a case which is weak or when a suspect maintains innocence, even after conviction. Unfortunately, incorrect forensic judgments are detected only when some other form of corroborating evidence, such as a DNA test or a clear alibi, provides convincing counterevidence to the accuracy of the judgment.

In recent years, the United States has seen a disturbing number of forensic judgments, and indeed the criminal convictions supported by these decisions, set aside based on evidence from belated DNA testing of convicted felons [59]. These cases have involved forensic examiners with expertise in wide variety of evidence types (e.g., face, fingerprint, and hair). As DNA testing has became more widely available and accessible to investigators and adjudicators, a steady stream of DNA-based exonerations have come to light in the United States, bringing with them an intense public discussion about the role of forensics in the justice system. This prompted the National Academy of Sciences in the United States to issue a report aimed at strengthening forensic sciences [42]. One recommendation of the report was a call for scientific research aimed at understanding the factors that impact accuracy in forensic analysis.

The goal of this chapter is to review what we know scientifically about the accuracy of forensic facial examiners. We begin with a brief statement of the nature of the facial identification problem solved by forensic examiners. The remainder of the chapter is organized as follows. First, we discuss the factors known to affect human face recognition accuracy in laboratory studies of people without training or professional experience in forensic face recognition. Second, we will review studies that compare the accuracy of professional forensic facial examiners to that of untrained participants. This literature is sparse, but growing, with several recent studies testing professional forensic examiners. These studies have expanded our knowledge of how professionals may differ, both qualitatively and quantitatively, from untrained individuals. In the third section of the chapter, we discuss advances in computer-based face recognition software and assess their strengths and weakness in comparison to both untrained and professional populations. We conclude with a road map for a way forward. We will argue that the route forward should include strategies for retaining the advantages and skills of forensic face examiners, while understanding the inherent limits of forensic facial identification science. We

will also consider the potential of combining human and computer-based biometric “systems” to their best advantage.

### **9.1.1 *The Problem***

The nature of the analysis presented by an expert typically involves an identity comparison made between two or more images and/or videos of faces. Often these images and videos are captured under photometric conditions (illumination, viewpoint, and distance) that vary substantially and across a variety of media types. Forensic facial examiners commonly deal with media types that include high quality passport images, drivers’ license photos, surveillance video taken at the scene of a crime, and image/video data taken by bystanders using handheld mobile devices.

In a perceptual or laboratory version of the task, participants are asked to match the identity of two individuals in pairs of images, usually presented simultaneously. In casework, as in laboratory studies, the response comes in the form of a simple judgment of whether the two (or more) images depict the same person or different people. Almost always, this judgment is accompanied by a rating of confidence or certainty in the judgment. In casework, the certainty rating provides a guide as to whether the judgment can be used as evidence in a legal setting. In laboratory studies, certainty ratings make the data amenable to standard signal detection analysis and to the construction of receiver operator characteristic (ROC) measures that can be compared across perceivers and conditions [36]. This type of analysis also facilitates comparisons between expert forensic facial examiners and untrained people tested in laboratory conditions and between computer-based face recognition systems and both types of participants.

## **9.2 Characteristics of Human Face Recognition Relevant for Forensics**

### **9.2.1 *Familiarity***

The most common face identification task that humans perform is not the image comparison task performed in forensic examination. Rather, in our daily lives, we recognize faces of colleagues, friends, and family by matching their face to a representation that is stored in our memory. Although an impressive computational feat, this skill does not require special training to perform accurately. Familiar face recognition is the primary route to person identification in our daily lives [75]. Indeed, through evolutionary and developmental pressures, the human brain has developed robust processes for this important task. As a result, people are excellent

at identifying familiar faces, enabling them to recognize high school classmates from yearbook photographs with very high accuracy, 15 years after graduating [3].

Our impressive abilities with *familiar* faces appear to stem from a representational system that is specifically tuned to face perception. Given our highly developed perceptual apparatus, human performance in this task has been exalted as the “gold-standard” for face identification, to which machine-based systems should aspire [61]. This is a sensible aim, given that human face recognition displays many desirable properties for a recognition system. For example, we are able to accurately identify familiar faces from very low quality closed caption television (CCTV) images [12]. Importantly, memory representations supporting familiar face recognition are highly tolerant of within-identity variation caused by changes in image-specific properties. This is particularly important for face recognition because of the relatively large within-identity variation (due to image- and person-based factors; see Sect. 9.2.2) in comparison to the subtle differences between faces that are diagnostic of identity. Being able to recognize familiar faces despite this very low signal-to-noise ratio is the core computational challenge of face recognition.

Importantly, our tolerance of image variation is strongly contingent on our familiarity with the face [8, 24]. This has been demonstrated consistently in studies of face identification using *unfamiliar* faces, which we discuss in more detail in the next section. For now, to demonstrate this important distinction and the challenge presented by unfamiliar faces, we invite the reader to estimate how many people are pictured in Fig. 9.1. Jenkins et al. [24] presented student participants with a similar array (containing pictures of Dutch celebrities) and asked them to sort the images into piles by identity; such that each pile contained a single person, with a unique



**Fig. 9.1** How many people are pictured in this array? (From [24]). See text for the correct answer

pile for each person. Students from the Netherlands, who were familiar with the faces, correctly answered that there were only two people present in the array. However, students from the United Kingdom, who were unfamiliar with the faces in the array, saw many more than two identities. These British participants sorted the photographs into an average of seven piles, indicating that they believed there were seven people in the array.

The Jenkins et al. [24] study demonstrates that face identification accuracy is strongly constrained by the degree of familiarity we have with a face. This is also true of the types of decisions typically made in forensic facial comparison; where two images are presented simultaneously and participants have to decide if the images are of the same person or two different people. When matching high quality images of unfamiliar faces, taken under very similar lighting conditions and on the same day, people make a large proportion of errors on this apparently straightforward task (e.g., Megreya and Burton [39], Burton et al. [11]). However, when the faces are familiar to participants, this task becomes trivial. In a series of studies, Clutterbuck and Johnston [14–16] compared performance in this task across faces of varying levels of familiarity. Their data show a graded improvement in face matching as faces become familiar, leading the authors to suggest that performance in pairwise matching provides a sensitive index of a person's familiarity with a face [15].

These findings are consistent with earlier work by Bruck et al. [10], who tested participants' ability to match yearbook photographs of classmates taken 25 years previously. This matching ability was better for classmates than for people who were not previously classmates. Remarkably, this was true despite the fact that the classmates had not seen the people in the photographs during the intervening 25 years. When images of faces were taken from different angles, match accuracy was adversely affected for control participants, but not for classmates. Therefore, the benefit of familiarity was remarkably sensitive and long lasting. The findings are also similar to those observed by Jenkins et al. [24], in showing that participants who were familiar with the faces were able to match identity across variation in the image-level qualities (in this case, head angle). This invariance is a key characteristic of familiar face representation (see also Jiang et al. [25]).

The ability of our visual systems to produce these higher-order representations is the main source of our impressive face recognition skills. Where these representations are not available, the power of human face recognition system is not fully exploited and performance is far less impressive. Unfortunately, in forensic practice and most other applied settings, the full benefits of this expert system are not available because the faces encountered are almost always *unfamiliar*. Because of the very clear advantages of familiarity, recent work has attempted to improve the accuracy of face matching by leveraging familiarity from multiple images of an individual [2, 13, 67]. In parallel to similar work to develop methods in machine learning (see Sect. 4.3), the hope is that methods that make best use of the image data for the purpose of face learning may benefit forensic practice in future.

It is important to note that the potential for leveraging familiarity in this way is limited to cases where multiple images of an individual are available.

Unfortunately, in most instances, forensic casework is based on comparison of single images. As a result, forensic methods tend to focus on developing strategies for one-to-one image comparison of unfamiliar faces. In such cases, the research summarized in this section should serve to underline that forensic examiners should not generalize the very accurate identification performance seen with familiar faces to the task performed in their daily work.

### 9.2.2 *Image and Demographic Factors*

Human face recognition has been studied for decades, primarily under laboratory conditions in which participants learn the identity of an individual from a single image and are tested subsequently with a different image. In addition, there are many studies in which people compare face identity perceptually in two images that are simultaneously available. From both recognition and perceptual matching studies, much is known about the factors that impact our ability to extract and compare identity information from face images. These factors can be divided into *stimulus factors*, *subject factors*, and *interactive factors*. Interactive factors result from specific combinations of stimulus factors and subject factors.

Stimulus factors encompass all of the ways that images of individuals can vary. There are two categories of stimulus factors. *Image-based factors* are the result of the way a particular image is taken. They include illumination, viewpoint (i.e., pose), distance, resolution, and other parameters inherent to the image capture process. *Person-based* factors are changes inherent to the person's appearance. These include changes in facial expression, facial hair, accessories (e.g., glasses, tattoos, make-up, etc.), as well as changes that occur on a longer time scale (aging, facial surgery, and weight gain/loss). *Categorical factors* refer to the sex, age, and race of the person to be recognized, as well as the “typicality” of the person relative to these reference groups. *Subject factors* pertain to characteristics of the perceiver, including their sex, race, and age. The only interactive factor we will consider is the “other-race effect”, which is the well-known finding that people recognize faces of their own race more accurately than faces of other races [37].

#### 9.2.2.1 *Stimulus Factors*

*Image-based factors*. To a first approximation, the story here is simple. People recognize faces best when the image they learn is captured under photometric conditions similar to the conditions under which the test image is captured. This is true for recognition over changes in illumination (e.g., Braje et al. [7], Hill and Bruce [22]) and viewpoint [47, 65]. In general, there is no “optimal view” for recognition, although several studies indicate that recognition is best from images that show a face in three-quarter profile (e.g., Logie et al. [31], O’Toole et al. [31, 47]). Notably, face recognition is extremely difficult when face images are

presented upside down [74]. It is also difficult when face images are presented in the photographic negative [20]—a manipulation that produces an image similar to what would be seen if a face were illuminated from below. The effects of image inversion and photographic negation suggest that human face processing operates in a mono-oriented fashion and with sensitivity that is not limited to the contours of the face image, but extends to the subtle shading variations that are disrupted with face inversion.

The rule that face recognition performance varies directly with the closeness of the viewing parameters between the learning and test images finds exception in the case of image resolution. Liu et al. [30] showed that identity matching is more accurate with a high-resolution and low-resolution image, than with two low-resolution images. They also showed that the recognition of high-resolution and low-resolution faces is generally equivalent when participants learned people from low-resolution gait videos. This suggests that the availability of a high-resolution image actually *improved* face-matching performance from a low-resolution image, perhaps by allowing for the creation of a more robust and generalizable representation from the high-resolution image.

*Person-based Factors.* It is tempting to conclude that person-based factors affect recognition in the same way as image-based factors—that is to say, when a person’s appearance is similar in two images (e.g., same age, same expression, same weight, and same make-up), recognition is more accurate than when appearance has changed in some way. Although this seems intuitively likely, and even obvious, remarkably few psychological studies have tested this question formally. Although systematic evaluations of these person-based factors have not been carried out, recent work shows a substantial drop-off in face-matching accuracy across relatively modest age spans. Megreya et al. [40] show that matching performance is reduced by around 20% when matching across age-differences of between one and two years, which is consistent with a more recent study (White et al. [69]; photo-to-photo test).

The sparseness of the literature on this topic is likely due the fact that most databases available for face recognition contain only one or two images of a person. These images are often taken over a relatively short time span, ranging from a few minutes up to a few months. In the last few years, far richer face databases, collected from web crawling, have become available. These databases now contain hundreds of images of the same person, taken over years and even decades. Recent use of these database resources for testing computer-based face recognition systems has contributed to the fast pace of progress in that field. These databases will likely become more prominent in psychological studies in the years to come, enabling more systematic assessments of the impact of person- and image-based factors on the accuracy of face identification in humans. In the meantime, photographic demonstrations of the variability in facial images of a single individual (see Fig. 9.1) provide compelling anecdotal evidence that person-based appearance changes (e.g., age, expression) make face identification more difficult—at least for unfamiliar faces.

*Categorical Factors.* Are some categories of faces easier to recognize than others? For standard demographic categories of faces, the answer to this question, based on what is currently known, is “no.” Although there have been limited tests to consider whether male versus female faces are more recognizable, findings of substantial differences in the literature are rare and inconsistent. This is also true when examining faces of different races. Faces are not differentially recognizable as a function of their race or ethnic heritage. Where age is concerned, the literature is sparse, but again, there is no indication that faces of particular age groups are more or less recognizable than faces of other-age groups [18, 71].

It would not be true, however, to say that all faces are equally recognizable. The primary face characteristic that predicts recognition success is its perceived typicality [29], such that the more similar a face is to the average, the less accurate recognition will be. This advantage for recognizing distinct or unusual faces helps to explain why face caricatures can be recognized easily. In caricatures, the distinctive or unusual features of a face are exaggerated. For instance, a large nose becomes even larger in a caricature. A paradox in understanding human face recognition abilities is that caricatures are highly recognizable, despite the degree of distortion of features and configural information introduced into the image. However, while caricatures improve familiar face recognition [57], recent attempts to improve unfamiliar face recognition have produced quite modest benefits. McIntyre et al. [38] asked if this same technique could be used to improve the veracity of photo-identification, and found evidence that participants showed enhanced ability to discriminate matching from non-matching photo-ID when images were subjected to subtle levels of caricature manipulation.

### 9.2.2.2 Subject Factors

The *perceiver* also introduces variability into the process of face recognition. Two effects are worth noting. First, girls and women are more accurate at face recognition than boys and men (cf. Herlitz and Lovén [21] for a review). This is not a large effect, but it is reasonably consistent. Second, and perhaps of more consequence, several studies have indicated that face recognition accuracy declines as we age [4, 18]. In particular, older people are more likely to falsely recognize an unfamiliar person as “known” than are younger people. Importantly however, this same effect of aging does not appear to hold for face identification tasks that do not involve memory, such as the image comparison tasks typical of forensic identification (see Burton et al. [11]).

### 9.2.2.3 Interactive Factors

In some cases, recognition accuracy varies as a combined function of the demographics of the perceiver and the person to be identified. The other-race effect—the finding that people recognize faces of their own race more accurately than faces of

other races [37]—is a well-known example of this kind of interaction. The phenomenon has been documented consistently in the literature and is associated with large effects. The general nature of this effect is clear and it has been found with diverse groups of participants, stimuli, and tasks (e.g., recognition: O’Toole et al. [46]; perceptual matching Phillips et al. [24, 55], Megreya et al. [41]; and categorization: O’Toole et al. [45]). For forensic examiners, this basic finding of an own-race advantage for recognition has far-reaching consequences that have not been explored.

Although less studied, a handful of recent papers have begun to consider whether an analogous “other-age” effect exists for face recognition [21]. This type of effect would also have implications for the accuracy of forensic facial examinations in particular situations where the examiner and face-to-be-examined are diverse in age. To date, although some studies have reported other-age effects in a range of tasks [1], others have not [71]. Moreover, recent work has indicated that findings of other-age effect might be contingent on experience [34] or the type of task being performed [35]. A definitive answer to this question requires more experimental data.

In summary, the factors that affect face recognition accuracy for untrained people include the familiarity of the face, stimulus factors that stem from the conditions under which an image is captured, subject factors (e.g., age, sex), and interactions between the stimulus and subject. In the next section, we consider the skills of expert facial examiners on the task of unfamiliar face image comparison.

### **9.3 Are Facial Image Comparison “Experts” More Accurate at Facial Image Comparison Than Untrained People?**

Despite the importance of knowing the answer to the question we pose in the heading of this section, professional facial forensic examiners have been tested directly in only a handful of scientific studies. Moreover, the few available studies have not provided converging evidence on the question of whether professionals are more accurate than naïve subjects. An early study by Wilkinson and Evans [72] (cf. also, Wilkinson and Evans [72]) tested two forensic examiners and found that they were superior to untrained participants on the task of matching identity between CCTV video and a small set of reference images. Participants in that study viewed a video showing a person’s head, either with or without a baseball cap, and were asked to determine whether the person also appeared in one of the reference images. Experts were more accurate both at detecting correct matches and rejecting false ones. Notably, when the entire head was visible, experts were better on all four measures of identification: hits, false alarms, correct rejections, and misses. One limitation of this study, however, is that only two examiners were tested. These examiners, who were also authors of the study, may not be representative of the overall population of experts.

In two other studies, trained participants fared no better than naïve participants when comparing identity in across a video and still images [9, 28]. Both of these studies tested a larger number of trained subjects, although the training and experience characteristics of the subjects differed. Lee et al. [28] tested students in a forensics course and Burton et al. [12] tested police officers with an average of 13.5 years of experience. These, and other, critical differences in the methods and tasks might account for the divergence of findings between these studies and Wilkinson and Evans [72, 73]. Though seemingly a minor change in the protocol, the task used by Burton et al. [12] required participants to view a series of videos prior to the identification task. This task involves more than perceptual matching. Instead, participants had to *remember* multiple individuals across a series of videos—a more difficult task than the perceptual matching task of Wilkinson and Evans [72]. Likewise, Lee et al. [28] tested participants with very poor quality videos. This is also more challenging than the task used in Wilkinson and Evans [72], albeit for different reasons. Image quality, as we shall see, is important in predicting the relative performance of experts and naïve observers. Notwithstanding, in these earlier studies it is not possible to know whether the different findings are due to the level of difficulty, the use of different test paradigms, or the different types of experience and training of the participants.

In a more naturalistic study, White et al. [69] tested passport officers and untrained students on their ability to match photo-ID cards to the card-bearers. Comparing photos to “live” people in this way mimics a ubiquitous task performed continuously at our national borders and in numerous other applications where an identity must be verified with an ID card for access to valuable resources (e.g., bank accounts) or entrance to restricted locations (e.g., embassies). In a first experiment, thirty passport officers with an average of over eight years experience falsely accepted “fraudulent” (i.e., nonmated) photo-IDs on 14% of trials. When participants presented a matching photo, these valid IDs were falsely rejected on 6% of trials. This substantial error rate is somewhat surprising given the task and the background of the participants. Notably, White et al. [67, 69] found no relationship between performance and years on the job, despite large individual differences in accuracy and a wide spread of experience levels.

In a second experiment, White et al. [67, 69] compared passport officers to untrained students on a photo-to-photo identity comparison test. Participants in this study viewed two images and were asked to determine whether the images showed the same person or different people. White et al. [67, 69] also compared accuracy for laboratory-style images taken two years apart and for image pairs consisting of a currently valid Photo-ID (of indeterminate age) and a recent high quality photograph. There was no overall difference in accuracy for the passport officers and students, for either the matching or non-matching identity trials. However, the type of photographs used for matching affected performance, with lower accuracy when participants matched official Photo-IDs to high quality recent images than when they matched laboratory images taken two years apart. In a final experiment, passport officers were tested with the Glasgow Face Matching Test (GFMT; Burton et al. [11])—a standardized test of photo-to-photo face-matching ability. Passport

officers' scores on the GFMT did not differ from existing normative scores from untrained participants.

It is worth noting, that the finding of comparable performance between passport officers and students should be considered in the context of the full range of performance duties assigned to the passport issuance officers in this study. Although clearly experienced in face matching, these face-matching tasks are performed in the context of many other eligibility checks involving biographical and other personal information in the passport application process. Thus, the primary focus of the job may not be the face identification task. Further, the training these passport officers received was limited to a short module provided as part of introductory training. This larger context issue applies also to the police officers tested by Burton et al. [12], and even more so, to the participants tested by Lee et al. [28]. It is possible, therefore, that advantages will be seen only in cases where face identification is done in a focused and deliberate way—such as in criminal investigations or for court testimony. Professionals who perform this type of task are likely to participate in intense and sustained training and to be cognizant of the importance of each and every decision.

To address this, two recent studies tested individuals who perform face identification in these more directed applications. In the first study, Norell et al. [43] tested the skills of 17 European forensic experts, using a facial image comparison task. These experts had experience in facial image comparison and have been called on to apply their knowledge to criminal casework for legal authorities. The primary goal of the study was to determine whether image quality affected the accuracy of the detectives in matching identity. Additionally, the study also included a comparison group of untrained university students, whose performance was compared with that of the experts.

For the experiment with trained examiners, Norell et al. [43] compiled a test of 30 one-to-one facial image comparisons. In each case, a reference image was paired with a test image. Reference images were high quality images taken under studio conditions, whereas test images were taken using a camera intended for surveillance video. The time lag between the reference and test images varied between several months to several years, thereby assuring some change in appearance (e.g., hair style, etc.) between two images that show the same identity. For images that showed different identities, “similar” people were paired to make the task challenging.

Image quality was varied for the test images by changing the zoom and distance parameters of the camera to blur the image. Norell et al. [43] tested identification with three image qualities, allowing the participants to directly examine the images, with no time constraints. An additional innovation of this study was to ask participants to respond on a 9-point *conclusion scale*, similar to the scales used in forensic facial image comparison casework in many European countries. This response procedure enabled participants to grade their responses according to nine levels of certainty, varying from *very strong support of the same source* (i.e., “the observations extremely strongly support that it is the same person”) to *very strong support of different source* (i.e., “the observations extremely strongly support that it

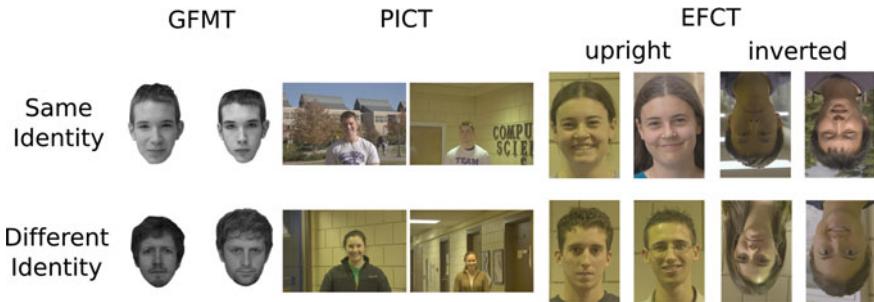
is not the same person”), with a mid-point indicating *no support for either conclusion* (“the observations support neither that it is the same person nor that it they are different persons”).

This conclusion scale allows, not only for the standard classes of errors, but also for a noncommittal response, which is neither correct nor incorrect. By standard classes of errors, we mean the following. For pairs that show the same identity, a *false negative* is defined as failing to affirm that the images are of the same person. For pairs of images that show different people, a *false positive* is defined as falsely judging the images to be of the same person. In forensic identification, it is often not possible to make an identity judgment based on the image evidence and so an important part of an examiners’ role is to assess whether the images are suitable for comparison (cf. Ulery et al. [66]).

Interestingly, Norell and colleagues found that noncommittal responses revealed differences between the experts and students. First, experts made fewer errors than students. Notably however, as image quality decreased, experts increased their use of the noncommittal response, whereas students did not. With high quality images, experts were considerably more accurate than students—more often determining that matching images depicted the same person. Students, on the other hand, may have hesitated to affirm the same identity in images that differed substantially. A particularly interesting finding was that the difference in errors for the students versus experts was smallest for the *lowest quality images*. This might suggest that experts are able to access more useful information from high quality images than untrained observers. It might also suggest that training may help experts to exercise more caution (i.e., by using the noncommittal response) when image quality is low.

Finally, the number of false positives is something that is generally seen as a serious problem in the forensics field, because in most cases, it means that an innocent person would be implicated as guilty. In Norell et al. [43], experts had fewer false positives than the untrained students, consistent with similar biases toward making “non-match” responses for fingerprint examiners [64, 66]. Interestingly, studies of radiologists screening x-ray images have shown opposite response bias in these experts, with false positive errors being made more frequently than “miss” errors (see e.g., Kundel et al. [27]). Encouragingly, this suggests that expert performance is calibrated to counteract the risk of costly errors.

A second study of face identification accuracy in forensic examiners has recently been carried out by White et al. [70]. They tested a large group of facial forensic examiners attending a Facial Identification Scientific Working Group meeting (FISWG; [www.fiswg.org](http://www.fiswg.org)) in May, 2014. This meeting was hosted at the FBI Academy in Quantico, Virginia, and was attended by an international group of forensic examiners who have very rigorous training and many years of experience in facial identification (7 years on average). To test the expertise of this group, White et al. created a battery of tests of pairwise facial image comparison (see Fig. 9.2). These tests were created to model the types of challenging decisions that might be encountered by forensic facial image examiners in their daily work, by selecting image pairs that were challenging to both machine recognition systems and also to



**Fig. 9.2** Examples of image pairs from face-matching tests in White et al. [70]. Examiners outperformed control participants on the Glasgow Face Matching Test (GFMT), Person Identification Challenge Test (PICT), and Expertise in Facial Comparison Test (EFCT). Images in the GFMT were all taken on the same day and under very similar lighting conditions, whereas images from the PICT and EFCT were captured in diverse ambient conditions and over a longer time period. For details see text below

untrained student participants. White and his colleagues also mapped the performance of this group to established normative performance using the GFMT [11]. Examiners performance on the three tests was compared to a control group of FISWG attendees who did not perform facial identification in their daily work, and to a group of untrained student participants.

The results of these tests showed a consistent ordering of performance across three groups with examiners more accurate than controls, and controls more accurate than students. Notably however, examiner performance was not perfect for any of the tests. Moreover, performance was quite variable within the examiner group; with some examiners performing extremely well and others performing much less impressively. Following from recent work showing that matching performance is improved by aggregating the responses across individuals [44, 68], White et al. [70] “fused” responses of the examiners and other groups by averaging match score ratings across examiners, controls, and students. Consistent with previous work, combining responses in this way produced near-perfect human performance on these very challenging tasks, and far better performance than had been achieved by computer algorithms on the same image pairs. In another recent study, student participants who performed the GFMT collaboratively showed similar gains in accuracy [17]. Evidence from that study also suggests that collaboration may carry an additional training benefit, by improving performance of those who initially performed poorly in the task.

These fusion results are encouraging because in forensic practice it is common for examiners to confer when making identity judgments. Therefore, better understanding of the group decision-making process in forensic face identification promises to produce substantial gains in accuracy and can inform the design of peer review processes in forensic identification.

In their study of forensic experts, White et al. [70] were also motivated to examine the nature of examiners' perceptual expertise. They therefore tested examiners' accuracy under different exposure durations and with different image orientations (cf., Yin [74]). First, the results indicated that examiners were more accurate both when making quick decisions (with 2 s limit) and when making slower decisions (with 30 s limit). However, their superiority was greatest when they were provided with the longer exposure time to reach their decision. This is evidence that examiners' superiority relied, to some extent, on the analytic and strategic analysis of images that is taught in training. Second, the data also showed that examiners were less impaired than controls or students when the images were presented upside down. The size of the "inversion impairment"—known in face recognition research as the face inversion effect (FIE)—is an important measure in the study of human face recognition performance. This is due to the fact that inversion impairs face recognition performance far more than object recognition. As a result, the FIE has been considered as an index of human "expertise" with faces. Because of this qualitative difference in examiner performance, we suggest that their superiority on the face comparison task relies on a type of expertise that is different from standard human abilities for faces. As noted, there is evidence to suggest that human face processing is specialized to operate in a mono-oriented fashion [74].

We speculate that the smaller inversion effect for examiners, relative to controls and students, may be because the examiners are treating the face images not, as faces *per se*, but as *images*. This account would explain why the expertise of the examiner group was more pronounced when the faces were presented upside down: training in forensic examination may have caused them to rely less on the face-specific processes that govern performance in the population at large, by revealing alternative routes to identification. For example, forensic examiners are taught to compare images for "unexplainable" and "exclusionary" differences (e.g., Scientific Working Group Imaging Technology [60]), which may reveal inconsistency in local details that are missed by untrained participants. For now this account is speculative, but it is nevertheless important to develop a theoretical understanding of the underpinnings of examiners' expertise in facial image comparison to inform training and recruitment of facial examiners in the future.

The FISWG test represents the first effort to apply a broad approach to testing the state-of-the-art in forensic face identification. However, it is important to note that the study has limits in the extent to which it can be applied as a performance benchmark for practitioners. First, because this was a perceptual test, it was not designed to recreate the normal working conditions of forensic examiners. The motivation of this research was to examine perceptual expertise in this special population, rather than to provide a rigorous evaluation of their working practices. Because of this, examiners were not provided with unlimited time or allowed access to digital tools that they would use to support decisions in their daily work. In future benchmarking exercises it will be important to administer tests in a manner that is more reflective of working practice, in order to establish best working practices and the limits of identification accuracy in facial forensic examination.

## 9.4 Can Computer-Based Face Identification Systems Address Weaknesses of the Forensic Examiner and the Forensic Examination Process?

There has been a decade-long effort to directly compare the accuracy of human and machine performance on face identification [51]. These human–machine comparisons have been carried out in conjunction with competitions organized and implemented by the National Institute of Standards and Technology (NIST), with performance measured on identity matching for pairs of still images and videos. The still image types have varied in these competitions from high quality images taken in a studio environment to low quality images taken with handheld cameras such as those found in cell phones. These facial image comparison tasks are quite similar to those given to forensic facial examiners. Indeed, the intended applications for machine-based face recognition include tasks currently done by human face recognition experts. Thus, understanding machine performance in the context of both trained and untrained humans will facilitate the design of human–machine processes for face identification.

One of the key goals of the NIST competitions has been to facilitate the advancement of automatic face recognition technology. For all participants (i.e., algorithms) in a competition, NIST reports the accuracy on a common set of face images. The reported accuracy allows the face recognition community to assess the performance of state-of-the-art algorithms on a controlled set of images/videos. Because the human face identification studies have been conducted in conjunction with the NIST competitions, human–machine benchmarks now exist for state-of-the-art algorithms on a variety of image/video types that span high quality studio images to handheld point and shoot camera data from mobile devices.

In what follows, we will first outline some key elements of the human–machine comparisons including the definition of an “unfamiliar face recognition” task for an algorithm. We also consider some critical parameters in measurement of human performance in this task. Next, we provide a brief synopsis of where machine performance stands relative to untrained and trained human perceivers.

### 9.4.1 *Unfamiliar Face Recognition Tasks for Machines*

As noted, the task done by most automatic face recognition systems is to determine whether two images depict the same person or different people. For face recognition algorithms to model “unfamiliar” face recognition (i.e., the task assigned typically to forensic face examiners), it is necessary to define what it means for a face to be familiar to an algorithm. Indeed, the notion of familiarity is conceptually different for machines than for humans. For humans, familiarity means experience with a wide variety of diverse images of a person that span the image- and person-based variability described in Sect. 9.2.2.1. For algorithms, there are four ways to produce

“familiarity effects.” All of these improve the performance of an algorithm based on gains derived from the algorithm’s “experience” with face identities and/or face images. In all but one case, the mechanisms behind the improvements are unrelated to the mechanism we assume for human familiarity. Notwithstanding, there is a relationship between the gains achieved with these nonhuman familiarity mechanisms and the benefits conferred by familiarity, as defined for humans.

First, an algorithm might become familiar with a person, if other images of the person are present in the database used to train the algorithm. Training is used in many face recognition algorithms to find “features” that are statistically reliable indicators of identity [19]. Training can also serve a variety of other functions involved in tuning algorithm parameters. When faces of the same identity are present in both the training and test set, there is a substantial increase in algorithm accuracy [33]. It is worth noting that this is an indirect definition of familiarity, because it does not entail providing the algorithm with additional specific and labeled information about individual people.

Second, consider the case where there are multiple images of a face in the set of images used as test images. Algorithm performance is increased in this case as well [56], because the existence of multiple test images increases the likelihood of finding a close match to the reference image.

The third way to achieve effects similar to familiarity is to “over-tune” an algorithm to a face dataset. Particular databases are often characterized by a common set of imaging parameters (viewpoints, illumination, resolution). When an algorithm is trained with a subset of images from a database, the machine becomes tuned to the characteristics of those images. This tuning can provide an accuracy advantage when the algorithm is tested with another subset of images from the same database. Importantly however, this is due to familiarity with the dataset as a whole and so nothing is learnt about individual identities. Instead, this familiarity is based on global characteristics of the training set, without propagating to the level of identity.

A careful look at the literature shows that algorithms tend to over-tune to the best available database, perhaps at the cost of thinking more generally about the problem of face identification. For publicly available datasets, there is always a competition to develop face recognition algorithms that are more accurate than the previous algorithm. As newer and more challenging databases are collected, over time, benchmark algorithm performance invariably reaches 95% or above—often crossing the 99% threshold. Although some of the progress in performance is linked to improvements in algorithmic techniques, it is likely that some progress is also due to over-tuning algorithm design to the data set. A face recognition algorithm consists of numerous components. The first part is selecting components that optimize performance on a particular data set. These components include pre-processing steps, feature extraction methods, feature normalization, and classifier selection. Another method is to select a training set tuned to the test data. For example, if a test set consists primarily of faces looking to the left, then a training set consisting of left-looking faces will optimize performance on the test set; however, the resulting algorithm will not necessarily generalize to frontal faces.

The effects of over-tuning may look like familiarity in that they improved algorithm performance via an understanding of the image variability that is present in the dataset. However, over-tuning to a single data set risks under-tuning to general image features that are not captured in that dataset. Ultimately, over-tuning is good, only if the database is highly diverse in terms of capturing image- and person-based face variability. If the database is not diverse, performance generalization to other kinds of images is typically poor.

A simple illustration of over-tuning can be seen in the other-race effect found for algorithms participating in the NIST Face Recognition Grand Challenge (FRCG) competition [52]. Algorithms from East Asian countries performed better on East Asian faces and algorithms from Western countries performed better on Caucasian faces. The likely reason for this other-race effect for algorithms is that algorithms may have over-tuned to faces from locally available data sets.

The fourth definition of algorithm familiarity is the closest to the human definition. This definition refers to the case in which an algorithm has access to multiple images and videos of a single person. This allows algorithms to use diverse images/videos to construct a unitary representation of a face that encompasses the variability of a person's appearance in different imaging conditions and under different person-based factors (aging, glasses, facial hair, etc.). The field is just beginning to go in this direction and we expect that the coming years will see computational examples that harness this variability to capture some basic elements of becoming familiar with a face. We return to this point in the Sect. 4.3.

#### ***9.4.2 Measuring Human Performance for Comparison with Machines***

Measuring human performance in a way that relates to algorithm performance requires researchers to make methodological choices that have implications for estimates of human versus machine standing. As noted, human and machine comparisons have been carried out regularly in the NIST tests, starting in 2007. At that time, the task was to compare images with controlled and uncontrolled illumination [50]. In the most recent test, image comparisons are made with image/video taken from a handheld mobile device camera data (Phillips et al. in press; cf. Phillips and O'Toole, 2014 for a review of these studies). Intervening studies have provided comparisons that track the progress of face recognition algorithms [48, 49].

In all of the NIST comparisons, human performance was measured with an *aggregation* method—a standard behavioral protocol for measuring human performance. In this method, each subject rates a pair of images or videos, with a comparison scale. If a subject rates  $m$  pairs and there are  $n$  subjects, this will produce  $nm$  ratings and an ROC is computed separately from each of the  $nm$  ratings. The aggregate ROC measures are then computed to produce an estimate of average human performance over a population of image pairs.

The human–machine comparisons in the NIST tests have spurred algorithm developers to regularly include man–machine comparisons for new face recognition algorithms. As such, other studies introducing automatic face recognition algorithms have followed suit. These algorithms have operated on large-scale, publicly available datasets such as Labeled Faces in the Wild (LFW) [23] and on other web-based training sets. The LFW dataset consists of still face images of celebrities and famous people downloaded from the World Wide Web. In measuring human performance on these datasets, the method applied involves an implicit fusion of human ratings that ultimately estimates human performance to be more accurate than the estimates made with the standard aggregation protocol. Generally, these methods have relied on large numbers of Amazon Mechanical Turk<sup>1</sup> workers who provide ratings of the likelihood that pairs of images show the same person.

More concretely, in the fusion method, all  $n$  subject ratings for each image or video pair are averaged. This results in  $m$  average ratings, one for each image or video pair. From the  $m$  average ratings a single ROC can be computed, which is used to represent human performance. In our experience, fusing human ratings in this way produces a significant increase in the estimate of performance accuracy over the aggregate method and estimates increase as the number of people rating the image pairs increases [50, 68, 70] (Phillips et al. accepted) [5]. Indeed, accuracy obtained by fusion methods often exceeds accuracy of the best performing individual in the group [68, 70].

Although both ways of estimating human performance are fundamentally sound, it is important to understand that estimates reached by the aggregate and fusion methods will not necessarily agree. Most importantly for benchmarking accuracy in forensic examination, the type of method used to compute human accuracy should correspond to the method used in practice; otherwise unrealistic estimates will be obtained. In situations where identification decisions are based on the decisions of single examiners, an aggregate method will produce most reliable measures. However, where the possibility exists for judgments to be made by multiple examiners, fusion offers a useful method for estimating the benefit of combining responses across groups of examiners (see White et al. [70]).

#### **9.4.3 Measuring Human Performance for Comparison with Machines**

How well do algorithms currently perform relative to humans? The answer to this question depends on the type or image or video available for identity matching. As of 2007, state-of-the-art systems surpassed untrained humans matching identity

---

<sup>1</sup>The identification of any commercial product or trade name does not imply endorsement or recommendation by NIST.

between high quality images, with one image taken under controlled illumination and the other taken under uncontrolled indoor illumination [50]. These image pairs varied in appearance and in facial expression. By 2012, the best algorithms surpassed humans on matching identity in images that varied also across indoor and outdoor illumination [48]. In that case, however, a stratification of the dataset base into three levels based on algorithm performance showed very strong machine advantages in the two easier partitions of the data, but equivalent performance for humans and machines on the most difficult data partition [53].

In a follow-up study, Rice et al. [58] isolated the “worst-case scenario” for the algorithms, finding a set of image pairs on which the algorithms failed 100% of the time. These image pairs showed examples of highly different images of the same person and highly similar images of different people. Humans performed quite accurately with this sample. The study demonstrated, however, that human performance was supported *entirely* by matching the identity using the body, rather than the face.

Most recently, Phillips et al. [54] replicated the Rice et al. [58] with humans and machines on images and video from the Point and Shoot Challenge (PaSC) dataset [6]. This dataset contains images and video taken by handheld cameras in unconstrained environments with both indoor and outdoor illumination. For images, human performance again was much better than the benchmark algorithm and improved as additional human rater responses were fused or averaged. For the video, beginning with the worst-case data for the benchmark algorithm, human performance again was much better than the benchmark algorithms on the most challenging cases, but was not as good as the best available algorithm on the less challenging case.

In expanding beyond the NIST tests, Kumar et al. [26] reported a human-machine comparison for a face verification algorithm based on face attributes. Both the algorithm and Mechanical Turk Workers matched image pairs from the LFW data set. The human score for each face-pair was the average of all Mechanical Turk Workers that rated the pair. Humans in this case fared better than the algorithm, though again it was clear, as in Rice et al. [58], that human performance was supported, at least in part, by information in the body and in the external face (e.g., hair, shoulders, etc.).

In 2014, the first two reports appeared showing better performance for algorithms than for people on the LFW database. The successful algorithms used deep learning [32, 62, 63]. Note that estimates of human accuracy were obtained by the fusion method, which is likely to have led to inflated estimates of human accuracy on this challenging task, pointing to an even more impressive result by the algorithm. Note also that the faces in LFW dataset are celebrities and famous people. From the human side, this takes us full circle, back to an understanding that humans in this task are doing *familiar* face identification, and so they are operating at their best. From the algorithm side, the fact that the faces are those of celebrities, (for whom diverse images are readily available), makes it possible that the algorithms may have over-tuned to the people in these images. The inclusion of multiple

images of people, even as part of the training, may represent a case where algorithm results reflect critical components of the mechanisms humans employ for recognizing familiar faces. We may, therefore, be on the cusp of performance parity for humans and machines. In other words, when powerful algorithms can utilize rich and diverse training data for the identities to be recognized, they may be able to achieve the high level of performance shown by humans recognizing the faces of people they know well.

In summary, computer-based face recognition has been at the level of human recognition of unfamiliar faces for some years now. Progress is being made now on the greater challenge of moving machines to performance levels that are close to humans recognizing familiar faces. Some part of this progress is due to the development of more powerful algorithms than those available previously. The remainder is due to the rich datasets available from the web for training and testing these algorithms.

To return to the question of whether forensic examiners perform more accurately than algorithms, we have to ask another question first. Do forensic facial examiners perform unfamiliar face recognition at the level at which most people perform familiar face recognition? We do not yet know the answer to this question yet, but future research should make the goal of answering this question a priority.

## 9.5 Discussion and Future Directions

In summary, we reviewed studies that speak to the question of how accurate forensic face examiners are relative to untrained humans. We began with evidence indicating that there are substantial qualitative and quantitative differences between familiar and unfamiliar face recognition. The former represents human face recognition skills at their best and is characterized by a high degree of accuracy and an ability to generalize recognition across a wide range of image- and person-based variability. Because unfamiliar faces do not benefit from the rich memory representations that are tolerant to these sources of variability, recognizing unfamiliar faces is a far greater challenge. This leaves unfamiliar face matching “exposed to the elements” such that even small changes in facial appearance can lead to substantial decreases in matching accuracy. Although our knowledge of human performance on this task is almost entirely based on performance of untrained novices, the image-bound nature of unfamiliar face identification is also likely to place an upper bound on human accuracy in forensic face identification.

Recent studies of forensic examiners have begun to test this hypothesis empirically and as a result some characteristics of examiner performance have emerged. First, quantitative differences show that the majority of examiners tested exceed performance of untrained novices. It is also apparent from these studies that this superiority varies as a product of the nature of image evidence and test protocol. For example, although examiners outperform novices overall, their superiority is tied to

image quality. When image quality is poor, examiners do not necessarily perform better than untrained novices on the task. This qualitative difference provides an interesting parallel to algorithm–novice comparisons, where algorithm superiority to humans is most pronounced in conditions with high quality image pairs. Examiners’ superiority is also most apparent when they are provided with longer times to study image pairs. This suggests that their superiority is driven by a close and careful analysis of features in facial images, which is consistent with the training that these examiners receive.

Despite these indications that forensic training is driving forensic examiners’ superior accuracy, it will be important in the future to establish directly the extent to which this superiority is caused by training, experience, and natural aptitude in face matching. This is important not only for developing training and recruitment of the next generation of forensic examiners, but also in courtroom settings, to inform judges and juries as to the reliability of identification evidence. To further this aim, in future benchmarking exercises, it will not only be necessary to assess the impact of these and other examiner-based factors in isolation, but also to explore their interactions with the image and subject-based factors described earlier in this chapter. Such interactions have served to reveal important differences in previous research, and have served to deepen scientific understanding of cognitive processes supporting expert forensic examination.

Finally, improved understanding of the interactions between human and machine processing perhaps hold most potential for improving forensic identification in the future. Algorithms, novices, and experts deal with the same challenges in identifying faces, but they come to the problem with different strengths and limitations. Determining optimal divisions of labor in forensic investigation and criminal trials presents an important challenge in the years ahead; we hope that computer engineers and psychologists rise to this challenge together.

## References

1. Anastasi JS, Rhodes MG (2005) An own-age bias in face recognition for children and older adults. *Psychon Bull Rev* 12:1043–1047. doi:[10.3758/BF03206441](https://doi.org/10.3758/BF03206441)
2. Andrews S, Jenkins R, Cursiter H, Burton AM (in press) Telling faces together: learning new faces through exposure to multiple instances. *Q J Exp Psychol*. doi:[10.1080/17470218.2014.1003949](https://doi.org/10.1080/17470218.2014.1003949)
3. Bahrick HP, Bahrick PO, Wittlinger RP (1975) Fifty years of memory for names and faces: a cross-sectional approach. *J Exp Psychol Gen* 104(1):54–75. doi:[10.1037/0096-3445.104.1.54](https://doi.org/10.1037/0096-3445.104.1.54)
4. Bartlett JC, Leslie JE (1986) Aging and memory for faces versus single views of faces. *Mem Cogn* 14:371–381
5. Best-Rowden L, Bisht S, Klontz JC, Jain AK (2014) Unconstrained face recognition: establishing baseline human performance via crowdsourcing. In: 2014 IEEE international joint conference on biometrics (IJCB)

6. Beveridge JR, Phillips J, Bolme DS, Draper B, Givens GH, Lui YM, Teli MN, Hao Z, Scruggs WT, Bowyer KW, Flynn PJ, Cheng S (2013) The challenge of face recognition from digital point-and-shoot cameras. In: IEEE sixth international conference on biometrics: theory, applications and systems (BTAS), 2013
7. Braje WL, Kersten DJ, Tarr MJ, Troje NF (1998) Illumination effects in face recognition. *Psychobiology* 26(4):371–380
8. Bruce V (1982) Changing faces: visual and non-visual coding processes in face recognition. *Br J Psychol* 73(1):105–116
9. Bruce V, Henderson Z, Greenwood K, Hancock PJB, Burton AM, Miller P (1999) Verification of face identities from images captured on video. *J Exp Psychol Appl* 5:339–360
10. Bruck M, Cavanagh P, Ceci SJ (1991) Fortysomething: recognizing faces at one's 25th reunion. *Mem Cogn* 19(3):221–228
11. Burton AM, White D, McNeill A (2010) The glasgow face matching test. *Behav Res Methods* 42:286–291
12. Burton AM, Wilson S, Cowan M, Bruce V (1999) Face recognition in poor-quality video: evidence from security surveillance. *Psychol Sci* 10:243–248
13. Burton AM, Kramer RS, Ritchie KL, Jenkins R (in press) Identity from variation: representations of faces derived from multiple instances. *Cogn Sci*. doi:[10.1111/cogs.12231](https://doi.org/10.1111/cogs.12231)
14. Clutterbuck R, Johnston RA (2002) Exploring levels of face familiarity by using an indirect face-matching measure. *Perception* 31(8):985–994
15. Clutterbuck R, Johnston RA (2004) Matching as an index of face familiarity. *Vis Cogn* 11 (7):857–869
16. Clutterbuck R, Johnston RA (2005) Demonstrating how unfamiliar faces become familiar using a face matching task. *Eur J Cogn Psychol* 17(1):97–116
17. Dowsett AJ, Burton AM (in press) Unfamiliar face matching: pairs out-perform individuals and provide a route to training. *Br J Psychol*. doi:[10.1111/bjop.12103](https://doi.org/10.1111/bjop.12103)
18. Fulton A, Bartlett JC (1991) Young and old faces in young and old heads: the factor of age in face recognition. *Psychol Aging* 6:623–630
19. Furl N, Phillips PJ, O'Toole AJ (2002) Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis. *Cogn Sci* 26(6):797–815
20. Galper RE, Hochberg J (1971) Recognition memory for photographs of faces. *Am J Psychol* 84(3):351–354
21. Herlitz A, Lovén J (2013) Sex differences and the own-gender bias in face recognition: a meta-analytic review. *Vis Cogn* 2013:1. doi:[10.1080/13506285.2013.823140](https://doi.org/10.1080/13506285.2013.823140)
22. Hill H, Bruce V (1996) Effects of lighting on the perception of facial surfaces. *J Exp Psychol Hum Percept Perform* 22:986–1004
23. Huang GB, Ramesh M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst
24. Jenkins R, White D, Montfort XV, Burton AM (2011) Variability in photos of the same face. *Cognition* 121(3):313–323. doi:[10.1016/j.cognition.2011.08.001](https://doi.org/10.1016/j.cognition.2011.08.001)
25. Jiang F, Blanz V, O'Toole AJ (2007) The role of familiarity in three-dimensional view-transferability of face identity adaptation. *Vis Res* 47(4):525–531
26. Kumar N, Berg AC, Belhumeur PN, Nayar SK (2009) Attribute and simile classifiers for face verification. In: 2009 IEEE 12th international conference on computer vision, pp 365–372
27. Kundel HL, Nodine CF, Conant EF, Weinstein SP (2007) Holistic component of image perception in mammogram interpretation: gaze-tracking study 1. *Radiology* 242(2):396–402
28. Lee WJ, Wilkinson C, Memon A, Houston K, Res M (2009) Matching unfamiliar faces from poor quality closed-circuit television (CCTV) footage: an evaluation of the effect of training on facial identification ability. *AXIS: Online J CAHId* 1(1):19–28
29. Light LL, Kayra-Stuart F, Hollander S (1979) Recognition memory for typical and unusual faces. *J Exp Psychol: Hum Learn Mem* 5:212–228

30. Liu CH, Seetzen H, Burton AM, Chaudhuri A (2003) Face recognition is robust with incongruent image resolution: relationship to security video images. *J Exp Psychol: Appl* 9:33–44
31. Logie RH, Baddeley AD, Woodhead MM (1987) Face recognition, pose and ecological validity. *Appl Cogn Psychol* 1(1):53–69
32. Lu C, Tang X (2014) Surpassing human-level face verification performance on LFW with GaussianFace. arXiv preprint [arXiv:1404.3840](https://arxiv.org/abs/1404.3840)
33. Lui YM, Bolme D, Phillips PJ, Beveridge JR, Draper B (2012) Preliminary studies on the good, the bad, and the ugly face recognition challenge problem. In: 2012 IEEE computer society conference computer vision and pattern recognition workshops (CVPRW), pp 9–16
34. Macchi Cassia V (2011) Age biases in face processing: the effects of experience across development. *Br J Psychol* 102:816–829. doi:[10.1111/j.2044-8295.2011.02046.x](https://doi.org/10.1111/j.2044-8295.2011.02046.x)
35. Macchi Cassia V, Proietti V, Gava L, Bricolo E (2015) Searching for faces of different ages: evidence for an experienced-based own-age detection advantage in adults. *J Exp Psychol Hum Percept Perform.* doi:[10.1037/xhp0000057](https://doi.org/10.1037/xhp0000057)
36. Macmillan NA, Creelman D (2005) Detection theory: a user's guide. Lawrence Erlbaum Associates
37. Malpass RS, Kravitz J (1969) Recognition for faces of own and other race faces. *J Pers Soc Psychol* 13:330–334
38. McIntyre AH, Hancock PJ, Kittler J, Langton SR (2013) Improving discrimination and face matching with caricature. *Appl Cogn Psychol* 27(6):725–734
39. Megreya AM, Burton AM (2006) Unfamiliar faces are not faces: evidence from a matching task. *Mem Cogn* 34:865–876
40. Megreya AM, Sandford A, Burton AM (2013) Matching face images taken on the same day or months apart: the limitations of photo ID. *Appl Cogn Psychol* 27(6):700–706
41. Megreya AM, White D, Burton AM (2011) The other-race effect does not rely on memory: evidence from a matching task. *Q J Exp Psychol* 64(8):1473–1483
42. National Research Council (2009) Strengthening forensic science in the United States: a path forward. The National Academies Press, Washington, DC
43. Norell K, Lathen KB, Bergstrom P, Rice A, Natu V, O'Toole A (2015) The effect of image quality and forensic expertise in facial image comparisons. *J Forensic Sci* 60:331–340
44. O'Toole AJ, Abdi H, Jiang F, Phillips PJ (2007) Fusing face-verification algorithms and humans. *IEEE Trans: Syst Man Cybern B* 37:1149–1155
45. O'Toole AJ, Deffenbacher KA, Peterson J (1996) An “other-race effect” for classifying faces by sex. *Perception* 25:669–676
46. O'Toole AJ, Deffenbacher KA, Valentín D, Abdi H (1994) Structural aspects of face recognition and the other-race effect. *Mem Cogn* 22:208–224
47. O'Toole AJ, Edelman S, Bühlhoff HH (1998) Stimulus-specific effects in face recognition over changes in viewpoint. *Vis Res* 38:2351–2363
48. O'Toole AJ, An X, Dunlop JP, Natu V, Phillips PJ (2012) Comparing face recognition algorithms to humans on challenging tasks. *ACM Trans Appl Percept* 9(4), Article 16
49. O'Toole AJ, Phillips PJ, Narvekar A (2008) Humans versus algorithms: comparisons from the face recognition vendor test 2006. In: Proceedings of the 8th IEEE international conference on automatic face and gesture recognition
50. O'Toole AJ, Phillips PJ, Jiang F, Ayyad J, Pénard N, Abdi H (2007) Face recognition algorithms surpass humans matching faces across changes in illumination. *IEEE: Trans Pattern Anal Mach Intell* 29(9):1642–1646
51. Phillips PJ, O'Toole AJ (2014) Comparison of human and computer performance across face recognition experiments. *Image Vis Comput* 32(1):74–85
52. Phillips PJ, Jiang F, Narvarkar A, Ayyad J, O'Toole AJ (2011) An other-race effect for face recognition algorithms. *ACM Trans Appl Percept* 8(2):ART14
53. Phillips PJ, Beveridge JR, Draper BA, Givens G, O'Toole AJ, Bolme DS, Dunlop J, Lui YM, Sahibzada H, Weimer S (2011) An introduction to the good, the bad, & the ugly face

- recognition challenge problem. In: IEEE international conference on automatic face and gesture recognition, pp 346–353
- 54. Phillips PJ, Hill MQ, Swindle JA, O'Toole AJ (2015) Human and algorithm performance on the PaSC face recognition challenge. In: IEEE seventh international conference on biometrics: theory, applications and systems (BTAS 2015)
  - 55. Phillips PJ, Jiang F, Narvekar A, Ayyad J, O'Toole AJ (2011) An other-race effect for face recognition algorithms. *ACM Trans Appl Percept* 8(2), ART 14
  - 56. Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Chang J, Hoffman K, Marques J, Min J, Worek W (2005) Overview of the face recognition grand challenge. In: IEEE computer society conference on computer vision and pattern recognition, 2005. CVPR 2005, vol 1, pp 947–954
  - 57. Rhodes G, Brennan S, Carey S (1987) Identification and ratings of caricatures: implications for mental representations of faces. *Cogn Psychol* 19(4):473–497
  - 58. Rice A, Phillips PJ, Natu V, An X, O'Toole AJ (2013) Unaware person recognition from the body when face identification fails. *Psychol Sci* 24(11):2235–2243
  - 59. Sacks MJ, Koehler JJ (2005) The coming paradigm shift in forensic identification science. *Science* 309:892–895
  - 60. Scientific Working Group Imaging Technology (2010) SWGIT guidelines for the forensic imaging practitioner section 6: guidelines and recommendations for training in imaging technology in the criminal justice system. Retrieved from: <https://www.swgit.org/documents/Current%20Documents>
  - 61. Sinha P, Balas B, Ostrovsky Y, Russell R (2006) Face recognition by humans: nineteen results all computer vision researchers should know about. *Proc IEEE* 94(11):1948–1962
  - 62. Sun Y, Wang X, Tang X (2014) Deep learning face representation from predicting 10,000 classes. In: Computer vision and pattern recognition meeting
  - 63. Taigman Y, Yang M, Ranzato MA, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. In: 2014 IEEE conference on computer vision and pattern recognition (CVPR), pp 1701–1708
  - 64. Tangen JM, Thompson MB, McCarthy DJ (2011) Identifying fingerprint expertise. *Psychol Sci* 22(8):995–997
  - 65. Troje NF, Bülthoff HH (1996) Face recognition under varying poses: the role of texture and shape. *Vis Res* 36(12):1761–1771
  - 66. Ulery BT, Hicklin RA, Buscaglia J, Roberts MA (2011) Accuracy and reliability of forensic latent fingerprint decisions. *Proc Natl Acad Sci* 108(19):7733–7738
  - 67. White D, Burton AM, Jenkins R, Kemp RI (2014) Redesigning photo-ID to improve unfamiliar face matching performance. *J Exp Psychol: Appl* 20(2):166–173
  - 68. White D, Burton AM, Kemp RI, Jenkins R (2013) Crowd effects in unfamiliar face matching. *Appl Cogn Psychol* 27(6):769–777
  - 69. White D, Kemp RI, Jenkins R, Matheson M, Burton AM (2014) Passport officers' errors in face matching. *PLoS ONE* 9(8):e103510. doi:[10.1371/journal.pone.0103510](https://doi.org/10.1371/journal.pone.0103510)
  - 70. White D, Phillips PJ, Hahn CA, Hill M, O'Toole AJ (2015) Perceptual expertise in forensic facial image comparison. In: *Proc R Soc B: Biological Sciences*, 282:1814–1822
  - 71. Wild HA, Barrett SE, Spence M, O'Toole AJ, Cheng Y, Brooke J (2000) Recognition and sex categorization of adults' and children's faces: examining performance in the absence of sex stereotyped cues. *J Exp Child Psychol* 77:269–291
  - 72. Wilkinson C, Evans R (2009) Are facial image analysis experts any better than the general public at identifying individuals from CCTV images? *Sci Justice* 49(3):191–196
  - 73. Wilkinson C, Evans R (2011) Corrigendum to “Are facial image analysis experts any better than the general public at identifying individuals from CCTV images?” [Sci Justice 49:191–196 (2009)]. *Sci Justice* 51(4):218–221
  - 74. Yin RK (1969) Looking at upside-down faces. *J Exp Psychol* 81(1):141–145
  - 75. Young AW, Hay DC, Ellis AW (1985) The faces that launched a thousand slips: everyday difficulties and errors in recognizing people. *Br J Psychol* 76(4):495–523

**Part III**

**Human Motion, Speech and Behavioral  
Analysis**

# Chapter 10

## Biometric Evidence in Forensic Automatic Speaker Recognition

Andrzej Drygajlo and Rudolf Haraksim

**Abstract** The goal of this chapter is to provide a methodology for calculation and interpretation of biometric evidence in forensic automatic speaker recognition (FASR). It defines processing chains for observed biometric evidence of speech (univariate and multivariate) and for calculating a likelihood ratio as the strength of evidence in the Bayesian interpretation framework. The calculation of the strength of evidence depends on the speaker models and the similarity scoring used. A processing chain chosen for this purpose is in the close relation with the hypotheses defined in the Bayesian interpretation framework. Several processing chains are proposed corresponding to the scoring and direct method, which involve univariate and multivariate speech evidence, respectively. This chapter also establishes a methodology to evaluate performance of a chosen FASR method under operating conditions of casework.

### 10.1 Introduction

Biometrics is the science of establishing identity of individuals based on their biological and behavioural characteristics [30]. Speaker recognition is a biometric modality that uses an individual's speech for distinguishing one person from another [24]. Forensics means the use of scientific methods and techniques in the establishment of facts or evidence in the court of law [28]. The role of forensic expert is the provision of information (factual or opinion) to help answer questions of importance to courts of law [27].

The purpose of forensic speaker recognition (FSR) is to provide information (e.g. non-categorical opinion) of evidential weight if speech in the questioned speaker recording originates from suspected speaker or not [8, 9]. The judge or the jury in the court uses such an opinion for their deliberations and decision [39].

---

A. Drygajlo (✉) · R. Haraksim

Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland  
e-mail: andrzej.drygajlo@gmail.com

Forensic automatic speaker recognition (FASR) is an established term used when automatic speaker recognition methods are adapted to and used within a forensic application [8, 11, 31].

Best-practice FASR is associated with the Bayesian interpretation framework [2, 3, 9, 21, 35, 40, 41]. In this framework, an opinion of evidential weight, based upon case specific hypotheses (propositions) and conditioning information (framework of circumstances) should be provided for use in court [15, 17]. If there are two competing hypotheses, then the odds form of Bayes' theorem can be used:

$$\frac{P(H_0|E, I)}{P(H_1|E, I)} = \frac{p(E|H_0, I)}{p(E|H_1, I)} \times \frac{P(H_0|I)}{P(H_1|I)}$$

posterior odds      likelihood ratio      prior odds

In FASR, the evaluative opinion of the forensic expert is a numerical statement of strength of evidence (likelihood ratio *LR*) of the observed biometric speech evidence (*E*) given two competing hypotheses  $H_0$  and  $H_1$  and background information *I* [9, 15].

The odds form of Bayes' theorem shows how new data, reported as likelihood ratio (*LR*), can be combined with prior knowledge about the case in order to arrive at posterior odds. Only the *LR* is the province of the forensic expert; the prior odds and posterior odds are the province of the court.

The numerator of the *LR*, i.e. likelihood  $p(E|H_0, I)$ , represents the degree of similarity of the evidence with respect to the suspect and the denominator of the *LR*, i.e. likelihood  $p(E|H_1, I)$ , represents the degree of typicality with respect to the relevant population. The *LR* is the ratio between these two degrees corresponding to non-categorical opinion in the form of numerical statement. With *LR* values larger than unity there is stronger support for the  $H_0$  hypothesis and with *LR* values smaller than unity there is stronger support for the  $H_1$  hypothesis.

The value of a likelihood ratio (*LR*) critically depends on the choices one makes for stating the observed speech evidence (*E*) and the hypotheses  $H_0$  and  $H_1$ , with corresponding models of within-speaker or same-speaker and between-speaker or different-speaker speech variability. It also depends on several aspects of the speech analysis, including the non-automatic aspects of the data pre-processing stage, the kind of features, feature models and similarity scoring used, as well as the databases employed in the process. Consequently, the methodology using the Bayesian interpretation framework concerns also performance evaluation of *LR* methods for method validation and case-specific evaluation purposes [34, 36, 42].

Therefore, in FASR three types of biometric measures should be provided:

- Observed biometric evidence of speech,
- Strength of observed biometric evidence (likelihood ratio),
- Performance evaluation measures of the strength of evidence.

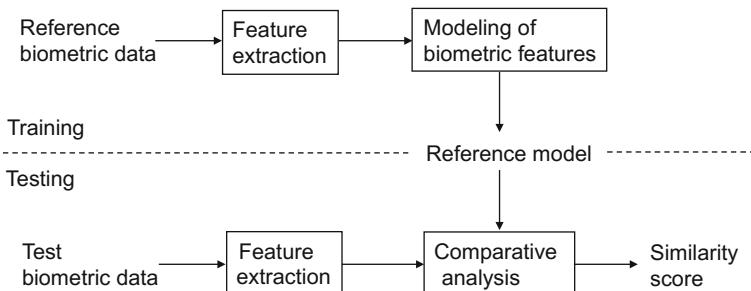
## 10.2 Biometric Evidence in FASR

Biometric recognition has essentially the same generic processing chain common across biometric modalities (Fig. 10.1) [30].

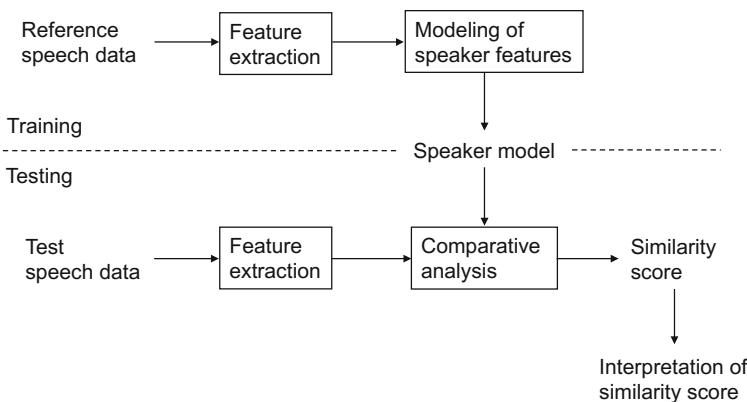
In automatic speaker recognition (ASR), the biometric recognition methodology concerns speech feature extraction, feature modelling (to create speaker models) and comparative analysis for similarity scoring (Fig. 10.2) [15].

In FASR, the observed biometric evidence ( $E$ ) of speech can be defined on the level of similarity scoring or feature extraction [4, 5, 16]. This chapter is focused on calculation of observed biometric evidence and its strength (likelihood ratio) using any feature extraction and feature modelling technique [29].

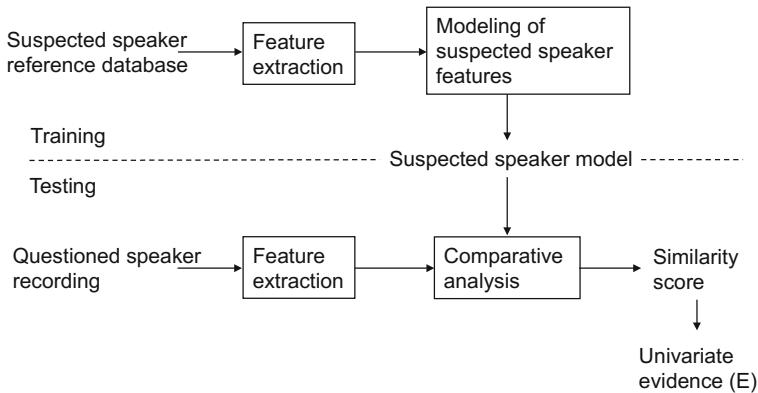
If the observed biometric evidence ( $E$ ) of speech is defined on the level of similarity scoring, it consists of single score representing the degree of similarity between speaker-dependent features extracted from the questioned speaker recording and speaker-dependent features extracted from the suspected speaker reference database, represented by this speaker model (Fig. 10.3) [13, 14].



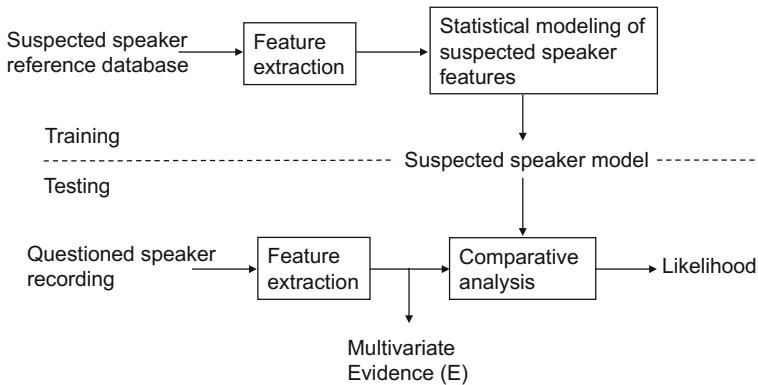
**Fig. 10.1** Generic processing chain for biometric recognition



**Fig. 10.2** Generic processing chain for automatic speaker recognition [15]



**Fig. 10.3** Processing chain for calculating univariate biometric speech evidence ( $E$ ) [15, 16]



**Fig. 10.4** Processing chain for calculating multivariate biometric speech evidence ( $E$ )

Automatic speaker recognition based on statistical modelling techniques has a useful property in that it directly returns a likelihood (as a score) of whether a speech signal (e.g. of questioned speaker recording) can correspond to the statistical model created for a speaker (e.g. suspected speaker in Fig. 10.4) [5, 16].

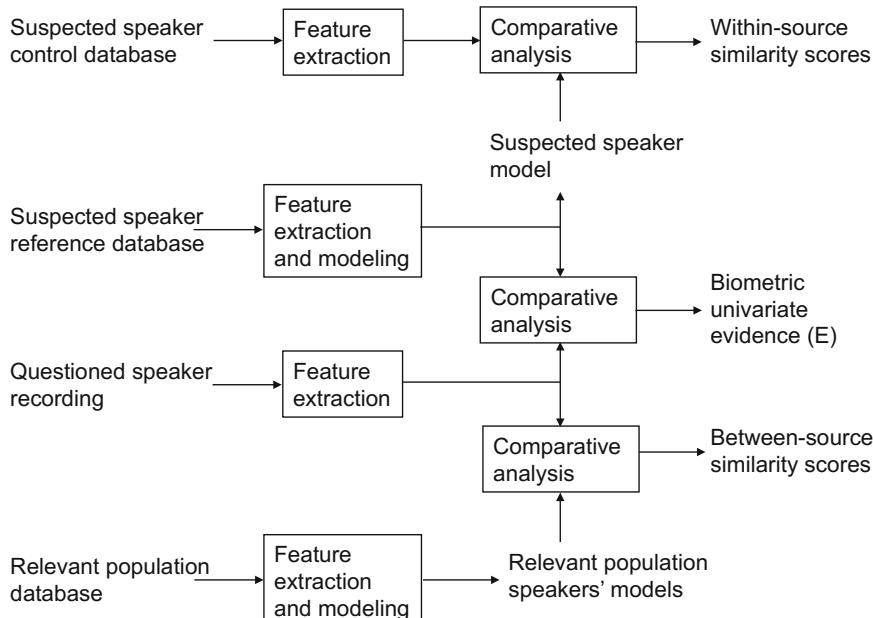
If statistical modelling techniques, which return likelihoods, are used in FASR, the observed biometric evidence is defined on the level of feature extraction as the ensemble of features extracted from the questioned speaker recording. Such an ensemble of features represents multivariate speech evidence ( $E$ ) [1, 5].

### 10.3 Calculation of Likelihood Ratio (*LR*)

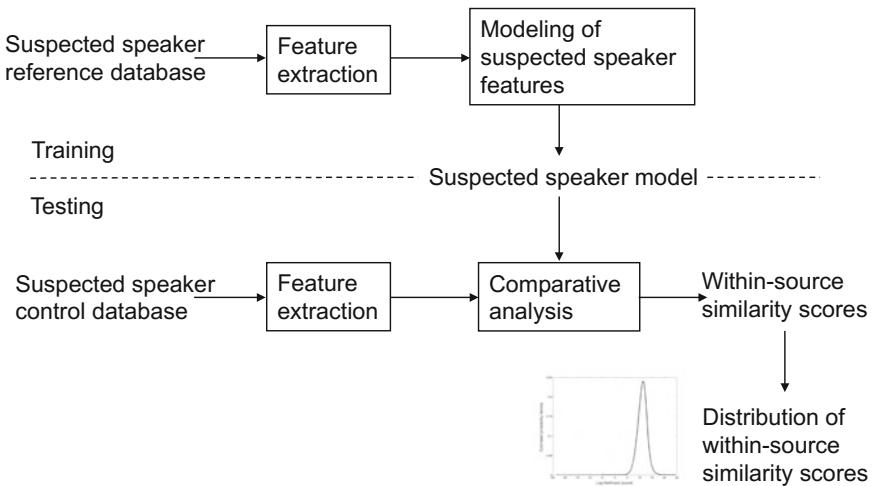
The calculation of a likelihood ratio (*LR*) depends on the speaker models and the similarity scoring used, as well as on the choice one makes for stating the hypotheses. There are two main methods commonly used for this purpose: the scoring method and the direct method [5, 12, 16].

#### 10.3.1 Scoring Method

The methodological approach based on the scoring method is independent of the features and speaker models chosen. This methodology needs a two-stage processing. The first stage consists of calculating scores using any type of feature vectors (multivariate data) as well as any of the speaker models (e.g. vector quantization (VQ), Gaussian mixture model (GMM), i-vector) and similarity scoring (Figs. 10.5 and 10.8) [10, 18, 32, 33]. The second stage transforms the obtained similarity scores into two univariate distributions, which are represented by probability density functions (Figs. 10.9 and 10.10). The values of these



**Fig. 10.5** Processing chain of feature extraction, feature modelling, comparative analysis and calculation of biometric univariate speech evidence as well as within-source and between source similarity scores for the scoring method [14, 16, 17]



**Fig. 10.6** Processing chain for calculating within-source similarity scores and their interpretation as a distribution [15, 16]

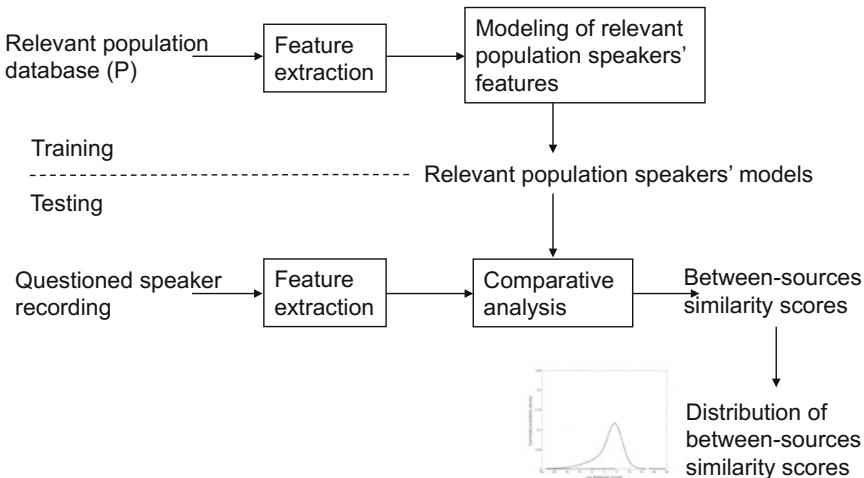
functions represent likelihoods of scores given chosen hypotheses. This two-stage processing can be done in many ways depending on the choices one makes for stating the hypotheses [4, 19, 22, 37].

If the hypotheses are stated as follows:

- $H_0$ —the suspected speaker is the source of the questioned recording,
- $H_1$ —the suspected speaker is not the source of the questioned recording,

the  $H_1$  hypothesis can be represented by the distribution of between-source similarity scores, which result from comparing the feature vectors or model of the questioned speaker with the ones of several other speakers from the relevant population database, and the  $H_0$  hypothesis can be represented by the within-source distribution of similarity scores as the result of comparing the feature vectors or model of the suspected speaker using its control database with the ones of the same speaker using its reference database (Figs. 10.5, 10.6 and 10.7) [14, 16, 17].

In this case, the strength of evidence is represented by the likelihood ratio ( $LR$ ) calculated as the ratio of likelihoods obtained from the distributions of within-source and between-source similarity scores for the single score  $E$  representing the value of the observed speech evidence (Figs. 10.9 and 10.10). This score ( $E$ ) is obtained by comparing the feature vectors or model of the questioned speaker recording with the ones of the suspected speaker using the suspected speaker reference database (Fig. 10.3).



**Fig. 10.7** Processing chain for calculating between-source similarity scores and their interpretation as a distribution [15, 16]

If the hypotheses are stated as follows:

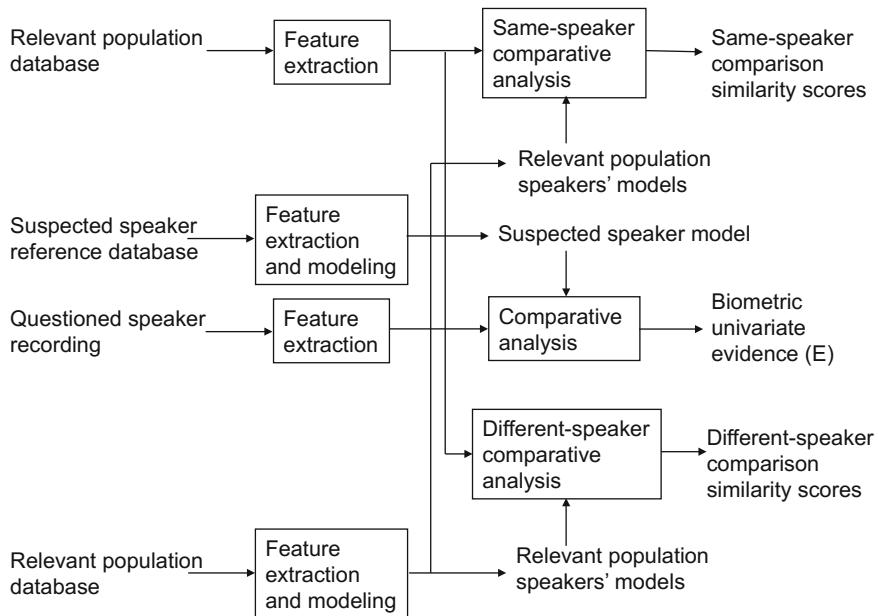
- $H_0$ —the suspected speaker recording and the questioned recording have the same source,
- $H_1$ —the suspected speaker recording and the questioned recording have different sources,

the  $H_0$  hypothesis can be represented by the distribution of similarity scores that result from same-speaker comparisons, and the  $H_1$  hypothesis can be represented by the distribution of similarity scores as the result of different-speaker comparisons using the relevant population database in both cases (Fig. 10.8) [23, 37].

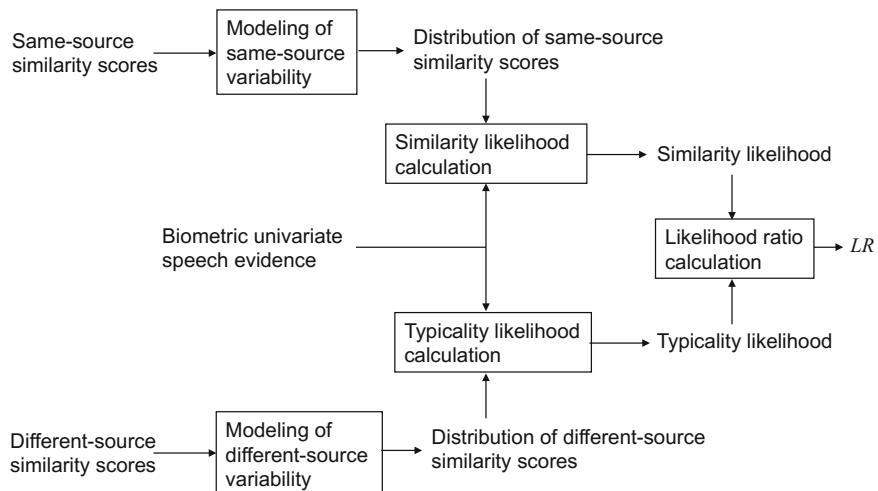
In this respect, the  $LR$ , as the strength of evidence, can be calculated by dividing the likelihoods from the distributions of the same-source and different-source similarity scores for the single score  $E$  representing the value of the observed speech evidence (Fig. 10.9). This score ( $E$ ) is obtained by comparing the feature vectors or model of the questioned speaker recording with the ones of the suspected speaker (Fig. 10.3).

The scoring method is illustrated in Fig. 10.10. The distributions of the same-source (within-source or same-speaker) and different-source (between-source or different-speaker) similarity scores constitute a common representation for any pair of  $H_0$  and  $H_1$  hypothesis used in this method.

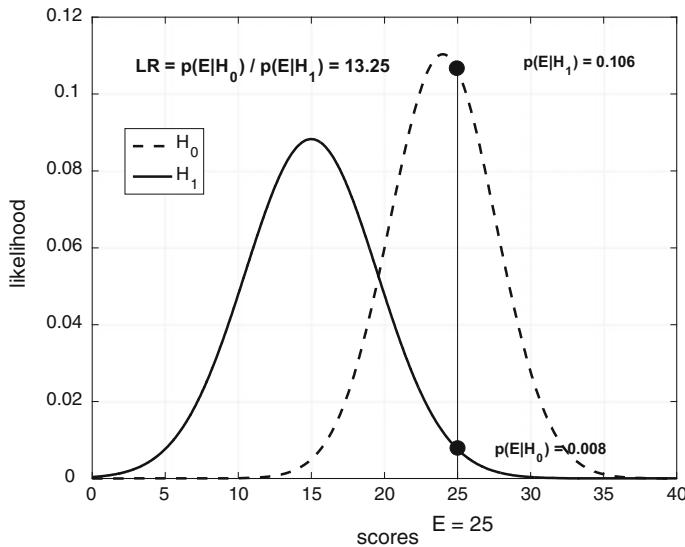
The conclusion that the forensic expert provides to the court shall be related to the assigned likelihood ratio  $LR$  (strength of evidence), the observable speech evidence  $E$  and the hypotheses  $H_0$  and  $H_1$  under consideration.



**Fig. 10.8** Processing chain of feature extraction, feature modelling, comparative analysis and calculation of biometric univariate speech evidence as well as same-speaker and different-speaker comparison similarity scores for the scoring method



**Fig. 10.9** Processing chain of the Bayesian interpretation framework of biometric univariate speech evidence for calculating the likelihood ratio ( $LR$ ) for the scoring method using the distributions of the same-source (within-source or same-speaker) and different-source (between-source or different-speaker) similarity scores [14, 16]



**Fig. 10.10** Illustration of the scoring method. In this example, the evidence score is  $E = 25$ . Dividing the likelihood value of  $H_0$  distribution (right side of figure) by the likelihood value of  $H_1$  distribution (left side of figure) for the observed speech evidence score  $E$  results in  $LR = 13.25$  [4, 31–33, 37]

Such a non-categorical opinion can be expressed, for example, as follows [19]:

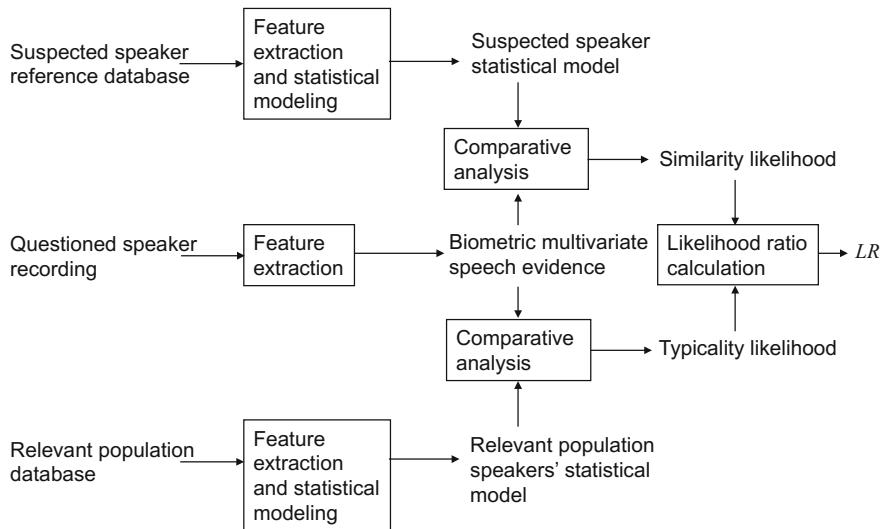
A likelihood ratio of 13.25 (obtained in Fig. 10.1), means that it is 13.25 times more likely to observe the speech evidence score ( $E$ ) given the hypothesis  $H_0$  (e.g., the suspected speaker recording and the questioned recording have the same source) than given the hypothesis  $H_1$  (e.g., the suspected speaker recording and the questioned recording have different sources).

The judge or the jury in the court uses such an opinion for their deliberation and decision.

### 10.3.2 Direct Method

The methodological approach based on the direct method needs a statistical model, which can compute a likelihood value when feature vectors are compared against such a model [4, 5, 7]. For example, GMMs or i-vectors combined with PLDA offer such a property [5, 6, 37, 42].

The direct method uses two databases: the suspected speaker reference database and the relevant population database. They can be used to create two statistical models, e.g., GMM 1—statistical model of the suspected speaker and GMM 2—statistical model of the relevant population. The universal background model



**Fig. 10.11** Processing chain of feature extraction, feature modelling, comparative analysis and calculation of biometric multivariate speech evidence as well as similarity and typicality likelihoods and likelihood ratio ( $LR$ ) for the direct method [16]

(UBM) trained with the relevant population database can also be used as GMM 2. In order to calculate the likelihood ratio, the multivariate evidence ( $E$ ) represented by the ensemble of feature vectors extracted from the questioned recording is compared to GMM 1 and GMM 2. The first comparison gives the similarity likelihood score (numerator of  $LR$ ) and the second one the typicality likelihood score (denominator of  $LR$ ). The  $LR$ s obtained with the direct method are often not well calibrated (Fig. 10.11).

The conclusion using the direct method, to be provided to the court by the forensic expert, can be stated in a way similar to the conclusion expressed in the scoring method.

## 10.4 Performance Evaluation

Evaluation of specific case findings concerns the likelihood ratio ( $LR$ ), which summarises the statement of the forensic expert in the casework, and its comparison with the likelihood ratios that can be obtained from the observed speech evidence under operating conditions of the case, on one hand when the hypothesis  $H_0$  is true and, on the other hand, when the hypothesis  $H_1$  is true [19].

Evaluation should be carried out using a relevant population database of the case. The intrinsic property of this database is the known ground truth regarding the

source of origin of each of the recordings. When the ground truth of the recordings related to the origin of the speakers is known, there are two types of trials, e.g.:

- Same-source (within-source or same-speaker) trials (SS trials)
- Different-source (between-source or different-speaker) trials (DS trials)

In the process of empirical evaluation on the relevant population database the FASR method outputs a *LR* for each of the SS and DS trials. Given the *LRs* and the ground truth regarding the SS and DS trials, it is possible to derive a number of performance characteristics and metrics. For evaluating the strength of evidence (*LR*), performance characteristics and metrics can be used together with the case-specific likelihood ratio  $LR_{\text{case}}$ .

### **10.4.1 Performance Characteristics and Metrics**

Performance characteristics (e.g. Tippett plots I and II as well as Empirical Cross Entropy (ECE) plots) describe support for the correct hypothesis of a *LR* method [17, 20, 36, 37].

Performance metrics provide a single numerical value that describes the performance in terms of, e.g. accuracy, discriminating power and calibration of the *LR* method (probabilities of misleading evidence (PMEH<sub>0</sub> and PMEH<sub>1</sub>), equal proportion probability (EPP) , log-likelihood-ratio cost (Clrr) [19, 37, 42].

A starting point for elaborating performance characteristics in FASR corresponds to histograms (count vs. *LR*) of the *LRs* for  $H_1$  and  $H_0$  hypotheses. In order to move from the histograms (count vs. *LR*) of the *LRs* for  $H_1$  and  $H_0$  hypotheses to cumulative distribution functions (probability vs. *LR*), they have to be approximated by probability density functions and normalised so that the area under each of the curves is equal to one. Following the normalisation of the *LR* distributions the cumulative distribution functions (CDFs) of the *LRs* can be computed. The cumulative distribution functions, unlike the normalised *LR* distributions are plotted as the probability on the y-axis, vs. the *LRs* on the x-axis. They represent the cumulative proportion of *LRs* less than or equal to the *LR* value on the x-axis for the  $H_1$  and  $H_0$  hypotheses.

#### **10.4.1.1 Performance Characteristics—Tippett Plots**

In this chapter, the case-specific evaluation is limited to Tippett plots, which correspond to cumulative distribution functions and are the most frequently used performance characteristics in FASR. There are two versions of these plots: Tippett plots I and II [17, 35].

Tippett plots I correspond to the inverse CDFs (iCDFs) of  $LR$ s for both  $H_0$  and  $H_1$  hypotheses. They represent the cumulative proportion of  $LR$ s greater than the  $LR$  value on the x-axis for the  $H_1$  and  $H_0$  hypotheses.

Tippett plots II are a combination of the CDF for the  $H_0$  hypothesis and the inverse CDF (iCDF) for the  $H_1$  hypothesis.

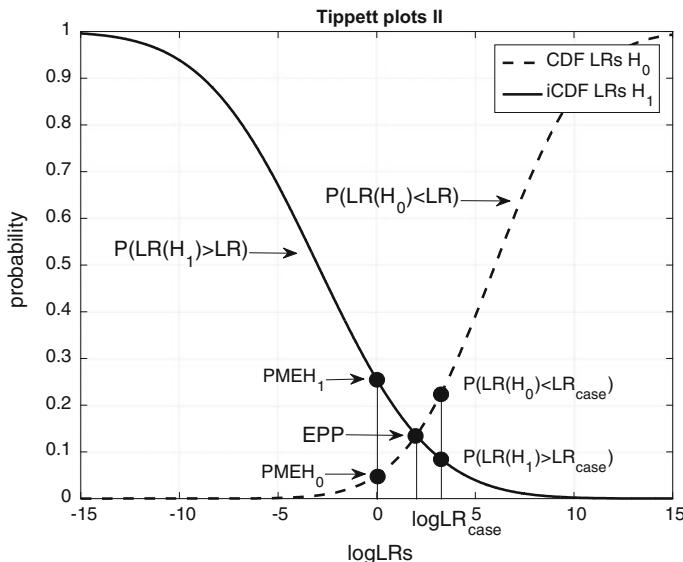
They represent:

- the cumulative proportion of  $LR$ s less than or equal to the  $LR$  value on the x-axis for the  $H_0$  hypothesis;  $P(LR(H_0) \leq LR)$ ,
- the inverse cumulative proportion of  $LR$ s greater than the  $LR$  value on the x-axis for the  $H_1$  hypothesis;  $P(LR(H_1) > LR)$ .

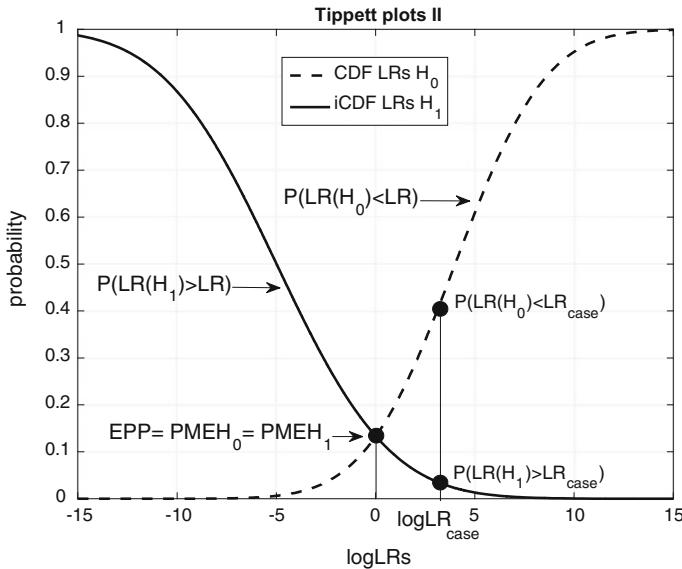
Figures 10.12 and 10.13 show examples of Tippett plots II. In these figures, the natural logarithm transformation ( $\log LR$ ) is applied to the  $LR$  values. The main advantage of the log transformation is the symmetrical scale with the centre of symmetry at  $\log LR = 0$ . The  $LR$ s supporting the hypothesis  $H_0$  tend to have positive values and the  $LR$ s supporting the hypothesis  $H_1$  tend to have negative values.

#### 10.4.1.2 Performance Metrics

Performance metrics related to the Tippett plots II are Probabilities of Misleading Evidence ( $\text{PMEH}_0$  and  $\text{PMEH}_1$ ) and Equal Proportion Probability (EPP). They are applied to measure accuracy and discriminating power of the  $LR$  method used in



**Fig. 10.12** Relation between case-specific  $LR$  ( $LR_{\text{case}}$ ) and Tippett plots II for non-calibrated  $LR$ s [19]



**Fig. 10.13** Relation between case-specific  $LR$  ( $LR_{case}$ ) and Tippett plots II for calibrated  $LRs$  [19]

casework, respectively [25]. Accuracy is a performance property, which presents the closeness of agreement between an assigned  $LR$  value and ground truth status of the hypothesis. Discriminating power is a performance property representing the capability of a given method to distinguish between same-source (SS) and different-source (DS) trials where different hypotheses are true.

Another metric that allows for combined measure of accuracy, discriminating power and calibration loss is the log-likelihood-ratio cost ( $Clr$ ) [37, 42].

Performance Metric 1—Probabilities of Misleading Evidence ( $PMEH_0$  and  $PMEH_1$ )

Probabilities of Misleading Evidence ( $PMEH_0$  and  $PMEH_1$ ) are associated with the Tippett plots representation. They are defined in the following way [19]:

- $PMEH_0$ : probability of misleading evidence in favour of the hypothesis  $H_1$ . The probability of all  $LRs$  that are smaller or equal to 1, knowing the  $H_0$  hypothesis is true  $PMEH_0 = P(LR(H_0) \leq 1)$ .
- $PMEH_1$ : probability of misleading evidence in favour of the hypothesis  $H_0$ . The probability of all the  $LRs$  that are bigger than 1, knowing the  $H_1$  hypothesis is true  $PMEH_1 = P(LR(H_1) > 1)$ .

Based on their definitions the PMEs can be seen as a measure of accuracy of a  $LR$  method used in casework (Figs. 10.12 and 10.13). The PMEs in these two

examples are:  $\text{PMEH}_0 = 0.05$  and  $\text{PMEH}_1 = 0.25$ , in Fig. 10.12 and  $\text{PMEH}_0 = \text{PMEH}_1 = 0.14$  in Fig. 10.13, in both examples at  $\log LR = 0$ .

### Performance Metric 2—Equal Proportion Probability (EPP)

The advantage of Tippett plots II is that at the intersection of the CDF and iCDF we find the Equal Proportion Probability (EPP) [35]. The corresponding EPP value is  $EPP = 0.14$  at  $\log LR = 2$  in Fig. 10.12 and  $EPP = 0.14$  at  $\log LR = 0$  in Fig. 10.13.

The EPP can be seen as a metric of discriminating power. The lower the EPP the better the discriminating capabilities of a FASR system, e.g. capacity of the system to discriminate between the SS and DS trials. If comparing different FASR methods or if comparing the same method based on data recorded in various conditions, the EPPs compare the systems based on the discriminating power [7, 26].

Tippett plots II also can serve as an indicator of miscalibration. For a perfectly calibrated system the EPP is equal to both types of misleading evidence ( $EPP = \text{PMEH}_0 = \text{PMEH}_1 = 0.14$ ) calculated for  $\log LR = 0$  (Fig. 10.13).

### Performance Metric 3—Log-Likelihood-Ratio Cost ( $Cllr$ )

Log-likelihood-ratio cost ( $Cllr$ ) is proposed as a metric of accuracy related to the average cost of a  $LR$  method used [37]:

$$Cllr = \frac{1}{2 \cdot N_{SS}} \sum_{i_{SS}} \log_2 \left( 1 + \frac{1}{LR_i} \right) + \frac{1}{2 \cdot N_{DS}} \sum_{j_{DS}} \log_2 (1 + LR_j),$$

where  $N_{SS}$  and  $N_{DS}$  are respectively the number of likelihood ratios in the SS and DS dataset. The indices  $i_{SS}$  and  $j_{DS}$  denote summing over the SS and DS likelihood ratios. As the accuracy of a speaker recognition system gets higher the  $Cllr$  tends towards zero. Following the pool-adjacent-violators (PAV) transformation of the  $\log LRs$  we can obtain the parameter  $Cllr^{\min}$  as discriminating power metric [42].

The calibration loss performance metric  $Cllr^{\text{cal}}$  is obtained by subtracting  $Cllr^{\min}$  from  $Cllr$ . In a well-calibrated system this difference is small [18, 37, 38, 43].

#### 10.4.2 Evaluation of Case-Specific Strength of Evidence

When applied to specific speaker recognition based casework, a FASR method allows a numerical statement about the strength of evidence ( $LR$ ) that the method can provide. Such casework can involve measuring calibration of likelihood ratios [38].

Tippett plots provide performance characteristics in terms of actual values of the *LRs* derived in the specific case for the relevant population of same-source and different-source trials under  $H_0$  and  $H_1$  hypotheses [6]. General information about the accuracy of the system used in the case can be given by probabilities of misleading evidence  $\text{PMEH}_0$  and  $\text{PMEH}_1$ , which represent accuracy metrics. The value of the discriminating power metric EPP (Equal Proportion Probability) , corresponding to the crossing point of  $H_0$  and  $H_1$  plots, shows the discrimination of the system used. Misalignment of this crossing point with  $\log LR = 0$  on the x-axis indicates that the system is non-calibrated.

Figure 10.12 shows an example of Tippett plots II for non-calibrated set of *LRs*. The intersection point between the plots of  $P(LR(H_1) > LR)$  and  $P(LR(H_0) \leq LR)$  does not occur at or near  $\log LR = 0$  on the x-axis but occurs at some distance from it (2 in this example).

$Cllr$  can be used to state the accuracy performance of the non-calibrated *LRs*.  $Cllr^{\min}$  and  $Cllr^{\text{cal}}$  can be used to confirm discrimination power and calibration loss of the system, respectively. Calibration loss is greater in a non-calibrated than a calibrated system.

An example of a case-specific  $LR_{\text{case}} = 30$  ( $\log LR_{\text{case}} = 3.4$ ) expressed as non-calibrated likelihood ratio is shown along with Tippett plots II in Fig. 10.12.

Calibrated *LR* is understood here as a likelihood ratio computed with a method aimed at improving the calibration metric. Some methods for calculating *LRs* produce well-calibrated results without the need for additional calibration.

Calibrated likelihood ratios are the most desirable strength-of-evidence statements and they are most compatible with the Bayesian interpretation framework. Calibrated likelihood ratios have a direct interpretation that can be reported and explained to the court. For example, (Fig. 10.13), if the likelihood ratio of a case has a value of 26.31 it is 26.31 times more likely to observe speech evidence supporting the  $H_0$  hypothesis than the  $H_1$  hypothesis given the questioned speaker and suspected speaker recordings, as well as the relevant population database.

An example of a case-specific  $LR_{\text{case}}$  expressed as calibrated likelihood ratio is shown along with Tippett plots II in Fig. 10.13.

Similarly to non-calibrated *LRs*, for calibrated *LRs* general information about the accuracy of the system used in the case can be given by probabilities of misleading evidence  $\text{PMEH}_0 = \text{PMEH}_1$ , which represent accuracy metrics. The value of discrimination power metric EPP (Equal Proportion Probability) , corresponding to the crossing point of  $H_0$  and  $H_1$  plots, shows the discriminating power of the system used.

Alignment of this crossing point with  $\log LR = 0$  ( $EPP = \text{PMEH}_0 = \text{PMEH}_1 = 0.14$ ) on the x-axis indicates that the system is calibrated.

As before with non-calibrated *LRs*,  $Cllr$  can be used to state the accuracy performance of the calibrated *LRs*.  $Cllr^{\min}$  and  $Cllr^{\text{cal}}$  can be used to confirm discrimination power and calibration loss of the system, respectively.

For calibrated *LRs*, if the case-specific  $LR_{\text{case}}$  falls on the right side of  $\log LR = 0$ , the case-specific speech evidence provides stronger support for the  $H_0$  hypothesis than the  $H_1$  hypothesis. In the example presented in Fig. 10.13, the

reported  $LR_{\text{case}}$  value is equal to 26.31 ( $\log LR_{\text{case}} = 3.27$ ). Such a value is reported to the court as the strength of evidence.

Within the performance characteristic (Tippett plots II) it is also possible to report what the probabilistic distance (PD) of case-specific  $LR_{\text{case}}$  is to the misleading evidence  $\text{PMEH}_0$  [19]. This distance can be calculated as  $P(LR(H_0) < LR_{\text{case}}) - \text{PMEH}_0$ . For the example of calibrated  $LR$ s PD = 0.42 – 0.03 = 0.39 (Fig. 10.13) and for non-calibrated case example PD = 0.23 – 0.05 = 0.18 (Fig. 10.12). It corresponds to the proportion of likelihood ratios  $LR(H_0)$  greater than  $LR = 1$  and smaller than  $LR_{\text{case}}$ .

It is furthermore possible to report the case-specific probabilistic error (PE) expressed as proportion of likelihood ratios  $LR(H_1)$  greater than the  $LR_{\text{case}}$  value  $P(LR(H_1) > LR_{\text{case}})$ . In calibrated case example (Fig. 10.13) PE = 0.03 and in non-calibrated one (Fig. 10.12) PE = 0.08.

An analogous set of values can be reported if  $LR_{\text{case}}$  falls on the left side of  $\log LR = 0$ .

## 10.5 Conclusion

In this chapter, a methodology for calculation and Bayesian interpretation of biometric evidence in forensic automatic speaker recognition (FASR) was provided. Processing chains for observed biometric evidence of speech and for calculating a likelihood ratio as the strength of this evidence were systematically developed. In FASR, the observed biometric evidence of speech can be defined on the level of similarity scoring (biometric univariate evidence) or feature extraction (biometric multivariate evidence). Based on this choice, calculation of a likelihood ratio can be performed by the “scoring method” or the “direct method”, respectively. Processing chains developed for this purpose are in close relation with the hypotheses defined in the Bayesian interpretation framework. Several proposed chains correspond to the scoring and direct method, which involve univariate and multivariate biometric speech evidence.

This chapter also established a methodology to evaluate performance of a chosen FASR method under operating conditions of casework. It was possible to derive a number of performance characteristics and metrics given a relevant population database, a FASR method, which outputs the likelihood ratio value for each of the same-source and different-source trials and the ground truth regarding these trials. The performance evaluation was focused mainly on performance characteristics of Tippett plots and performance metrics such as probabilities of misleading evidence, equal proportion probability and log-likelihood-ratio cost. All these performance characteristics and metrics were combined together for evaluation of the case-specific strength of biometric evidence. Examples for non-calibrated and calibrated likelihood ratios were presented.

This chapter aimed at developing universal methodology, to be used for any specific speech data of the case, that provides a coherent way of presenting recorded

speech as observed biometric evidence, calculation of its strength (likelihood ratio) and evaluation of the FASR method used for this calculation.

The methodology developed in this chapter has been adapted by the same authors to “Methodological Guidelines for Best Practice in Forensic Semiautomatic and Automatic Speaker Recognition” elaborated in the framework of ENFSI FSAAWG project “Methodological guidelines for semiautomatic and automatic speaker recognition for case assessment and interpretation”, chaired by Andrzej Drygajlo.

## References

1. Aitken CGG, Lucy D (2004) Evaluation of trace evidence in the form of multivariate data. *Appl Stat* 53:109–122
2. Aitken CGG, Taroni F (2004) Statistics and evaluation of evidence for forensic scientists, 2nd edn. Wiley, Chichester
3. Aitken CGG, Roberts P, Jackson G (2010) Fundamentals of probability and statistical evidence in criminal proceedings. Guidance for judges, lawyers, forensic scientists and expert witnesses. Practitioner guide no 1, Royal Statistical Society
4. Alexander A (2005) Forensic automatic speaker recognition using Bayesian interpretation and statistical compensation for mismatched conditions. PhD dissertation, EPFL Lausanne
5. Alexander A, Drygajlo A (2004) Scoring and direct methods for the interpretation of evidence in forensic speaker recognition. In: Proceedings of the International Conference on Spoken Language Processing (ICSLP), Jeju, Korea, pp 2397–2400
6. Alonso Moreno V, Drygajlo A (2012) A joint factor analysis model for handling mismatched recording conditions in forensic automatic speaker recognition. In: Proceedings of the International Conference on Biometrics (ICB 2012), New Delhi, pp 484–489
7. Arcienega M, Alexander A, Zimmermann P, Drygajlo A (2005) A Bayesian network approach combining pitch and spectral envelope features to reduce channel mismatch in speaker verification and forensic speaker recognition. In: Proceedings of INTERSPEECH 2005, Lisbon, pp 2009–2012
8. Campbell JP, Shen W, Campbell WM, Schwartz R, Bonastre J-F, Matrouf D (2009) Forensic speaker recognition: a need for caution. *IEEE Signal Process Mag* 26:95–103
9. Champod C, Meuwly D (2000) The inference of identity in forensic speaker recognition. *Speech Commun* 31:193–203
10. Dehak N, Kenny P, Dehak R, Dumouchel P, Ouellet P (2011) Front-end factor analysis for speaker verification. *IEEE Trans Audio Speech Lang Process* 19:788–798
11. Drygajlo A (2007) Forensic automatic speaker recognition. *IEEE Signal Process Mag* 24:132–135
12. Drygajlo A (2009) Statistical evaluation of biometric evidence in forensic automatic speaker recognition. In: Gerardts ZJ, Franke KY, Veenman CJ (eds) Computational forensics. Springer, Berlin, pp 1–12
13. Drygajlo A (2009) Forensic evidence of voice. In: Li SZ (ed) Encyclopedia of biometrics. Springer, Berlin, pp 1388–1395
14. Drygajlo A (2011) Voice: Biometric analysis and interpretation of. Wiley Encyclopedia of Forensic Science. Accessed 15 Dec 2011. doi:[10.1002/9780470061589.fsa1034](https://doi.org/10.1002/9780470061589.fsa1034)
15. Drygajlo A (2012) Automatic speaker recognition for forensic case assessment and interpretation. In: Neustein A, Patil HA (eds) Forensic speaker recognition. Law enforcement and counter-terrorism. Springer, Berlin, pp 21–39

16. Drygajlo A, (2014) From speaker recognition to forensic speaker recognition. In: Cantoni V, Dimov D, Tistarelli M (eds) Biometric authentication: first international workshop, BIOMET 2014, Sofia, Bulgaria, Revised Selected Papers. Springer, Berlin, pp 93–104
17. Drygajlo A, Meuwly D, Alexander A (2003). Statistical methods and Bayesian interpretation of evidence in forensic automatic speaker recognition. In: Proceedings of EUROSPEECH 2003, Geneva, pp 689–692
18. Drygajlo A, Ugnat L (2012) Comparative evaluation of calibrated deterministic and statistical models for forensic automatic speaker recognition systems. Presentation at the European Academy of Forensic Science Conference (EAFS 2012), The Hague
19. Drygajlo A, Jessen M, Gfroerer S, Wagner I, Vermeulen J, Niemi T (2015) Methodological guidelines for best practice in forensic semiautomatic and automatic speaker recognition including guidance on the conduct of proficiency testing and collaborative exercises. ENFSI, Verlag für Polizeiwissenschaft, Frankfurt
20. Evett IW, Buckleton JS (1996) Statistical analysis of STR data. In: Carracedo A, Brinkmann B, Bär W (eds) Advances in forensic haemogenetics, vol 6. Springer, Berlin, pp 79–86
21. Evett IW (1998) Towards a uniform framework for reporting opinions in forensic science casework. *Sci Justice* 38:198–202
22. Gonzalez-Rodriguez J, Drygajlo A, Ramos-Castro D, Garcia-Gomar M, Ortega-Garcia J (2006) Robust estimation, interpretation and assessment of likelihood ratios in forensic speaker recognition. *Comput Speech Lang* 20:331–355
23. Gonzalez-Rodriguez J, Rose P, Ramos D, Toledano DT, Ortega-Garcia J (2007) Emulating DNA: rigorous quantification of evidential weight in transparent and testable forensic speaker recognition. *IEEE Trans Audio Speech Lang Process* 15:2104–2115
24. Hansen JHL, Taufiq H (2015) Speaker recognition by machines and humans. *IEEE Signal Process Mag* 32:74–99
25. Haraksim R (2014) Validation of likelihood ratio methods used in forensic evidence evaluation: Application in forensic fingerprints. PhD dissertation, University of Twente, Enschede
26. Haraksim R, Ramos D, Meuwly D, Berger CEH (2015) Measuring coherence of computer-assisted likelihood ratio methods. *Forensic Sci Int* 249:123–132
27. Jackson G, Jones S, Booth G, Champod C, Evett I (2006) The nature of forensic science opinion—a possible framework to guide thinking and practice in investigations and in court proceedings. *Sci Justice* 46:33–44
28. Jackson G, Aitken C, Roberts P (2015) Case assessment and interpretation of expert evidence. Guidance for judges, lawyers, forensic scientists and expert witnesses. Practitioner guide no 4
29. Kinunen T, Li H (2010) An overview of text-independent speaker recognition: from features to supervectors. *Speech Commun* 52:12–40
30. Li SZ, Jain A (eds) (2015) Encyclopedia of biometrics, 2nd edn. Springer, US
31. Meuwly D (2001) Reconnaissance automatique de locuteurs en sciences forensiques: l'apport d'une approche automatique. PhD dissertation, University of Lausanne
32. Meuwly D, El-Maliki M, Drygajlo A (1998) Forensic speaker recognition using Gaussian Mixture Models and a Bayesian framework. In: COST-250 workshop on speaker recognition by man and by machine: directions for forensic applications, Ankara, pp. 52–55
33. Meuwly D, Drygajlo A (2001) Forensic speaker recognition based on a Bayesian framework and Gaussian Mixture Modelling (GMM). In: Proceedings of ODYSSEY 2001, Crete, pp 145–150
34. Meuwly D, Haraksim R, Ramos D (2016) A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation. To appear in *Forensic Science International*
35. Morrison GS (2009) Forensic voice comparison and the paradigm shift. *Sci Justice* 49:298–308
36. Morrison GS (2010) Forensic voice comparison. In: Freckelton I, Selby H (eds) Expert evidence (Chapter 99). Thomson Reuters, Sydney

37. Ramos-Castro D (2007) Forensic evaluation of the evidence using automatic speaker recognition systems. PhD dissertation, Universidad Autónoma de Madrid
38. Ramos D, Gonzalez-Rodriguez J (2013) Reliable support: Measuring calibration of likelihood ratios. *Forensic Sci Int* 230:156–169
39. Robertson B, Vignaux GA (1995) Interpreting evidence. Evaluating forensic science in the courtroom. Wiley, Chichester etc.
40. Rose P (2002) Forensic speaker identification. Taylor & Francis, London
41. Rose P (2006) Technical forensic speaker recognition: evaluation, types and testing of evidence. *Comput Speech Lang* 20:159–191
42. Van Leeuwen DA, Brümmer N (2007) An introduction to application-independent evaluation of speaker recognition systems. In: Müller C (ed) *Speaker classification I: fundamentals, features, and methods*. Springer, Berlin, pp 330–353
43. Van Leeuwen D, Brümmer N (2013) The distribution of calibrated likelihood-ratios in speaker recognition. In: Proceedings of INTERSPEECH 2013, Lyon, pp 1619–1623

# **Chapter 11**

## **On Using Soft Biometrics in Forensic Investigation**

**Paulo Lobato Correia, Peter K. Larsen, Abdenour Hadid,  
Martin Sandau and Miguel Almeida**

**Abstract** This chapter addresses the usage of biometric recognition tools in the context of forensic investigations. In particular, the authors are concerned with the extraction of evidence from video sequences captured by surveillance cameras. In such scenarios many of the biometric traits traditionally used for recognition purposes, such as fingerprints, palmprints or iris, are not available. Therefore, the focus is on the extraction of soft biometrics, which encompasses personal characteristics used by humans to recognize or help to recognize an individual. This work starts by reviewing how forensic casework relying on surveillance video information is conducted nowadays. Then, a software platform, BioFoV, is proposed to automate many of the required procedures and including some initial implementation of soft biometric extraction tools. Furthermore, novel biometric methods to analyze human gait and facial traits are described and experimentally validated as a demonstration of future perspectives in soft biometrics.

### **11.1 Introduction**

Forensic investigations rely on the application of a broad spectrum of sciences to answer questions of interest to a legal system, eventually in relation to a crime or to a civil action. One of the main goals is often to identify people involved in some action.

---

P.L. Correia (✉) · M. Almeida

Instituto de Telecomunicações, Instituto Superior Técnico, Universidade de Lisboa,  
Torre Norte, 10-15, 1049-001 Lisbon, Portugal  
e-mail: plc@lx.it.pt

P.K. Larsen  
University of Copenhagen, Copenhagen, Denmark

A. Hadid  
University of Oulu, Oulu, Finland

M. Sandau  
Danish Institute of Fire and Security Technology, Hvidovre, Denmark

Biometric researchers, on the other hand, have done much research on the recognition of people, but often without targeting or having access to data from cases found in real forensic scenarios. The EU COST Action 1106 “Integrating Biometrics and Forensics for the Digital Age” was established to bring these two communities together to foster cooperation and new research, and the work reported here has been carried out in this context.

The present work addresses the analysis of surveillance videos captured in crime scenes, notably in view of the extraction of soft biometrics that can provide evidence to be used as part of a forensic investigation.

Soft biometrics can be understood as traits that can provide information useful for recognition purposes, even if in some cases it may not be enough to, on its own, establish identity [23]. Soft biometrics are also often related to traits that humans use to recognize their peers, whose degree of distinctiveness and permanence may not always be the same [8]. Examples of soft biometric traits include gender, ethnicity, age, height, weight, body geometry, gait, scars, marks, or tattoos, among others. In some contexts also the information about clothing or other accessories can be used for recognition purposes.

This chapter presents one an effort made toward easing the usage of soft biometrics in forensic investigations. It starts by overviewing some of the procedures often adopted in forensic case investigations. Then a software platform developed to support forensic investigators is proposed, into which a first set on modules requested by a law enforcement agency were included. The rest of the chapter is devoted to present further soft-biometrics research work conducted by the authors that can benefit forensic researchers.

## 11.2 Forensic Case Work as It Is Performed Today

### 11.2.1 *Forensic Image Analysis at Present*

Much forensic case work is based on the same methods as finger print examination, where forensic experts for many decades have examined two fingerprints for similarities and differences in the patterns of the skin, known as minutiae. The more similarities the examiner can find in the fingerprints, the more confident can the expert be in stating that the fingerprints belong to the same person.

The same principle is used in facial image comparison and as well as comparison of soft biometrics, which include bodily traits, such as birthmarks, scars, tattoos, gait, and height measurements based on pictures (known as photogrammetry—this subject will be described later in this chapter). The principle still is: the more similar traits found, the stronger is the support for perpetrator and suspect having congruent identity.

Minutiae are very characteristic traits, such as a special scar, while other traits are more general and can be said to distinguish between groups of people. Often it



**Fig. 11.1** Example of real-case where ankle bending angle is used as soft biometric

is a subjective judgement of the expert how characteristic a given soft biometric trait is. An example of two traits of gait, notably bodily proportions and feet rotation angle, are illustrated in Fig. 11.1. Here it can be seen that the suspect (to the left) has similar bodily proportions as the perpetrator and also a markedly outward rotated feet. These traits can be categorized as quite common to a lot of people. Nevertheless, there will also be many subjects not sharing these traits. More notably the trait marked with “b”—an “outward bent” left ankle—indicate that both suspect and perpetrator have the same anomaly. This trait will provide strong support for perpetrator and suspect having congruent identity.

How strong support a given trait provides is often based solely on the expert experience, due to the lack of databases. A reason for this could be that it is time consuming and challenging to build up databases and forensic scientists often do not have the expertise and/or are too busy with case work.

However, the biometric is contributing in this direction with the creation of databases, for instance for tattoos, gait and facial images.

If the forensic examination can be based on empirical data, then it is possible to base the statement on uncertainty expressions such as likelihood ratios giving a statement of how likely it is that a given trace originates from the suspect rather than from any other subject in the population [34]. Likelihood ratios have been implemented in forensic courtrooms, for instance through DNA evidence.

In fact, for some traits it may be simpler to establish databases to calculate uncertainty measures, like for a DNA profile, for fingerprint minutiae or for the measurement of heights because it is quite clear how to find differences within the traits objectively. However, for some other traits, e.g., as obtained from images, it

can be harder, because they depend on the perception of the observer [4]. Different observers will, for instance, have different perceptions of when and in which degree a nose seen in profile should be labeled as curved, crooked or bent.

### ***11.2.2 Presentation of Findings in Court***

All biometrics traits found to be in support or against congruent identity are summarized by the expert in a statement to court using a conclusion scale.

One of the simpler scales has been used by fingerprint experts based on the number of matching minutiae found [24]:

- Identification
- No conclusion
- Different marks

Other scales used are versions where the “no conclusion” statement has been replaced with different levels on support for congruent identity such as this one first used by the Danish police to compare tool marks and then modified by the Unit of Forensic Anthropology, University of Copenhagen to use for comparison of a given perpetrator and suspect(s):

- (a) Identification
- (b) Strong support for congruent identity
- (c) Moderate support for congruent identity
- (d) Inconclusive
- (e) Limited support for congruent identity
- (f) Not congruent identity

Both these scales are based on the expert knowledge. In the last decades, there has been an ongoing discussion in the forensic community whether and in which degree statements should be based on uncertainty statements or expert knowledge. “The expert knowledge side” argues that it is possible to positively identify a perpetrator and future research should point toward clarifying the error ratio associated with the identification. Contrary “the uncertainty statement side” would say that is not possible to positively identify a given suspect as being the perpetrator so research should be aimed toward expressing how probable it is that the suspect indeed is the perpetrator compared to any other subject in a relevant population.

A conclusion scale used by the Digital Image Working Group (DIWG) for proficiency tests in facial image comparison (Fig. 11.2) in the European Network of Forensic Scientists (ENFSI) tries to combine the two domains.

When using this conclusion scale, the forensic expert is encouraged to express an expert based opinion, which also can be used in the uncertainty domain. In this way, this scale can be used to express an “experience based uncertainty measure”.

+5	The results of the examination extremely strongly support that the same person is depicted in the questioned and the reference image. <i>(The results are extremely more probable if the main hypothesis is true compared to if the alternative hypothesis is true. <math>1\ 000\ 000 \leq V</math>)</i>
+4	The results of the examination strongly support that ... <i>(The results are far more probable if the main hypothesis is true compared to if the alternative hypothesis is true. <math>10\ 000 \leq V &lt; 1\ 000\ 000</math>)</i>
+3	The results of the examination support that ... <i>(The results are much more probable if the main hypothesis is true compared to if the alternative hypothesis is true. <math>100 \leq V &lt; 10\ 000</math>)</i>
+2	The results of the examination give moderate support that ... <i>(The results are more probable if the main hypothesis is true compared to if the alternative hypothesis is true. <math>10 \leq V &lt; 100</math>)</i>
+1	The results of the examination give weak support that ... <i>(The results are slightly more probable if the main hypothesis is true compared to if the alternative hypothesis is true. <math>2 \leq V &lt; 10</math>)</i>
0	The results of the examination support neither ... nor ... <i>(The results are equally probable if the main hypothesis is true compared to if the alternative hypothesis is true. <math>1/2 &lt; V &lt; 2</math>)</i>
-1	The results of the examination give weak support that it <u>is not</u> ... <i>(The results are slightly more probable if the alternative hypothesis is true compared to if the main hypothesis is true. <math>1/10 &lt; V \leq 1/2</math>)</i>
-2	The results of the examination give moderate support that it <u>is not</u> ... <i>(The results are more probable if the alternative hypothesis is true compared to if the main hypothesis is true. <math>1/100 &lt; V \leq 1/10</math>)</i>
-3	The results of the examination support that it <u>is not</u> ... <i>(The results are much more probable if the alternative hypothesis is true compared to if the main hypothesis is true. <math>1/10\ 000 &lt; V \leq 1/100</math>)</i>
-4	The results of the examination strongly support that it <u>is not</u> ... <i>(The results are far more probable if the alternative hypothesis is true compared to if the main hypothesis is true. <math>1/1\ 000\ 000 &lt; V \leq 1/10\ 000</math>)</i>
-5	The results of the examination extremely strongly support that it <u>is not</u> ... <i>(The results are extremely more probable if the alternative hypothesis is true compared to if the main hypothesis is true. <math>V \leq 1/1\ 000\ 000</math>)</i>

**Fig. 11.2** Conclusion scale used in proficiency tests in facial image comparison. Note that it is possible both to give expert based opinions and uncertainty based measures with this scale

If the expert does not find he/she has the empirical data to express such a ratio, then there is the possibility to perform a one-to-one comparison between the suspect and perpetrator and then use an experience based scale to express a degree of support for or against congruent identity between perpetrator and suspect. If it is questioned in court how many other people could be the perpetrator, the answer should be that this is not known. In this context, the evidence cannot stand for itself but should be considered as a part of an evidence chain.

### ***11.2.3 Directions of Further Research***

It has been shown that, statements based on a forensic expert's experience tend to express stronger support for congruence or non-congruence for identity than statements based on empirical data [47]. This encourages the movement from forensic work based on experience to work based on the empirical data indicating that collecting and building databases must be of primary focus.

The problem of traits which are difficult to label or quantify, such as the profile of the nose, could possibly be addressed by using soft biometric pairwise comparisons where observers use pairwise comparisons to judge which of two persons, e.g., has the most crooked or straight nose, and then rank a group of people by the given trait [44]. The new challenge is to ask the right questions for each pairwise comparison.

However, methods such as pairwise comparisons still need an observer to make the judgements. Another interesting direction of research is to develop automatic recognition systems so the human bias is bypassed. There has been progress made in this direction, for instance in the field of fingerprint examination.

The field of biometrics is tailor made for research in these directions. However, there are some challenges to address before biometric research can be easily transformed into techniques for forensic casework. Some of the problems are that biometric research traditionally does not answer forensic questions and some terms are used differently in the two communities. As an example, the term “recognition rate” is widely used in biometric research to describe how good a given technique is at identifying the right subject in a database. In forensics, “recognition” is used when a witness recognizes a subject, and it is never known whether the perpetrator is present in the database used. One of the ways to bridging the gap would naturally be to find ways to express result of research as uncertainty measures instead.

### 11.3 A Software Platform to Support Forensic Investigations: BioFoV

A forensic expert practitioner needs to handle a diverse set of tools, often not available in the same integrated software package. BioFoV aims to provide a forensic casework tailor-made software, based on state of art knowledge in fields such as soft biometrics, gait recognition, photogrammetry, etc. Forensic experts, police officers and others working with forensic video analysis can also benefit from more “intelligent” video handling tools, such as automatic event detection, to help search the relevant video footage prior to the actual case work.

Software packages aimed at forensic video and image analysis have been developed, such as Forevid [20] or the tools developed by Ocean Systems [39], although more directed either for video capture and processing or enhancement of the material, not fully addressing the above-mentioned challenges. A tool for forensic gait analysis has recently been developed [22] and automatic facial algorithms are valuable tools for the forensic expert [27]. Including such tools into a common and extensible software platform would allow narrowing down the software tools a forensic video analyzer needs to master, and, more importantly, would support the combination of results from different algorithms to enhance overall forensic analysis results.

This book chapter proposes such a software platform, BioFoV, resulting from the cooperation between a biometrics research group and an experienced group of forensic practitioners. The objective is to provide an extensible tool, useful to forensic investigators of different backgrounds. By automating a set of time consuming and laborious tasks, it is expected to enable a deeper case analysis, requiring less effort from the investigator, while allowing to focus the effort on those case aspects that require expert insight, contributing to an overall quality increase.

Two key characteristics of this proposal are: (i) the usage of open software, to avoid complex licensing; and (ii) its extensibility, to allow the inclusion of new modules adding functionality, as well as the possibility to include alternative or improved versions for existing modules when new relevant research emerges making new developments available to the community, instead of remaining in a hard-disk once a research project is completed.

BioFoV is developed in C++, using Qt [43] for the graphical user interface (GUI) and the OpenCV library [41] for supporting computer vision and image processing functionalities. With this option, the same software can be compiled for different platforms. It has been successfully tested on GNU/Linux (Debian and Ubuntu) and Microsoft Windows. An OS X build should also be easily compiled on the target platform. All developed software is Free and Open Source, therefore the final product can be shipped without restrictive licenses or extra fees, allowing free sharing amongst those wanting to use and improve it.

### 11.3.1 User Interface

The GUI allows users to take advantage of the implemented functionalities. It supports the selection of input video contents, its display, information about the current user selections and, to execute the required module functionalities.

The GUI main window—see Fig. 11.3—is organized into four main areas: (i) top menu, where all functionalities are listed in sub-menus; (ii) input and result tabs, where inputs and outputs are listed; (iii) details area, where extra information about what is selected or being played back can be shown to the user; (iv) playback area, supporting user interaction with video or still images, via mouse input, and supporting video actions like seeking, zooming, pausing, or taking snapshots.

The present GUI is structured to support improvements, as further feedback from practitioners is gathered on the best layout for user interaction.

### 11.3.2 Modules

The motivation behind the BioFoV software platform development is the functionality to be offered to forensic researchers. The current version includes a set of general purpose modules, expected to be used by most practitioners, such as camera calibration, or event detection. Also a couple of more specific modules have been included, notably for the computation of soft biometrics such as bodily and joint angle measurements, or facial detection, as examples of possible feature extraction modules.

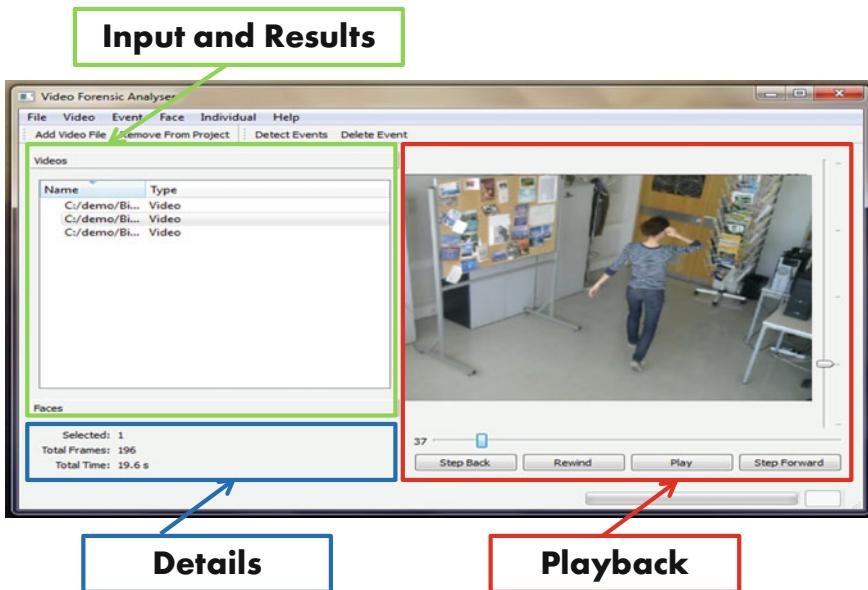
A set of other useful functionalities are included such as: (i) print, to enable quick physical access to the displayed image; (ii) export, allowing extracted features to be bulk exported to separate image files; or (iii) a tool for exporting video events that were automatically detected.

To illustrate BioFoV modules' functionality, sample results are included considering input videos acquired with a Logitech c270 webcam, with  $640 \times 480$  pixel resolution, at 10 frames per second, from which videos were captured using ffmpeg and encoded with an H.264 codec.

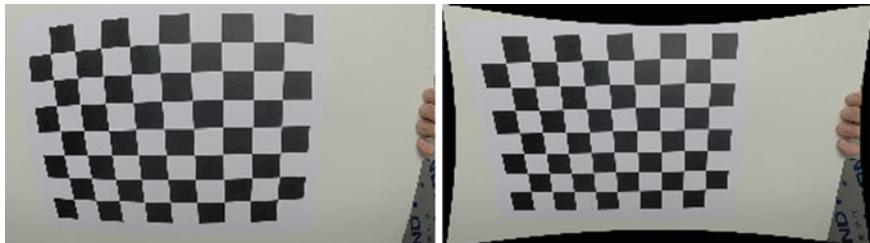
#### 11.3.2.1 Camera Calibration

Even the best video cameras distort the captured images, requiring a calibration step to compensate for distortions and to support accurate measurements. By recording a video of a symmetrical chess pattern with the same camera, the calibration parameters can be calculated once and saved to a file. Later it can be imported for usage with footage captured by the same camera.

The present implementation is based on the OpenCV functions *findChessboardCorners* and *calibrateCamera*, the latter being based on [6, 51]. The



**Fig. 11.3** Graphical User Interface of the proposed BioFoV platform



**Fig. 11.4** Camera calibration example: original (*left*) and calibrated (*right*) images

algorithm is applied to patterns captured from surveillance videos, which sometimes can be blurry, affecting the precision of the calibration. For this reason, the pattern should be moved slowly and under good illumination conditions while performing the calibration. In order to capture the edge distortions properly, the pattern was moved along the areas captured near the image border. Camera calibration results are illustrated in Fig. 11.4.

### 11.3.2.2 Event Detection

When a surveillance video has to be analysed thoroughly, a lot of man time needs to be spent visualizing it, even if nothing relevant is happening for most of the time. This module allows a faster analysis of long videos, automatically looking for the

occurrence of relevant events. The present implementation trims the video parts that show no movement, keeping the user from having to watch the entire video, which sometimes can be several days long. An adaptive background subtraction model is used, checking each video frame for significant changes [53]. The user can adjust a parameter to control the amount of change considered significant enough for selecting frames for further analysis.

As an example, a 2 days, 11 h and 30 min sample video was captured in an office environment. BioFoV identified a set of events considered relevant with a total length of 17 min (210 times shorter). These results were manually validated by an operator as not missing any of the relevant events taking place in the area under surveillance. This illustrates the saving of a considerable amount of time that the forensic expert can use to a better purpose.

### 11.3.2.3 Re-Projected Image Plane Measurements

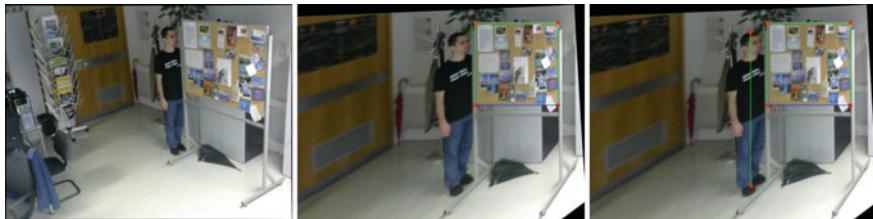
BioFoV enables the computation of soft biometrics such as heights, lengths or angles, by supporting the measurement of distances in a re-projected image plane. Prior to any measurement, the camera has to be calibrated.

An image transformation tool allows an image to be re-projected to a vertical plane visible in a captured frame. This allows, for example, the comparison of heights or widths between known static planar objects and those of items of interest, or to estimate soft biometrics such as bodily measurements of individuals in the re-projected plane. The implemented tools support the measurement of: (i) linear distance between two points; (ii) vertical distance between two points; (iii) horizontal distance between two points; (iv) angles.

It should be noticed that bodily measurements of individuals computed using this tool, for instance the height, is precise only under very strict conditions, such as: (i) the individual standing straight; (ii) being in the same vertical plane as the one used for re-projection; (iii) the point on the top of the head being well defined, as well as its vertical projection on the floor. These conditions are seldom met at the same time. Alternative measurements may be considered, as the height to the eyes, the length of a limb, or the angle between leg and foot in a given part of the gait cycle. As discussed in the previous chapter, the best option can be to compare measurements obtained in similar poses.

A measurement example is presented in Fig. 11.5, where measurement conditions are close to ideal. The image is re-projected into the plane of the board, which is aligned with the person. The board height is known to be 91 cm. Using, a previously calibrated camera, a height measurement value for the person of 1.68 m was obtained, while the person height is in fact 1.72 meters. The difference is mainly due to the imprecision of manually marking the top and bottom points of the person image.

To give the user more freedom and avoid the restriction of the individual having to stand in the same plane as the referential a more complete spatial measurement module needs to be developed, notably to allow, from several post event photos of



**Fig. 11.5** Height measurement with a reference: calibrated frame (*left*); re-projected frame (*middle*); height measurement (*right*)

the same scene, eventually captured from different angles, to reconstruct points, edges, planes, and volumes that can be used as measurement reference. From these, virtual references can be constructed and overlaid with the video, allowing bodily measurements in arbitrary locations in the scene, thus avoiding the coplanar restriction imposed by the previously described image measurement module. A step in this direction is presented in the following chapter.

#### 11.3.2.4 Feature Extraction—Face Detection Example

Feature extraction modules enable the automatic detection and extraction of a selected feature, from a set of videos or from previously detected events. Additional modules can be added accommodating the desired type of analysis.

In the present BioFoV version a feature extraction module dedicated to frontal face detection has been included. The feature extraction module is the implementation of a Haar Feature-based Cascade Classifier [33, 48] and implemented in OpenCV. With this module the user can easily create a facial database of people detected in the video.

#### 11.3.3 How to Get BioFoV

A first release of the BioFoV platform code and an issue tracking tool are freely available online, open to anybody who wants to contribute or use it. The reference for the repository is <https://www.github.com/BioFoV/BioFoV>.

Code is documented using Doxygen and a tutorial for implementing a simple module will be made available, allowing anyone with programming knowledge to integrate new modules and share them with all BioFoV users. The management of the project, and code review of any new feature submission, as well as tracking of bug reports will be ensured by the current team to allow tight quality control on the program releases.

## 11.4 Applications of 3D Markerless Motion Capture in Forensic Gait Analysis

Forensic analysis of gait and bodily dimensions suffers from a high resource demand and the bias caused by the subjective judgements provided by the experts [30], as previously mentioned. Fully automatic gait recognition algorithms like [1, 13] may cope with this in the future. However, these kinds of algorithms are infeasible for forensic analysis at the current stage.

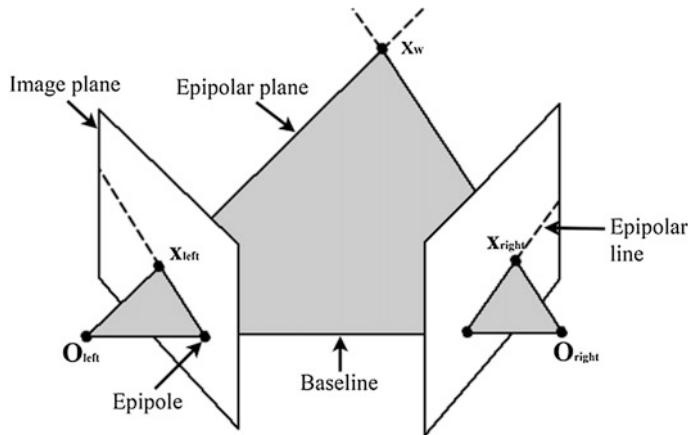
Comparison of gait and bodily dimensions obtained from different image perspectives further challenges the expert's judgement power. The BioFoV's re-projection tool can compensate for this to some extent, but no single view method has proven to be inherently view invariant and thus 3D CCTV may be a solution in the future.

3D imaging can be obtained by active or passive methods. Active methods provide depth perception using the relative position of light emitters and photo or imaging sensors, while passive methods uses the relative positions of the photo or imaging sensors exclusively. Active methods embrace Time-of-Flight (ToF) [50] and triangulation methods based on shape from shading [21], structured light or pseudo-random patterns [49], while passive methods embrace the triangulation methods based on either stereo vision or shape from silhouette (SFS) [3]. The advantages of the active systems are that they are less dependent on ambient illumination from external light sources and the texture of the object, but the emission of light limits the effective range and are often inapplicable under exposure of direct sun light. Passive methods can typically obtain a larger effective range farther out assuming the external light conditions, the focal length and the distance between the cameras are appropriately configured. However, oppose to active systems, SFS methods are highly dependent on high contrast between object and background and stereo vision methods require locally uniqueness in the texture of surfaces.

This chapter describes how 3D measurements and 3D reconstructions can be obtained with stereo vision and how this can be applied to create a laboratory setup feasible for full 3D reconstructions of human gait. The laboratory description is followed by a study demonstrating the discriminatory power of gait features conducted in the proposed 3D laboratory. Lastly, you will find a presentation of the latest work on conducting 3D CCTV for field environments.

### 11.4.1 Accurate 3D Imaging of Human Gait and Bodily Dimensions

To obtain measurements in 3D from images, the pin-hole model is applied to describe the geometric relations between the object and its projection into the image plane. The parameters required to use the model embrace the intrinsic camera



**Fig. 11.6** Sketch of a point  $X_w$  in world space and its projection into the *left* and *right* image planes as the points  $x_{left}$  and  $x_{right}$ , respectively. The optical centers of the cameras are defined as  $O$  and these are separated by a baseline. Together with  $X_w$  the optical centers define the epipolar plane, which intersection with the image planes define the epipolar lines

parameters including the focal length, principal point, pixel aspect ratio, and lens distortion and the extrinsic parameters describing the location and the orientation of the cameras.

Given two cameras with different viewpoints, stereo matching algorithms can be applied to identify point correspondences between the images automatically and knowing the epipolar geometry 3D world coordinates can be computed by triangulation [19]. Point matching algorithms also reduce the search of corresponding points using the epipolar geometry as one point in one image is constrained to lie on the epipolar line in the other image, as illustrated in Fig. 11.6.

The laboratory setup was custom built and consisted of eight monochromatic  $2048 \times 2048$  pixel cameras mounted pair wise with a vertical baseline in the corners of the laboratory. The field of view covered roughly  $2.0 \times 2.0 \times 2.0 \text{ m}^3$ .

Point matching and triangulation was obtained using the Patch based Multi-View Stereo algorithm [11]. The PMVS algorithm finds point correspondences in a three-step procedure embracing matching, expanding and filtering. First image features based on Harris corners [18] and Difference of Gaussians (DoG) [36] are matched by a correlation along the epipolar lines, then an expansion step and a filtration step are iteratively executed to create a denser point cloud and to filter the outliers by constraining the visibility consistency and the point spread. A sequence of 3D reconstructions from a normal gait trial are illustrated in Fig. 11.7.

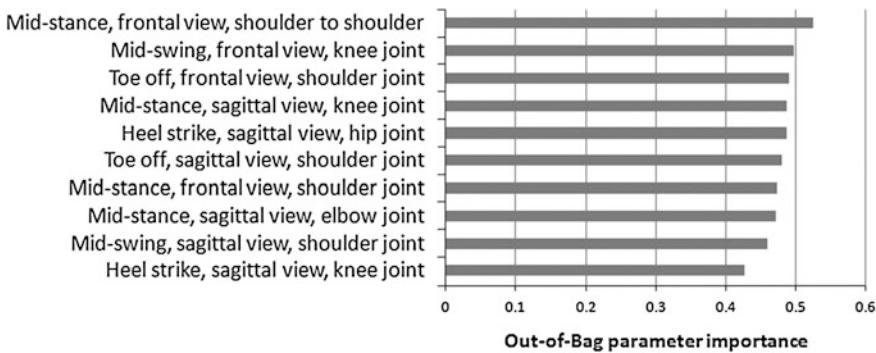


**Fig. 11.7** Side view of multiple 3D point clouds from a person performing a gait cycle

#### **11.4.2 Using Gait Kinematics and Random Forests for Recognition**

Recognition based on the kinematics from the upper and lower extremities has been tested on 16 slim suited participants who were annotated by the same expert. Random forests [7] was applied for automatic recognition as this allows estimation of a certainty score for each recognition using the distribution of the votes among the trees and the importance of the parameters can be quantified by the out-of-bag variable importance [7], which provides substantial information for the decision maker. Five trials from two different days were applied as test and training data, respectively. The training data were applied to train the random forest, whereas recognition was performed on the test data. The random forest was built with 250 trees as the out-of-bag recognition error hereby reached convergence.

The ranking of the parameter importance listed in Fig. 11.8 shows that shoulder-to-shoulder angles, shoulder joint angles and the knee joint angles were among the most important parameters and both the frontal and the sagittal parameters seemed equally important for recognition. However, the importance of



**Fig. 11.8** The top 10 of out-of-bag parameter importance (i.e., the out-of-bag error increase as a fraction of the error before the parameter was randomly permuted)

the kinematic parameters and the view angle depended highly on the phase in the gait cycle, which should be considered during comparisons of gait patterns.

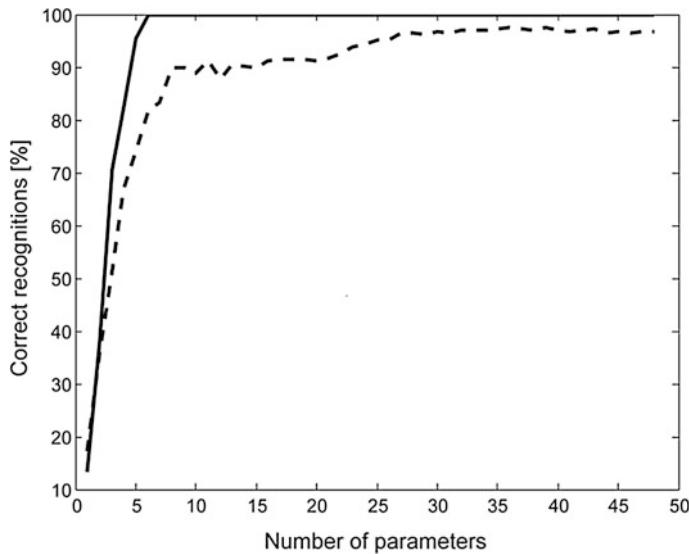
The correct recognition rate as a function of the number of parameters included in the recognition is illustrated in Fig. 11.9. Parameters were included according to the ranking of the parameter importance listed in Fig. 11.8 in decreasing order. The results show that single parameters have a relatively weak discriminatory power with 13% to 17% correct recognitions. With inclusion of six parameters, all participants were correctly recognized based on five trials and the correct recognition rate for one gait trial converged around 27 parameters with over 96% correct recognitions.

The results of the study emphasize the importance of considering a series of gait trials rather than a single trial and to consider the kinematics of both the upper and lower extremities.

Current study was conducted with a controlled experimental setup and a group of participants with limited variations in terms of bodily dimensions. Challenges related to gender, adiposity, race, clothing, lighting conditions, etc. were therefore not considered in the study and these parameters have to be addressed in future studies to reflect the practical issues regarding surveillance.

#### 11.4.3 3D Surveillance and Future Perspectives in Gait Recognition

In the previous sections, we have shown the discriminatory power of gait and bodily dimensions and how 3D imaging can provide an improved alternative for biometric measurements compared to 2D images. However, the 3D imaging method presented so far is only feasible for laboratory environments. Therefore, the research community is currently testing other 3D acquisition systems suited for

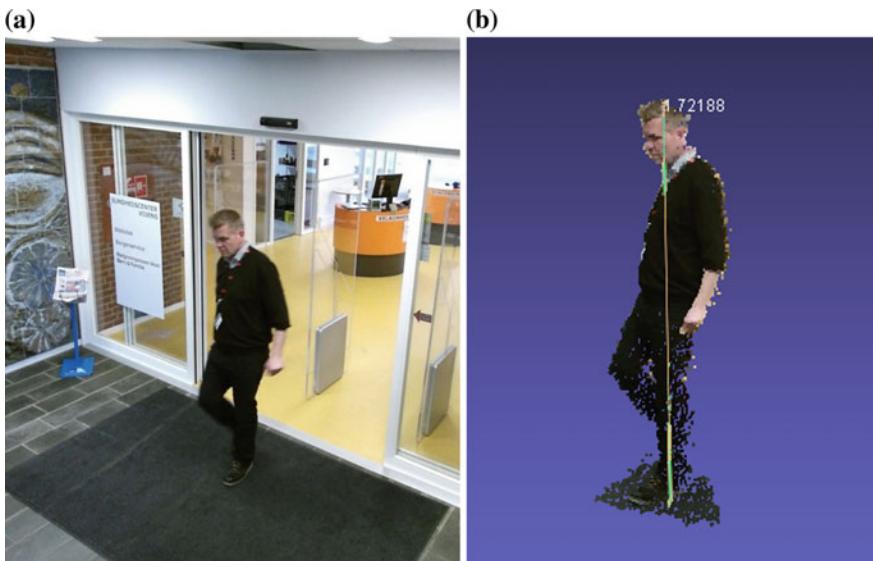


**Fig. 11.9** Correct recognition rate as function of number of random forest parameters. Parameters were included according to their importance in decreasing order. The *dashed line* and the *solid line* are recognition based on one and five gait trials, respectively

biometric analysis in field environments. So far, a low cost system consisting of a Kinect v2,<sup>1</sup> a mini computer and a NAS server, which enables acquisition of 3D point clouds with an effective range between 2 and 5 m. The Kinect obtains 3D reconstructions using the ToF technology and thus it is less sensitive to textures on clothes and skin but suffers from the exposure of direct sunlight. Nevertheless, the Kinect v2 has several advantages as it is already calibrated such that the 3D point cloud is scaled in known metrics and is capable of capturing RGB images so the 3D point cloud can be textured with ‘true’ colors, which may be helpful for the forensic expert. Proof of concept experiments are currently in progress in various indoor environments as illustrated in Fig. 11.10. However, further scientific work should also validate the accuracy of the 3D reconstructions obtained with the system.

---

<sup>1</sup>Microsoft, Redmond, Washington, USA.



**Fig. 11.10** **a** RGB image obtained from the Kinect and **b** body height measurement obtained from the corresponding 3D point cloud rotated to side view

## 11.5 Extraction of Soft Biometrics from Facial Images

A significant number of soft biometric traits can be extracted from face images and facial movements. This generally includes gender recognition (i.e., man vs. woman), age categorization (e.g., child, youth, adult, middle-age, and elderly) and ethnicity classification (e.g., Asian, Caucasian and African). Other soft biometric traits that can be extracted from faces include eye movements, kinship information (i.e., verifying whether two persons are from the same family or not), skin and eye color, the periocular region of a face image, facial tattoo, facial scars and moles, etc.

Though there has been a great deal of progress in face analysis in the last two decades, most of the work has mainly focused on face detection and recognition problems, rather than on soft biometrics. Consequently, the design of algorithms that are effective in discriminating between males and females, or classifying faces into different age and ethnic categories remain open areas of research. Findings from human perception studies, suggested that peculiar texture patterns, sparsely located on the face image, are often used by humans for a rapid identification of faces. The same approach demonstrated to be applicable in forensic cases, as well as other applications. Such isolated patterns, which may correspond either to a skin defect (such as a scar or a mole), or to a specific facial area (eyebrow, chin) with a distinctive appearance, often do not allow for an accurate classification but can be exploited in conjunction with other techniques to improve the recognition performance. For instance, Klare et al. [26] proposed an approach, where human

describable face attributes are exploited to perform face identification in criminal investigations. The extracted facial attributes, such as eyebrows, chin, and eyes shape are compared with the same attributes encoded in hand drawn police sketches. The experimental analysis demonstrated its applicability in forensic identification of suspects.

### ***11.5.1 Extracting Gender from Face Images***

First attempts of using computer vision based techniques to gender classification started in early 1990s. Since then, a significant progress has been made and several approaches have been reported in literature. Fundamentally, the proposed techniques differ in (i) the choice of the facial representation, ranging from the use of simple raw pixels to more complex features, such as Gabor responses, and in (ii) the design of the classifier, ranging from the use of nearest neighbor (NN) and fisher linear discriminant (FLD) classifiers to artificial neural networks (ANN), support vector machines (SVM), and boosting schemes. For instance, Moghaddam and Yang [37] used raw pixels as inputs to SVMs while Baluja and Rowley [2] adopted AdaBoost to combine weak classifiers, constructed using simple pixel comparisons, into a single strong classifier. Both systems showed good classification rates of above 90% but under controlled settings. An empirical comparative analysis on gender classification approaches can be found in [35] whereas a recent survey on the topic can be found in [38]. Note that there is no public database specifically designed for gender recognition evaluation. Most of the recent works have been evaluated on the Images of Groups database [12], which is a collection of face images of groups of people from Flickr. This database has been mainly designed for unconstrained demographic classification (age and gender).

### ***11.5.2 Age Classification from Facial Images***

Automatic age classification aims to assign a label to a face regarding the exact age (age estimation) or the age category (age classification) it belongs to. This is challenging problem because the appearance of a particular face varies due to changes in pose, expressions, illumination, and other factors such as make-up, occlusions, image degradations caused by blur and noise, etc. In addition to these difficulties, which are shared with the standard problems of face recognition, aging is a very complex process that is extremely difficult to model: a group of people of the same age may look very different depending on, for example, environment, lifestyle, genes, etc. Thus, deriving a universal age classification model is troublesome. Many face image representations for age estimation have been studied, such as anthropometric models, active appearance models (AAM), aging pattern subspace, and age manifold. An extensive review of age representation methods can

be found in [9]. Regarding age classification schemes, the existing methods are based on either pure classification or regression analysis. Perhaps, among the pioneering studies on age classification are those proposed by Kwon and da Vitoria Lobo [28], Lanitis et al. [29], and Guo et al. [14]. Although relatively successful in some scenarios (e.g., high-quality and occlusion-free images, neutral facial expression), most existing methods tend to suffer under uncontrolled settings as noted in [9]. The most commonly used public face databases for evaluating age classification are MORPH [45] and FG-NET [42] databases and the more challenging Images of Groups [12] database.

### ***11.5.3 Ethnicity Classification from Facial Images***

While gender and age recognition have been explored by many other researchers, automatic ethnicity classification problem has received relatively far less attention despite the potential applications. This is perhaps due to the ambiguity and complexity in defining and describing different ethnic groups. The terms “race” and “ethnicity” are sometimes used interchangeably although they refer to biological and sociological factors, respectively. Generally, race refers to a person’s physical appearance or characteristics, while ethnicity is more viewed as a culture concept, relating to nationality, rituals and cultural heritages, or even ideology [10]. The most commonly encountered and accepted racial groups are African/African American, Caucasian, East Asian, Native American/American Indian, Pacific Islander, Asian Indian, and Hispanic/Latino. A very recent survey on extracting race from face images can be found in [10]. Most of the proposed methods are evaluated on general face databases, which are not intentionally designed for race classification.

### ***11.5.4 Experimental Analysis on Extracting Facial Soft Biometrics from Videos***

The way a person is moving his/her head and facial parts (such as the movements of the mouth when a person is talking) defines so called facial dynamics and characterizes personal behaviors. In this section, we provide an experimental analysis on the use of facial dynamics for extracting soft biometrics from video sequences. Two baseline approaches based on LBP features [40] and support vector machines (SVM) are implemented and discussed. The first approach is using only static images and thus ignoring the facial dynamics while the second approach uses spatiotemporal representation thus combining facial structure and dynamics. The aim of the experiments is to evaluate the benefit of incorporating the facial dynamics. The choice of adopting the LBP approach [40] is motivated by the recent

success of using it for combining appearance and motion for face and facial expression recognition [15] and for dynamic texture recognition [52]. We describe below the two experimental approaches and then report the results on gender classification, age estimation and ethnicity classification.

#### **11.5.4.1 Static Image-Based Approach**

Given a target face video sequence, a straightforward approach to extract facial soft biometrics is to analyze each frame in the video and then combine the results through majority voting which consists of determining the gender (or age or identity) in every frame and then fusing the results. Basically, for efficient representation, each facial image (frame) is divided into several local regions from which LBP histograms are extracted and concatenated into an enhanced feature histogram. Then, the results are presented to an SVM classifier for recognition. Finally, the recognition scores over the face sequence are fused using majority voting. In such an approach, only static information is used while the facial dynamics are discarded.

#### **11.5.4.2 Spatiotemporal-Based Approach**

For spatiotemporal representation, volume LBP (VLBP) operator has been introduced in [52] and successfully used for combining appearance and motion for face and facial expression recognition [15] and for dynamic texture recognition [52]. The idea behind VLBP is very simple. It consists of looking at a face sequence as a rectangular prism (or volume) and defining the neighborhood of each pixel in three-dimensional space (X, Y, T) where X and Y denote the spatial coordinates and T denotes the frame index (time). Then, similarly to LBP in spatial domain, volume textons can be defined and extracted into histograms. Therefore, VLBP combines structure and motion together to describe the moving faces.

Each face sequence is divided into several overlapping rectangular prisms of different sizes, from which local histograms of VLBP code occurrences are extracted. To combine facial structure and dynamics, VLBP features are first extracted from the face sequences and feature selection is performed using AdaBoost. The result is then fed to an SVM for classification. In such an approach, both static facial information and facial dynamics are used.

#### **11.5.4.3 Experiments on Gender Recognition**

The static image-based and the spatiotemporal-based approaches are first applied to the problem of gender recognition from videos. Three different publicly available video face databases (namely CRIM (<http://www.crim.ca/>), VidTIMIT [46] and Kanade et al. [25]) are considered. They contain a balanced number of male's and

female's sequences and include several subjects moving their facial features by uttering phrases, reading broadcast news or expressing emotions. The datasets are randomly segmented to extract over 4,000 video shots of 15–300 frames each. From each shot or sequence, the eye positions are automatically detected from the first frame. The determined eye positions are then used to crop the facial area in the whole sequence. Finally, the resulted images are scaled into  $40 \times 40$  pixels. For evaluation, a fivefold cross-validation test scheme is adopted by dividing the 4,000 sequences into five groups and using the data from four groups for training and the left group for testing. This process is repeated five times and we report the average classification rates. When dividing the data into training and test sets, we explicitly considered two scenarios. In the first one, a same person may appear in both training and test sets with face sequences completely different in the two sets due to facial expression, lighting, facial pose, etc. The goal of this scenario is to analyze the performance of the methods in determining the gender of familiar persons seen under different conditions. In the second scenario, the test set consists only of persons who are not included in the training sets. This is equivalent to train the system on one or more databases and then do evaluation on other (different) databases. The goal of this scenario is to test the generalization ability of the methods to determine the gender of unseen persons. Table 11.1 summarizes the gender classification results using the two approaches (static image based and spatiotemporal based) in both scenarios (familiar and unfamiliar). We can notice that both methods gave better results with familiar faces than unfamiliar ones. This is not surprising and can be explained by the fact that perhaps the methods did not rely only on gender features for classification but may also exploited information about face identity. For familiar faces, the combination of facial structure and dynamics yielded in perfect classification rate of 100%. This proves that the system succeeded in learning and recognizing the facial behaviors of the subjects even under different conditions of facial expression, lighting and facial pose. For unfamiliar faces, the combination of facial structure and dynamics yielded in classification rate of about 83% which is still encouraging although the best result for unfamiliar faces is obtained using the static image-based approach (without facial dynamics). This may indicate that incorporating motion information with facial appearance was useful for only familiar faces but not with unfamiliar ones. More detailed experiments and results can be found in [16].

**Table 11.1** Gender classification results on test videos of familiar and unfamiliar subjects using static image-based and spatiotemporal-based methods

Method	Gender classification rates	
	Familiar subjects (%)	Unfamiliar subjects (%)
Static image-based	94.4	90.6
Spatiotemporal-based	100	82.9

#### 11.5.4.4 Experiments on Age Estimation

To study whether facial dynamics may enhance the automatic age estimation performance, a set of experiments using the static image-based and spatiotemporal-based approaches is performed. Five age classes are considered as follows: child = 0–9 years old; youth = 10–19; adult = 20–39; middle age = 40–59 and elderly = above 60. Then, a novel classification scheme based on a tree of four SVM classifiers is built. The first SVM classifier is trained to learn the discrimination between child class and the rest. If the target face is assigned into the child category, then the classification is completed. Otherwise, the second SVM classifier is examined to decide whether the face belongs to the Youth category or not. If not, the third SVM is examined and so on. The static image-based and spatiotemporal-based approaches are applied to age estimation from videos. For evaluation, a set of video sequences (mainly showing celebrities giving speeches in TV programs and News) is collected from Internet. The videos of unknown individuals (especially children), are manually labeled using our (human) perception of age. Then, the videos are randomly segmented to extract about 2,000 video shots of about 300 frames each. In the experiments, we adopted a tenfold cross-validation test scheme by dividing the 2,000 sequences into 10 groups and using the data from nine groups for training and the left group for testing. We repeated this process 10 times and we report the average classification rates. The performances of both static image-based and spatiotemporal-based approaches are shown in Table 11.2. From the results, we can notice that both methods did not perform very well and this somehow confirms the difficulty of the age estimation problem. Interestingly, the static information-based method significantly outperformed the spatiotemporal-based method (i.e., combination of face structure and dynamics). This might be an indication that facial dynamics is not useful for age estimation. However, due to the challenging nature of the age estimation problem, it is perhaps too early to make such a conclusion and more investigations are needed to study the integration of facial dynamics and facial structure for age estimation.

#### 11.5.4.5 Experiments on Ethnicity Classification (Asian Versus Non-Asian)

Similarly to the previous experiments on gender recognition, the two experimental approaches are also applied to ethnicity classification from videos. Because of lack of ground truth data for training, only two ethnic classes (namely Asian and non-Asian) are considered. The same set of 2,000 video shots previously used in the experiments on age estimation is also considered here for ethnicity classification

**Table 11.2** Average age classification rates

Method	Average age classification rates (%)
Static image-based	77.4
Spatiotemporal-based	69.2

**Table 11.3** Average ethnicity classification rates using static image-based and spatiotemporal-based methods

Method	Ethnicity classification rates (%)
Static image-based	97.0
Spatiotemporal-based	99.2

tests. A manual labeling yielded in 81% of non-Asian and 19% of Asian data samples (i.e., video shots). For evaluation, we also adopted a fivefold cross-validation test scheme. The performances of both static image-based and spatiotemporal-based approaches are shown in Table 11.3. From the results, we can notice that both approaches perform quite very well but the spatiotemporal-based method (i.e., combination of face structure and dynamics) slightly outperforms the static image-based method (using only facial structure). This is somehow surprising because one may not expect better results using spatiotemporal methods for ethnicity classification.

#### 11.5.4.6 Discussion

To gain insight into the use facial dynamics in extracting face soft biometric traits, we considered two approaches to face analysis from videos using LBP features and SVMs, and reported experimental results on gender classification, age estimation and ethnicity determination. The experiments on age estimation pointed out that combining face structure and dynamics does not enhance the performance of static image-based automatic systems. Our experiments also showed that incorporating motion information with facial appearance for gender classification might be only useful for familiar faces but not with unfamiliar ones (while the psychological and neural studies indicated that facial movements do contribute to gender classification in the human visual system). Finally, our experiments on the ethnicity classification problem yielded in quite surprising results indicating some relative usefulness of facial dynamics in ethnicity classification.

The primary goal of the above experiments is to provide the reader with clear case studies and examples on using facial dynamics for extracting facial soft biometric traits from videos. Note that the reported results may be specific to the methods and test material used in the experiments. Therefore, it is perhaps early to make final conclusions on the role of facial dynamics in extracting facial soft biometrics.

While many other works on extracting facial soft biometrics reported accuracy of above 90%, most of these studies have utilized face images captured in rather controlled sensing and cooperative subject scenarios [17]. However, in many real-word applications such as video surveillance, the available face images of a person of interest are most likely to be captured under unconstrained and un-cooperative scenarios. Most of the proposed techniques have been evaluated on general face databases, which are not intentionally designed for the problem in

hands. Thus, there is clear lack of benchmark databases and protocols to evaluate the progress in extracting soft biometric traits. Recently, efforts have been made within the BeFIT project (<http://www.fipa.cs.kit.edu/befit/>) by proposing standardized datasets and protocols for evaluating different face analysis tasks. We approve this valuable initiative, which allows a fair comparison and an easy reproduction of research results.

## 11.6 Conclusions

Today, biometric measurements in CCTV are used in a limited number of cases compared to the number of cameras installed in urban environments, despite the methods being known for decades and the resulting evidences being fully usable in court. The limited applications partly rely on a high resource demand of skilled experts and partly on the typically limited image quality and the challenging perspectives, which reduce the strength of evidence [5, 32, 31].

To meet these challenges in future, the research community has developed a new software platform which makes various video analysis tools and photogrammetric tools easier to apply in a user-friendly manner. Furthermore, a study on applying 3D imaging for analysis of human gait has been conducted, which showed that gait recognition can be conducted with high reliability by using only six carefully selected kinematic parameters. Therefore, the current research of the community focuses on developing new 3D imaging methods suitable for CCTV in the fields. Finally, whenever faces are visible in surveillance videos, this information should be taken into account in the recognition process. In unconstrained scenarios, using low resolution unaligned images, traditional face recognition algorithms may not be used, but several soft biometrics may still be considered, for instance exploring facial dynamics.

## References

1. Ariyanto G, Nixon M (2012) Marionette mass-spring model for 3D gait biometrics. In: Proceedings of the International Conference on Biometrics (ICB), pp 354–359
2. Baluja S, Rowley H (2007) Boosting sex identification performance. Int J Comput Vision 71:111–119
3. Baumgart BG (1974) Geometric modeling for computer vision. Stanford University
4. Biber K (2009) Visual jurisprudence: the dangers of photographic identification evidence. CJM 78:35–37
5. Bouchrika I, Goffredo M, Carter J, Nixon M (2011) On using gait in forensic biometrics. J Forensic Sci 56(4):882–889
6. Bouguet J-Y (2011) MATLAB calibration tool. [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/). Accessed 26 Mar 2015
7. Breiman L (2001) Random forests. Mach Learn 45:5–32

8. Dantcheva A, Velardo C , D'angelo A, Dugelay JL (2011) Bag of soft biometrics for person identification: new trends and challenges. *Multimed Tools Appl* 51(2):739–777
9. Fu Y, Guo G, Huang TS (2010) Age synthesis and estimation via faces: a survey. *IEEE Trans Pattern Anal Mach Intell (T-PAMI)*, 32(11):1955–1976
10. Fu S, He H, Hou Z (2014) Learning race from face: a survey. *IEEE Trans Pattern Anal Mach Intell (TPAMI)*
11. Furukawa Y, Ponce J (2010) Accurate, dense and robust multi-view stereopsis, vol 32
12. Gallagher AC, Chen T (2009) Understanding images of groups of people. In: Proceedings of IEEE CVPR
13. Guan Y, Li C-T (2013) A robust speed-invariant gait recognition system for walker and runner identification. In: The 6th IAPR international conference on biometrics: IAPR, pp 1–8
14. Guo G, Mu G, Fu Y, Huang T (2009) Human age estimation using bio-inspired features. In: CVPR'09, pp 112–119
15. Hadid A, Pietikäinen M, Li SZ (2007) Learning personal specific facial dynamics for face recognition from videos. In: IEEE international workshop on analysis and modeling of faces and gestures (in conjunction with ICCV 2007), pp 1–15
16. Hadid A, Pietikäinen M (2008) Combining motion and appearance for gender classification from video sequences. In: 19th international conference on pattern recognition (ICPR 2008), p 4
17. Han H, Jain AK (2014) Age, Gender and Race Estimation from Unconstrained Face Images, MSU Technical report (2014): MSU-CSE-14-5
18. Harris C, Stephens M (1988) A combined corner and edge detector. In: Alvey vision conference: Manchester, UK, p 50
19. Hartley R, Zisserman A (2003) Multiple view geometry in computer vision, 2nd edn. Cambridge University Press
20. Hautamaki S (2011) Forevid: an open source software for forensic video analysis. MSc Thesis, Tampere University of Technology
21. Horn BKP (1970) Shape from shading: a method for obtaining the shape of a smooth opaque object from one view
22. Iwama H, Muramatsu D, Makihara Y, Yagi Y (2012) Gait-based person-verification system for forensics. In: 2012 IEEE 5th international conference on biometrics: theory, applications and systems (BTAS), Sept 2012, pp 113–120
23. Jain A, Dass S, Nandakumar K (2004) Soft biometric traits for personal recognition systems. In: Proceedings of the international conference on biometric authentication, ICBA, LNCS 3072, pp 731–738
24. Jain A, Russ A (2015) Bridging the gap: from biometrics to forensics. *Philos Trans Roy Soc B*
25. Kanade T, Cohn JF, Tian Y (2000) Comprehensive database for facial expression analysis. In: IEEE international conference on automatic face and gesture recognition, pp 46–53
26. Klare B, Klum S, Klontz J, Taborsky E, Akgul T, Jain AK (2014) Suspect identification based on descriptive facial attributes. In: Proceedings of the international joint conference on biometrics
27. Klontz JaC, Jain AK (2013) A case study of automated face recognition: the Boston marathon bombings suspects. *IEEE Comput* 46(11):91–94
28. Kwon YH, da Vitoria Lobo N (1994) Age classification from facial images. In: CVPR'94, pp 762–767
29. Lanitis A, Taylor C, Cootes T (2002) Toward automatic simulation of aging effects on face images. *TPAMI* 24(4):442–455
30. Larsen PK, Hansen L, Simonsen EB, Lynnerup N (2008) Variability of bodily measures of normally dressed people using PhotoModeler® Pro 5. *J Forensic Sci* 53:1393–1399
31. Larsen PK, Lynnerup N, Henriksen M, Alkjær T, Simonsen EB (2010) Gait recognition using joint moments, joint angles, and segment angles. *J Forensic Biomech* 1:7
32. Larsen PK, Simonsen EB, Lynnerup N (2008) Gait analysis in forensic medicine. *J Forensic Sci* 53:1149–1153

33. Lienhart R, Maydt J (2002) An extended set of haarlike features for rapid object detection. In: Proceedings of the international conference on image processing, Rochester, USA
34. Lucy D (2005) Introduction to statistics for forensic scientists. Wiley
35. Makinen E, Raisamo R (2008) An experimental comparison of gender classification methods. *Pattern Recogn Lett* 29(10):1544–1556
36. Marr D, Hildreth E (1980) Theory of Edge Detection. *Proc Roy Soc B: Biol Sci* 207:31
37. Moghaddam B, Yang M-H (2002) Learning gender with support faces. *IEEE Trans Pattern Anal Mach Intell* 24(5):707–711
38. Ng CB, Tay YH, Goi B-M (2012) Recognizing human gender in computer vision: a survey. *PRICAI* 335–346
39. Ocean Systems (2015) Ocean systems forensic video and image analysis solutions. <http://www.oceansystems.com>. Accessed 26 Mar 2015
40. Ojala T, Pietikäinen M, Mäenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 24:971–987
41. OpenCV (2015) Open source computer vision and machine learning software library. <http://www.opencv.org>. Accessed 26 Mar 2015
42. Panis G, Lanitis A (2014) An overview of research activities in facial age estimation using the FG-NET aging database. In: International ECCV workshop on soft biometrics
43. Qt (2015) Cross-platform application and UI framework. <http://www.qt.io/>, Accessed 26 Mar 2015
44. Reid DA, Nixon MS, Stevenage SV (2013) Soft biometrics; human identification using comparative descriptions. *IEEE Trans Biom Compnd Pattern Anal Mach intell.* 36(6)
45. Ricanek K, Tesafaye T (2006) MORPH: a longitudinal image database of normal adult age-progression. In: Proceedings of FG
46. Sanderson C, Paliwal KK (2003) Noise compensation in a person verification system using face and multiple speech feature. *Pattern Recogn* 36(2):293–302
47. Thornton J, Peterson J (2002) The general assumptions and rationale of forensic identification. In: Modern scientific evidence: the law and science of expert testimony, vol 3. West Publishing Company
48. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition, 2001. CVPR 2001, vol 1, pp I-511–I-518
49. Will PM, Pennington KS (1971) Grid coding: a preprocessing technique for robot and machine vision. *Artif Intell* 2:319–329
50. Xu Z, Schwarte R, Heinol H-G, Buxbaum B, Ringbeck T (1998) Smart pixel: photonic mixer device (PMD); new system concept of a 3D-imaging camera-on-a-chip. In: Eheung EHM (ed) Proceedings: M2VIP '98; Nanjing, China, 10–12 Sept 1998, Hong Kong. pp 259–64
51. Zhang Z (2000) A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell* 22(11):1330–1334
52. Zhao G, Pietikäinen M (2007) Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans Pattern Anal Mach Intell* 29(6):915–928
53. Zivkovic Z (2004) Improved adaptive Gaussian mixture model for background subtraction. In: Proceedings of the 17th international conference on pattern recognition, 2004. ICPR 2004, Aug 2004, vol 2, pp 28–31

# Chapter 12

## Locating People in Surveillance Video Using Soft Biometric Traits

Simon Denman, Michael Halstead, Clinton Fookes and Sridha Sridharan

**Abstract** Semantic descriptions are a commonly used and very natural way for people to describe one another. Descriptions comprising details of clothing types and colours, skin and hair colour, gender and build are very effective ways to communicating an approximate appearance; however such descriptions are not easily utilised within intelligent video surveillance systems, as they are difficult to transform into a representation that can be utilised by computer vision algorithms. In this chapter, we will investigate two recent approaches to using these semantic, soft biometric descriptions to automatically locate people in surveillance imagery. We present the strengths and weaknesses of each, and discuss their suitability for real-world deployment and how they may be further improved.

### 12.1 Introduction

When describing the appearance of a person, we commonly use a set of semantic traits. This includes the use of details such as the colour of clothing, skin or hair, the type and pattern (or lack of pattern) of the clothing being worn, and the gender, build or stature of the person being described. Traits such as these can be viewed as a set of soft biometrics, and are widely used in soft biometric systems [6, 17].

In law enforcement and security situations, soft biometrics such as these are commonly used to describe a suspect or person of interest (i.e. ‘a tall, middle aged male

---

S. Denman (✉) · M. Halstead · C. Fookes · S. Sridharan  
Image and Video Research Laboratory, Queensland University of Technology,  
2 George Street, Brisbane 4000, Australia  
e-mail: s.denman@qut.edu.au

M. Halstead  
e-mail: m.halstead@qut.edu.au

C. Fookes  
e-mail: c.fookes@qut.edu.au

S. Sridharan  
e-mail: s.sridharan@qut.edu.au



**Fig. 12.1** A closed circuit television image of the London bombing suspects taken from [3]. The low resolution of the image and small size of the suspects faces makes identification using traditional biometrics such as face recognition extremely challenging

wearing a red hoodie and black jeans'). Due to the ease and speed in which these descriptions can be relayed and understood by human operators, soft biometrics are part of a shifting focus in the surveillance community.

However locating a person from such a description is still a significant challenge. Typically, it is a manual task that involves either combing through hours of CCTV footage, or having officers on the ground looking. The recent Boston Marathon bombing provides an example of how challenging this task is, as it took the FBI 3 days to search CCTV footage based on eye-witness reports and release suspect photographs to the public [19].

State-of-the-art identification currently still relies heavily on classical biometric systems like facial recognition, however these may fail in a surveillance environment due to factors including low-resolution cameras and relative distance from the camera, deliberate and accidental target occlusions, and various lighting conditions. Although carrying the limitations of a lack of permanence and limited individual discriminability, soft biometrics remain detectable at distance from low-resolution cameras (commonly seen in CCTV networks) where classical biometrics may fail, as witnessed in Fig. 12.1, depicting the London bombers.

Anthropometric measurements can be viewed as a forerunner to soft biometrics. Lucas and Henneberg [18] compared Anthropometric methods of body measurement to facial recognition, and demonstrated their effectiveness as a means of subject identification. When identifying or redetecting a subject of interest in real-world surveillance, the outward appearance is generally dominated by the appearance of

clothing, as illustrated in Fig. 12.1. Jaha and Nixon [16] have shown the efficacy of these traits for subject identification, and Scoleri et al. [24] have shown that while clothing may distort body measurements, accurate data can still be obtained in their presence, further improving the ability for soft biometrics to perform identification tasks.

Recently, spurred on by developments in soft biometrics and attribute matching [29], a number of automated approaches to locate people in video from semantic descriptions have been proposed [8, 22]. The development of such techniques has the potential to greatly reduce the manual labour required to locate people, while also improving security. To date, approaches have operated in one of two ways, either as

1. a database indexing approach [14, 22, 27, 30], where people are first detected (i.e. using a person detection approach [4, 13, 21]) in a video feed, and their attributes are subsequently extracted and stored in a database for later search and retrieval;
2. a stimulus-driven approach [5, 8, 9, 15], where the image is searched directly using a template or model that describes the intended subject.

However, despite the growing interest in this area, existing approaches are limited in their utility. In particular, a reliance on person detection to first locate people in a scene limits a technique's ability to operate in crowded and complex scenes. While the use of person detection as a preprocessing step may initially seem intuitive, it was shown in [8] that the use of person detection actually reduces the performance of a semantic person search system. The complex surveillance environment in which the search takes place is also a challenging environment for person detection, and both false alarms and missed detections are highly common, complicating the search process. Indexing approaches are also limited by their design, which restricts their use to post-event analysis tasks as they are unable to search live data. Finally, the use of simplistic methods to represent the query limits the power of the search, and leads to a variety of matching errors including regions of the background being detected as the person of interest.

Other key considerations for any semantic search include the need to model uncertainty: both within the incoming video data and the provided semantic query. For instance, a shirt that is described by a witness as ‘red’ may actually be ‘orange’; but due to local lighting conditions in the scene this same garment may actually appear to be ‘brown’ in CCTV footage. Similar concerns exist with other traits, including the type of clothes being worn, and the height or build of the person of interest.

In this chapter, we discuss two recent approaches to locate people in surveillance footage using soft biometric traits. The first uses a simplistic region-based representation to describe a person, and a particle filter to search and track them in video [15]. The second uses a channel representation to model the person’s appearance [9], which is then incorporated into a single object tracking approach [12] to locate and track the target subject. Both of these approaches are stimulus driven (i.e., they do not require the use of person detection and are able to operate in a live setting), and both use dominant torso and leg colours, height and clothing type to describe a per-

son's appearance. We compare the performance of these two approaches, and discuss their strengths, limitations and suitability for use in large surveillance installations.

The remainder of this chapter is structured as follows: Sect. 12.2 presents an overview of recent research in the area; Sect. 12.3 discusses the traits used by the two systems and how they are represented; Sects. 12.4 and 12.5 outline the approaches of [15] and [9] respectively; Sect. 12.6 outlines the data and evaluation protocol used in this chapter; Sect. 12.7 presents experimental results and discussion; and Sect. 12.8 concludes this chapter.

## 12.2 Prior Work

In surveillance and security two distinct tracks of personal identification can be considered: person re-identification and person search (semantic search). In each case, the primary goal is to locate and tag each instance of a target subject's appearance in a video, based on the subject's soft biometric signature.

While the re-identification [1, 10] of a target is a common method of target search, in many law enforcement situations visual enrolment of the subject (i.e. enrolment based on an acquired image) is not possible and a description is required. Of primary concern in description-based target searches is limiting the impact of the ‘semantic gap’, Reid et al. [23] outline methods of achieving this in their study of comparative soft biometric labels, reducing the uncertainty and subjectiveness witnessed in categorical labelling.

While several approaches have been proposed to locate a person from a semantic query, the majority of these approaches are detection based [14, 22, 27, 30], in that they first require the person (or in the case of [14], the vehicle) to be detected using a computationally expensive object detection routine (such as [4, 13]). Following this detection, traits are extracted and stored in a database, allowing later processes to search for an entity within this index.

A variety of traits are used including clothing colour [22, 27, 30], gender and the presence of luggage [27], and facial attributes (i.e. bald, hair, hat) [30]; while [14] incorporates details on the vehicle size, its location and direction of travel. While promising results are obtained for all systems, the evaluations of these approaches are piecemeal, and a rigorous evaluation using a standard database is not performed. Furthermore, these approaches are all reliant on object detection, limiting utility in crowded environments. In an unconstrained surveillance environment, person detection remains a challenging problem with state-of-the-art approaches [21] still prone to false alarms and missed detections, while also being computationally expensive.

In contrast to detection and indexing-based approaches, the techniques proposed in [5, 8] are designed to work with a live video feed. Motivated by a desire to reduce confrontations between rival sporting fans, [5] developed a system to detect situations where supporters of one team are located near supporters of a second, based on known colour quantities (i.e. jersey colours) within a crowded scene. While not spe-

cific to a single individual, the system did have some success in accurately gauging when two possibly hostile groups were converging on each other.

Denman et al. [8] sought to interrogate the scene directly to locate people who matched a given description. This technique allowed for a query to be searched for within a live video feed without the use of detection routines. Colour (torso and leg) and height features are used to construct a simplistic avatar, which the system subsequently uses to search the video feed. A particle filter is used to detect and track the subject, with the avatar being compared to a region described by each particle to determine the similarity of the particle to the target. This approach [8], although moderately effective, also has several limitations, largely arising from the use of a small number of traits and the manner in which the avatar is constructed. The simplistic nature of the avatar, which effectively models the person as a series of regions which are assumed to be dominated by the target clothing colours, leads to the incorrect localisation of several subjects due to factors including incorrect trait attachment (torso as legs, and background as torso) and cases where the background is confused for an individual (i.e. the background is the same approximate colour as the target).

In this chapter, we examine two recent approaches to the problem of locating a person in a scene from a semantic query, both of which search the scene directly rather than using person detection to first localise targets. The approach of Halstead et al. [15] is an extension of [8], which overcomes some of the problems relating to incorrectly localising targets in background regions. We also discuss the approach of Denman et al. [9], who propose using a channel representation (CR) [11] to model the search query. In this approach the target query is represented by a multi-dimensional image, where each channel corresponds to a particular colour band, or texture type.

### 12.3 Modelling Traits

A variety of traits are available for use when searching for a person from a description. In theory, any trait that can be observed from CCTV footage is suitable, and the database of [15] used in this chapter includes traits that cover the primary and secondary torso and leg colours, clothing texture and type, age, height, build, skin and hair colour, gender and the presence of luggage.

The two systems presented in this chapter only use a subset of these traits

- Torso and leg primary colour, categorised into the 11 culture colours (black, blue, brown, gray, green, orange, pink, purple, red, yellow, white) [2];
- Clothing type, categorised into long sleeve and other for torso garments, and long pants and other for leg garments (note that [15] uses only the leg clothing type, while [9] uses both legs and torso);
- Height (in metres), incorporated as either a height range (i.e. average height: 1.7–1.9 m), or directly as a value (i.e. 1.76 m). The database of [15] defines 5 height ranges: very short, short, average, tall, very tall.

These five traits are denoted as  $T_{\text{colour}}$  and  $L_{\text{colour}}$  for the torso and leg colour;  $T_{\text{type}}$  and  $L_{\text{type}}$  for the torso and leg clothing type; and  $H$  for the subject height;

Colour models are trained for each of the target colours and skin (i.e. 12 colours in total) using the colour patches provided by [15]. GMMs with up to 12 components (determined using the BIC) are trained in the LAB colour space. A confusion matrix is also computed using the test patches provided in [15], which is used by the second system (see Sect. 12.5). No normalisation or illumination correction is made when training the models, and each model relies solely on the range of colour snippets available, and the expected illumination invariance achieved through using the LabCie colour space.

Height is catered for using camera calibration [28], allowing templates and/or regions of interest to be scaled to the appropriate real-world size.

## 12.4 Locating People Using a Region-Based Approach

The first approach we examine is the region-based approach of [15]. This approach builds on the earlier work of [8], including provision for leg clothing type and a method to model the uncertainty present in the colour traits. Like [8], the search process is driven by a particle filter.

### 12.4.1 Search Query Formulation

As outlined in Sect. 12.3, colour selection is categorised into the 11 culture colours while also incorporating skin, and a GMM is trained for each colour. A particle filter drives the search, and each particle describes a candidate location (x and y position, height and aspect ratio), and particle heights and aspect ratios are distributed within a range set according to the target query (i.e. when searching for a ‘tall’ person, the height will be uniformly distributed between 1.7 and 1.9 m).

The approach compares torso and leg colour similarity by comparing fixed regions within the bounding box described by each particle to the target colours (see [8]). Leg clothing type is incorporated by allowing the size of the lower region (i.e. the region that contains the legs) to vary according to type of clothing expected. However the region-based approach makes a similar consideration of torso clothing types difficult, and thus torso clothing is not considered.

### 12.4.2 Searching for a Target

The approach has four stages: particle filtering, trait similarity calculation, trait weight filtering and finally particle weighting. The particle filter is initialised with a



**Fig. 12.2** An image to be searched, and a particle location (*red bounding box*) are shown in (a). Given this particle location, regions of interest for the torso (b) and legs (c) are defined. Within each of these regions, the motion (d) and colour (e) are considered and used to determine how well the candidate bounding box matches the query (note that when considering the colour, a soft decision approach is used rather than the hard decision indicated by (e))

randomly generated 500 particle set, and three iterations are performed per frame to avoid convergence on incorrect subjects.

The similarity of the primary torso colour and primary leg colour are computed using a method similar to that outlined in [8]. A similarity score ( $S_t$  for the torso, and  $S_l$  for the legs) is produced based on the normalised probability of the target colour appearing in the desired area, such that for the torso,

$$S_t = \frac{\sum_{x,y \in t} P_{x,y,c} \times Mo_{x,y}}{\sum_{x,y \in t} Mo_{x,y}}, \quad (12.1)$$

where the summation of the likelihood,  $P_{x,y,c}$  of the target colour,  $c$  appearing at each pixel location,  $x, y$ , is completed over the limits of the torso region,  $t$ .  $Mo_{x,y}$  indicates the presence of motion at  $x, y$ , and is set to  $Mo_{x,y} = 1$  if motion is present, and  $Mo_{x,y} = 0.5$  otherwise, resulting in regions that contain motion (and are thus more likely to contain the target) receiving a higher weight, and regions that are likely to correspond to the background (and are thus less likely to match the target colour) being diminished.  $S_l$  is calculated in a similar manner. Torso and leg regions are selected as in [8], such that areas likely to correspond to the head and feet (and thus likely to have a different colour) are discarded. Figure 12.2 displays an overview of the two primary regions of interest for which the torso and leg similarity scores are calculated.

A key consideration when comparing traits is their reliability. To incorporate this, the quality of each trait is incorporated such that observations that have a lower quality receive a diminished weight, helping the particle filter converge on areas of greater quality and confidence.

To model the quality of the trait, a combination of motion segmentation (the approach of [7]) and a pixel-wise histogram that captures the colour distribution

at each pixel over time is used. The use of motion segmentation allows uncertainty around the location to be captured (confidence is reduced when there is little to no motion), while the colour history allows the confidence of the query to be assessed (a query colour that is the same as the dominant historic background colour has greater uncertainty).

The motion segmentation quality component,  $Q_{Mo}$ , is calculated in a two-step process. Initially the probability of motion within the bounded region is obtained  $Mo$ . If this probability is above 0.5,

$$Q_{Mo} = 1 - \log_2(Mo), \quad (12.2)$$

is used to calculate the final quality score for motion segmentation. However if  $Mo < 0.5$ , the quality of the motion segmentation component is set to zero.

A pixel-wise histogram model that captures individual pixel colour classification over a period of time is used to capture uncertainty associated with the colour query. Additional captured video sequences are classified using the trained GMM's to capture the distribution of colours at each individual location over time to create the histogram model,  $V_{PH}$ . The colour quality,  $Q_{vp}$ , is then calculated as follows:

$$Q_{PH} = \frac{\sum_{x,y \in r} 1.0 - V_{PH}(x, y, c)}{R}, \quad (12.3)$$

where  $r$  is the region of interest (i.e. the target torso or leg region),  $c$  is the target colour and  $R$  is the size of the region,  $r$ . Thus,  $Q_{PH}$  represents the proportion of the background that has historically been a different colour to the target colour.

The two quality components are then simply combined by calculating their geometric mean,

$$Q = \sqrt{Q_{Mo} * Q_{PH}}. \quad (12.4)$$

Quality measures are calculated for both the torso and leg regions ( $Q_t$  and  $Q_l$  respectively), and are subsequently used to scale the similarities ( $S_t$  and  $S_l$  for the torso and legs respectively). The similarities are then combined using the geometric mean.

### 12.4.3 Assessing Clothing Type

A limitation of the approach in Sect. 12.4.2 is that for torso and leg regions, when the subject is wearing clothing that does not cover the entire region (i.e. shorts rather than long trousers), a significant portion of the region will be skin colour which will falsely diminish the weight. To help alleviate this problem, the bounds of the target leg region are modified to remove areas that do not represent the article of clothing. Two categories of clothing are considered: ‘long’ clothes that will cover the bulk of the leg region, and all others.

The motivation behind this is to rebound the leg region by decreasing the total area, and in turn increasing the desired target colour presence in the target region. The asymmetry driven chromatic and spatial operators described in [10] are used (which were proposed to segment a body into head, torso and legs and find the dominant plane of symmetry in the head and leg regions), and applied to a skin mask,  $Sk$ , rather than a motion mask. The chromatic operator measures the difference in two adjoining colour patches, while the spatial operator measures the difference in two adjoining mask regions. The skin mask is computed as follows:

$$Sk_{x,y} = P_{x,y,c} \times Mo_{x,y}, \quad (12.5)$$

where  $P_{x,y,c}$  is the probability of a given colour,  $c$ , in this case skin colour; and  $Mo_{x,y}$  is a scaling factor based on the motion at pixel  $x, y$ , such that  $Mo_{x,y} = 1$  if motion is present, otherwise  $Mo_{x,y} = 0.5$ . As with the region similarity computation in Eq. 12.1, pixels that contain motion are emphasised as they are less likely to represent background regions.

The chromatic,  $C(i, \delta)$ , and spatial,  $S(i, \delta)$  operators from [10] are then applied as follows to rebound the region of interest,

$$A_i = argmin((1 - C(i, \delta)) + S(i, \delta)) \quad (12.6)$$

where  $A_i$  corresponds to the image index associated with the minimum score achieved through the summation of the spatial and chromatic operators resulting in a newly bounded leg region, and is intended to represent the lower boundary between the item of clothing (i.e. shorts) and skin.

## 12.5 Searching Using a Channel Representation

The recent single object tracking approaches of [12, 25] have demonstrated state-of-the-art performance using a simple distribution field [25] or channel representation [12]. Furthermore, the template used by these approaches is well suited to being generated from a description rather than an image patch. The nature of these template representations is such that they model an estimated appearance, incorporating uncertainty about the exact colour, or location. In this section, we outline a recent approach [9] that uses a channel representation to model a subjects appearance.

### 12.5.1 Generating an Avatar

An avatar is generated from a set of characteristics that describe the target subject. In this approach, the traits outlined in Sect. 12.3 are used. To generate an avatar, a set of learned components that relate to the selected traits are summed. The average

appearances for different body regions and clothing types are learned using a set of manually annotated silhouettes (see Sect. 12.6 for details). Silhouettes are resized to the same height, and edges are zero padded to the same width. Following this, all examples belonging to a given class (i.e. torso regions for ‘long sleeve shirt’) are used to learn the appearance by simply averaging the examples,

$$A_c = \frac{1}{N} \sum_n^N a_c(n), \quad (12.7)$$

where  $A_c$  is the average appearance of component  $c$ ;  $a_c(n)$  is the  $n$ th example image; with  $N$  examples in total. This approach is used to learn the appearance of the two torso clothing types and corresponding torso skin regions; the two leg clothing types and leg skin regions; and the head skin regions.

The learned appearances and colour models are then used to generate an avatar that describes the target subject’s appearance. An occupancy mask that indicates the likelihood of a person being at a location is generated by combining the learned silhouettes for the target modes,

$$A_{skin} = A_{ts,T_{type}} + A_{ls,L_{type}} + A_h, \quad (12.8)$$

$$A_{torso} = A_{tc,T_{type}}, \quad (12.9)$$

$$A_{legs} = A_{lc,L_{type}}, \quad (12.10)$$

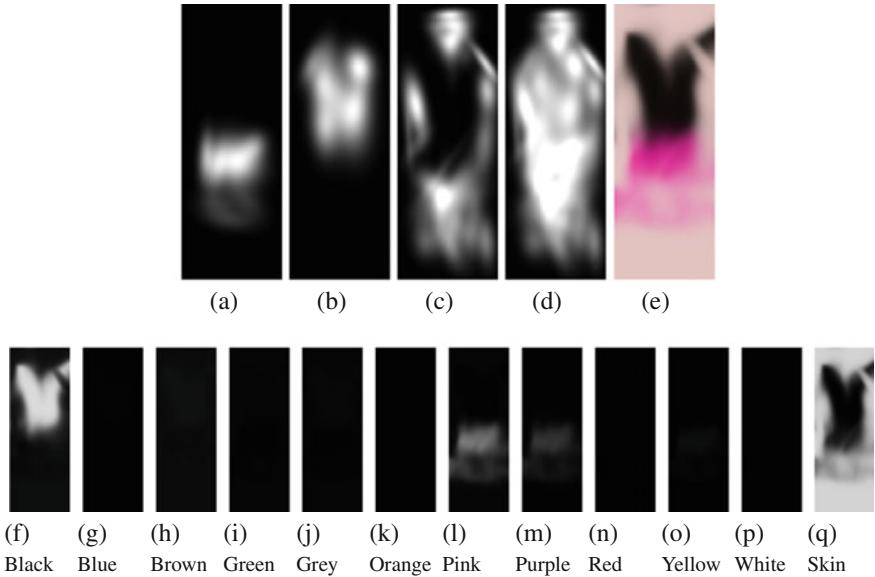
where  $A_{ts,T_{type}}$  and  $A_{tc,T_{type}}$  are the appearance of the torso skin ( $ts$ ) and clothing ( $tc$ ) regions for the selected torso type,  $T_{type}$ ;  $A_{ls,L_{type}}$  and  $A_{lc,L_{type}}$  are the appearance of the leg skin ( $ls$ ) and clothing ( $lc$ ) regions for the selected leg type,  $L_{type}$ ; and  $A_h$  is the average appearance of the head.

A set of colour masks,  $C_n$ , that correspond to a channel representation are then generated as follows:

$$\begin{aligned} c(x, y, n) &= A_{torso}(x, y) \times P(n|T_{colour}) \\ &\quad + A_{legs} \times P(n|L_{colour}) + A_{skin} \times P(n|S_{colour}), \end{aligned} \quad (12.11)$$

$$C_n(x, y) = \frac{c(x, y, n)}{A_{torso}(x, y) + A_{legs}(x, y) + A_{skin}(x, y)}, \quad (12.12)$$

where  $C_n(x, y)$  is a pixel,  $x, y$ , in the  $n$ th channel (i.e. representing the  $n$ th colour) of the channel representation;  $T_{colour}$ ,  $L_{colour}$  and  $S_{colour}$  are the selected torso, leg and skin colours;  $P(n|T_{colour})$  is the likelihood of colour  $n$  being the target colour,  $T_{colour}$ , and is estimated using the earlier computed confusion matrix. The channel representation is normalised using the sum of the masks (i.e.  $A_{torso}(x, y) + A_{legs}(x, y) + A_{skin}(x, y)$ ) to ensure that it sums to 1. This process will generate a mask for each colour, with the mask’s content determined by the likelihood of the colour being



**Fig. 12.3** An example avatar for the query ‘Tall, pink shorts, black short sleeve shirt’: **a–c** show the masks for the leg clothing, torso clothing and skin regions respectively. These masks are generated based on the type of clothing selected, and are combined to create **(d)** and **(e)**. **d** shows the overall occupancy mask, and **e** shows the expected colour at each location. Note that the most likely colour for every pixel is given in **(e)**, even though some pixels are very unlikely to belong to the person as shown by **(d)**. The images in the *second row* show the channels for the 12 colours. The uncertainty that exists in the channel representations can be seen in the weights of the *pink* and *purple* channels

observed at each pixel. This likelihood is driven by two factors: the colour in the target query, and the confusability of that colour.

Finally, a combined mask that represents the likelihood of each pixel belonging to the person is created,

$$A_P = \min(1.0, A_s + A_{tc,T_{type}} + A_{lc,L_{type}}), \quad (12.13)$$

where *min* is an operator that takes the minimum of two values. Importantly in this combined mask, normalised masks for the components are not used to ensure that pixels that are unlikely to contain the target have a low weight. This mask is used as an auxiliary channel in the CR to weight the contribution of each pixel to the overall similarity. An example avatar is shown in Fig. 12.3.

### 12.5.2 Searching for a Target

A particle filter-like approach is used to search for the target. A small number of particles are created at random locations within the field of view, and the single object tracking approach of [12] is used to refine the location of each. The similarity of each refined particle location to the target model and a stored background model is computed, based on which the particle set is refreshed and the current position of the target is estimated.

Input images are transformed into a channel representation where the channels are given by the culture colour space,

$$I_n(x, y) = P(n|I(x, y)), \quad (12.14)$$

where  $I_n$  is the  $n$ th channel of the channel representation;  $P(n|I(x, y))$  is the likelihood of the pixel  $I(x, y)$  being the  $n$ th colour, and is computed using a lookup table that contains the likelihoods for the whole colour space computed using the learned GMMs.<sup>1</sup> As per [12], each channel is then convolved with a Gaussian filter.

A gradient descent search is then performed to find the particle location  $(X_p, Y_p)$  within a local region that minimises

$$S_F = \sum_{n,x,y}^{N,X,Y} |C_n(X_p + x, Y_p + y) - I_n(x, y)| \times A_P(x, y), \quad (12.15)$$

where  $S_F$  is the similarity of the foreground to the template;  $C_n(x, y)$  is the  $n$ th channel of the template (i.e. search query) and  $X$  and  $Y$  are the width and height of the template, and  $N$  is the number of channels. Note that  $A_P$  is used to weight the summation such that pixels that are more likely to belong to the person are given a greater weight. Although this search will converge on an optimal location, it is only guaranteed to be locally optimal [25]. To overcome this, multiple particles are refined using this search process, after-which particles are filtered in preparation for the next input frame.

As the particles move during the search process, a dense collection of particles is not needed. Following the updating of particle locations, a two-stage filtering process is applied, where first particles which record a poor match are removed, after-which particles which are located nearby another are removed.

One of the limitations observed in [8] is that locations in the background can be well matched to the target. To help overcome this, a universal background model style approach is used. An average channel representation for the background which is progressively updated every frame is stored,

$$B_n(x, y, t) = \alpha \times B_n(x, y, t - 1) + (1 - \alpha) \times I_n(x, y), \quad (12.16)$$

---

<sup>1</sup>The lookup table is pre-computed for computational efficiency.

where  $B_n(x, y, t)$  is the  $n$  channel in the background channel representation at time  $t$ ; and  $\alpha$  is the learning rate. Using this stored background model, the similarity of the template to the target location in the background is computed,

$$S_B = \sum_{n,x,y}^{N,X,Y} |B_n(X_p + x, Y_p + y) - T_n(x, y)| \times A_P(x, y), \quad (12.17)$$

and used to determine if the particle is a better match to the background or the foreground,

$$S_R = S_B / S_F. \quad (12.18)$$

If  $S_R \leq 1$ , the particle is discarded, as it is more like the background than the foreground. Remaining particles are then filtered such that if two particles lie within a distance,  $d$ , the particle with the lower value of  $S_R$  is removed. The location of the target subject is given by the particle that yields the highest value of  $S_R$ .

### 12.5.3 Compensating for Scale

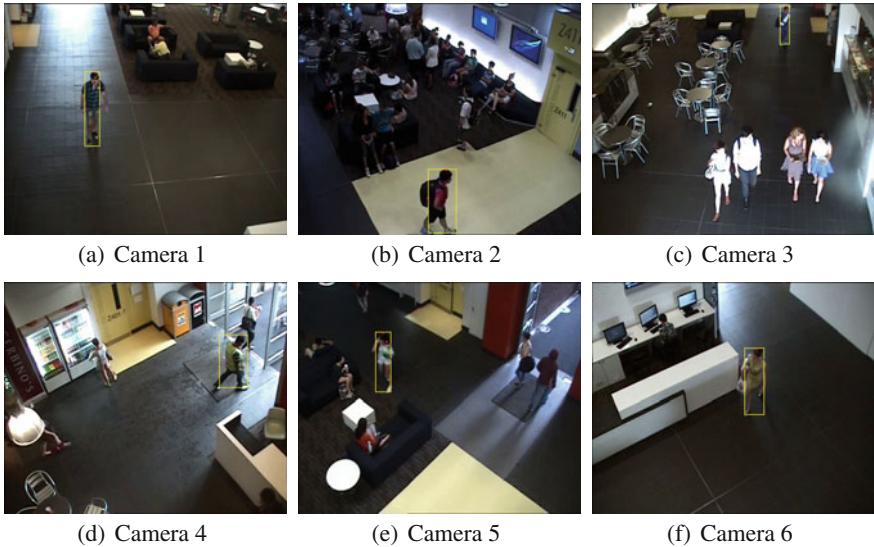
An additional parameter for the search query is the target height. This is incorporated using camera calibration information (Tsai's camera calibration [28], information is provided with the dataset of [15]). A resolution parameter,  $R$ , in pixels per meter, is used to generate the template and resized target patches when searching. The target template is set to a height (in pixels) of  $R \times H$ , where  $H$  is the target query height in metres. When comparing the template to an image, the local region around the initial estimated position (i.e.  $X_p, Y_p$ ) is resized to the same resolution.

## 12.6 Database and Evaluation Protocol

### 12.6.1 Data

We use the database and evaluation protocol outlined in [15]. This database consists of 110 video clips with semantic queries and the correct match annotated at every frame. Example images from the database are shown in Fig. 12.4.

Avatar models are trained using a set of person images extracted from the background videos in [15] (i.e. they do not overlap with the queries), and are originally presented in [9]. In total, images of 103 people from the same 6 camera network are used. An example of the annotation is shown in Fig. 12.5.

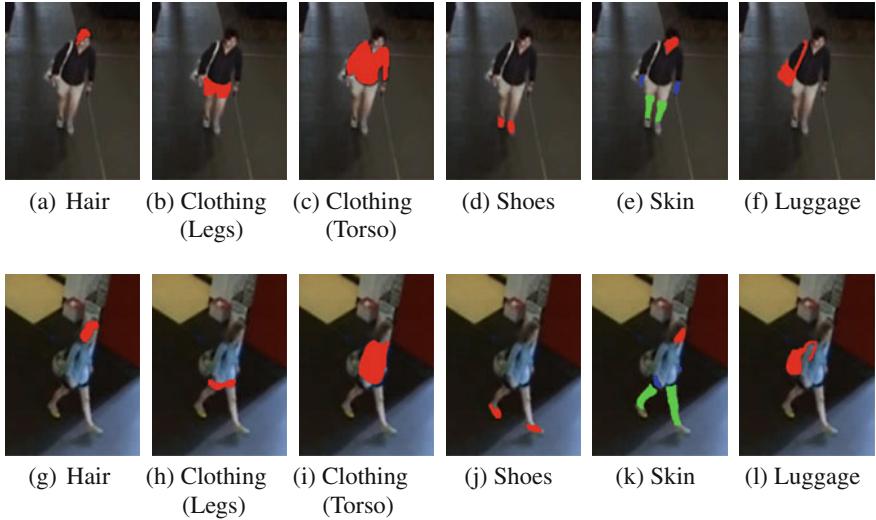


**Fig. 12.4** A sample image from each camera, each showing an image from a different sequence. The target queries for the six images are: **a** Sequence *C1-BlackBlueStripe-BlueJeans*: 15–35-year-old short Caucasian male of average build with *brown hair*, wearing a *blue striped short sleeve shirt*, *plain blue shorts* and *brown shoes*, and carrying *luggage*; **b** Sequence *C2-RedShirt-BlackShorts*: 25–45-year-old very short Asian female of large build with *dark hair*, wearing a *red short sleeve shirt*, *plain black shorts* and *white shoes*, and carrying *luggage*; **c** Sequence *C3-DiagonalTop-HorizontalPants*: 15–35-year-old short Asian male of average build and height with *dark hair*, wearing a *green and white striped short sleeve shirt*, *blue and white striped trousers* and *white shoes*, and carrying *luggage*; **d** Sequence *C4-FluroYellowTop-BluePants*: a 15–35-year-old *dark skinned* and *dark haired* male of average height and large build, wearing a *yellow and grey striped sleeveless shirt*, *plain blue trousers* and *white shoes*; **e** Sequence *C5-LightGreenShirt-WhiteShorts*: a 15–35-year-old short Caucasian female of very slim build with *brown hair*, wearing *green sleeveless top*, *white shorts* and *white shoes*, and carrying *luggage*; **f** Sequence *C6-YellowWhiteSpotDress*: a 15–35-year-old short Caucasian female of slim build with *brown hair*, wearing a *yellow and white spotted dress*, and carrying *luggage*. Images also show the annotated *feet*, *waist*, *shoulder*, *neck* and *head* positions, and the *target bounding box* (yellow)

### 12.6.2 Evaluation Protocol

Performance is evaluated by computing the localisation accuracy for all frames with annotated ground truth. A ground truth bounding box is constructed from the annotated points. Left and right bounds are taken from the feet, shoulder, waist or neck annotations, while the head and feet provide the top and bottom bounds. If any of these bounds were not located then that frame is deemed to be un-annotated, and is not used in the evaluation.

The bounding box computed by the proposed algorithm is compared to the ground truth using,



**Fig. 12.5** Example of the annotation of two subjects into their constituent parts. Note that in the skin mask we annotate the head (red), arms (blue) and legs (green) separately

$$L_{x,i} = \frac{D_{x,i} \cap GT_{x,i}}{L_{x,i} \cup GT_{x,i}}, \quad (12.19)$$

where  $D_{x,i}$  is the output of the algorithm for subject  $x$ , frame  $i$ , and  $GT_{x,i}$  is the corresponding ground truth bounding box; to measure the localisation accuracy.

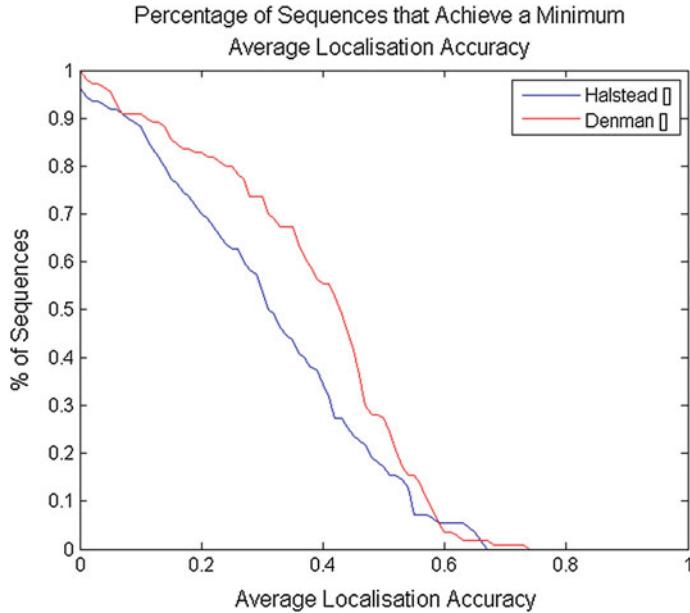
The accuracy over an entire sequence is given by the average accuracy for each frame in the sequence,

$$L_x = \frac{1}{N} \sum_{i=1}^{i=N} L_{x,i}, \quad (12.20)$$

where  $N$  is the total number of frames for subject  $x$ ; and the overall performance is given by the average of each sequence accuracy,

$$L = \frac{1}{M} \sum_{x=1}^{x=M} L_x, \quad (12.21)$$

where  $M$  is the total number of subjects (110). Sequences are not weighted by their length when computing the average to ensure that all queries are given equal weight, and long sequences do not dominate the overall performance metric.



**Fig. 12.6** The percentage of sequences above a given accuracy level

## 12.7 Results

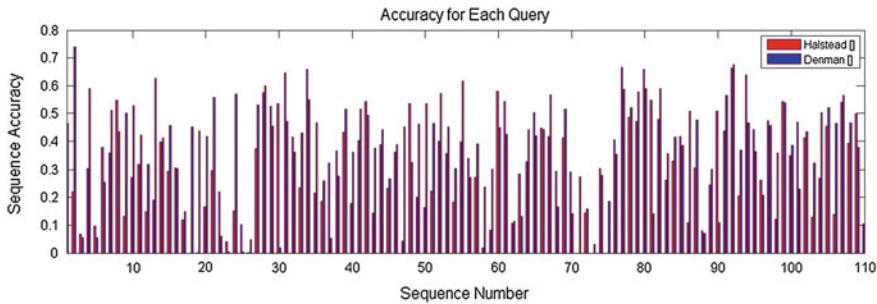
Performance of the two approaches are evaluated on the dataset of [15]. For the approach of [15] (see Sect. 12.4 for details) we use the recommended parameters of 500 particles and 3 iterations of the particle filter. For the approach of [9] (see Sect. 12.5 for details) we use the suggested default parameters of 20 particles,  $R = 20$  (i.e. 1 meter is scaled to 20 pixels), and Gaussian blur kernel width and standard deviation to 3 and 1.5 respectively.

Figure 12.6 shows the number of sequences that achieve above a given threshold and it can be clearly seen that [9] is able to better locate the majority of queries. For instance, 56% of queries are located with an average accuracy greater than 0.4 by [9], while only 37% can be located with the same level of accuracy for [15].

However, despite the approach of [9] performing better across the majority of sequences, there are a number of sequences for which [15] performs best, as shown by Fig. 12.7.

Figures 12.8 and 12.9 show example output for a variety of queries from the systems of [15] and [9] respectively. It can be seen that in the majority of cases, the approach of [9] outperforms that of [15], however there are some exceptions.

In Sequences 12 and 70, [15] confuses the target subject with the background. The use of a background model in a UBM-style configuration by [9] helps suppress matches to background regions, improving performance in situations such



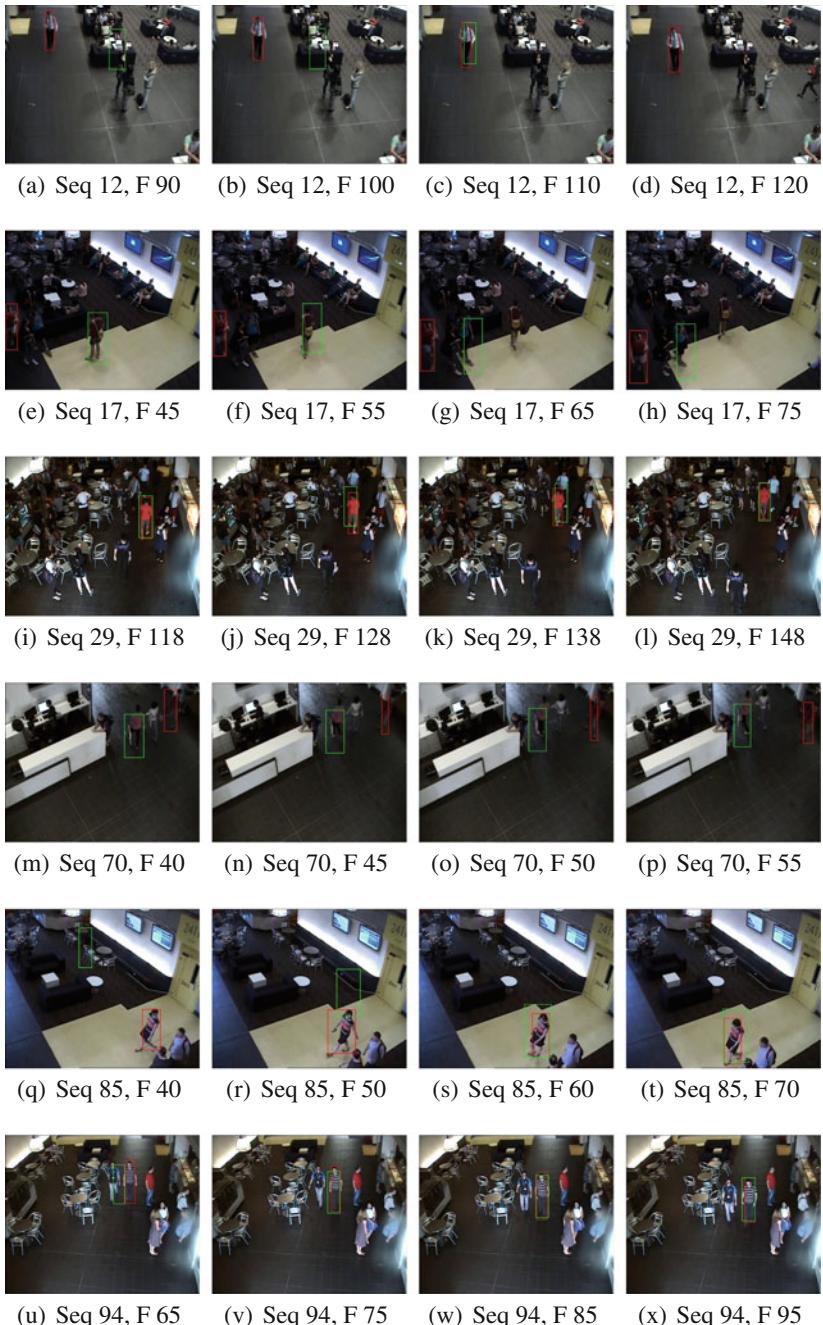
**Fig. 12.7** The percentage of sequences above a given accuracy level

as this. In Sequence 17, [15] incorrectly detects a person of similar appearance as the target. In this case, the richer query used by [9] helps, as it better enforces the spatial constraints regarding the relative positions of the colours.

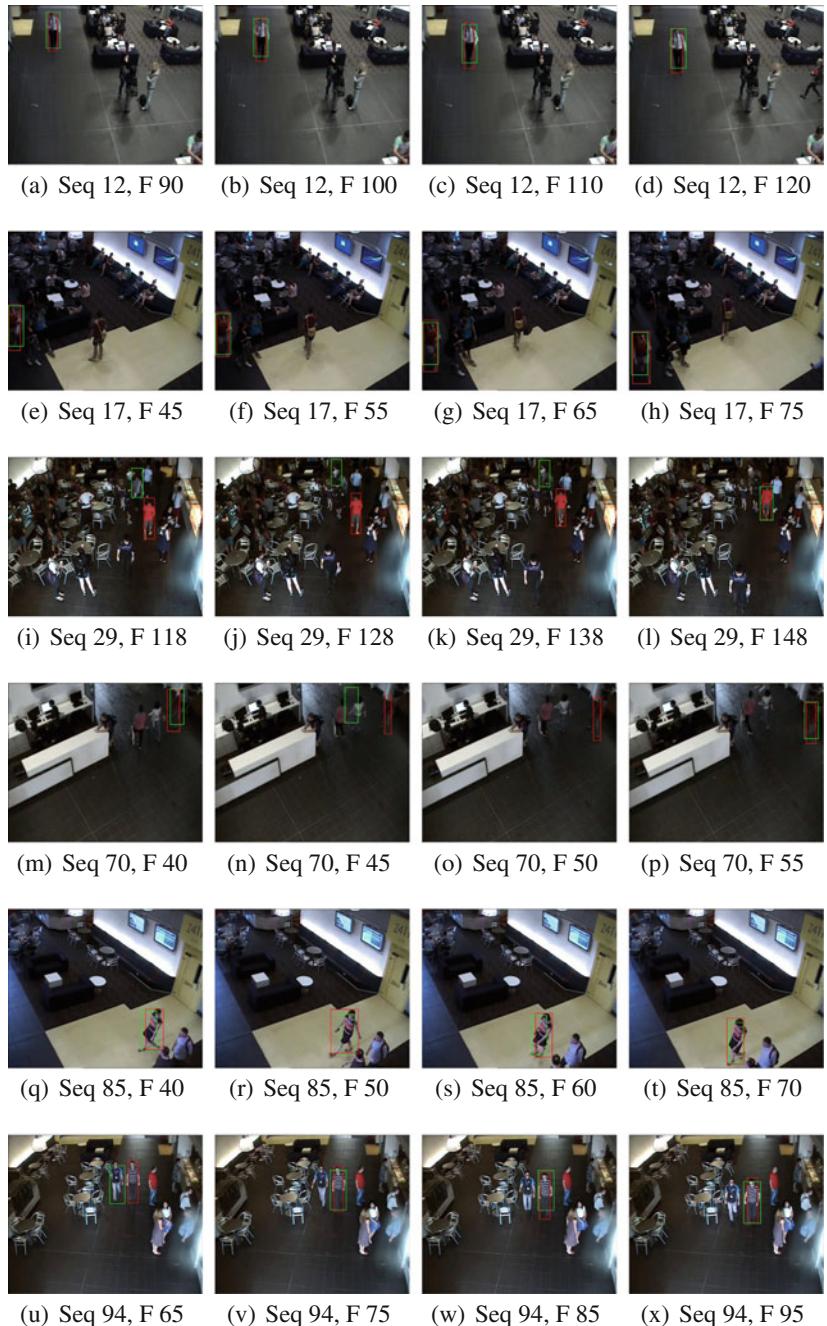
In Sequence 29, we observe that [15] achieves much better performance, with [9] only localising the subject towards the very end of the sequence. In this sequence, the subject's shirt is specified as being 'orange', however the shirt is predominately classified as being 'pink'. The CR used by [9] is intended to allow for this sort of mis-match, however it does so using a confusion matrix. In this case, the confusion matrix records that orange is predominately confused with red, and so there is very little expectation of observing pink; resulting in a poor match with the CR. This, combined with the prevalence of other grey regions (the targets shorts colour) leads to misclassification. Ultimately, the simpler matching method used by [15] is better able to deal with this form of ambiguity.

In the remaining two sequences (Sequences 85 and 94), both approaches manage to detect the target with a reasonable degree of success. In Sequence 85 [15] initially confuses the target with the background causing a drop in performance, while in Sequence 94 [9] does not localise the subject as accurately, although both systems occasionally switch tracking to the second person that the target is walking beside. In both these sequences (85 and 94), the target subject is wearing patterned clothing (a striped shirt in both cases), highlighting a weakness with the approaches. Each system is effectively searching for a single colour within both the torso and leg regions. In sequence 94, this leads to a second person wearing a similar coloured shirt (although one that does not contain the desired pattern) being detected by mistake.

Overall, the approach of [9] performs better for most sequences, however it is less flexible in the way that it manages ambiguous colour matches. This may in part be aided using more data to train the confusion matrix, however there are still likely to be situations that arise where colour mis-match occurs due to unusual lighting conditions, or similar environmental effects. Both approaches are also impacted by the presence of clothing textures (i.e. stripes or checks), which reduce matching accuracy. Ultimately, the use of more traits such that colour is less important is one way to address this.



**Fig. 12.8** Example frames from six sequences showing performance of [15]. Ground truth locations are marked with a *red bounding box*, while the estimated location is marked with a *green box*. For the 6 sequences, sequence accuracies of 0.1918, 0.0, 0.5356, 0.0, 0.1100 and 0.4411 are achieved



**Fig. 12.9** Example frames from six sequences showing performance of [9]. Ground truth locations are marked with a *red bounding box*, while the estimated location is marked with a *green box*. For the six sequences, sequence accuracies of 0.6266, 0.4528, 0.0190, 0.2757, 0.5072 and 0.3633 are achieved

### 12.7.1 Computational Efficiency and Scalability

Comparing the runtimes of both approaches, [9] is able to execute faster running at 5.59 frames per second compared to 3.23 frames per second for [15] running on identical hardware.<sup>2</sup> It should be noted that neither approach has been optimised for speed, and performance gains are possible for both approaches.

The main computational process for both systems is the calculation of particle similarity. For [15], this process is greatly faster for each particle due to the simple nature of the comparison (a summation of likelihoods over a region). [9] doesn't compute a simple comparison, rather each particle is subject to a search process which itself involves several comparisons, and each comparison is the sum of a difference over a large number of channels, making the overall process for an individual particle much more demanding. However, the search process in [9] that allows particles to refine their own position means that an order of magnitude fewer particles are needed. This, combined with the fact that motion segmentation is not required leads to faster run times. For both approaches, parameters can be adjusted to further improve speed. Obviously particle numbers can be reduced in both approaches, and in the case of [9] the resolution parameter,  $R$ , can also have a significant impact on processing speed. By reducing  $R$  to 10, which as shown by [9] has only a slight impact on accuracy, throughput can be roughly increased by 25%. Similar gains could be made for [15] by simply reducing image resolution.

One key point of difference between the two system is the use of motion segmentation. While [15] could be run without motion segmentation, it does greatly benefit from it, and would further benefit from more accurate and reliable segmentation. This use of motion segmentation becomes potentially problematic when dealing with very large camera networks, as an up-to-date background model must be maintained for all cameras. Although [9] also maintains a background model of sorts, it simply requires an average frame which can be very quickly computed using a moving average. As such, while [15] may be faster for a single field of view, across a large network for 10s or even 100s of cameras, an approach such as [9] that does not require the continual update of a background model may be preferable.

## 12.8 Conclusions and Future Work

In this chapter, we have examined two approaches to locate a person in video from a semantic query. Both approaches use clothing colour and height information, and incorporate clothing type to varying degrees. The channel representation-based approach of [9] is shown to outperform the region-based approach of [15], however both approaches have their strengths and weaknesses.

A common limitation of the approaches is the small number of traits they consider, with other traits such as gender, clothing texture (or more sophisticated cloth-

---

<sup>2</sup>both approaches are running on a single core of an Intel Xeon E5-2670 processor.

ing type), hair and skin colour and the presence of luggage all traits that may warrant inclusion. For traits that impact upon the silhouette of the subject (i.e. gender, clothing types), it is clear how these could be easily incorporated into the channel representation approach of [9] by simply learning richer appearance models to guide the generation of the channel representation. The incorporation of such traits into [15] is less clear cut, however [15] is better suited to the addition of traits relating to clothing texture. Texture requires the analysis of a region (i.e. it cannot be determined from a single pixel), which is well aligned with [15]'s region-based approach. However within the CRs used by [9] it would require the use of a sliding window, which is potentially computationally demanding.

Ultimately, both approaches are limited in how they fuse data. [9] relies on the CR to model uncertainty, and thus the possible errors need to be first learnt; while [15] modifies trait weights based on motion segmentation data and the observed background colour. There is obvious scope to incorporate a more robust approach such as Dempster-Shafer theory [20, 26] into the approach of [15], which would better model the uncertainty observed and help improve performance.

## References

1. Bak S, Charpiat G, Corvee E, Bremond F, Thonnat M (2012) Learning to match appearances by correlations in a covariance metric space. In: ECCV. Springer, pp 806–820
2. Berlin B, Kay P (1969) Basic color terms: their universality and evolution. University of California Press, Berkeley
3. CBC: London police investigation timeline. CBC News (2005)
4. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: CVPR, pp 886–893
5. D'Angelo A, Dugelay JL (2010) Color based soft biometry for hooligans detection. In: ISCAS, pp 1691–1694
6. Dantcheva A, Velardo C, D'Angelo A, Dugelay JL (2011) Bag of soft biometrics for person identification: New trends and challenges. Multimedia Tools Appl 51(2):38
7. Denman S, Fookes C, Sridharan S (2009) Improved simultaneous computation of motion detection and optical flow for object tracking. In: DICTA, Melbourne, Australia
8. Denman S, Halstead M, Bialkowski A, Fookes C, Sridharan S (2012) Can you describe him for me? a technique for semantic person search in video. In: DICTA, pp 1–8
9. Denman S, Halstead M, Sridharan S, Fookes C (2015) Searching for semantic person queries using channel representations. In: IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE
10. Farenzena M, Bazzani L, Perina A, Murino V, Cristani M (2010) Person re-identification by symmetry-driven accumulation of local features. In: CVPR, pp 2360–2367
11. Felsberg M (2012) Adaptive filtering using channel representations. In: Mathematical methods for signal and image analysis and representation. Springer, London, pp 31–48
12. Felsberg M (2013) Enhanced distribution field tracking using channel representations. In: International conference computer vision workshops. IEEE, pp 121–128
13. Felzenszwalb P, McAllester D, Ramanan D (2008) A discriminatively trained, multiscale, deformable part model. In: CVPR
14. Feris R, Siddique B, Zhai Y, Petterson J, Brown L, Pankanti S (2011) Attribute-based vehicle search in crowded surveillance videos. In: ACM international conference on multimedia retrieval, p 18

15. Halstead M, Denman S, Fookes C, Sridharan S (2014) Locating people in video from semantic descriptions: a new database and approach. In: ICPR
16. Jaha E, Nixon M (2014) Soft biometrics for subject identification using clothing attributes. In: 2014 IEEE international joint conference on biometrics (IJCB), pp 1–6. doi:[10.1109/BTAS.2014.6996278](https://doi.org/10.1109/BTAS.2014.6996278)
17. Jain AK, Dass SC, Nandakumar K (2004) Soft biometric traits for personal recognition systems. In: International conference on biometric authentication, pp 717–738
18. Lucas T, Henneberg M (2015) Comparing the face to the body, which is better for identification? *Int J Legal Med* 1–8. doi:[10.1007/s00414-015-1158-6](https://doi.org/10.1007/s00414-015-1158-6)
19. McKelvey T, Dailey K (2013) Boston marathon bombings: how notorious bombers got caught. *BBC News Mag*
20. Nguyen Thanh K, Denman S, Sridharan S, Fookes C (2014) Score-level multibiometric fusion based on dempster-shafer theory incorporating uncertainty factors. *IEEE Trans Hum Mach Syst*
21. Paisitkriangkrai S, Shen C, Hengel AVD (2013) Efficient pedestrian detection by directly optimize the partial area under the roc curve. In: ICCV, pp 1057–1064
22. Park U, Jain A, Kitahara I, Kogure K, Hagita N (2006) Vise: Visual search engine using multiple networked cameras. *ICPR* 3:1204–1207
23. Reid D, Nixon M, Stevenage S (2014) Soft biometrics; human identification using comparative descriptions. *IEEE Trans Pattern Anal Mach Intell* 36(6):1216–1228. doi:[10.1109/TPAMI.2013.219](https://doi.org/10.1109/TPAMI.2013.219)
24. Scoleri T, Lucas T, Henneberg M (2014) Effects of garments on photoanthropometry of body parts: application to stature estimation. *Forensic Sci Int* 237(0):148.e1–148.e12. doi:[10.1016/j.forsciint.2013.12.038](https://doi.org/10.1016/j.forsciint.2013.12.038). <http://www.sciencedirect.com/science/article/pii/S0379073814000073>
25. Sevilla-Lara L, Learned-Miller E (2012) Distribution fields for tracking. In: CVPR. IEEE, pp 1910–1917
26. Shafer G (1976) Mathematical theory of evidence. Princeton University Press
27. Thornton J, Baran-Gale J, Butler D, Chan M, Zwahlen H (2011) Person attribute search for large-area video surveillance. In: IEEE international conference on technologies for homeland security (HST), pp 55–61
28. Tsai RY (1986) An efficient and accurate camera calibration technique for 3d machine vision. In: CVPR, pp 364–374
29. Turakhia N, Parikh D (2013) Attribute dominance: what pops out. In: ICCV. IEEE, pp 1225–1232
30. Vaquero DA, Feris RS, Tran D, Brown L, Hampapur A, Turk M (2009) Attribute-based people search in surveillance environments. In: WACV

# **Chapter 13**

## **Contact-Free Heartbeat Signal for Human Identification and Forensics**

**Kamal Nasrollahi, Mohammad A. Haque, Ramin Irani  
and Thomas B. Moeslund**

**Abstract** The heartbeat signal, which is one of the physiological signals, is of great importance in many real-world applications, for example, in patient monitoring and biometric recognition. The traditional methods for measuring such this signal use contact-based sensors that need to be installed on the subject's body. Though it might be possible to use touch-based sensors in applications like patient monitoring, it will not be that easy to use them in identification and forensics applications, especially if subjects are not cooperative. To deal with this problem, recently computer vision techniques have been developed for contact-free extraction of the heartbeat signal. We have recently used the contact-free measured heartbeat signal, for biometric recognition, and have obtained promising results, indicating the importance of these signals for biometrics recognition and also for forensics applications. The importance of heartbeat signal, its contact-based and contact-free extraction methods, and the results of its employment for identification purposes, including our very recent achievements, are reviewed in this chapter.

### **13.1 Introduction**

Forensic science deals with collecting and analyzing information from a crime scene for the purpose of answering questions related to the crime in a court of law. The main goal of answering such questions is identifying criminal(s) committing the crime. Therefore, any information that can help identifying criminals can be

---

K. Nasrollahi · M.A. Haque (✉) · R. Irani · T.B. Moeslund  
Visual Analysis of People (VAP) Laboratory, Aalborg University, Aalborg, Denmark  
e-mail: mah@create.aau.dk

K. Nasrollahi  
e-mail: kn@create.aau.dk

R. Irani  
e-mail: ri@create.aau.dk

T.B. Moeslund  
e-mail: tbm@create.aau.dk

useful. Such information can be collected from different sources. One such a source, which has a long history in forensics, is based on human biometrics, i.e., human body or behavioral characteristics that can identify a person, for example, DNA [1], fingerprints [2, 3], and facial images [4]. Besides human biometrics, there is another closely related group of human features/characteristics that cannot identify a person, but can help the identification process. These are known as soft biometrics, for instance, gender, weight, height, gait, race, and tattoo.

The human face, which is of interest of this chapter, is not only used as a biometric, but also as a source for many soft biometrics, such as gender, ethnicity, and facial marks and scars [5, 6, 7]. These soft biometrics have proven great values for forensics applications in identification scenarios in unconstrained environments wherein commercial face recognition systems are challenged by wild imaging conditions, such as off frontal face pose and occluded/covered face images [8, 7]. The mentioned facial soft-biometrics are mostly based on the physical features/characteristics of the human. In this chapter, we look into heartbeat signal which is a physiological feature/characteristic of the human that similar to the mentioned physical ones can be extracted from facial images in a contact-free way, thanks to the recent advances in computer vision algorithms. The heartbeat signal is one of the physiological signals that is generated by the cardiovascular system of the human body. The physiological signals have been used for different purposes in computer vision applications. For example, in [9] electromyogram, electrocardiogram, skin conductivity and respiration changes and in [10] electrocardiogram, skin temperature, skin conductivity, and respiration have been used for emotion recognition in different scenarios. In [11] physiological signals have been used for improving communication skills of children suffering from Autism Spectrum Disorder in a virtual reality environment. In [12–14] these signals have been used for stress monitoring. Based on the results of our recent work, which are reviewed here, the heartbeat signal shows promising results to be used as a soft biometric.

The rest of this chapter is organized as follows: first, the measurements of heartbeat signal using both contact-based and contact-free methods are explained in the next section. Then, employing these signals for identification purposes is discussed in Sect. 13.3. Finally, the chapter is concluded in Sect. 13.4.

## 13.2 Measurement of Heartbeat Signal

The heartbeat signal can be measured in two different ways: contact-based and contact-free. These methods are explained in the following subsections.

### 13.2.1 Contact-Based Measurement of Heartbeat Signal

The heartbeat signal can be recorded in two different ways using the contact-based sensors:

- By monitoring electrical changes of muscles during heart functioning, by a method that is known as Electrocardiogram which records ECG signals.
- By listening to the heart sounds during its functioning, by a method that is known as Phonocardiogram which records PCG signals.

The main problem of the above-mentioned methods is obviously the need for the sensors to be in contact (touch) with the subject's body. Depending on the application of the measured heartbeat signal, this requirement may have different consequences, for example, for

- Constant monitoring of patient, having such sensors on the body may cause some skin irritation.
- Biometric recognition, the subjects might not be cooperative to wear the sensors properly.

Therefore, contact-free measurement of heartbeat signals can be of great advantage in many applications. Thanks to the recent advances in computer vision techniques, this has been possible recently to measure heartbeat signals using a simple webcam. Methods developed for this purpose are reviewed in the following subsection.

### ***13.2.2 Contact-Free Measurement of Heartbeat Signal***

The computer vision techniques developed for heartbeat measurement mostly utilize facial images. The reason for this goes back to this fact that heart pulses generate some periodic changes on the face, as well as other parts of the human body. However, since the human face is mostly visible, it is usually this part of the body that has been chosen by the researchers to extract the heartbeats from. The periodic changes that are caused by the heartbeat on the face are of two types:

- Changes in head motion which is a result of periodic flow of the blood through the arteries and the veins for delivering oxygenated blood to the body cells.
- Changes in skin color which is a result of having a specific amount of blood under the skin in specific periods of time.

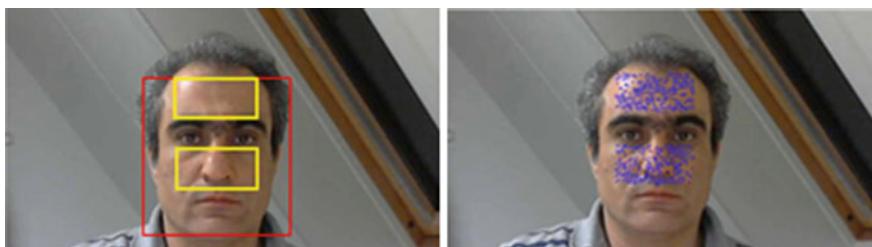
None of these two types of changes are visible to the human eyes, but they can be revealed by computer vision techniques, like Eulerian magnification of [15, 16]. The computer vision techniques developed for heartbeat measurement are divided into two groups, depending on the source (motion or color) they utilize for the measurement. These two types of methods are reviewed in the following subsections.

### 13.2.2.1 Motion for Contact-Free Extraction of Heartbeat Signal

The first motion-based contact-free computer vision method was just recently released by [17]. This system utilizes the fact that periodic heart pulses, through aorta and carotid arteries, produce periodic subtle motions on the head/face which can be detected from a facial video. To do that, in [17] some stable facial points, known as good features to track, are detected and tracked over time. The features they track are located on the forehead area and the region between the mouth and then nose are as these areas are less affected by internal facial expressions and their moments should thus be from another source, i.e., heart pulses. Tracking these facial points' results in a set of trajectories, which are first filtered by a Butterworth filter to remove the irrelevant frequencies. The periodic components of these trajectories are then extracted by PCA and considered as the heartbeat rate. Their system has been tested on video sequences of 18 subjects. Each video was of resolution of  $1280 \times 720$  pixels, at a frame rate of 30 with duration of 70–90 s.

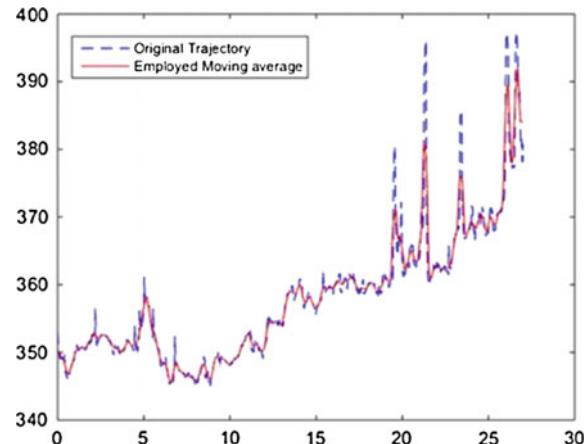
To obtain the periodicity of the results of the PCA (applied to the trajectories), in [17] a Fast Fourier Transform (FFT) has been applied to the obtained trajectories from the good features to track points. Then, a percentage of the total spectral power of the signal accounted for by the frequency with the maximal power and its first harmonic is used to define the heartbeat rate of the subject [17]. Though this produces reasonable results when the subjects are facing the camera, it fails when there are other sources of motion on the face. Such sources of motion can be for instance, changes in facial expressions and involuntary head motion. It is shown in [18] that such motions makes the results of [17] far from correct. The main reason is because the system in [17] uses the frequency with the maximal power as the first harmonic when it estimates the heartbeat rate. However, such an assumption may not always be true, specifically when the facial expression is changing [18]. To deal with these problems [18] has detected good features to track Fig. 13.1 (right) from the facial regions shown in Fig. 13.1 (left).

Then, [18] tracks the good features to track to generate motion trajectories of these features. Then, it replaces the FFT with a Discrete Cosine Transform (DCT), and has employed a moving average filter before the Butterworth filter of [17]. Figure 13.2 shows the effect of the moving average filter employed in [18] for



**Fig. 13.1** The facial regions (yellow areas in the *left* image) that have been used for detecting good feature to track (blue dots in the *right* image) for generating motion trajectories in [18]

**Fig. 13.2** The original signal (red) versus the moving averaged one (blue) used in [18] for estimating the heartbeat rate. On x and y-axis are the time (in seconds) and the amplitude of a facial point (from (good features to track)) that has been tracked, respectively



reducing the noise (resulting from different sources, e.g., motion of the head due to facial expression) in the signal that has been used for estimating the heartbeat rate.

Experimental results in [18] show that the above-mentioned simple changes have improved the performance of [18] compared to [17] in estimating the heartbeat signal, specifically, when there are changes in facial expressions of the subjects. The system in [18] has been tested on 32 video sequences of five subjects in different facial expressions and poses with duration about 60 s.

To the best of our knowledge, the above two motion-based systems are the only two methods available in the literature for contact-free measurement of the heartbeat rate using motion of the facial features. It should be noted that these systems do not report their performance/accuracy on estimating the heartbeat signal, but do so only for the heartbeat rate. The estimation of heartbeat signals are mostly reported in the color-based systems which are reported in the next subsection.

### 13.2.2.2 Color for Contact-Free Extraction of Heartbeat Signal

Using expensive imaging techniques for utilizing color of facial (generally skin) regions for the purpose of estimating physiological signal has been around for decades. However, the interest in this field was boosted when the system of [19] reported its results on video sequences captured by simple webcams. In this work, [19], Independent Component Analysis (ICA) has been applied to the RGB separated color channels of facial images, which are tracked over time, to extract the periodic components of these channels. The assumption here is that periodic blood circulation makes subtle periodic changes to the skin color, which can be revealed by Eulerian magnification of [15]. Having included a tracker in their system, they have measured heartbeat signals of multiple people at the same time [19]. Furthermore, it has been discussed in [19] that this method is tolerant towards motion of the subject during the experiment as it is based on the color of the skin. They

have reported the results of their systems on 12 subjects facing a webcam that was about 0.5 m away in an indoor environment. Shortly after in [20] it was shown that besides heartbeat, the methods of [19] can be used for measuring other physiological signals, like respiratory rate.

The interesting results of the above systems motivated others to work on the weakness of those systems, which had been tested only in constrained conditions. Specifically,

- In [21], it has been discussed the methods of [19, 20] are not that efficient when the subject is moving (questioning the claimed motion tolerance of [19, 20]) or when the lightning is changing, like in an outdoor environment. To compensate for these they have performed a registration prior to calculating the heartbeat signals. They have tested their system in an outdoor environment in which the heartbeat signals are computed for subjects driving their vehicles.
- In [22] auto-regressive modeling and pole cancelation have been used to reduce the effect of the aliased frequency components that may worsen the performance of a contact-free color-based system for measuring physiological signals. They have reported their experimental results on patients that are under monitor in a hospital.
- In [23] normalized least mean square adaptive filtering has been used to reduce effect of changes in the illumination in a system that tracks changes in color values of 66 facial landmarks. They reported their experimental results on the large facial database of MAHNOB-HCI [24] which is publicly available. The reported results show that the system of [23] outperforms the previously published contact-free methods for heartbeat measurement including the color-based methods of [19, 20] and the motion based of [17].

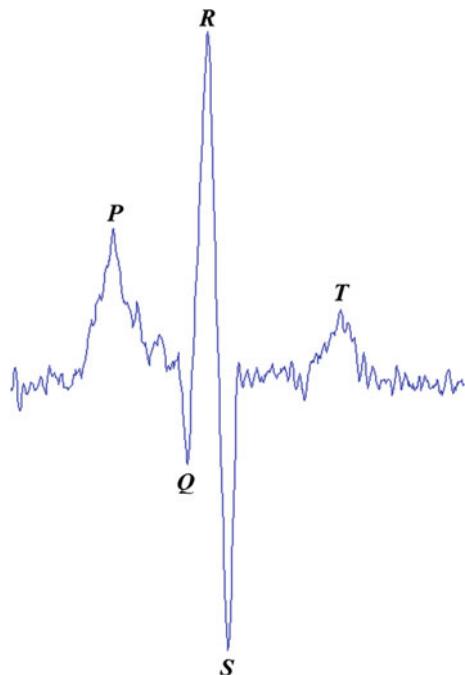
### 13.3 Using Heartbeat Signal for Identification Purposes

Considering the fact the heartbeat signals can be obtained using the two different methods explained in the previous section, different identification methods have been developed. These are reviewed in the following subsections.

#### 13.3.1 *Human Identification Using Contact-Based Heartbeat Signal*

The heartbeat signals obtained by contact-based sensors (in both ECG and PCG forms) have been used for identification purposes for more than a decade. Here we only review those methods that have used ECG signals as they are more common than PCG ones for identification.

**Fig. 13.3** A typical ECG signal in which the important key points of the signal are labeled



There are many different methods for human identification using ECG signals: [25–42], to mention a few. These systems have either extracted some features from the heart signal or have used the signal directly for identification. For example, it has been discussed in [25] that the heart signal (in an ECG form) composes of three parts: a P wave, a QRS complex, and a T wave (Fig. 13.3). These three parts and their related key points (P, Q, R, S, and T, known as fiducial points) are then found and used to calculate features that are used for identification, like the amplitude of the peaks or valleys of these point, the onsets and offsets of the waves, and the duration of each part of the signal from each point to the next. Similar features have been combined in [27] with radius curvature of different parts of the signal. It is discussed in [36] that one could ultimately extract many fiducial points from an ECG signal, but not all of them are equally efficient for identification.

In [28, 29, 33, 35, 37] it has been discussed that fiducial points detection based methods are very sensitive to the proper detection of the signal boundaries, which is not always possible. To deal with this, in

- [28] the ECG signals have directly been used for identification in a Principal Component Analysis (PCA) algorithm.
- [29] the ECG signal has been first divided into some segments and then the coefficients of the Discrete Cosine Transform (DCT) of the autocorrelation of these segments have been obtained for the identification.

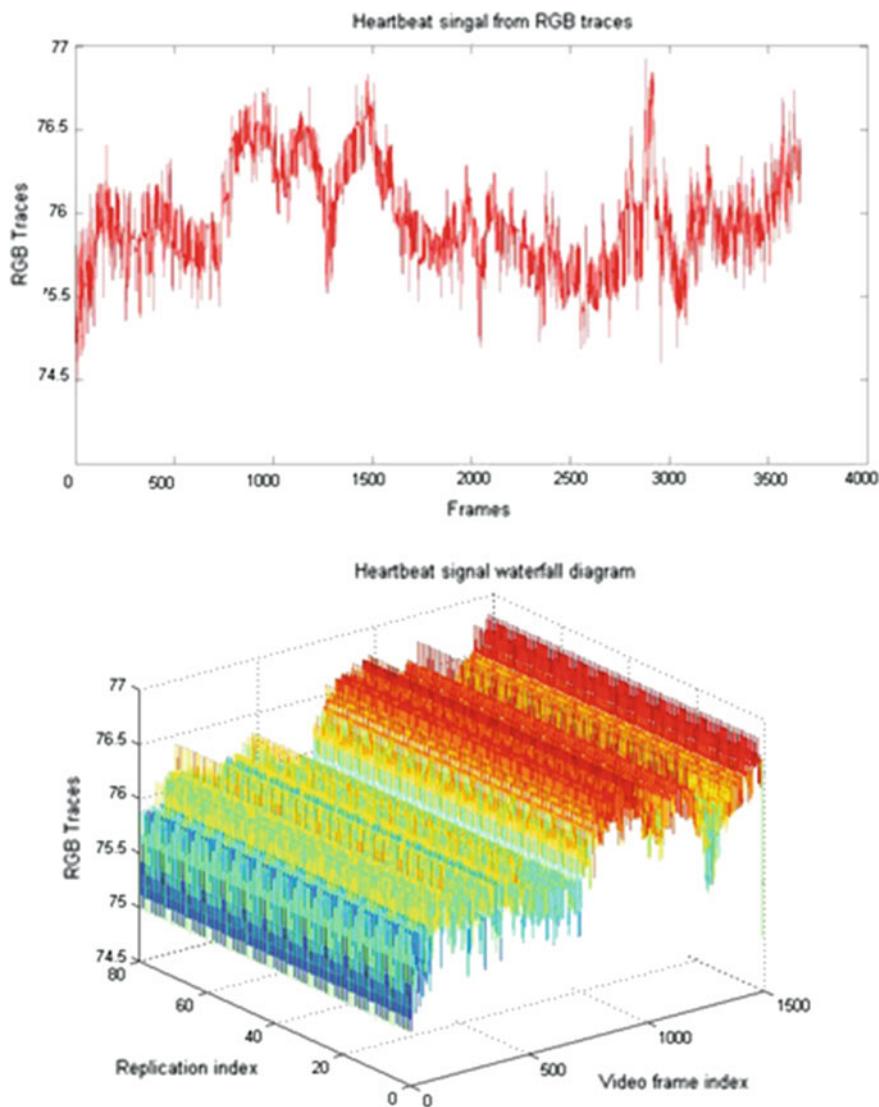
- [33] a ZivMerhav cross parsing method based on the entropy of the ECG signal has been used.
- [35, 38, 41] the shape and morphology of the heartbeat signal have been directly used as feature for identification.
- [37] sparse representation of the ECG signal has been used for identification.
- [31, 32, 34] frequency analysis methods have been applied to ECG signals for identification purposes. For example, in [31, 34] wavelet transformation has been used and it is discussed that it is more effective against noise and outliers.

### **13.3.2 Human Identification Using Contact-Free Heartbeat Signal**

The above-mentioned systems use contact-based (touch-based) sensors for measuring heartbeat signals. These sensors provide accurate, however sensitive to noise, measurements. Furthermore, they suffer from a major problem: these sensors need to be in contact with the body of the subject of interest. As mentioned before, this is not always practical, especially in identification context if subjects are not cooperative. To deal with this, we in [43] have developed an identification system that uses the heartbeat signals that are extracted using the contact-free technique of [20]. To the best of our knowledge, this is the first system that uses the contact-free heartbeat signals for identification purposes. In this section, we review the details of this system and its findings.

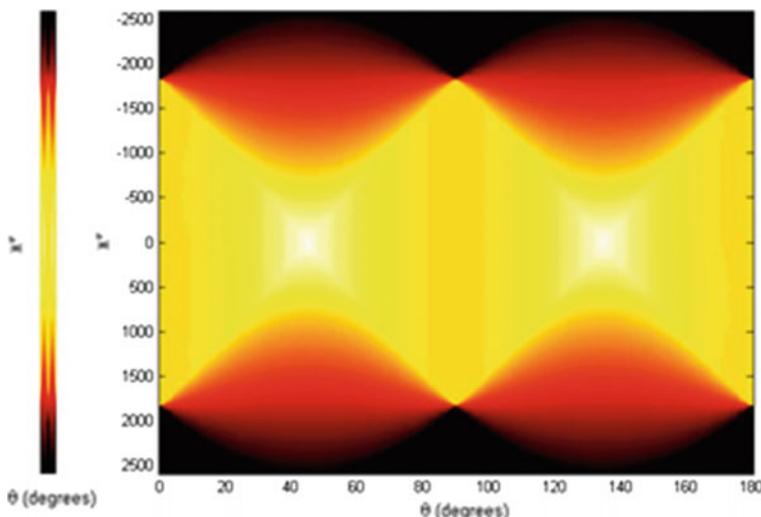
Having obtained the heartbeat signals, in form of RGB traces, from facial images using the method of [20] in [43] first a denoising filter is employed to reduce the effect of the external sources of noise, like changes in the lightning and head motions. To do that, the peak of the measured heartbeat signal is found and used to discard the outlying RGB traces. Then, the features that are used for the identification are extracted from this denoised signal. Following [44] the features that are used for identification in [43] are based on Radon images obtained from the RGB traces. To produce such images from the RGB traces, in [43] first the tracked RGB traces are replicated to the same number as the number of the frames that is available in the video sequence that is used for the identification, to generate a waterfall diagram. Figure 13.4 shows an example of such a waterfall diagram obtained for a contact-free measured heartbeat signal.

The Radon image, which contains the features that are going to be used for the recognition in [43], is then generated from the waterfall by applying a Radon transform to the waterfall diagram. Figure 13.5 shows the obtained Radon image from the waterfall diagram of Fig. 13.4. The discriminative features that are used for identification purposes from such a Radon image are simply the distance between every two possible pixels of the image.



**Fig. 13.4** A contact-free measured heartbeat signals (on the top) and its waterfall diagram (on the bottom)

The experimental results in [43] on a subset of the large facial database of MAHNOB-HCI [24] shown in Table 13.1 indicate that the features extracted from the Radon images of the heartbeat signals that were collected using a contact-free computer vision technique carries some distinctive properties.



**Fig. 13.5** The Radon image obtained from the waterfall diagram of Fig. 13.4. The discriminative features for the identification purposes are extracted from such Radon images [43]. The image on the *right* is the zoomed version of the one on the *left*

**Table 13.1** The identification results using distance features obtained from Radon images of contact-free measured heartbeat signals from [43] on a subset of MAHNOB-HCI [24] containing 351 video sequences of 18 subjects

Measured parameter	Results (%)
False positive identification rate	1.28
False negative identification rate	1.30
True positive identification rate	98.70
Precision rate	80.86
Recall rate	80.63
Specificity	98.63
Sensitivity	97.42

## 13.4 Discussions and Conclusions

If the heartbeat signals are going to be used for identification purposes, they should serve as a biometric or a soft-biometric. A biometric signal should have some specific characteristics, among others, it needs to be

- Collectible, i.e., the signal should be extractable. The ECG-based heartbeat signal is obviously collectible, though one needs to install ECG sensors on the body of subjects, which is not always easy, especially when subjects are not cooperative.

- Universal, i.e., everyone should have an instance of this signal. An ECG heartbeat signal is also universal as every living human has a beating heart. This also highlights another advantage of heartbeat signal which liveness. Many of the other biometrics, like face, iris, and fingerprints, can be spoofed by printed versions of the signal, and thus need to be accompanied by a liveness detection algorithm. Heartbeat signal however does not need liveness detection methods [33] and is difficult to disguise [27].
- Unique, i.e., instances of the signal should be different from one subject to another.
- Permanent, i.e., the signal should not change over time [45].

Regardless of the method used for obtaining the heartbeat signal (contact-free or contact-based), such a signal is collectible (according to the discussions in the previous sections) and obviously universal. The identification systems that have used the contact-based sensors (like ECG), [25–42] have almost all reported recognition accuracies that are more than 90% on datasets of different sizes from 20 people in [25] to about 100 subjects in the others. The lowest recognition rate, 76.9%, has been reported in [32]. The results here are however reported on a dataset of 269 subjects for which the ECG signals have been collected in different sessions. Some of the sessions are from the same day and some are from different days. If both the training and the testing samples are from the same day (but still different sessions), the system report 99% recognition rate, but when the training and testing data is coming from different sessions recorded in different days, the recognition rate drops to 76.9% for the rank-1 recognition, but still as high as 93.5% for rank-15 recognition. This indicates that an ECG-based heartbeat signal might be considered as a biometric, though there is not that much study on the permanent dimensions of such signals for the identification purposes, reporting their results on very large databases.

On the other hand, due to the measurement techniques that contact-free methods provide for obtaining the heartbeat signals, the discriminative properties of contact-free obtained heartbeat signals is not as high as their peer contact-based ones. However, it is evident from the results of our recent work, which was reviewed in the previous section, that such signals have some discriminative properties that can be utilized for helping identification systems. In another words, a contact-free obtained heartbeat signal has the potential to be used as soft-biometric.

To conclude, the contact-free heartbeat signals seem to have promising applications in forensics scenarios. First of all, because according to the above discussions they seem to carry some distinguishing features, which can be used for identification purposes. Furthermore, according to the discussion presented in [17] the motion-based methods, which do not necessarily need to extract these signals from a skin region, can extract the heartbeat data from the hairs of the head, or even from a masked face. This will be of great help in many forensics scenarios, if one could extract a biometric or even a soft-biometric from a masked face because in many such scenarios the criminals are wearing a mask to hide their identity from, among others, surveillance cameras.

## References

1. Ehrlich D, Carey L, Chiou J, Desmarais S, El-Difrawy S, Koutny L, Lam R, Matsu-daira P, McKenna B, Mitnik-Gankin L, ONeil T, Novotny M, Srivastava A, Streechon P, Timp W (2002) MEMS-based systems for DNA sequencing and forensics. In: Proceedings of IEEE Sensors, vol 1, pp 448–449
2. Lin WS, Tjoa SK, Zhao HV, Liu KJR (2009) Digital image source coder forensics via intrinsic fingerprints. *IEEE Trans Inf Forensics Secur* 4(3):460–475
3. Roussev V (2009) Hashing and data fingerprinting in digital forensics. *IEEE Secur Priv* 8 (2):49–55
4. Peacock C, Goode A, Brett A (2004) Automatic forensic face recognition from digital images. *Sci Justice* 44(1):29–34
5. Jain AK, Unsang P (2009) Facial marks: soft biometric for face recognition. In: 16th IEEE international conference on image processing (ICIP), pp 37–40
6. Unsang P, Jain AK (2010) Face matching and retrieval uses soft biometrics. *IEEE Trans Inf Forensics Secur* 3:406–415
7. Han H, Otto C, Liu X, Jain A (2014) Demographic estimation from face images: human versus machine performance. *IEEE Trans Pattern Anal Mach Intell* 99
8. Jain AK, Klare B, Unsang P (2011) Face recognition: some challenges in forensics. In: 2011 IEEE international conference on automatic face gesture recognition and workshops (FG 2011), pp 726–733
9. Wagner J, Jonghwa K, Andre E (2005) From physiological signals to emotions: implementing and comparing selected methods for feature extraction and classification. In: IEEE international conference on multimedia and expo, pp 940–943
10. Li L, Chen J (2006) Emotion recognition using physiological signals. *Adv Artif Reality Tele-Existence Lect Notes Comput Sci* Springer 4282:437–446
11. Kuriakose S, Sarkar N, Lahiri U (2012) A step towards an intelligent Human Computer Interaction: Physiology-based affect-recognizer. In: 4th international conference on intelligent human computer interaction (IHCI), pp 1–6
12. Liao W, Zhang W, Zhu Z, Ji Q (2005) A real-time human stress monitoring system using dynamic Bayesian network. In: IEEE computer society conference on computer vision and pattern recognition—workshops
13. Zhai J, Barreto A (2006) Stress detection in computer users through non-invasive monitoring of physiological signals. *Biomed Sci Instrum* 42:495–500
14. Barreto A, Zhai J, Adjouadi M (2007) Non-intrusive physiological monitoring for automated stress detection in human-computer interaction. *Human-Comput Interact Lect Notes Comput Sci* Springer 4796:29–38
15. Liu C, Torralba A, Freeman WT, Durand F, Adelson EH (2005) Motion magnification. *ACM Trans Graph* 24(3):519–526
16. Wu HY, Rubinstein M, Shih E, Guttag J, Durand F, William TF (2012) Eulerian video magnification for revealing subtle changes in the world. In: Proceedings of SIGGRAPH ACM transactions on graphics, vol 31, no 4
17. Balakrishnan G, Durand F, Guttag J (2013) Detecting pulse from head motions in video. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp 3430–3437
18. Irani R, Nasrollahi K, Moeslund TB (2014) Improved pulse detection from head motions using DCT. In: 9th International Conference on Computer Vision Theory and Applications, vol 3, pp 118–124
19. Poh MZ, McDuff DJ, Picard R (2010) Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt Express* 18:10762–10774
20. Poh MZ, McDuff DJ, Picard R (2011) Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Trans Biomed Eng* 58:7–11

21. Sarkar A, Abbott AL, Doerzaph Z (2014) Assessment of psychophysiological characteristics using heart rate from naturalistic face video data. In: IEEE international joint conference on biometrics (IJCBI)
22. Tarassenko L, Villarroel M, Guazzi A, Jorge J, Clifton DA, Pugh C (2014) Non-contact video-based vital sign monitoring using ambient light and autoregressive models. *Physiol Meas* 35:807–831
23. Li X, Chen J, Zhao G, Pietikainen M (2014) Remote heart rate measurement from face videos under realistic situations. In: IEEE conference on computer vision and pattern recognition (CVPR), pp 4264–4271
24. Soleymani M, Lichtenauer J, Pun T, Pantic M (2012) A multimodal database for affect recognition and implicit tagging. *IEEE Trans Affect Comput* 3(1):42–55
25. Biel L, Pettersson O, Philipson L, Wide P (2001) ECG analysis: a new approach in human identification. *IEEE Trans Instrum Measur* 50(3):808–812
26. Hoekema R, Uijen GJH, van Oosterom A (2001) Geometrical aspects of the interindividual variability of multilead ECG recordings. *IEEE Trans Biomed Eng* 48(5):551–559
27. Israel SA, Irvine JM, Cheng A, Wiederhold MD, Wiederhold BK (2005) ECG to identify individuals. *Pattern Recogn* 38(1):133–142
28. Wang Y, Plataniotis KN, Hatzinakos D (2006) Integrating analytic and appearance attributes for human identification from ECG signals. In: Biometrics symposium: special session on research at the biometric consortium conference
29. Plataniotis KN, Hatzinakos D, Lee JKM (2006) ECG biometric recognition without fiducial detection. In: Biometrics symposium: special session on research at the biometric consortium conference
30. Singh YN, Gupta P (2008) ECG to individual identification. In: 2nd IEEE international conference on biometrics: theory, applications and systems
31. Fatemian SZ, Hatzinakos D (2009) A new ECG feature extractor for biometric recognition. In: 2009 16th international conference on digital signal processing
32. Odinaka I, Po-Hsiang L, Kaplan AD, O'Sullivan JA, Sirevaag EJ, Kristjansson SD, Sheffield AK, Rohrbaugh JW (2010) ECG biometrics: a robust short-time frequency analysis. In: IEEE international workshop on information forensics and security (WIFS)
33. Coutinho DP, Fred ALN, Figueiredo MAT (2010) One-lead ECG-based personal identification using Ziv-Merhav cross parsing. In: 20th international conference on pattern recognition (ICPR), pp 3858–3861
34. Can Y, Coimbra MT, Kumar BVKV (2010) Investigation of human identification using two-lead Electrocardiogram (ECG) signals. In: Fourth IEEE international conference on biometrics: theory applications and systems (BTAS)
35. Islam MS, Alajlan N, Bazi Y, Hichri HS (2012) HBS: A novel biometric feature based on heartbeat morphology. *IEEE Trans Inf Technol Biomed* 16(3):445–453
36. Tantawi M, Revett K, Tolba MF, Salem A (2012) A novel feature set for deployment in ECG based biometrics. In: 7th international conference on computer engineering systems (ICCES), pp 186–191
37. Wang J, She M, Nahavandi S, Kouzani A (2013) Human identification from ECG signals via sparse representation of local segments. *IEEE Sign Proces Lett* 20(10):937–940
38. Fratini A, Sansone M, Bifulco P, Romano M, Pepino A, Cesarelli M, D'Addio G (2013) Individual identification using electrocardiogram morphology. In: 2013 IEEE international symposium on medical measurements and applications proceedings (MeMeA), pp 107–110
39. Rabhi E, Lachiri Z (2013) Biometric personal identification system using the ECG signal. In: Computing in cardiology conference (CinC), pp 507–510
40. Ming L, Xin L (2014) Verification based ECG biometrics with cardiac irregular conditions using heartbeat level and segment level information fusion. In: IEEE international conference on acoustics, speech and signal processing (ICASSP), pp 3769–3773
41. Lourenco A, Carreiras C, Silva H, Fred A (2014) ECG biometrics: A template selection approach. In: IEEE international symposium on medical measurements and applications (MeMeA)

42. Nomura R, Ishikawa Y, Umeda T, Takata M, Kamo H, Joe K (2014) Biometrics authentication based on chaotic heartbeat waveform, In: Biomedical engineering international conference (BMEiCON)
43. Haque MA, Nasrollahi K, Moeslund TB (2015) Heartbeat signal from facial video for biometric recognition. In: Proceedings of 19th Scandinavian conference on image analysis
44. Hegde C, Prabhu HR, Sagar DS, Shenoy PD, Venugopal KR, Patnaik LM (2011) Heartbeat biometrics for human authentication. *SIViP* 5(4):485–493
45. Van de Haar H, Van Greunen D, Pottas D (2013) The characteristics of a biometric. In: Information security for South Africa

**Part IV**

**Statistical Analysis of Forensic  
Biometric Data**

# Chapter 14

## From Biometric Scores to Forensic Likelihood Ratios

Daniel Ramos, Ram P. Krish, Julian Fierrez and Didier Meuwly

**Abstract** In this chapter, we describe the issue of the interpretation of forensic evidence from scores computed by a biometric system. This is one of the most important topics into the so-called area of forensic biometrics. We will show the importance of the topic, introducing some of the key concepts of forensic science with respect to the interpretation of results prior to their presentation in court, which is increasingly addressed by the computation of likelihood ratios (LR). We will describe the LR methodology, and will illustrate it with an example of the evaluation of fingerprint evidence in forensic conditions, by means of a fingerprint biometric system.

### 14.1 Likelihood Ratio Framework for Evidence Evaluation

The evaluation of the relationship between two pieces of evidence at judicial trials has been the subject of discussion in the past years [1]. Here, the problem is to give a value to a comparison of a trace specimen of unknown origin (for instance a finger-mark revealed in a crime scene, or a wire tapping involving an incriminating conver-

---

D. Ramos (✉) · R.P. Krish · J. Fierrez  
ATVS - Biometric Recognition Group, Escuela Politecnica Superior,  
Universidad Autonoma de Madrid, Calle Francisco Tomas y Valiente 11,  
28049 Madrid, Spain  
e-mail: daniel.ramos@uam.es

R.P. Krish  
e-mail: ram.krish@uam.es

J. Fierrez  
e-mail: julian.fierrez@uam.es

D. Meuwly  
Netherlands Forensic Institute, Laan van Ypenburg 6, 2497GB  
The Hague, The Netherlands  
e-mail: d.meuwly@utwente.nl

D. Meuwly  
University of Twente, Drienerlolaan 5, 7522NB Enschede, The Netherlands  
e-mail: d.meuwly@utwente.nl

sation) with a reference specimen of known origin (for instance, a fingerprint from a suspect, or some recordings of a known individual). From a formal logical perspective [2], the given value should represent the degree of support of the comparison to any of the *propositions* (also called *hypotheses*) involved in the trial. Examples of simple hypotheses might be “the trace and the reference specimens originated from the same source” or “the trace and the reference specimens originated from different sources”, but more complex hypotheses can be considered [2]. In some sense, the value of the evidence represents the strength of the link between the trace and the reference specimen in the context of the propositions considered.

Evidence evaluation using a Bayesian probabilistic framework has been proposed in recent years as a logical and appropriate way to report evidence to a court of law [3]. In Europe, there have been initiatives to foster this approach, some of them in response of notorious reluctance to the use of statistics in courts [4]. These have been the main reason leading to the release of a Guideline [5] for the expression of conclusions in evaluative forensic reports. This Guideline is proposed by the European Network of Forensic Science Institutes (ENFSI), an organization that includes almost all the main forensic laboratories in Europe.<sup>1</sup> According to this Guideline, a Bayesian framework for forensic evaluative reports is recommended for all disciplines and laboratories within ENFSI. Under this Bayesian approach, a likelihood ratio (LR) is computed to represent the value of the evidence, and to be reported to a court of law (mainly in the form of a verbal equivalent). This framework clearly complies with the requirements of evidence-based forensic science [1]: it is scientifically sound (transparent procedures, testability, formal correctness), and clearly separates the responsibilities of the forensic examiner and the court.

The increasing establishment of this Bayesian evaluative framework has motivated the convergence of pattern recognition and machine-learning approaches to yield probabilistic outputs in the form of likelihood ratios. A common architecture for this considers two steps: first, the computation of a discriminating score between two specimens, trace and reference (e.g., a fingermark in the crime scene and an exemplar fingerprint from a known suspect), which can be obtained from a standard biometric system; and second, the transformation of the score into a likelihood ratio [6–9]. This architecture is especially suited for biometric systems, where the output of a conventional biometric system is typically expressed as a score, even though it is used as *black-box* technology. Therefore, the score is most of the times a necessary intermediate step to the likelihood ratio.

### 14.1.1 Challenges in LR-Based Evidence Evaluation

Despite its advantages, the computation of likelihood ratios still presents important challenges. We enumerate the most important as follows.

---

<sup>1</sup><http://www.enfsi.eu/>.

First, the typical scenario in forensic science involves data presenting diverse and unfavorable conditions, which means that automatic comparisons between the specimens will result in a challenging problem. Efforts to model or compensate the effects of these adverse conditions in likelihood ratio computation should be improved. Some works such as [10] have contributed to evaluate the impact of this problem. Moreover, integration of advanced machine-learning algorithms (like in [11, 12]) for the compensation of adverse conditions into forensic evaluation helps in this direction. However, adverse condition compensation still remains a challenge.

Second, in forensic science the databases are difficult to obtain and to use, even for research purposes. This is because, although there is plenty of forensic data in some disciplines (e.g., large fingerprint databases), there are legal, privacy and interoperability issues that hamper the use of this data by academic and research institutions. This leads to two opposite situations: either the databases are big when there is access to the data, and therefore big-data solutions are a challenge to face; or the databases are highly scarce, and the use of robust models is necessary. Data scarcity has been tackled by different techniques as in [13, 14]. However, to our knowledge, evidence evaluation models have not been adapted to big-data scenarios to handle big databases when possible, which represents a loss of information in these scenarios. Other lines of research have proposed the use of simulated forensic data in order to prevent the problem of data scarcity [15]. The involvement of simulated data is a big improvement against data scarcity situations, but the testing of the validity of simulated databases for the operational use of systems in a real setup is still controversial.

Third, although likelihood ratio computation methods are becoming more and more popular, the validation of those methods for its use in forensic casework is still not standardized. Even if likelihood ratios are computed to evaluate the links between evidential materials, this does not guarantee that they will be able to be integrated into a Bayesian decision framework to ultimately allow a fact finder to do optimal decisions. In this sense, the measurement of the performance characteristics that a likelihood ratio model should manifest is of paramount importance. Recent work has shown that one of the most important characteristic that forensic likelihood ratios should present is the so-called calibration [9]. This is a property of a set of likelihood ratios, by which the LR is itself a measure of evidential weight. This leads to the property that “The LR of the LR is the LR”, meaning that the LR is interpreting the evidence with the best possible probabilistic meaning in terms of Bayes decisions [16]. Therefore, computing likelihood ratios is not enough, they should also be the best calibrated as possible. There are current efforts of the forensic community in order to establish formal frameworks for the validation of likelihood ratio models [9, 17, 18], but research is still needed. Also, a framework for the validation of likelihood ratio methods has been recently published [19].

Fourth, evidence evaluation in complex cases is still problematic. Probabilistic graphical models, particularly Bayesian networks [20], have been proposed to address those situations. However, this emerging field is an active area of research in forensic science. More efforts are needed in order to provide forensic examiners with appropriate tools in operational scenarios, especially if those models are to be learned from data.

## 14.2 Case Assessment and Interpretation Methodology

A milestone in the use of the LR methodology in Europe was the Case Assessment and Interpretation (CAI) methodology developed by the Forensic Science Service (FSS) in the late 1990s [2]. This was the result of the efforts of the now closed Forensic Science Service of the United Kingdom, in order to homogenize and make more agile the relationship between courts and forensic service providers (e.g., police forces or other public or private forensic laboratories). An ultimate aim is the use of a logical methodology to avoid pitfalls of reasoning and fallacies. The methodology has been described in several papers during the end of the twentieth century, remarkably [2, 21, 22], and serves as the core of likelihood ratio-based evidence interpretation.

There are several characteristic features of the CAI methodology, which we summarize below.

- Full integration of the LR methodology into the forensic evidence evaluation process. In this sense, all the elements typical from LR evidence evaluation are present, namely the evidence, propositions, probabilistic reasoning, etc.
- A particular emphasis is put in the definition of the propositions in a given case, which have to be informed by the circumstances of the case themselves. Thus, the relationship between the court and the forensic science provider should be essential in order to define the propositions. Issues like the definition of the population considered to model the alternative proposition, the specificity of the propositions with respect to the population, the suspect and the trace, or the selection of the most appropriate database to address the propositions [23], are of particular importance.
- A hierarchy of propositions [21] is introduced in order to address the forensic casework in the most appropriate manner with respect to the information in the case. In this sense, there are three basic levels in the hierarchy: *source level*, where the source which originated the trace(s) is considered; *activity level*, where the activities from which originate the traces are under discussion; and *offence level*, that focuses on the question whether the activity from which originate the traces is an infraction. Depending on the question asked by the requester/fact finder and on the information in the case available to the forensic scientist, it is possible to escalate the inference a to higher level, but in most cases the forensic examiner is requested to report at source level, reason for which most of the effort to produce increasingly robust models has been focused on source level. Nevertheless,

nowadays there is a push towards the use of activity-level propositions in casework (even in the ENFSI Guideline for evaluative reports [5]), although LR models for activity or offence levels are mostly in a research stage.

- Case pre-assessment is encouraged by the model. Under this concept, a preliminary LR value is reported prior to the case itself, in order to indicate what would be the expected outcome of the forensic analysis by the examiner. This helps to focus the expectations of the fact finder, and has important implications regarding the efficiency of resources in a case.

### 14.3 Evidence Evaluation with Likelihood Ratios

The LR framework for interpretation of the evidence represents a mathematical and logical tool in order to aid in the inference process derived from the analysis of the evidence. In this methodology, the objective of the forensic scientist is computing the likelihood ratio (LR) as a degree of support of one proposition versus its alternative [3, 24].

The LR framework is stated as follows. Consider a forensic case. There is a forensic evidence  $E$ , which contains the specimens to compare in a forensic case, namely, in a fingerprint case, a *recovered* fingermark of unknown origin and a *reference* fingerprint (namely the *exemplar*) whose origin is known to be a given suspect involved in the case. In this context, the unobserved variable of interest is the true proposition  $H$  with values  $\{H_p, H_d\}$ , where  $H_p$  and  $H_d$  are the possible relevant propositions defined in the case, according to the CAI methodology. As mentioned before, the definition of  $H_p$  and  $H_d$  varies in each case. A possible definition at the source level could be as follows:

- $H_p$ : The origin of the fingermark (trace) and the fingerprint (reference) is the same finger of one single donor.  
 $H_d$ : The origin of the fingermark (trace) and the fingerprint (reference) are fingers from two different donors.

$H_p$  is typically called *prosecution* proposition, whereas  $H_d$  is referred to as *defense* proposition. This is due to the fact that alternative, and mutually exclusive propositions arise naturally in an adversarial trial system like in the UK, where the CAI methodology was developed. Other propositions can be addressed, and variation of their statement can lead to a radically different selection of databases for LR computation [23], and even to different likelihood ratio models [25]. Therefore, care should be taken in order to clearly and appropriately define the propositions in a case.

In Bayesian decision theory, decisions are made considering the probability distribution of the variable of interest (in this case, the proposition variable  $H$ ), given all the observed information. In a forensic case, this can be represented as  $P(H = H_p | E, I)$  and  $P(H = H_d | E, I)$ , or simply  $P(H_p | E, I)$  and  $P(H_d | E, I)$ , where  $I$  is the background information available in the case not related to the evidence  $E$ , as defined by the CAI methodology.  $H_p$  and  $H_d$  are in most cases mutually

exclusive. Then, Bayes' theorem [3] relates probabilities before and after evidence analysis.

$$P(H_p | E, I) = \frac{P(E | H_p, I) \cdot P(H_p | I)}{P(E | I)} \quad (14.1)$$

In terms of interpretation, it is useful to use ratios of probabilities. Then, Eq. 14.1 becomes

$$\frac{P(H_p | E, I)}{P(H_d | E, I)} = LR \cdot \frac{P(H_p | I)}{P(H_d | I)} \quad (14.2)$$

$$LR = \frac{P(E | H_p, I)}{P(E | H_d, I)} \quad (14.3)$$

In Eq. 14.2, we can distinguish the following:

1. The prior probabilities  $P(H_p | I)$  and  $P(H_d | I)$ , which are province of the fact finder and should be stated assuming only the background information ( $I$ ) in the case [24].
2. The LR (Eq. 14.3), assigned or computed by the forensic practitioner [3].

A critical point in the application of the LR methodology is the selection of proper databases to address the propositions and also the trace material, for example the language of the trace will determine the language of the speech databases used in the case, but also from the definition of the propositions themselves. We will address this issue in the example below, but many works in the literature give recommendations on how to select these databases, both in fingerprints [26] or in forensic science in general [23].

This LR-based framework for interpretation presents many advantages

- It allows forensic practitioners to evaluate and report a meaningful value for the weight of the evidence to the court, with a universal interpretation, allowing for the combination of results across disciplines when the same propositions are considered [5, 24].
- The role of the examiner is clearly defined, leaving to the court the task of using prior judgments or costs in the decision process.
- Probabilities can be interpreted as degrees of belief [27], allowing the incorporation of subjective opinions as probabilities in the inference process in a clear and scientific way.

The LR value has an interpretation as a *support* to a previously stated opinion, due to the analysis of the evidence  $E$ . In other words

- If the  $LR > 1$  the evidence will support that  $H = H_p$ , i.e., the prosecutor proposition.

- If the  $LR < 1$  the evidence will support that  $H = H_d$ , i.e., the defense proposition.

Moreover, the value of the LR represents the *degree of support* of the evidence to one value of  $H$  against the other. For instance,  $LR = 3$  means that “the evidence supports  $H = H_p$  against  $H = H_d$  with a degree of 3 versus 1”. Therefore, a single LR value has a *meaning* by itself, as opposed to a biometric score, that may have only meaning if compared to a reference threshold or another set of scores.

It is important to note that the LR *supports* an opinion about  $H$ , but the LR *is not* an opinion about  $H$ . Opinions about  $H$  are represented as probabilities of propositions, or in our binary case, their ratios. Therefore, it is not possible to make a decision about the value of  $H$  based solely on the value of the LR, because decisions will be taken from posterior probabilities, not only from degrees of support.

## 14.4 Interpreting Biometric System Scores with Likelihood Ratios

According to [6, 7], in biometrics all the information that the systems yield about the propositions after observing  $E$  is in most cases condensed into a so-called *score*, a single number which is an observation of a random variable  $S$ , and must contain as much information as possible about  $H$ . Therefore, the interpretation of the evidence using biometric systems requires that the score will be first computed by the system, yielding the particular value  $S = s$ , and then the score is interpreted using a likelihood ratio. This leads to the following expression:

$$LR = \frac{P(E|H_p, I)}{P(E|H_d, I)} = \frac{P(s|H_p, I)}{P(s|H_d, I)} \quad (14.4)$$

Moreover, most biometric scores are continuous, and in that case the ratio of probabilities becomes a ratio of probability density functions, yielding

$$LR = \frac{P(s|H_p, I)}{P(s|H_d, I)} = \frac{p(s|H_p, I)}{p(s|H_d, I)} \quad (14.5)$$

Thus, the LR value (Eq. 14.5) is the quotient of two probability densities. On the one hand, the probability density function (pdf)  $p(S|H_p, I)$  in the numerator in Eq. 14.3 is known as the intra-variability distribution. Its evaluation in the particular value of the score  $S = s$  gives a measure of the probability density of observing the evidence under  $H_p$ . On the other hand, the pdf  $p(S|H_d, I)$  in the denominator is known as the inter-variability distribution, and its evaluation in the particular value of

the score  $S = s$  gives a measure of the probability density of observing the evidence under  $H_d$ .<sup>2</sup>

The aim of LR methods with biometric score-based systems is to provide a model that transforms scores into LR values in a case. Moreover, the resulting LR values should present adequate performance in order to correctly aid the decisions of fact finders.

## 14.5 LR Computation Methods from Biometric Scores

In this section, some of the most common algorithms for LR computation from biometric scores are described.

### 14.5.1 Generating Training Scores

The main commonality of all the methods described in this section is that they need two proposition-dependent sets of *training* scores, namely  $\mathbf{S}_p$  and  $\mathbf{S}_d$ . These sets and some of the issues associated to them are described as follows.

- The set  $\mathbf{S}_p = \left\{ s_p^{(1)}, \dots, s_p^{(N_{pt})} \right\}$  consists of  $N_{pt}$  scores computed assuming that  $H = H_p$ . Therefore, the selection of data to compute the scores in  $\mathbf{S}_p$  has to be done accordingly to the definition of the propositions. As  $H_p$  proposition typically assumes that the trace and reference specimens in the case come from the same person, the  $\mathbf{S}_p$  consists of *same-source* scores.<sup>3</sup> However, the rest of information in  $H_p$  can be determinant in order to select the database to generate those same-source scores. For instance, if the particular suspect in the case is included in the propositions (e.g., “the trace was left by Mr. Dean Keaton”), then the propositions will be *person-specific* or *source-specific*, and the database to generate  $\mathbf{S}_p$  should include specimens coming from the particular donor, because in many biometric traits each person has a particular behavior regarding their score distribution [28]. On the other hand, *person-generic* or *source-generic* propositions (e.g., “the trace and the reference samples come from the same source”) would allow the use of any same-source score from other people, since there is no knowledge of a particular subject. Another example of the influence of propositions in the model for LR computation is related to the definition of suspect-based or finger-based proposition for fingerprint interpretation [25].

---

<sup>2</sup>The background information about the case  $I$  will be eliminated from the notation for the sake of simplicity hereafter. It will be assumed that all the probabilities defined are conditioned to  $I$ .

<sup>3</sup>Here we work at the source level, and therefore *same-source* scores refer to scores generated from two biometric specimens coming from the same source. They are what in biometric authentication terminology are called *genuine* scores.

- The set  $\mathbf{S}_d = \left\{ s_d^{(1)}, \dots, s_d^{(N_{dt})} \right\}$  consists of  $N_{dt}$  scores computed assuming that  $H = H_{dt}$ . For the computation of these scores, several things should be taken into account. First,  $H_d$  typically assumes that the questioned materials were not generated by the suspect in the case, but other person. Therefore, the scores in  $\mathbf{S}_d$  will essentially be *different-source* scores,<sup>4</sup> since the case always considers the donor reference specimens as part of the evidence. Second, the way in which these scores are generated is critical, since the selection of different strategies to obtain  $\mathbf{S}_d$  might lead to different LR values. Also, theoretical issues should be taken into account regarding these strategies (for a discussion about this, see [29]). Last, but not least, the determination of the population of sources of the test specimen must be handled with care. The key point is that the population must be seen as the set of potential donors of the test specimen, considering the definition of the proposition and the information about the case that is relevant and available to the forensic examiner.

An important remark is in order here. The aim of the  $\mathbf{S}_p$  and  $\mathbf{S}_d$  score sets is to represent the variation of  $S$  conditioned to the propositions. As  $S$  is the variable representing the score to be obtained from the evidence in the case, the conditions in the forensic case should be preserved for all comparisons in  $\mathbf{S}_p$  and  $\mathbf{S}_d$ . For instance, if the evidence consists of a degraded, partial fingermark and a fingerprint from a ten-print card of a known suspect, all the scores in  $\mathbf{S}_p$  and  $\mathbf{S}_d$  should be generated from comparisons of degraded fingermarks and fingerprints from ten-print cards, in the conditions as similar as possible to those in the case. An exception would be if the conditions do not affect the behavior of the scores at all, but this rarely happens in real forensic scenarios.

Moreover, the scores in  $\mathbf{S}_p$  and  $\mathbf{S}_d$  should represent all possible sources of variability in  $S$ . Therefore, the use of models of variability is essential in order compute better likelihood ratios. Good examples exist in the literature of the use of variability models to compute the LR [26], or to compensate this variability at the level of the biometric score [11, 12].

## 14.5.2 Common Methods for Score-Based LR Computation

### 14.5.2.1 Generative Assignment of Probability Densities

*LR* computation in forensic biometrics has been classically performed by the use of generative techniques modeling the hypotheses-conditional distribution of the scores variable  $S$ . This is the approach already presented in [30], and has been followed in subsequent works in the literature. Under this approach, the objective is assigning

---

<sup>4</sup>Here we work at the source level, and therefore *different-source* scores refer to scores generated from two biometric specimens coming from different sources. They are what in biometric authentication terminology are called *impostor* scores.

the likelihoods  $p(S|H_p)$  to the training scores  $\mathbf{S}_p$ , and  $p(S|H_d)$  to  $\mathbf{S}_d$ . Then, the ratio of the particular value of these densities at  $S = s$  will be the LR value.

Assigning  $p(S|H_p)$  and  $p(S|H_d)$  implies the selection of a proper model. The most straightforward choice for biometric scores could be the Gaussian distribution, obtained via Maximum Likelihood from the training set of scores. However, this requires the distributions involved to present a good fitting with Gaussian probability density functions, which is not typically the case. Fortunately, some score normalization techniques such as T-Norm tend to generate Gaussian distributions for scores when  $H_d$  is true [31]. Other approaches for generative ML fitting includes the use of Kernel Density Functions [30, 32], Gaussian Mixture Models [32] and other parametric distributions [18, 42].

#### 14.5.2.2 Logistic Regression

Logistic regression is a well-known pattern recognition technique widely used for many problems including fusion [33, 34] and more recently likelihood ratio computation [7, 35]. The aim of logistic regression is obtaining an affine transformation (i.e., shifting and scaling) of an input dataset in order to optimize an objective function. Let  $\mathbf{S}_f = \{s_f^{(1)}, s_f^{(2)}, \dots, s_f^{(K)}\}$  be a set of scores from  $K$  different biometric systems. The affine transformation performed by the logistic regression model can be defined as

$$f_{lr} = \log \left( \frac{P(H_p | \mathbf{S}_f, I)}{P(H_d | \mathbf{S}_f, I)} \right) = a_0 + a_1 \cdot s_f^{(1)} + a_2 \cdot s_f^{(2)} + \dots + a_K \cdot s_f^{(K)} \quad (14.6)$$

This leads to the following *logistic regression model*:

$$P(H_p | \mathbf{S}_f, I) = \frac{1}{1 + e^{-f_{lr}}} = \frac{1}{1 + e^{-\log(LR) - \log(O(H_p))}} \quad (14.7)$$

where  $O(H_p)$  determines the prior odds in favor of  $H_p$ .

The weighting terms  $\{a_0, a_1, a_2, \dots, a_K\}$  can be obtained from a set of training data with optimization procedures found in the literature.<sup>5</sup> Moreover, by training the weights for some given simulated value of the prior odds, and removing the influence of that value of the prior odds after computing  $f_{lr}$ , likelihood ratios are obtained.

Notice that logistic regression can be used for computing likelihood ratios from a single biometric score ( $K = 1$ ), but also to perform fusion and LR computation simultaneously (when  $K > 1$ ) [34]. This fact, joined to the good behavior that logistic

---

<sup>5</sup>Typical implementations used in biometrics include toolkits like FoCal or BOSARIS, which can be found in <http://niko.brummer.googlepages.com>.

regression presents in most situations, have made this LR computation algorithm one of the most popular ones.

#### 14.5.2.3 Pool Adjacent Violators (PAV)

Another approach to score-to-LR transformation has been proposed by the use of the Pool Adjacent Violators (PAV) algorithm [7]. The PAV algorithm transforms a set of scores into a set of LR values presenting optimal calibration. However, it is only possible to apply an optimal PAV transformation if the ground-truth labels of the propositions for each score in the set are known. As suggested in [8], a PAV transformation can be trained on the set of training scores  $S_p$  and  $S_d$ , and then apply the trained transformation to a score in a forensic case. Although a straightforward use of PAV leads to a non-invertible transformation, several *smoothing* techniques can be applied to PAV in order to keep it monotonically increasing. For instance, adding a small slope to the function defining the PAV transformation will lead to an invertible transformation. Interpolating with linear, quadratic or splines approaches are also possible smoothing schemes.

## 14.6 Performance Measurement of LR Methods

As it was previously stated, the issue of performance measurement of LR methods is paramount in order to achieve validation of forensic interpretation prior to its use in casework [18]. In this section, we describe some of the performance metrics adequate for LR-based forensic interpretation.

At the source level, performance measurement is typically carried out in an empirical way. In order to measure the performance of a LR method, a test set of LR values should be generated by comparisons of specimens from a biometric database using that LR method. These comparisons should be in fact simulated cases, where the conditions of specimens should be similar to the conditions of the case scenario whose performance is to be measured. This would lead to  $N_p$  LR values computed when  $H_p$  is true and  $N_d$  LR values computed when  $H_p$  is true.

A solution to measure the performance of likelihood ratio values has been proposed in [7] for speaker recognition, and has been dubbed *log-likelihood-ratio cost* ( $C_{llr}$ ). Later, it has been used in many other fields in forensic sciences [14, 17, 36, 37].  $C_{llr}$  is defined as follows:

$$C_{llr} = \frac{1}{2 \cdot N_p} \sum_{i_p} \log_2 \left( 1 + \frac{1}{LR_{i_p}} \right) + \frac{1}{2 \cdot N_d} \sum_{j_d} \log_2 (1 + LR_{j_d}) \quad (14.8)$$

The indices  $i_p$  and  $j_d$  respectively denote summing over the LR values of the simulated cases where each proposition is respectively true.

An important result is derived in [7], where it is demonstrated that minimizing the value of  $C_{llr}$  also encourages to obtain reduced Bayes decision costs for all possible decision costs and prior probabilities [38]. This property has been highlighted as extremely important in forensic science [9]. Moreover, in [7], the Pool Adjacent Violators algorithm is used in order to decompose  $C_{llr}$  as follows:

$$C_{llr} = C_{llr}^{min} + C_{llr}^{cal} \quad (14.9)$$

where

- $C_{llr}^{min}$  represents the *discrimination cost* of the LR method, and it is due to non-perfect discriminating power.
- $C_{llr}^{cal}$  represents the *calibration cost* of the system.

$C_{llr}$  is a scalar measure of performance of LR values, the lower its value the better the performance. Another useful measure of performance, with interpretation in terms of information loss, is the Empirical Cross-Entropy (ECE), which is a generalization of  $C_{llr}$ , as follows:

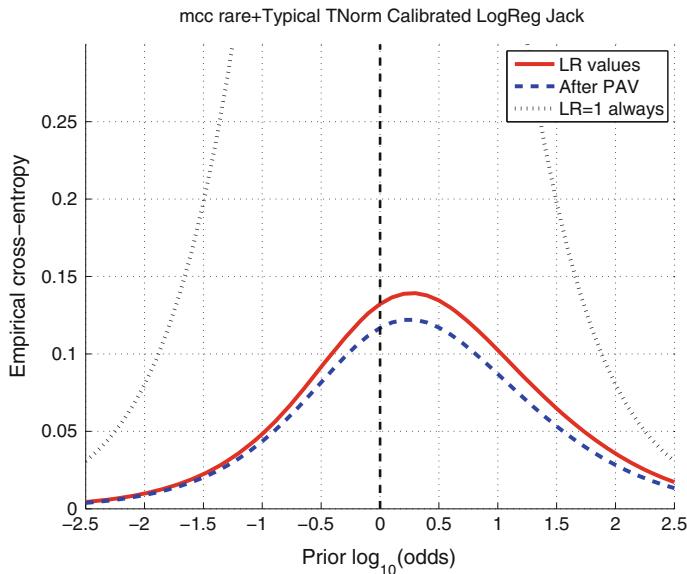
$$\begin{aligned} ECE = & - \frac{P(H_p | I)}{N_p} \sum_{i_p} \log_2 P(H_p | s_{i_p}, I) \\ & - \frac{P(H_d | I)}{N_d} \sum_{j_d} \log_2 P(H_d | s_{j_d}, I), \end{aligned} \quad (14.10)$$

where  $s_{i_p}$  and  $s_{j_d}$  denote the scores from trace and reference specimens in each of the simulated cases, where either  $H_p$  or  $H_d$  is respectively true.

As it happens with  $C_{llr}$ , ECE can be additively decomposed also using the PAV algorithm into  $ECE = ECE^{min} + ECE^{cal}$ . This leads to  $ECE^{min}$  measuring information loss due to bad discriminating power, and  $ECE^{cal}$  measuring information loss due to miscalibration.

As it can be seen, ECE is dependent of the prior probabilities both explicitly and through  $P(H_p | s_{i_p}, I)$  and  $P(H_d | s_{j_d}, I)$ . Thus, ECE can be represented in a prior-dependent way. This has been proposed to be done by a so-called ECE plot [17], which shows three comparative performance curves together (Fig. 14.1)

- solid, red curve: accuracy. This curve is the ECE of the LR values in the validation set, as a function of the prior log-odds. The lower this curve, the more accurate the method. This curve shows the overall performance of the LR method;
- dashed, blue curve:  $ECE^{min}$ . This curve is the ECE of the validation set of LR values after the application of the PAV algorithm. This shows the best possible ECE in terms of calibration, and it is a measure of discriminating power;



**Fig. 14.1** Example of ECE plot

- dotted curve: neutral reference. It represents the comparative performance of a so-called *neutral LR method*, defined as the one which always delivers  $LR = 1$  for each forensic case simulated in the set of LR values. This neutral method is taken as a *floor of performance*: the accuracy should always be better than the neutral reference. Therefore, the solid curve in an ECE plot should be always lower than the dotted curve, for all represented values of the prior log-odds (the names *floor* and *ceiling* are the opposite of the usual physical connotations but are chosen to represent the lowest and highest levels of performance).

A free Matlab<sup>TM</sup> software package to draw ECE plots can be found in the following webpage: <http://arantxa.ii.uam.es/~dramos/software.html>.

## 14.7 Computing LR Values from Biometric Scores: An Example with Forensic Fingerprint Recognition

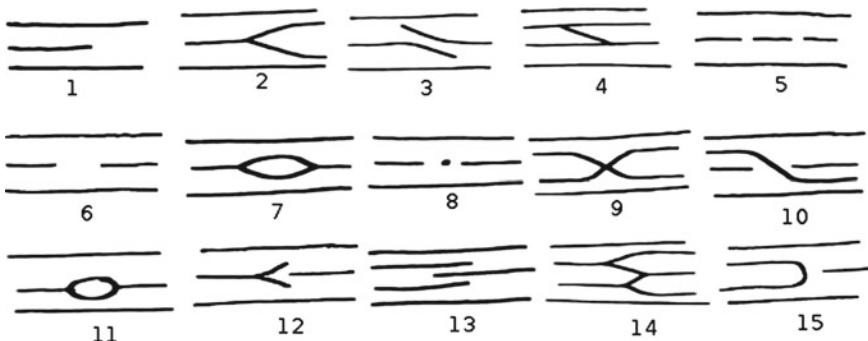
In this section, we illustrate the process of computing forensic LR values from biometric fingerprint scores. We use a database collected from real cases, and in collaboration with Spanish Guardia Civil. Also, we use a state-of-the-art fingerprint system based on Minutiae Cylinder Code [39–41].

Evidence evaluation in fingerprints by the use of LR has been recently proposed in remarkable works like in [26] for minutiae configurations extracted manually from

forensic examiners. However, other models based on the use of AFIS scores to compute likelihood ratio values can be found in [42], and more recently [18]. The reasons of modeling AFIS scores are diverse. On the other hand, it may give a complementary information to other methods more based on the direct statistical modeling of the minutiae extracted by the examiners. On the other hand, it allows the use of powerful systems to extract the information of identity, after which a likelihood ratio model performs the interpretation with the less loss of information possible [6].

### 14.7.1 Database and Statistics

The database used in this work was obtained from Guardia Civil, the Spanish law enforcement agency. The Guardia Civil database (GCDB) is a realistic forensic fingerprint casework database. Apart from having typical minutiae feature types (*ridge-endings*, *bifurcations*), GCDB also comprises rare minutiae types like *fragments*, *enclosures*, *dots*, *interruptions*, etc. [43]. A comprehensive list of rare minutia features used by Guardia Civil are shown in Fig. 14.2 and the corresponding minutiae type names are listed in Table 14.1.



**Fig. 14.2** Minutia types used by Guardia Civil. Names corresponding to individual minutia type numbers can be found in Table 14.1

**Table 14.1** List of minutiae types used by Guardia Civil. Numbering with respect to Fig. 14.2

No	Minutiae type	No	Minutiae type	No	Minutiae type
1	Ridge ending	6	Interruption	11	Circle
2	Bifurcation	7	Enclosure	12	Delta
3	Deviation	8	Point	13	Assemble
4	Bridge	9	Ridge crossing	14	M-structure
5	Fragment	10	Transversal	15	Return

**Table 14.2** The probability of occurrence and the entropy-based weights for the minutiae types present in the 268 fingermarks of GCDB. The numbers correspond to minutiae types in Fig. 14.2

No	Minutiae type	Probability ( $p_i$ )	Weight ( $w_i = -\log_{10} p_i$ )
1	Ridge-ending	0.5634	0.2492
2	Bifurcation	0.3620	0.4413
3	Deviation	0.0015	2.8294
4	Bridge	0.0024	2.6253
5	Fragment	0.0444	1.3523
6	Interruption	0.0021	2.6833
7	Enclosure	0.0204	1.6896
8	Point	0.0036	2.4492
10	Transversal	0.0003	3.5284

GCDB used in this work consists of 268 fingermark and reference fingerprint images and minutiae sets. All the minutiae in the fingermark images were manually extracted by forensic examiners of Guardia Civil. The corresponding mated minutiae in the reference fingerprints were also manually established. This includes the typical (ridge-endings and bifurcations) minutiae and the rare minutiae. The minutiae in the reference fingerprints were combined with the minutiae extracted by Neurotechnology VeriFinger algorithm, in order to generate a complete set of minutiae for the reference fingerprint. For the fingermark, the minutiae will be the ones marked by the examiners. The average number of minutiae in the fingermarks was 13 and that of tenprints was 125.

The original fingermark minutiae sets provided by Guardia Civil and the post-processed VeriFinger generated minutiae sets are used in all our experiments. To represent some rare minutiae, multiple points were needed. For example, to represent a *deviation* two points are needed (see type 3 in Fig. 14.2), and to represent an *assemble* three points are needed (see type 13 in Fig. 14.2). Whenever multiple points are needed to represent a rare minutia, we mapped them to a single point representation by taking the average of locations and orientations of all points.

From the 268 fingermark minutiae sets, we estimated the probability of occurrence ( $p_i$ ) of various minutiae types. The probability ( $p_i$ ) and the entropy-based weights ( $w_i = -\log_{10} p_i$ ) for each minutiae type present in GCDB are listed in Table 14.2. In the 268 fingermarks of GCDB, we noticed only seven types of rare minutiae features. They are listed in Table 14.2. Other rare minutiae types are not found in the current database used in this study, because they did not appear in the whole database.

### 14.7.2 Biometric System

The system used to compare the minutiae was based on Minutiae Cylinder Code (MCC) representation, also extensively presented in another chapter of this book, deeply described in [39–41].<sup>6</sup> It is not the aim to deeply describe the score computation system in this chapter, because we aim at the likelihood ratio computation process. Details about the algorithm can be found in the relevant references [39–41].

In order to exploit the information in the rare minutiae features in the GCDB, the minutiae included in those rare points are part of the features directly added to the ridge endings and bifurcations. Thus, everything together has been used to feed the MCC system. Therefore, the scores obtained by the system include information from both typical and rare minutiae.

Finally, a T-Norm stage has been performed in order to align and normalize the output different-source scores of the system. Test-Normalization, or T-Norm [44] has been used to perform score normalization. In order to do that, a set of different-source fingerprints, namely a *cohort* of different sources, is needed. From those so-called T-Norm scores, the mean and the standard deviation  $\mu_{Tnorm}$  and  $\sigma_{Tnorm}$  are computed. The T-Norm technique is then applied to a particular score computed form a given fingermark query as follows:

$$s_{Tnorm} = \frac{s_{raw} - \mu_{Tnorm}}{\sigma_{Tnorm}} \quad (14.11)$$

Thus, T-Norm performs query-dependent score normalization, and the result is the alignment of the query-dependent different-source score distributions for all comparisons in the particular set of scores.

Thus, this normalization technique compensates variability in the scores due to the recovered fingermark. The T-Norm cohort in this experiment has been selected from the same Guardia Civil database that has been used to simulate real forensic fingermark-reference fingerprint comparisons, and therefore the results may be overstated in terms of performance. However, for the sake of illustration on the computation of likelihood ratios, this is not a problem.

It has been reported that the different-source scores of a biometric system tend to be more Gaussian after the application of T-Norm [31]. Therefore, we will assume that a Gaussian model will be appropriate for the MCC scores after T-Norm is applied.

---

<sup>6</sup>We have used the implementation of this score computation system provided by the authors.

### 14.7.3 Methodology and Proposed LR Methods

This section proposes several methods for likelihood ratio computation using scores from the MCC algorithm with the Guardia Civil database (GCDB) described in Sect. 14.7.1.

#### 14.7.3.1 Definition of Propositions

According to the methodology of CAI, the first step to compute likelihood ratios is to establish the propositions considering the information present in the case. The *simulated cases* that we are going to conduct here consist of the comparison of one fingermark and one reference fingerprint. Both fingermark and reference fingerprint come from GCDB. The scores used to train the models for the LR computation are the rest of scores in the GCDB generated from individuals different from the donors of the fingermark and the reference fingerprint. In this way, the models are trained with scores not used in the case, and the data handling is honest in the sense of the performance measurement.

According to this setup, there are several observations that are in order:

- The information in the case is almost non-existent. We only have the images of the fingermark and the reference fingerprint, and therefore no assumption can be done about the donors of fingermark and reference (e.g. ethnicity, gender, etc.). This only allows generic propositions about the populations involved.
- The trace and reference specimens are pseudonymised because the metadata of the donors are not necessary.
- We only have a single same-source comparison for each subject in the database. Therefore, it is impossible for us to focus in source-specific models, because there are no additional data available to model the particular behavior of their scores in comparison to the whole population of scores.
- There is no information whatsoever about the relevance of the donor of fingermark and reference fingerprint with respect to the action in the crime scene, or even more with respect to any offense. Therefore, only propositions at source level can be addressed.
- Because of the way it was built, we assume that all fingermarks in the GCDB dubbed as different in the ground-truth labels are generated by different people. It is assumed also in the corresponding reference fingerprint. Therefore, in this database it will be equivalent to talk about donors as about fingers, since different fingerprints will definitely belong to different donors (and not to different fingers of the same donor).

Under these premises, we decide to state source level, person-generic and general-population propositions for this case. Therefore, we have the following propositions:

$H_p$ : The origin of the fingermark (trace) and the fingerprint (reference) is the same finger of the donor.

$H_d$ : The origin of the fingermark (trace) and the fingerprint (reference) are fingers from two different donors.

This definition of the propositions implies that, for a forensic case involving the comparison of a fingermark and its corresponding reference fingerprint, the scores needed to train the LR model should be generated with comparisons of fingermarks and reference fingerprints without the constrain of belonging to a particular individual. This implies that more scores will be typically available to train the models, therefore improving their statistical robustness. On the other hand, the use of person-generic propositions inevitably implies an important loss of information in cases where the identity of the individual is known, as it is typical in court. However, for this example we will consider this person-specific scenario because of the limitations of the GCDB, as explained above.

#### 14.7.4 Likelihood Ratio Models

As example in this chapter, we will compare the performance of the following common models for likelihood ratio computation.

- Pool Adjacent Violators.
- Gaussian-ML.
- Logistic regression.

#### 14.7.5 Experimental Results

##### 14.7.5.1 Experimental Protocol

The experimental protocol has been designed in order to simulate a real forensic scenario where fingermarks are compared with reference fingerprints using typical minutia features and also rare minutia features.

In our experiments, we used the Guardia Civil database (as described in Sect. 14.7.1), because it is the only forensic fingerprint database which contains rare minutiae, as it has been previously described. Since the GCDB is limited in size, a cross-validation strategy has been followed in order to optimally use the data without using the same dataset to train and test the LR models proposed. This cross-validation strategy is described as follows: for each same-source comparison of a fingermark and a reference fingerprint, the scores to train the LR model for that particular comparison will consist of all the scores generated with the GCDB, except those generated with either the fingermark or the reference fingerprint involved in the case. Therefore, the separation between the fingermark and reference fingerprint and the individuals in the training database is guaranteed.

This cross-validation strategy has many advantages in the sense of the optimal usage of the available database. However, it also presents the disadvantage that the conditions of the training scores matches the conditions of the fingermark and reference fingerprint under comparison to a higher degree than in a potential real case. Thus, the results presented here could be overstated in terms of performance. However, due to the limitation of the database, and also because the aim of the work is to show how to apply the methodology, we consider it appropriate to use this protocol.

Notice that this cross-validation strategy does not only guarantee that the data used to train and test the models are different. Moreover, it also guarantees that the T-Norm scores generated with the cohort are not present in the training database. This is because the T-Norm cohort scores must be generated with the scores of the query fingermark, which will be not present in the training database. Therefore, the situation is realistic in the sense of the data handling to normalize the scores and also to train the LR models.

#### ***14.7.6 Results on the Comparison of LR Computation Methods***

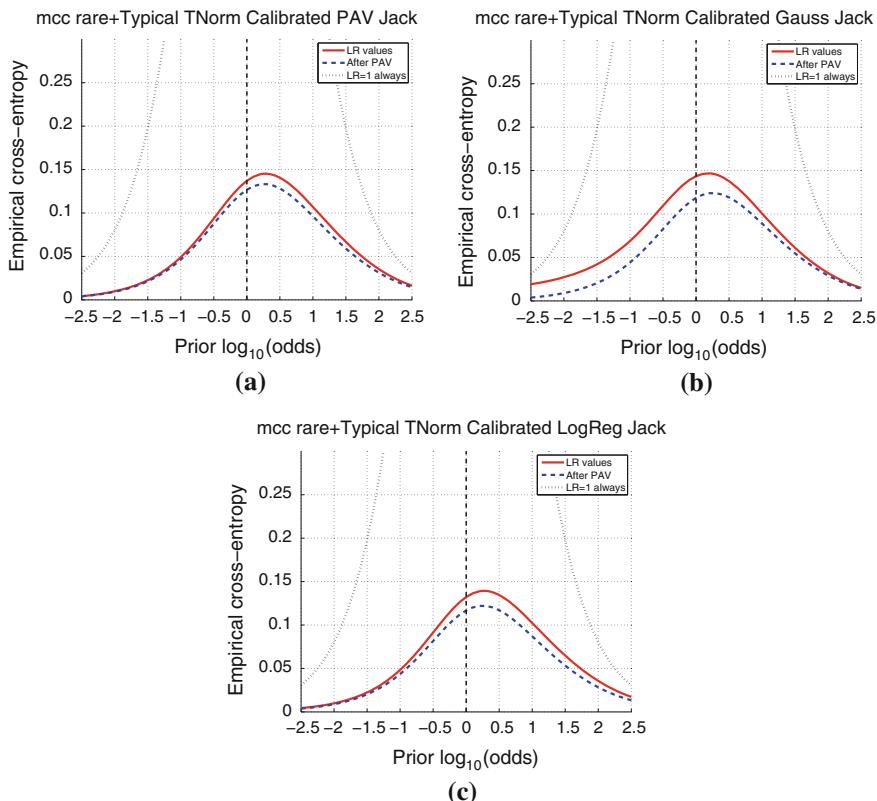
In this section, we compare all the proposed LR computation methods not only from the perspective of the discriminating power, but also with respect to the calibration loss. Thus, accuracy as the sum of both performance measures will allow us to select the best choice for LR computation. From Fig. 14.3, it is seen that the logistic regression model presents the best accuracy (red solid curve), and therefore is the best of the three methods proposed.

We now analyze calibration (separation between red and blue curves) more deeply. It is generally seen in Fig. 14.3 that the calibration loss is better for PAV and logistic regression methods rather than for the Gaussian method. Thus, both methods are apparent good options for LR computation. Regarding the Gaussian-ML method, it is seen that the calibration performance is worse than for PAV or logistic regression, especially in the region of low prior odds. As a possible explanation, although T-Norm different-source scores tend to be Gaussian when they are pooled among all queries, it is not the case for the same-source scores, and this makes the same-source distribution to seriously diverge from Gaussianity even after T-Norm is applied.

An additional warning is in order here. The cross-validation procedure to train LR models and to select T-Norm scores implies a scenario with low dataset shift between training and testing data. In a forensically realistic setup, where dataset shift between training and testing data can be severe, the performance of LR methods that excessively fits the training data can seriously degrade. On the other hand, it is known in pattern recognition that models with lower complexity are more robust to this effect to avoid overfitting. Therefore, the much lower complexity of logistic regression with respect to PAV indicates that the former can be potentially more robust to overfitting

and dataset shift than the latter in forensically realistic conditions. Due to this reason, logistic regression is more preferable to PAV computation method in this scenario.

As a conclusion of this section, the calibration loss represents a low percentage of all the loss of accuracy for logistic regression and PAV LR computation methods, in this order. This makes the overall performance of logistic regression superior, which among other reasons makes it the best choice. On the other hand, Gaussian-ML presents poorer calibration loss, sometimes presenting worse performance than the neutral reference, which makes it less recommendable for the score computation systems proposed in this chapter.



**Fig. 14.3** ECE plots showing performance of LR methods with T-Normed MCC scores. The different LR methods are PAV **(a)**, Gaussian-ML **(b)** and Logistic regression **(c)**

## 14.8 Conclusions

In this chapter, we have described the methodology for computing likelihood ratios from biometric scores, with an example of fingerprint recognition in forensic conditions using rare minutiae. This has allowed the interpretation of the evidence from the fingermark-to-reference-fingerprint comparisons simulating a real forensic case by the use of a cross-validation strategy with the GCDB. Several LR methods have been proposed and compared in terms of discriminating power and calibration performance. These results clearly show that the proposed methods present far better performance than the neutral reference and therefore are useful for forensic interpretation.

## References

1. Saks MJ, Koehler JJ (2005) The coming paradigm shift in forensic identification science. *Science* 309(5736):892–895
2. Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998) A model for case assessment and interpretation. *Sci Justice* 38:151–156
3. Aitken CGG, Taroni F (2004) Statistics and the evaluation of evidence for forensic scientists. Wiley, Chichester
4. Berger CA, Champod JS, Curran C, Dawid J, Kloosterman AP (2011) Expressing evaluative opinions: a position statement. *Sci Justice* 51:1–2. Several signatories
5. Willis S (2015) ENFSI guideline for the formulation of evaluative reports in forensic science. Monopoly Project MP2010: the development and implementation of an ENFSI standard for reporting evaluative forensic evidence. Technical report, European Network of Forensic Science Institutes
6. Ramos D (2007) Forensic evaluation of the evidence using automatic speaker recognition systems. PhD thesis, Depto. de Ingenieria Informatica, Escuela Politecnica Superior, Universidad Autonoma de Madrid, Madrid, Spain. <http://atvs.ii.uam.es>
7. Brümmer N, du Preez J (2006) Application independent evaluation of speaker detection. *Comput Speech Lang* 20(2–3):230–275
8. van Leeuwen D, Brümmer N (2007) An introduction to application-independent evaluation of speaker recognition systems. In: Müller C (ed) Speaker classification. Lecture notes in computer science/Artificial intelligence, vol 4343. Springer, Heidelberg, Berlin, New York
9. Ramos D, Gonzalez-Rodriguez J (2013) Reliable support: measuring calibration of likelihood ratios. *Forensic Sci Int* 230:156–169
10. Zadora G, Ramos D (2010) Evaluation of glass samples for forensic purposes—an application of likelihood ratio model and information-theoretical approach. *Chemometr Intell Lab Syst* 102:62–63
11. Li P, Fu Y, Mohammed U, Elder J, Prince SJD (2010) Probabilistic models for inference about identity. *IEEE Trans Pattern Anal Mach Intell (PAMI)* 34(1):144–157
12. Dehak N, Kenny P, Dehak R, Dumouchel P, Ouellet P (2010) Front-end factor analysis for speaker verification. *IEEE Trans Audio Speech Lang Process* 19(4):788–798
13. Villalba J, Brümmer N (2011) Towards fully Bayesian speaker recognition: integrating out the between-speaker covariance. Proceedings of the 12th annual conference of the international speech communication association, Interspeech 2011. Florence, Italy, pp 505–508
14. Zadora G, Martyna A, Ramos D, Aitken C (2014) Statistical analysis in forensic science: evidential values of multivariate physicochemical data. Wiley

15. Rodriguez CM, de Jongh A, Meuwly D (2013) Introducing a semiautomatic method to simulate large numbers of forensic fingermarks for research on fingerprint identification. *J Forensic Sci* 57(2):334–342
16. van Leeuwen DA, Brümmer N (2013) The distribution of calibrated likelihood-ratios in speaker recognition. arXiv preprint
17. Ramos D, Gonzalez-Rodriguez J, Zadora G, Aitken C (2013) Information-theoretical assessment of the performance of likelihood ratio models. *J Forensic Sci* 58:1503–1518
18. Haraksim R, Ramos D, Meuwly D, Berger CE (2015) Measuring coherence of computer-assisted likelihood ratio methods. *Forensic Sci Int* 249:123–132
19. Meuwly D, Ramos D, Haraksim R (in press) A guideline for the validation of likelihood ratio methods used for forensic evidence evaluation. *Forensic Sci Int*. [10.1016/j.forsciint.2016.03.048](https://doi.org/10.1016/j.forsciint.2016.03.048)
20. Taroni F, Aitken C, Garbolino P, Biedermann A (2006) Bayesian networks and probabilistic inference in forensic science. Wiley
21. Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA (1998) A hierarchy of propositions: deciding which level to address in casework. *Sci Justice* 38(4):231–239
22. Evett IW, Jackson G, Lambert JA (2000) More on the hierarchy of propositions: exploring the distinction between explanations and propositions. *Sci Justice* 40(1):3–10
23. Champod C, Evett IW, Jackson G (2004) Establishing the most appropriate databases for addressing source level propositions. *Sci Justice* 44(3):153–164
24. Evett IW (1998) Towards a uniform framework for reporting opinions in forensic science casework. *Sci Justice* 38(3):198–202
25. Neumann C, Evett IW, Skerrett JE, Mateos-Garcia I (2011) Quantitative assessment of evidential weight for a fingerprint comparison I: generalisation to the comparison of a mark with set of ten prints from a suspect. *Forensic Sci Int* 207:101–105
26. Neumann C, Evett I, Skerrett JE (2012) Quantifying the weight of evidence from a forensic fingerprint comparison: a new paradigm. *J R Stat Soc Ser A: Stat Soc* 175(2):371–415
27. Taroni F, Aitken CGG, Garbolino P (2001) De Finetti's subjectivism, the assessment of probabilities and the evaluation of evidence: a commentary for forensic scientists. *Sci Justice* 41(3):145–150
28. Doddington G, Liggett W, Martin A, Przybocki M, Reynolds DA (1998) Sheeps, goats, lambs and wolves: a statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. In: Proceedings of ICSLP
29. Hepler AB, Saunders CP, Davis LJ, Buscaglia J (2011) Score-based likelihood ratios for handwriting evidence. *Forensic Sci Int* 219(1–3):129–140
30. Meuwly D (2001) Reconnaissance de Locuteurs en Sciences Forensiques: L'apport d'une Approche Automatique. PhD thesis, IPSC-Universite de Lausanne
31. Navratil J, Ramaswamy G (2003) The awe and mystery of T-Norm. In: Proceedings of ESCA European conference on speech, communication and technology, EuroSpeech, pp 2009–2012
32. Gonzalez-Rodriguez J, Fierrez-Aguilar J, Ramos-Castro D, Ortega-Garcia J (2005) Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems. *Forensic Sci Int* 155(2–3):126–140
33. Pigeon S, Druyts P, Verlinde P (2000) Applying logistic regression to the fusion of the NIST'99 1-speaker submissions. *Digit Signal Process* 10(1):237–248
34. Brümmer N, Burget L, Cernocky J, Glembek O, Grezl F, Karafiat M, van Leeuwen DA, Matejka P, Schwartz P, Strasheim A (2007) Fusion of heterogeneous speaker recognition systems in the STBU submission for the NIST speaker recognition evaluation 2006. *IEEE Trans Audio Speech Signal Process* 15(7):2072–2084
35. Gonzalez-Rodriguez J, Rose P, Ramos D, Toledoano DT, Ortega-Garcia J (2007) Emulating DNA: rigorous quantification of evidential weight in transparent and testable forensic speaker recognition. *IEEE Trans Audio Speech Signal Process* 15(7):2072–2084
36. Vergeer P, Bolck A, Peschier LJ, Berger CE, Hendriks JN (2014) Likelihood ratio methods for forensic comparison of evaporated gasoline residues. *Sci Justice* 56(6):401–411

37. Morrison GS (2009) Likelihood-ratio-based forensic speaker comparison using parametric representations of vowel formant trajectories. *J Acoust Soc Am* 125:2387–2397
38. Duda RO, Hart PE, Stork DG (2001) Pattern classification. Wiley
39. Cappelli R, Ferrara M, Maltoni D (2010) Minutia cylinder-code: a new representation and matching technique for fingerprint recognition. *IEEE Trans Pattern Anal Mach Intell* 32:2128–2141
40. Cappelli R, Ferrara M, Maltoni D (2010) Fingerprint indexing based on minutia cylinder code. *IEEE Trans Pattern Anal Mach Intell* 33:1051–1057
41. Ferrara M, Maltoni D, Cappelli R (2012) Noninvertible minutia cylinder-code representation. *IEEE Trans Inf Forensics Secur* 7:1727–1737
42. Egli N (2009) Interpretation of partial fingermarks using an automated fingerprint identification system. PhD thesis, Institute de Police Scientifique, Ecole de Sciences Criminelles
43. Santamaria F (1955) A new method of evaluating ridge characteristics. *Fingerprint Ident Mag*
44. Auckenthaler R, Carey M, Lloyd-Tomas H (2000) Score normalization for text-independent speaker verification systems. *Digit Signal Process* 10:42–54

## Chapter 15

# Dynamic Signatures as Forensic Evidence: A New Expert Tool Including Population Statistics

Ruben Vera-Rodriguez, Julian Fierrez and Javier Ortega-Garcia

**Abstract** This chapter presents a new tool specifically designed to carry out dynamic signature forensic analysis and give scientific support to forensic handwriting examiners (FHEs). Traditionally FHEs have performed forensic analysis of paper-based signatures for court cases, but with the rapid evolution of the technology, nowadays they are being asked to carry out analysis based on signatures acquired by digitizing tablets more and more often. In some cases, an option followed has been to obtain a paper impression of these signatures and carry out a traditional analysis, but there are many deficiencies in this approach regarding the low spatial resolution of some devices compared to original offline signatures and also the fact that the dynamic information, which has been proved to be very discriminative by the biometric community, is lost and not taken into account at all. The tool we present in this chapter allows the FHEs to carry out a forensic analysis taking into account both the traditional offline information normally used in paper-based signature analysis, and also the dynamic information of the signatures. Additionally, the tool incorporates two important functionalities, the first is the provision of statistical support to the analysis by including population statistics for genuine and forged signatures for some selected features, and the second is the incorporation of an automatic dynamic signature matcher, from which a likelihood ratio (LR) can be obtained from the matching comparison between the known and questioned signatures under analysis. An example case is also reported showing how the tool can be used to carry out a forensic analysis of dynamic signatures.

---

R. Vera-Rodriguez (✉) · J. Fierrez · J. Ortega-Garcia

ATVS - Biometric Recognition Group, Escuela Politecnica Superior, Universidad Autonoma de Madrid, Calle Francisco Tomas y Valiente 11, 28049 Madrid, Spain  
e-mail: ruben.vera@uam.es

J. Fierrez  
e-mail: julian.fierrez@uam.es

J. Ortega-Garcia  
e-mail: javier.ortega@uam.es

## 15.1 Introduction

Forensic handwriting examiners (FHEs) have been carrying out studies about the authorship of handwritten signatures for court cases for over a century [1]. The great majority of works in the forensic field relates to offline signature analysis [2–7]. With the rapid evolution of technology, which allows the acquisition of dynamic signatures from tablets and smartphones, applications are spreading in the commercial sector to facilitate payments and also in banking to facilitate the digital storage of all the signed paperwork. Therefore, FHEs are being required to provide forensic evidence to determine the authenticity of handwritten signatures written on digitizing tablets [8], which can provide a static image of the signature but also, and most importantly, contain the dynamic information of at least the X and Y spatial coordinates over time.

Signature dynamics can be further processed to provide features such as the signing velocity, acceleration, and other stroke information along the signing trajectory. However, there are very few research works in the field of dynamic signature for forensic examinations [9–11]. The majority of relevant literature regarding dynamic signature analysis is in the field of biometric recognition [12], which make use of algorithms such as Hidden Markov Models [13, 14] or Dynamic Time Warping [15, 16].

In the last years there have been competitions on forensic signature verification for both offline and on-line signatures for automatic systems organized within the International Conference on Document Analysis and Recognition and (ICDAR) and the International Conference on Frontiers in Handwriting Recognition (ICFHR) from 2009 to date. It is interesting to note that in ICFHR 2010 offline signature competition, results of FHEs were also given allowing a comparison of performance of both automatic systems and FHEs [17]. It is also worth noting that in real practice FHEs carry out a multi-class problem classifying the signatures under analysis into genuine, forged, disguised (written by the authentic reference author, where he has deliberately tried to make the signatures look like a forgery, normally with the purpose of denying the signature at a later stage) or as inconclusive. On the other hand, automatic systems normally perform a two class problem deciding whether or not the given signature belongs to a referenced author. Results showed that different FHEs perform very differently, with some FHEs having very little opinion errors while some others many, and no correlation was observed with the experience in years of the FHEs or with the time (hours) they took to carry out the analysis (proficiency tests). Also, a significant percentage of FHEs decisions were inconclusive (between 30–50% of the datasets considered).

There are some commercially available tools for dynamic signature analysis (e.g., TOPAZ SigCompare<sup>1</sup> or KOFAX FraudOne<sup>2</sup>), which provide very limited functionalities to carry out a forensic analysis. This paper introduces e-BioSign, a new tool specifically designed to carry out forensic analysis of dynamic handwritten sig-

<sup>1</sup><http://www.topazsystems.com/sigcompare.html>, accessed April 2015.

<sup>2</sup><http://www.kofax.com/products/kofax-signature-solutions/kofax-fraudone>, accessed April 2015.

natures in order to facilitate the work of FHEs and give scientific support to their conclusions. In this sense, a survey of the methodology employed by the FHEs has been conducted and included in the functionalities of the tool. Together with these functionalities, e-BioSign tool also allows the measurement of dynamic information contained in the signatures, not taken into account normally by FHEs. With the use of dynamic signatures there is additional information available which can be used to carry out a more comprehensive and reliable forensic analysis.

Additionally, e-BioSign tool includes two important functionalities. On the one hand, it gives statistical support to the FHEs for some selected parameters such as the duration, fluency or level of tremor of the signatures. Population distributions for these parameters were computed for genuine and forged signatures allowing to position the questioned and known signatures under analysis on these distributions and extract some conclusions with statistical support. On the other hand, a dynamic signature verification system is included, from which a Likelihood Ratio (LR) can be obtained from the matching comparison between the signatures under analysis which is complementary to the analysis carried out by the FHE.

The remainder of the paper is organized as follows. Section 15.2 describes the traditional forensic practice followed to carry out the analysis of dynamic signatures. Section 15.3 describes e-BioSign tool with all its functionalities including the description of the statistical analysis carried out on a set of selected features in order to give statistical support to this forensic tool. Section 15.4 reports a Case Study using the tool for some genuine and forged dynamic signatures, and finally, Sect. 15.5 draws the final conclusions. This chapter is based on an extension of the previous work [18].

## 15.2 Forensic Practice for Signature Analysis

As mentioned, traditionally the practice of FHEs has been mainly concerned with the analysis of paper-based (offline) signatures. In order to carry out an analysis regarding the authorship of a questioned signature FHEs normally use some kind of variant of the following protocol.<sup>3</sup>

The first requirement is to have an appropriate set of signatures to perform the analysis, otherwise it wouldn't be possible to obtain convincing conclusions. Therefore, FHEs can ask the person whose signature is being investigated to provide a set of signatures (around 20) in order to have some samples produced with natural fluency. If the questioned signature was produced a considerable time before, then FHEs try to find some examples of contemporary genuine signatures.

Then, the analysis is performed taking into account aspects such as the **composition** of the signature (with or without name, surname, if legible, presence of flourish, etc.), **location** regarding other text or box (if it is close to the text or box on

---

<sup>3</sup>Based on published documentation from the Spanish Guardia Civil [4], the Netherland Forensic Institute [7] and the Victoria Police Forensic Services Centre (Australia) [3].

the right, left, etc.), **direction** (inclination of the written part regarding the horizontal, also the flourish), **written part** (FHEs carry out a comparison letter by letter), **flourish** (initial and final points and their direction), **fluency** and **pressure**. Even if in offline signature analysis fluency and pressure can not be measured as accurate as with dynamic signatures, this dynamic information is considered as an important and discriminative factor and it is estimated by analyzing the width of the stroke or the groove left in the paper.

Some important aspects taken into account by FHEs to detect forged signatures are the following: in general the forger is only able to focus in one of the two main aspects required to obtain a good quality forgery: (i) precision in the production of the signature (size, proportion and shape), or (ii) written fluently. Therefore, in general the forgeries can be precise regarding the appearance but not fluent, or written fluently but imprecise. Other signs to detect forgeries are changes in velocity in different strokes, tremors, monotonous pressure, traces of practice or guiding lines, unnatural pen lifts, corrections, etc. Also, the complexity of the signature is an important aspect to take into account as complex signatures are much harder to be forged.

### 15.3 e-BioSign Tool

This section describes the main functionalities of e-BioSign tool, which is a tool designed to be used by FHEs to carry out the analysis of dynamic signatures and give scientific support to their forensic reports. This first version of the tool has been developed under Matlab GUI interface, but a second version of the tool as an independent application is under development. The most important functionalities of this tool are:

- Several signatures can be loaded and visualized simultaneously (i.e., reference signatures and the signature under analysis).
- Signatures can be normalized in the spatial and time domains.
- Strokes can be manually selected for further analysis (to measure dimensions, angles, etc.).
- Statistical analysis of a selection of parameters can be conducted positioning the signatures under analysis in a population distribution.
- Automatic signature verification provides a matching score to complement the analysis of the FHE.

Next, the functionalities of e-BioSign Tool are described. We have divided these functionalities in four main modules.

### ***15.3.1 Module 1: Signatures Loading and Normalization***

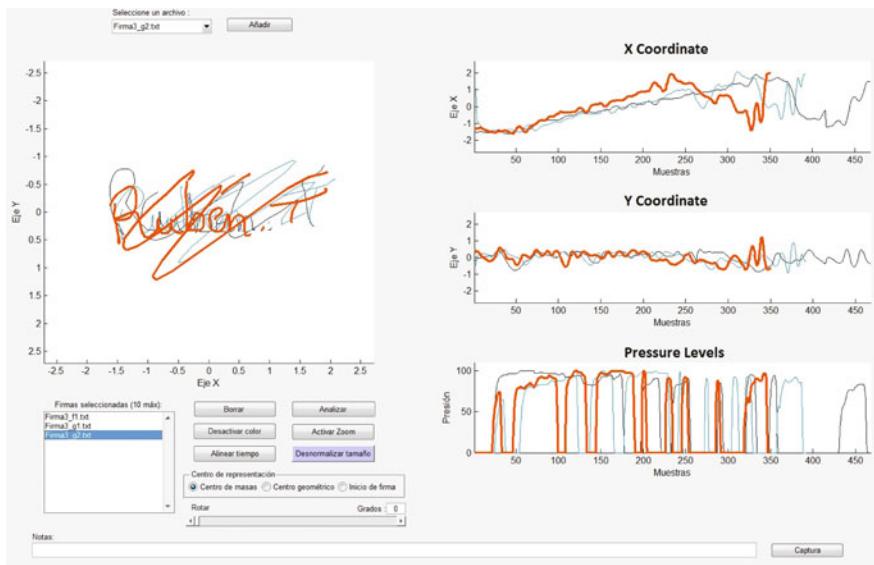
Module 1 allows to load several signatures for further analysis. The signatures can be visualized simultaneously, i.e., both the spatial image of the signature and the dynamic information of the X and Y coordinates and pressure. This is very useful as questioned and known signatures can be visualized at the same time allowing to analyze similarities and dissimilarities. Figure 15.1 shows a screenshot of Module 1 of e-BioSign tool with three signatures loaded, two of them genuine and one forgery.

When loading the signatures the information regarding frequency sampling (in Hz) and spatial resolution (pixel per inch) needs to be entered in a pop up window. In the example shown in Fig. 15.1a it is interesting to see how the two genuine signatures (orange and blue) have similar time duration, while the forgery (black) has a longer duration. In Module 1, it is also possible to normalize the loaded signatures both in the spatial domain and in the time domain. In the spatial domain, three position normalizations are possible considering different reference points: (i) center of mass, (ii) geometric center, or (iii) beginning of the signatures. A size normalization can be also applied maintaining the aspect ratio of input signatures. In the time domain, the signatures can be resampled to have the same time length. Figure 15.1b shows the same three example signatures shown in Fig. 15.1a after time normalization. In this case, it is possible to see how the two genuine signatures provide a good match in X, Y and pressure values, while there are more dissimilarities (especially in the pressure) regarding the forged signature.

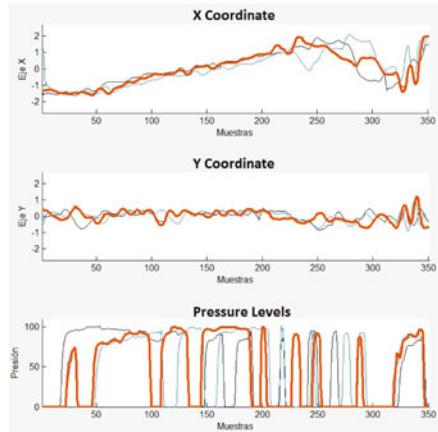
### ***15.3.2 Module 2: Individual Signature Analysis and Stroke Selection***

Module 2 allows to analyze the input signatures independently, and also to select strokes from each signature for further analysis. In order to analyze each signature, it is possible to reproduce the realization of the dynamic information of the signature, both in the spatial and time domains, with or without considering the pen-up dynamics. The pen-up dynamic information can be also visualized in the spatial representation. This is very interesting as this information can be very discriminative. Also, the pressure level information of each point can be incorporated in the visualization through a color map. Figure 15.2 shows a screenshot of Module 2 with one signature, in which the signature is represented with a color map based on the pressure values, and also the pen-up information is visible (in pink).

This module also allows to select strokes from the signature for a more detailed analysis. The strokes can be selected both by choosing initial and final points in the spatial representation of the signature, or using sliding bars in the time representation.

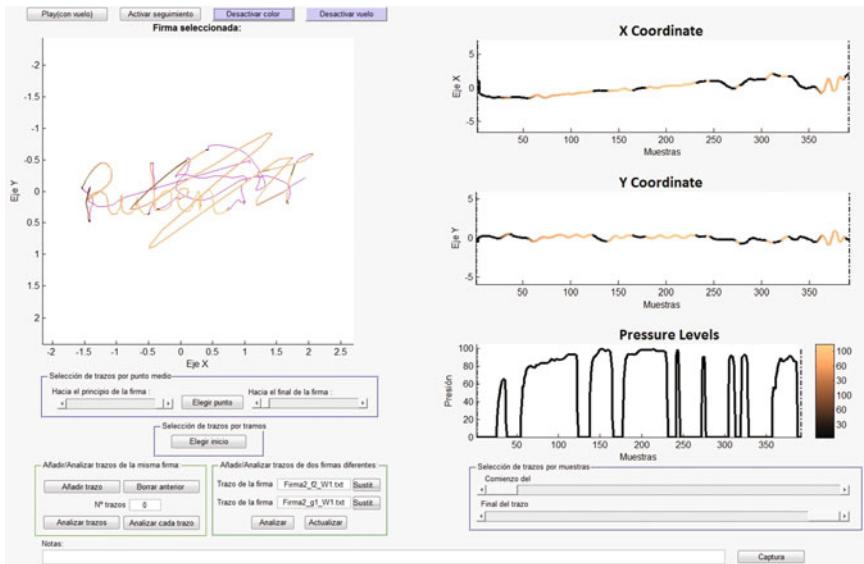


(a)



(b)

**Fig. 15.1** **a** Screenshot of e-BioSign tool Module 1, which allows to load several signatures and carry out a joint analysis. **b** Same signals normalized in time



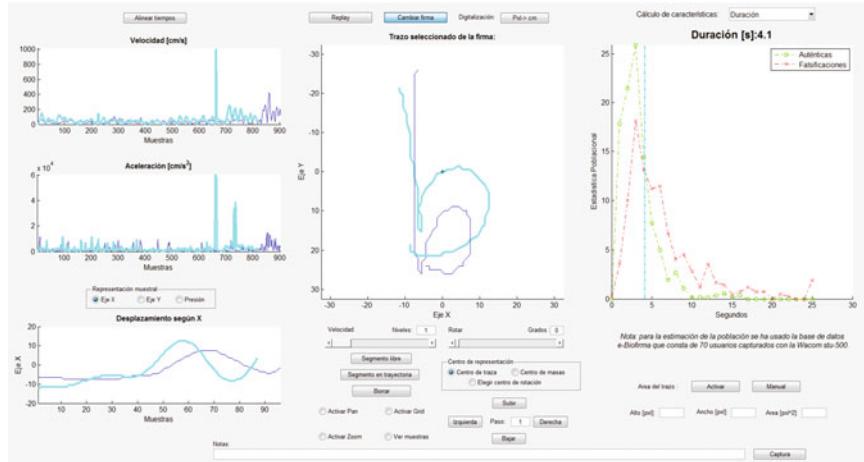
**Fig. 15.2** Screenshot of e-BioSign tool Module 2, which allows to analyse signatures individually reproducing the dynamic of the signature and showing the pen-up information. Strokes can be selected for further analysis

### 15.3.3 Module 3: Strokes Analysis

Module 3 allows to carry out a more detailed analysis of the selected strokes from the signatures. It is worth noting that the whole signature can be also selected as one stroke. Figure 15.3 shows a screenshot of Module 3. On the left part, it is possible to visualize the dynamics of the velocity and acceleration, and below again the X and Y coordinates and pressure. The analysis here can be conducted on single or multiple strokes at the same time, from one or more different signatures.

In the middle part of Fig. 15.3 there are some additional functionalities: it is possible to rotate the stroke regarding the center of representation chosen (geometric center, center of mass or any other fixed points), the stroke thickness can also be selected, it is possible to zoom in and out the stroke and also the real sample points of the signature can be visualized. Moreover, this module allows to take measurements of the length (both in pixels and cm) and the angle of any segment with respect to the horizontal line (in degrees).

Module 3 also allows to carry out a statistical analysis of some features automatically extracted from the signatures, as can be seen on the right part of Fig. 15.3. The idea is to provide the forensic expert with a population distribution of genuine and forged signatures for a selection of features together with the actual value of these features for the signatures at hand. For the initial release of e-BioSign Tool five global features have been selected. Three of them are common in feature based



**Fig. 15.3** Screenshot of e-BioSign tool Module 3. This module allows to carry out a detailed analysis of the selected strokes, it also allows to position the selected strokes (or signatures) in a population distribution of genuine and forged signatures for five selected features

dynamic signature recognition systems [14, 19]: total duration of the signature, average velocity and average acceleration. The other two parameters are commonly used in offline signature forensic analysis [2–6]: time fluency and spatial tremor. These two parameters are normally considered as good indicators to discriminate between genuine and forged signatures.

The **time fluency** of the signature is related to the number of samples with very low velocity in X and Y coordinates. Therefore, the time fluency was calculated following Eq. 15.1.

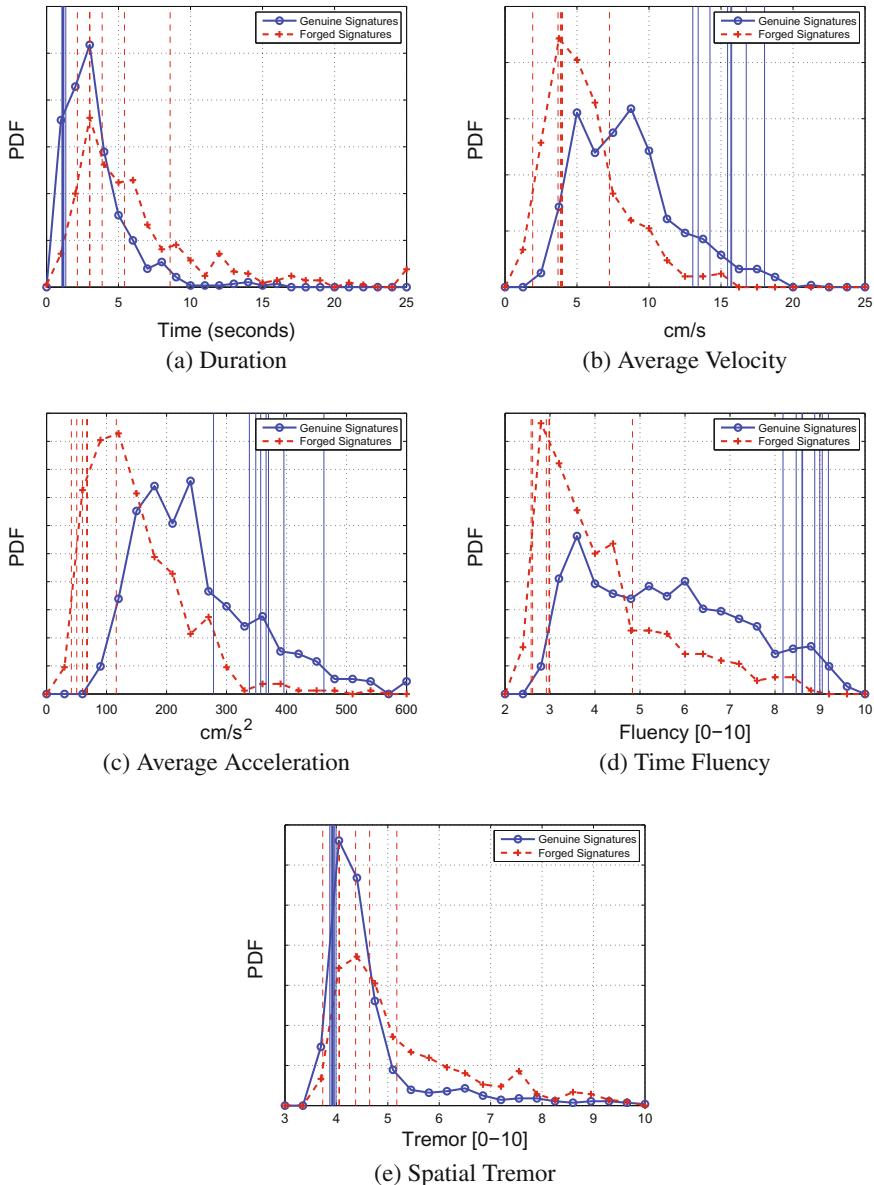
$$Fluency = \frac{-(N_{Vx} + N_{Vy})}{N} \quad (15.1)$$

where  $N_{Vx}$ ,  $N_{Vy}$  and  $N$  correspond respectively to the number of samples with velocity in X or Y ( $Vx$  or  $Vy$ ) equal or less than a threshold, which was set empirically to value 1 for obtaining Fig. 15.4, and  $N$  is the total number of time samples. The fluency is finally normalized in the range [0, 10] using an hyperbolic tangent normalization [20]:

$$Fluency' = \frac{10}{2} \left\{ \tanh(1.5(\frac{Fluency - \mu}{\sigma})) + 1 \right\} \quad (15.2)$$

where  $\mu$  and  $\sigma$  are the mean and standard deviation of a set of signatures used exclusively to carry out the data normalization.

The **spatial tremor** present in the signatures can be due to low confidence in the realization of the (possibly forged) signature. The level of tremor of a signature was obtained using the Euclidean distance between the X and Y time functions of



**Fig. 15.4** Frequency histograms for genuine and forged signatures for the five selected parameters: **a** signature duration, **b** average velocity, **c** average acceleration, **d** time fluency, and **e** spatial tremor. Vertical lines show the positioning of the genuine and forgery signatures for a given user of the database

the signature to analyse and a smoothed version of them. This smoothed signature is obtained using a low pass Gaussian filter. Finally, these distance values for the tremor were also normalized to the [0, 10] range using the hyperbolic tangent method similar as before, adjusting the values of  $\mu$  and  $\sigma$  accordingly for this case.

Figure 15.4 shows the population statistics for the five selected features. These distributions were obtained for a subset of e-BioSign database [21] which is comprised of 70 users signing on a Wacom STU-530 tablet. This database was comprised of eight genuine signatures per user and six skilled forgeries collected in two different sessions with a time gap of at least three weeks. The skilled forgeries signatures were performed in two different ways, in the first session forgers were allowed to visualize a recording of the dynamic realization of the signature to forge for a few times, while in the second session, a paper with the image of the signatures to forge was placed over the device and they can trace the lines to perform the forgery.

For each of the graphs shown in Fig. 15.4 we also show the position of the eight genuine and 6 forgery signatures with vertical lines for one of the users (as an example of use) inside the population distribution. This can help the FHEs to analyze if the questioned signatures are within the distribution of genuine signatures in general and for that user in particular. In the examples shown, it can be seen that for that particular user genuine and forgery signatures are well separated, especially for the average velocity, average acceleration and time fluency.

In a future release of e-BioSign Tool, in order to provide the FHEs with statistical support for these five parameters, apart from plotting population distributions, a Likelihood Ratio (LR) will be also provided for each parameter, which would be calculated from a matching score using the signatures under analysis, and using a LR model trained on a reference database (or a subset of it).

#### **15.3.4 Module 4: Automatic Signature Verification**

An additional functionality of e-BioSign Tool is Module 4, which is an automatic signature matcher. With this matcher, a questioned signature can be compared to a number of known or reference signatures to obtain a matching score.

The automatic signature matcher is based on a selection of time functions extracted from the signature such as the X and Y coordinates, the pressure, velocity, etc. Then, Dynamic Time Warping (DTW) is used to compare the similarity between the selected time functions extracted from the signatures [22]. The matching scores are then converted to likelihood ratios (LR) as commonly done in the forensic community [23–25]. In this case, e-BioSign database is used to train a likelihood ratio model following a logistic regression approach. As stated in [26], the aim of logistic regression is to obtain an affine transformation of an input dataset to generate an output value which optimizes an objective function. It may be demonstrated that logistic regression leads to LR values with low calibration loss from a training score set. In our case, Bosaris Toolkit [27] has been used.

In a first release of this tool, a person-generic LR model is considered, therefore obtaining same-source scores and different-source scores (in this last case comparing genuine signatures with skilled forgeries signatures). In future releases, functionalities to select a group of specific users by age, hand they use to sign, complexity of the signature, etc., will be provided in order to obtain more meaningful LR models for the particular case to study.

## 15.4 Case Study

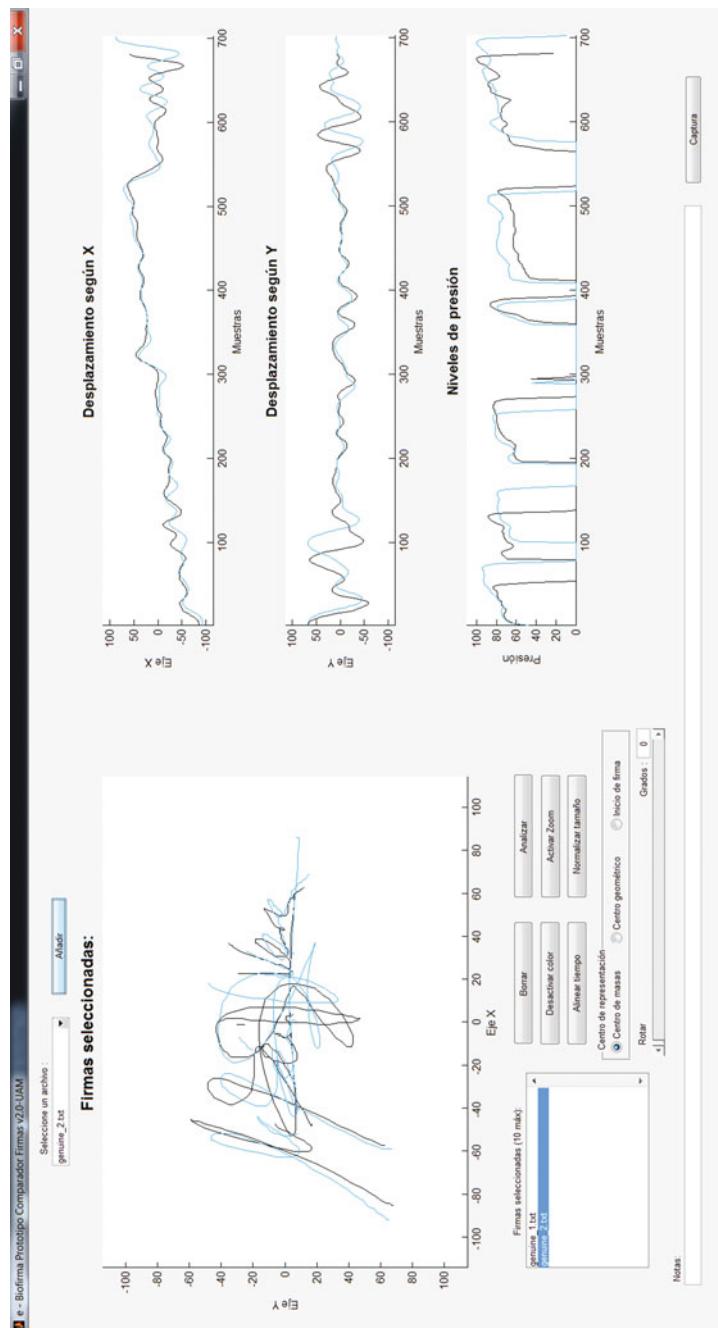
In this section we describe a case study for e-BioSign Tool. In this example, we analyze two genuine signatures from a given user and two skilled forgeries. Signatures are acquired using a Wacom STU-530 device. The two skilled forgeries are performed in different ways, for the first one the forger places a printed copy with the signature to forge on top of the device and traces the lines to perform the forgery (for the remaining of the analysis we refer to it as *traced forgery*). For the second one, the forger carries out the forgery directly on the device after practicing a few times on a paper (we refer to it as *natural forgery*). In both cases, the forger is allowed to visualize a recording of the dynamic realization of the signature to forge for a few times in order to obtain good quality forgeries.

This case example is not meant to be an exhaustive forensic analysis of the signatures as the one that would be done by a FHE. The purpose is to show with real examples how the functionalities of e-BioSign Tool can be used to analyze and measure the similarities and dissimilarities of dynamic signatures.

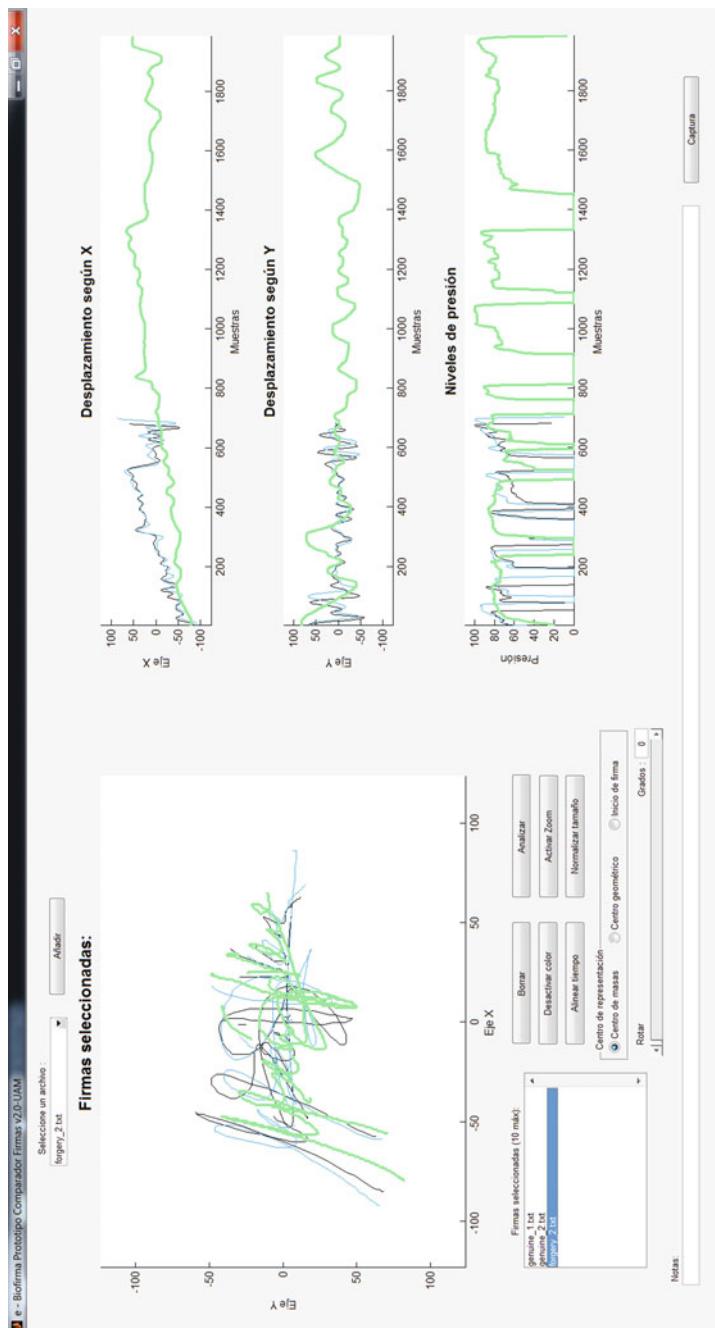
First, signatures are loaded for their analysis on e-BioSign Tool. Figure 15.5 shows the two genuine signatures (Module 1). They are both normalized in the spatial domain by the center of mass. Both signatures have a similar time duration and a striking similar shape of their X, Y and pressure temporal values, even not being aligned in time. This shows that this user is very consistent performing his signature. There are some parts (beginning and end) with some time misalignments which can be better analyzed with functionalities present in Module 3 of the tool.

Figure 15.6 shows the same two genuine signatures and the traced forgery signature. As can be seen, the traced forgery has a much longer time duration with more than double number of samples (almost 2,000 samples which correspond to 10 s as the Wacom STU-530 was configured to sample at 200 Hz, compared to less than 700 samples for both genuine signatures), which can be a sign that the signature has been performed by a forger. Figure 15.7 shows the three signatures aligned in time, with the forgery having longer strokes in general (particularly at the flourish) and also shorter duration pen-ups.

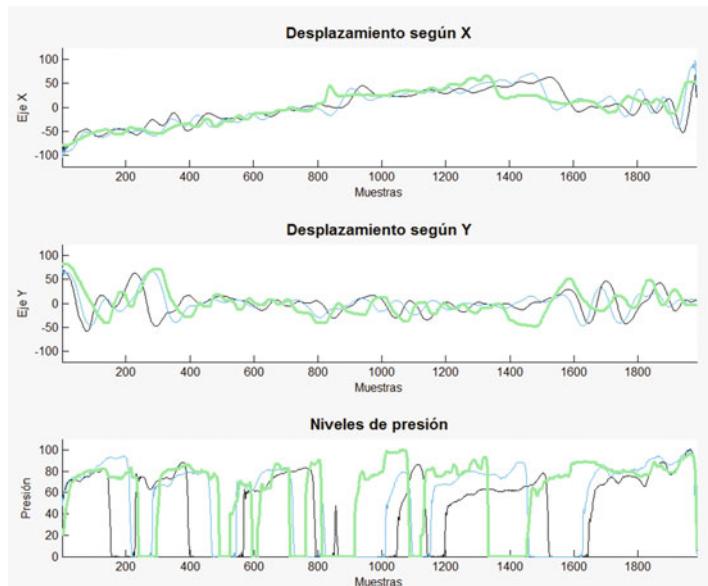
Figure 15.8 shows the same two genuine signatures as previously and the natural forgery signature. As can be seen, this forgery contains a similar number of time samples compared to the two genuine and in general shows a similar shape. We will describe the subsequent analysis steps using these three signatures with the remaining modules of the tool.



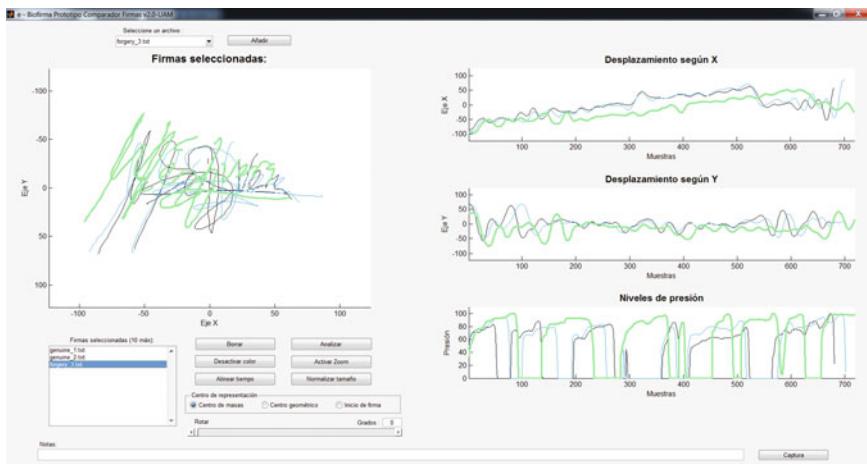
**Fig. 15.5** Examples of two genuine signatures from the same user (reference known samples)



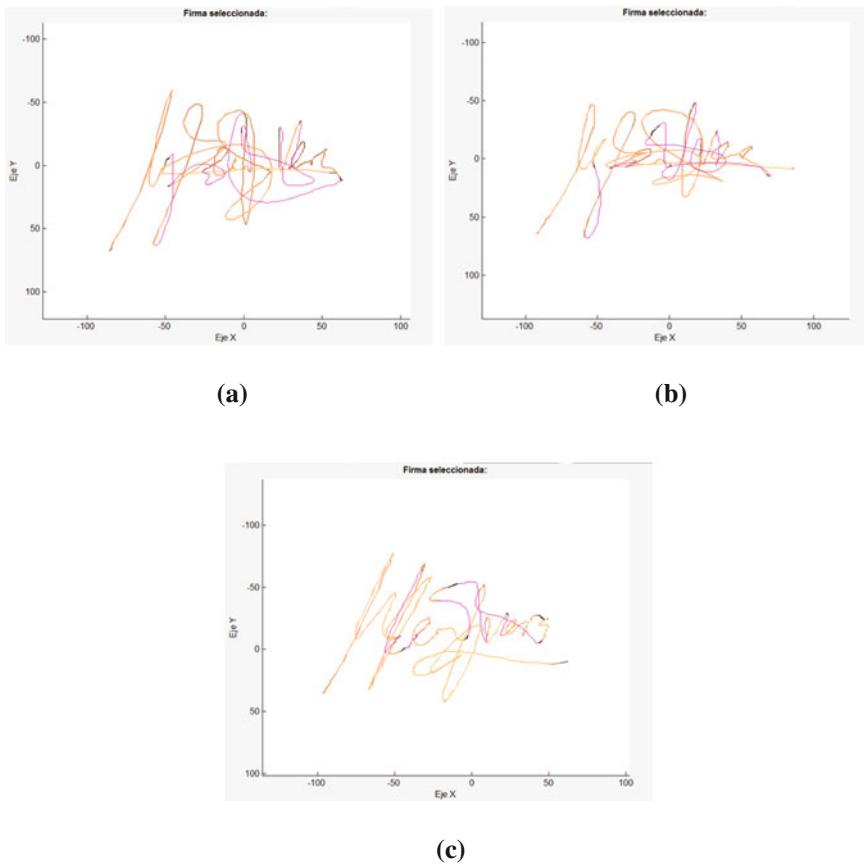
**Fig. 15.6** Examples of two genuine signatures and one traced forgery signature



**Fig. 15.7** Same signals as Fig. 15.6 but normalized in time



**Fig. 15.8** Examples of two genuine signatures and one natural forgery



**Fig. 15.9** Module 2. Spatial information (X and Y coordinates) also showing the pen-up information for **a** and **b** genuine signatures, and **c** forged signature

Module 2 of e-bioSign Tool is used to carry out a comparative general analysis of the signatures and to select some strokes for further analysis. Figure 15.9 shows the spatial information (X and Y coordinates) for the three signatures under analysis (the two genuine and the natural forgery). Also, the pen-up information is shown in pink color. Having a general look at the three signatures, there are some features quite different between the two genuine signatures and the forgery. For example, some letters have different shapes, or the last line of the flourish is placed just below the name letters for the genuine signatures, while it is placed much lower for the forgery. Additionally, Module 2 allows to visualize the direction of the trajectory of the signatures, indicating the beginning and end and connecting the different strokes as can be visualized in Fig. 15.10. With this functionality is easy to check that the signatures under analysis follow the same trajectory pattern.

**Fig. 15.10** Module 2.  
Trajectory information for  
one of the genuine signatures

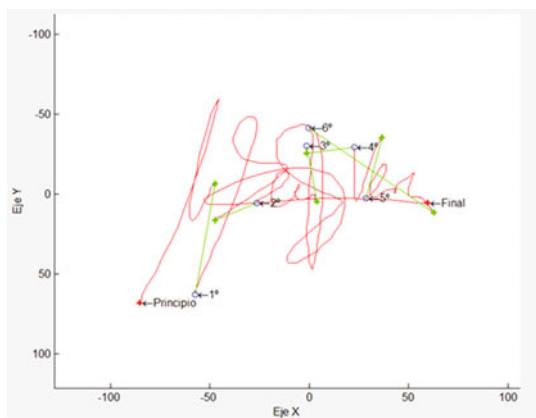
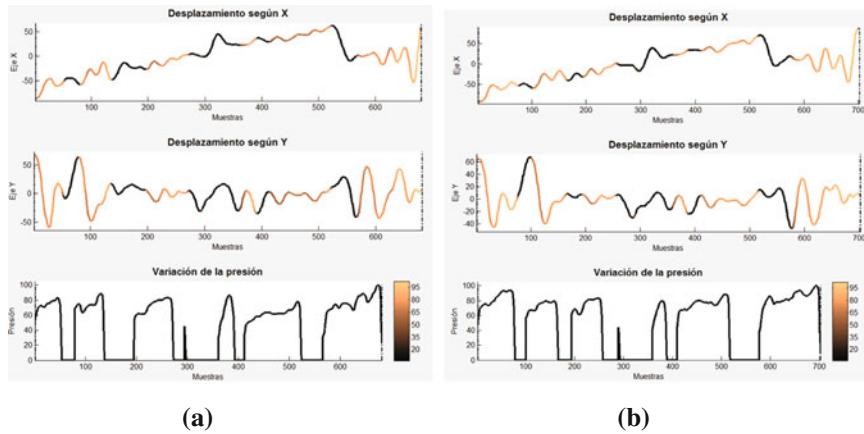


Figure 15.11 shows the X and Y coordinates and pressure information along the time axis for the same three signatures. Having a close analysis to Fig. 15.11a, b, the two genuine signatures have very similar temporal information, following the same pattern for the X, Y and pressure axes. However, although there are some general similarities with the remaining signature (c), a close analysis reveals many differences, such as: the accent mark (samples around 300 for the two genuine and around 400 for the forgery) has a longer duration and higher pressure value compared to the two genuine; in general the pressure information presents higher values (for example see the last stroke). Also, the last part of the forged signature (Fig. 15.11c) (samples between 500 and 650) presents a completely different pressure pattern compared to the two genuine. Moreover, there are some strokes with very different pattern for the X and Y coordinates (the second stroke, for example). Some of these strokes of the signatures are selected for further analysis using Module 3 of the tool.

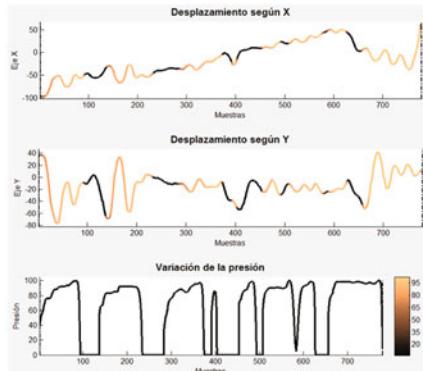
Module 3 of e-bioSign Tool is now used to carry out a more detailed analysis for some of the strokes selected with Module 2. Figure 15.12 shows the spatial and temporal information (X, Y, pressure, velocity and acceleration) that can be visualized in Module 3 for a comparison of the flourish stroke for the two genuine signatures (a) and one of the genuines and the forgery (b). These strokes have been aligned spatially for a better comparison with a shift in X and Y coordinates and rotation (for which the tool incorporates semi-automatic functionalities).

Figure 15.12a shows that even though the two flourish strokes do not have exactly the same spatial shape, the time functions follow a very similar pattern, with only a small difference in Y coordinates for samples between 60 and 90. It is possible to see how the velocity and acceleration functions follow a very similar pattern with just small time misalignments. On the other hand, Fig. 15.12b shows the same for one genuine signature and the forgery under analysis. In this case the differences between the two signatures are very significant both in spatial shape and for all the time functions.



(a)

(b)

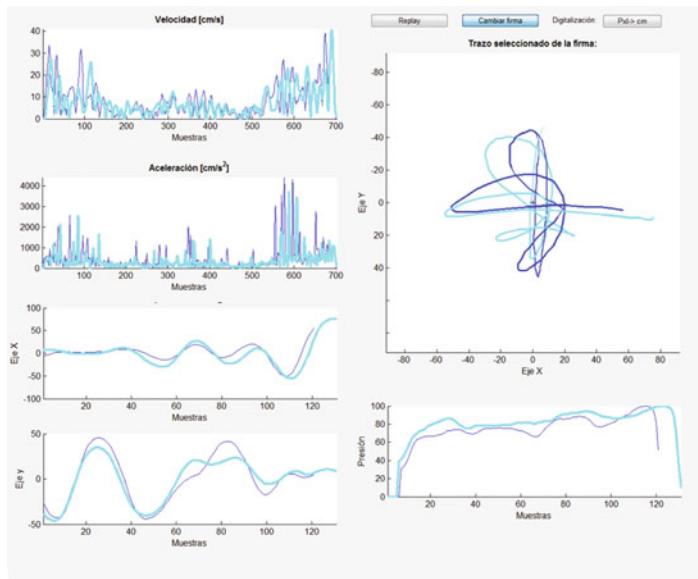


(c)

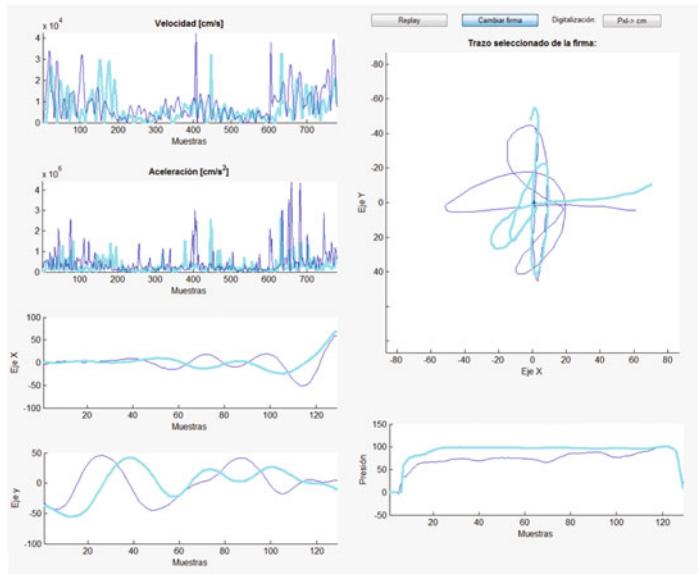
**Fig. 15.11** Module 2. Temporal X, Y and pressure coordinates information for **a** and **b** genuine signatures, and **c** forged signature

For completeness, we also show in Fig. 15.13 the pen-up information that can be visualized (in red) with Module 3. In the case shown, following the trajectory of the pen-up information we can observe that the accent mark can be seen that is performed following a different trajectory for the genuine signature (a) and the forgery. This difference, although small, can be very significant for a FHE to give an opinion on these signatures.

Module 3 also allows to carry out a statistical analysis of five parameters extracted automatically. The values of these parameters for the four signatures considered in this case example are shown in Table 15.1. It is worth noting that in a real case analysis carried out by a FHE, a larger number of reference or known signatures should be considered (if possible) in order to carry out the analysis with higher statistical

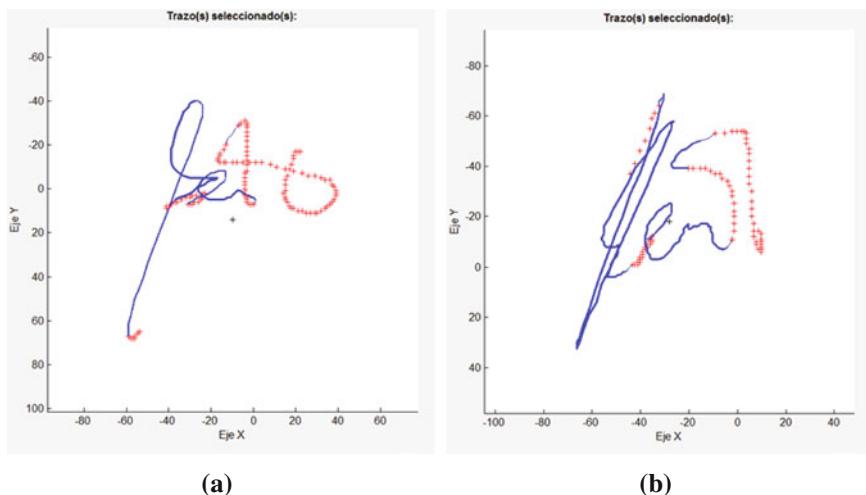


(a)



(b)

**Fig. 15.12** Module 3. Comparison of the flourish stroke with their spatial and time information (X, Y, pressure and acceleration). **a** Genuine-genuine comparison and **b** genuine-forgery comparison



**Fig. 15.13** Module 3. Comparison of a stroke for one **a** genuine and **b** the natural forgery signatures with their spatial information. Also pen-up samples are shown for a more complete analysis

**Table 15.1** Values of the five statistical parameters obtained for the four signatures under analysis. Time fluency and spatial tremor are normalized to the range [0–10]

	Genuine 1	Genuine 2	Traced forgery	Natural forgery
Duration (s)	3,4	3,5	9,9	3,9
Average velocity (cm/s)	8,4	7,7	2,6	6,4
Average acceleration (cm/s <sup>2</sup> )	302,9	273,6	64,7	228,4
Time fluency	6,3	5,8	2,8	5,2
Spatial tremor	3,8	3,7	5,2	3,8

significance. The values shown in Table 15.1 can be visualized on top of the population distributions as per Fig. 15.4. As can be seen in the table, the two genuine signatures obtain similar values for the five parameters considered. Also, the natural forgery obtains similar values for these five parameters with just a bit longer duration, lower velocity, lower acceleration, lower fluency value and similar tremor value, but not very significant difference with just two reference signatures to compare. On the other hand, the traced forgery obtains very different values for the five parameters, which would indicate with a higher confidence this signature was produced by a different person.

## 15.5 Conclusions

This paper has described a new tool e-BioSign specifically designed to carry out dynamic signature forensic analysis and give scientific support to FHEs. This tool allows to analyse the traditional information taken into account by FHEs to carry out the analysis for paper-based signatures, and also permits to exploit the dynamic information contained in signatures acquired from digitizing tablets which can be very discriminative. As mentioned in Sect. 15.2 it is very difficult to perfectly forge a signature, good forgeries normally either have a similar appearance and shape but are not written fluently, or the opposite. With the analysis of both spatial and time information for dynamic signatures using e-BioSign, we believe it will be easier for FHEs to detect forged signatures. Additionally, the tool incorporates two important functionalities, the first is the provision of statistical support to the analysis by including population statistics for genuine and forged signatures for some selected features (signature duration, average velocity and speed, time fluency, and spatial tremor), and the second is the incorporation of an automatic signature matcher which can provide a matching score between the known and questioned signatures under analysis.

For future work, this tool will be provided to FHEs for testing and a new version will be released based on their suggestions. Also, LR values will be provided for the signatures under analysis for the five statistical parameters considered. FHEs will be able to select a particular subset of the reference dataset (based on the gender, hand used for writing, age, etc.) in order to obtain more meaningful LR values for the particular cases under analysis.

**Acknowledgements** This work has been supported by project CogniMetrics TEC2015-70627-R (MINECO/FEDER) and in part by Cecabank e-BioFirma2 Contract.

## References

1. Huber RA, Headrick AM (1999) Handwriting identification: facts and fundamentals. CRC Press
2. Found B, Dick D, Rogers D (1994) The structure of forensic handwriting and signature comparisons. *Int J Speech Lang Law Forensic Linguist* 1:183–196
3. Found B, Rogers D (1999) Documentation of forensic handwriting comparison and identification method: a modular approach. *J Forensic Doc Examination* 12:1–68
4. de la Uz Jimenez J (2013) Manual de Grafistica. Tirant Lo Blanch
5. Vinals F (2008) Boletin Electronico num. 8, tech. rep., Instituto de Ciencias del Grafismo
6. Galende-Diaz JC, Gomez-Barajas C (2008) En busca de la falsedad documental: La figura del perito caligrafico. In: Proceedings of the VII Jornadas Cientificas Sobre Documentacion Contemporanea (1868–2008), pp 193–231, Univ. Complutense de Madrid
7. Alewijnse L (2013) Forensic signature examination. In: Tutorial at international workshop on automated forensic handwriting analysis (AFHA)
8. Harralson HH (2012) Forensic document examination of electronically captured signatures. *Digit Evid Elec Signat L Rev* 9:67–73

9. Ahmad SMS, Ling LY, Anwar RM, Faudzi MA, Shakil A (2013) Analysis of the effects and relationship of perceived handwritten signature's size, graphical complexity, and legibility with dynamic parameters for forged and genuine samples. *J Forensic Sci* 58(3):724–731
10. Mohammed MCLA, Found B, Rogers D (2011) The dynamic character of disguise behavior for text-based, mixed, and stylized signatures. *J Forensic Sci* 56
11. Franke K (2009) Analysis of authentic signatures and forgeries. In: Geradts Z, Franke K, Veenman C (eds) Computational forensics. Lecture notes in computer science, vol 5718. Springer, Berlin, pp 150–164
12. Fierrez J, Ortega-Garcia J (2008) Handbook of biometrics, ch. On-line signature verification. Springer, pp 189–209
13. Fierrez J, Ortega-Garcia J, Ramos D, Gonzalez-Rodriguez J (2007) HMM-based on-line signature verification: feature extraction and signature modeling. *Pattern Recognit Lett* 28:2325–2334
14. Martinez-Diaz M, Fierrez J, Krish RP, Galbally J (2014) Mobile signature verification: feature robustness and performance comparison. *IET Biometrics* 3:267–277
15. Houmani N, Mayoue A, Garcia-Salicetti S, Dorizzi B, Khalil M, Moustafa M, Abbas H, Muramatsu D, Yanikoglu B, Kholmatov A, Martinez-Diaz M, Fierrez J, Ortega-Garcia J, Alcob JR, Fabregas J, Faundez-Zanuy M, Pascual-Gaspar J, Cardeoso-Payo V, Vivaracho-Pascual C (2012) Biosecure signature evaluation campaign (BSEC2009): evaluating online signature algorithms depending on the quality of signatures. *Pattern Recognit* 45:993–1003
16. Martinez-Diaz M, Fierrez J, Hangai S (2009) Encyclopedia of biometrics, ch. Signature matching. Springer
17. Malik M, Liwicki M, Dengel A, Found B (2013) Man vs. machine: a comparative analysis for forensic signature verification. In: Proceedings of the biennial conference of the international graphonomics society
18. Vera-Rodriguez R, Fierrez J, Ortega-Garcia J, Acien A, Tolosana R (2015) e-BioSign tool: towards scientific assessment of dynamic signatures under forensic conditions. In: Proceedings of the IEEE international conference on biometrics: theory, applications and systems (BTAS), (Washington), Sept 2015
19. Richiardi J, Katabdar H, Drygajlo A (2005) Local and global feature selection for on-line signature verification. In: Proceedings of the eighth international conference on document analysis and recognition, vol 2, pp 625–629, Aug 2005
20. Jain A, Nandakumar K, Ross A (2005) Score normalization in multimodal biometric systems. *Pattern Recognit* 38(12):2270–2285
21. Vera-Rodriguez R, Tolosana R, Ortega-Garcia J, Fierrez J (2015) e-BioSign: stylus- and finger-input multi-device database for dynamic signature recognition. In: Proceedings of the 3rd international workshop on biometrics and forensics (IWBFF), (Norway). IEEE Press, March 2015
22. Tolosana R, Vera-Rodriguez R, Ortega-Garcia J, Fierrez J (2015) Preprocessing and feature selection for improved sensor interoperability in online biometric signature verification. *IEEE Access* 3:478–489
23. Gonzalez-Rodriguez J, Fierrez-Aguilar J, Ramos-Castro D, Ortega-Garcia J (2005) Bayesian analysis of fingerprint, face and signature evidences with automatic biometric systems. *Forensic Sci Int* 155:126–140
24. Morrison GS (2011) Measuring the validity and reliability of forensic likelihood-ratio systems. *Sci Justice* 51(3):91–98
25. Ramos D, Gonzalez-Rodriguez J, Zadora G, Aitken C (2013) Information-theoretical assessment of the performance of likelihood ratio computation methods. *J Forensic Sci* 58(6):1503–1518
26. Gonzalez-Rodriguez J, Rose P, Ramos D, Toledano DT, Ortega-Garcia J (2007) Emulating DNA: rigorous quantification of evidential weight in transparent and testable forensic speaker recognition. *IEEE Trans Audio Speech Lang Process* 15:2104–2115
27. Brummer N, de Villiers E (2011) The BOSARIS toolkit user guide: theory, algorithms and code for binary classifier score processing. Tech. Rep, Agnitio

**Part V**  
**Ethical and Legal Issues**

# Chapter 16

## Ethics and Policy of Forensic Biometrics

Emilio Mordini

**Abstract** Ethical issues raised by forensic biometrics partly overlap with general ethical implications of biometrics. They include issues related to collecting, processing, and storing, personal data, privacy, medical information, and respect for body integrity, risks of misuse and subversive use, and respect for human dignity. There are, however, also ethical issues specifically raised by forensic biometrics. One of them is particularly intriguing. It concerns the nature of biometric evidence and to what extent biometric findings could be accepted as an evidence in court. At a first glance, this problem could seem purely legal, without major ethical implications. Yet, at a deeper analysis, it turns out to have significant ethical components. I will focus on them and on some recent policy developments in this field.

Ethical issues raised by forensic biometrics partly overlap with general ethical implications of biometrics [1]. They include issues related to collecting, processing, and storing, personal data [2], privacy [3], medical information and respect for body integrity [4], risks of misuse and subversive use [5], respect for human dignity [6]. There are, however, also ethical issues specifically raised by forensic biometrics. One of them is particularly intriguing. It concerns the nature of biometric evidence and to what extent biometric findings could be accepted as an evidence in court. At a first glance, this problem could seem purely legal, without major ethical implications. Yet, at a deeper analysis, it turns out having significant ethical components. I will focus on them and on some recent policy developments in this field.

---

E. Mordini (✉)

Responsible Technology SAS, 12 rue de la Chaussee d'Antin, 75009 Paris, France  
e-mail: emilio.mordini@rtexpert.com

## 16.1 Biometric Evidence in Law Enforcement and Criminal Justice

The history of biometrics largely coincides with the history of its forensic applications, which include applications used for crime prevention, crime investigation, and administration of justice. Also prejudicial scientific discourses applied to criminology, such as physiognomy and phrenology, owed mostly to biometrics their temporary good scientific reputation (incidentally, this is likely to be one the reasons why biometrics have had later on such a bad press among human rights advocates and ethicists).

In early 1900s, law enforcement agencies started collecting fingerprints from convicted criminals and suspected individuals, and rather soon, they created vast fingerprint libraries. In parallel, the refinement of methods for latent fingerprints retrieval made biometric identification a fundamental tool for investigation at the crime scene. Finally, fingerprints (and later on, palm prints and impressions of bare soles) were accepted as admissible evidence into courts in most jurisdictions. Since then, biometrics (chiefly fingerprint) have been increasingly used for revealing or confirming the identity of people involved in the criminal investigation or in the judicial process (including victims); for establishing whether an accused person was at the scene of a crime or used an item that was used in perpetration of the crime; whether other person was at the scene of the crime or touched an item that was used in perpetration of the crime; whether an alleged victim was at some place that was consistent with the prosecution; whether a witness was at a place it is claimed he was. Biometrics have been also used to impeach the credibility or integrity of a suspect or a witness or a victim, based upon criminal records that show his prior history. Into courts, biometrics provided also a certain, additional, “scientific objectivity” to “traditional” (e.g., eyewitness) circumstantial evidence.

The way of assessing of biometric findings varies among jurisdictions, notably between common law and civil law systems. There are, however, also important similarities. As it happens with most scientific evidences,<sup>1</sup> usually biometric findings are not considered immediate evidence, but they have to be presented in court by one or more qualified experts who provide their testimony. Their opinions are assessed by the judge, who takes the final decision about whether biometric findings are admissible as an evidence in that specific case. In some jurisdictions (especially in the civil law area) the law establishes the minimum number of biometric details that should be analyzed in order to produce a positive identification; in other jurisdictions (especially in the common law area), the judge simply assesses experts' qualification and credibility. In both cases, the judge asks the expert to state whether a given biometric finding allows recognizing an individual, say, whether it matches with any recorded biometrics collected in the past, or with biometrics

---

<sup>1</sup>“Scientific evidence” in court is an evidence that is inferred from a known fact by using the scientific method.

gathered from any relevant individual involved in the judicial procedure. Experts are expected to answer yes or no.

In conclusion, the main events occurring under the heading of forensic biometrics include (1) the preventive activity of police and other law enforcement agencies, which collect biometrics from convicted criminals and suspected individuals in order to monitor them and prevent crime; (2) if a crime occurs, biometrics could be collected on the crime scene and in other relevant contexts, in order to ascertain the identify of supposed victims, suspected criminals, and alleged witnesses; (3) finally biometric findings are usually brought into court where the judge (after hearing experts and parties) decides whether, and to what extent, these findings could be considered an evidence and will contribute to form the judicial decision.

## 16.2 Digital Biometrics

This scenario is changing with the arrival of new digital biometrics. New digital biometrics (automated biometrics) entered into the market in the late 1970s, chiefly thanks to the development of new sensors, capable of capturing a vast array of different inputs. Sensors are devices that turn various inputs—generated by the interaction between them and physical object—into electrical outputs, whose electrical magnitude is proportional to the magnitude of the initial signal. Through the repetitive measurement of the electric output at certain intervals of time, the magnitude of the voltage is turned into a proportional number, which is further encoded into a binary number, or a gray code, or still in other ways. This allows representing bodily attributes (including those that were traditionally considered qualitative, such as skin color, or walking style) as measurable physical properties.<sup>2</sup> Two or more individuals could be subsequently compared in quantitative terms to ascertain whether they share the same properties as far as specific bodily attributes are concerned. From such a comparison, one could deduce whether two (apparent) individuals are actually the same individual, whose attributes were captured in different fractions of time, say, one could identify the individual (in the literally sense of assessing whether there is only one individual, considered under different accounts, instead of two or more distinct individuals). The main conceptual differences between pre-digital and digital biometrics is that digitalization allows performing the comparison (1) always in quantitative terms, avoiding qualitative assessment; (2) automatically, by exploiting ad hoc computer algorithms; (3) on a huge number of bodily attributes, unthinkable in the pre-digital era.

---

<sup>2</sup>Physical properties are discrete elements that can be put in bi-univocal correspondence with a set of numbers. There are seven base physical properties, Length, Mass, Time, Electric Current, Temperature, Amount of Substance, and Luminous Intensity. Biometric sensors measure one or more of these properties.

Absolute identity is logically and practically impossible. It does not take a Heraclitus, to realize that all bodily attributes, even the most stable and persistent, change (maybe slightly) over time, and conditions of data collection change, as well as sensor sensitivity. In other words, biometric features vary, sometime degrade by themselves; moreover, they also vary in the way in which they are presented to sensors. Finally, sensor precision is limited and may vary at different sites and according to different conditions of usage. This implies that it is impossible that two or more data sets captured by a sensor and turned into digits might exactly match, digit by digit. In other words, digital biometric recognition is always by approximation, and it inevitably implies some errors. This could be mitigated by creating, storing and comparing normalized biometric samples, called “templates”. Yet errors cannot be eliminated because of their systematic nature. Accordingly, one of the main features to be considered in any given biometric system is always its error rate. One speaks of “false rejection”, or “false negative”, when the system fails to recognize someone; and “false acceptance”, or “false positive”, when the system recognizes someone erroneously. The ratio between false negatives and the total examined population is called “specificity”; the ratio between false positives and the total examined population is called “sensitivity”. Specificity indicates the system ability to discriminate between different individuals; sensitivity indicates the system ability to detect all searched individuals. In principle, sensitivity and specificity are independent, in practice; there is a tradeoff, such that they are almost inversely proportional to one another. Also pre-digital biometrics could not avoid systematic errors, but their degree of uncertainty have been never studied in rigorous probabilistic terms, as it is today with new digital biometrics.

If only biometric features perfectly matching, point by point, led to recognition, the system would never recognize anyone because such a perfect identity can never be obtained. Consequently, the system tolerance level must be tuned, that is to say, the system must be “told” within what confidence interval it should consider two different biometric sets as though they were identical. In practice, this means that engineers have to decide when the gap between two biometric series can be considered negligible. Rather intuitively, the narrower the confidence interval is, the higher are the probabilities that the recognition is accurate, say, there will be less false positive. Yet, with a too narrow confidence interval, the system would increase the risk of failing to recognize the same individual presented twice, because the gap between his biometric features, taken in different moments and circumstances, could fall outside the confidence interval. Larger confidence intervals would mitigate this risk, but they would increase false acceptance (someone confused with someone else). It is important to emphasize that there is not a “right” confidence interval, but it depends on the specific context and application. Ultimately, the decision on the confidence interval depends on the policy adopted by system administrators.

### 16.3 Probabilistic Biometrics in Court

Past debates on forensic biometrics chiefly focused on conditions of admissibility of biometric evidence and qualifications of expert witnesses [7]. Today, legal experts, jurists, and scholars find increasingly problematic the probabilistic nature of biometric identification. I will only hint at the main terms of this debate, which is richer and full of nuances, because it is not the focus of my article, although it is a necessary premise to my argument.

Current debate on probabilistic biometrics in court is driven by two main facts. The first is a 1993 decision of the US Supreme Court,<sup>3</sup> which introduced new standards for scientific evidence. Although this decision was directly relevant only to the US legal system, its philosophy is considered a benchmark and has deeply influenced many other legal systems, including those belonging to the civil law area. In this sentence the Supreme Court ruled that a scientific evidence could be admitted into court only if it respects the following five criteria: (1) the evidence “can be (and has been) tested” using a scientific method; (2) such a method has “been subjected to peer review and publication”; (3) it is known the “potential rate of error” of the method in question; (4) the “existence and maintenance of standards controlling the technique’s operation”; (5) the “general acceptance” of the technique within the relevant scientific community. Would biometric evidence survive *Daubert* criteria? This has been the core discussion, developed first within the US forensic community and then among scholars and legal experts belonging to other jurisdictions and legal schools. In particular, some scholars [8, 9] have argued that the scientific soundness and reliability of expert-testimony-based biometrics is definitely suboptimal. This was emphasized also by a few clamorous cases of error, for instance the one that occurred in the wake of Madrid terrorist attack [10].

The second driver of the current debate is a technological driver. With the rapid development of new digital biometrics (and a vast array of new biometric applications, exploiting modalities whose existence were not even imaginable in the pre-digital era) the issue of probabilistic identification has become paramount. Scholars [11, 12] have advocated a more mindful, and scientifically refined, approach to forensic biometric by adopting a probabilistic mindset. For instance, Champod et al. [13] have suggested that, instead of posing the naïf question whether a biometric feature is *identical* to another, a biometric examiner should ask to himself “*Given the detail that has been revealed and the comparison that has been made, what inference might be drawn in relation to the propositions that I have set out to consider?*” (p. 25).

The two discussions are clearly intertwined, as *Daubert* criteria should be applied to new digital biometrics, and digital biometrics should come to terms to more stringent legal criteria for admissibility of scientific evidence [14]. The interplay between the notion of forensic evidence and the probabilistic nature of

---

<sup>3</sup>Daubert v. Merrell Dow Pharm. Inc., 727 F. Supp. 570, 575 (S.D. Cal. 1989).

biometrics is the core of the next chapters, in which I will present my central argument.

## 16.4 Law Enforcement and Administration of Justice: Two Different Missions

In the first chapter of this paper, I listed the main activities included under the heading of forensic biometrics, say, anticipation of crime, criminal investigation, and administration of justice. Although practically and theoretically distinct, these three main activities belong to a unique cycle, whose languages must be interoperable if the cycle works. Two of these activities (anticipation of crime and criminal investigation) are carried out by law enforcement agencies; the third (administration of justice) is up to the judicial system. Law enforcement and judicial systems are thus distinct but interoperable. Their missions are different, yet in democratic and liberal societies, they have been developed rather harmonically in order to compensate each other.

Law enforcement's overarching mission is to prevent, and, in case, repress crime. To be sure, crime prevention includes many other activities, which are not up to law enforcement agencies, such as education, social policies, urban management, and so. Yet, there is an important aspect of crime prevention that is up to law enforcement authorities, that is to say, pre-crime (anticipative) investigation. Anticipative investigation [15] aims to prevent crime by identifying potential criminals and future criminal events, and by taking action accordingly. Ultimately, anticipative investigation is based on the assessment and management of risks, its goal is to prevent risks or, at least, to mitigate them. While anticipative investigation is proactive, criminal investigation is reactive, it follows crime and it aims to ascertain facts, identify, and prove the guilt of one or more suspects. Ultimately, its goal is to repress crime by bringing the suspects into the judicial system. Overall, anticipative and criminal investigations do not aim to discover truth, to pursue justice, or to achieve fairness—at least as their primary goal—rather they simply aim to reduce the rate of crimes in society. Unpleasant though it may sound, law enforcement must be ruled by the “principle of suspicion” (of course opportunely mitigated), that is to say, anybody could be in principle suspected.

The judicial system aims instead to administer the justice. What is “justice”? Lay citizens often think that making justice means to discover factual truth about a crime or a civil case. Yet legal scholars know that in court factual truth is hardly at stake. Actually, the business of a court of justice is not to discover the truth, as Oxford professor and distinguished English legal scholar, Sir Frederick Pollock (1845–1937) put it,

Its real business is to pronounce upon the justice of particular claims, and incidentally to test the truth of assertions of fact made in support of the claim in law, provided that those assertions are relevant in law to the establishment of the desired conclusions [16]

In other words, the goal of the judicial procedure is not to ascertain the “truth” in trivial sense, rather to identify the “legal truth”. The “legal truth” is the truth as it emerges from the judicial hearing [17]. It may or may not corresponds to the totality of facts. Even when it does not correspond to facts, the “legal truth” is still practically effective, say, it produces effects, as it is showed, for instance, by the principle that a defendant cannot be tried twice on the same charge following a legitimate court decision. The notion of “legal truth” is better understood by considering it in parallel with another notion, which is foundational for the whole Western legal system, the principle of presumption of innocence. Presumption of innocence is a legal principle adopted in criminal cases. In criminal cases, the gap between “factual” and “legal” truths could produce unacceptable outcomes; presumption of innocence mitigates this risk. Its original formulation dates back to Roman law and reads “*Ei incumbit probatio qui dicit, non qui negat*” (the burden of proof is on he who declares, not on he who denies), also “*in dubio pro reo*” (when in doubt, for the accused).<sup>4</sup>

In modern terms, presumption of innocence is usually expressed by saying that a defendant is innocent until proven guilty,

In a criminal case the truth of facts against the accused must be established “beyond a reasonable doubt,” and in certain civil cases, e.g., where punitive damages may be awarded for fraud, the truth of facts against the defendant must usually be shown by “a clear and convincing preponderance of the evidence”. The more that is at stake, e.g., criminal blame, or punitive damages, the higher the standard of proof. In an ordinary civil case involving an ordinary claim for damages, the facts against the defendant need only be shown by a “balance of probabilities”, a significantly lower standard of truth. Thus, depending on the relevant standard of truth, the very same evidence would warrant a finding of truth in one type of case but not in another. Thus, truth varies with standards of proof, and standards of proof vary with what is at stake. Yet, as indicated, there are good reasons for these variations in standards of truth. In criminal cases for example, we accept a higher risk of erroneous acquittals in order to minimize the risk of erroneous convictions. Moreover, this may have the effect of increasing the total number of erroneous verdicts. Our tolerance for the risk of divergence, here, goes up the more that is at stake [17]

Presumption of innocence, which was initially a legal procedural rule, has become one of the main principles of Western legal systems, a constitutional principle in many democracies,<sup>5</sup> and a fundamental human right. The Universal Declaration of Human Rights, Article 11, reads

Everyone charged with a penal offence has the right to be presumed innocent until proved guilty according to law in a public trial at which he has had all the guarantees necessary for his defense

Similarly Article 6.2 of the Convention for the Protection of Human Rights and Fundamental Freedoms of the Council of Europe reads

<sup>4</sup>“*Digesta seu Pandectae 22.3.2*” (<http://webu2.upmf-grenoble.fr/Haiti/Cours/Ak/Corpus/d-22.htm>).

<sup>5</sup>Presumption of innocence is mentioned for instance by Italian, French, German, Brazilian, Canadian, Russian constitutions. It is not explicitly mentioned by the US Constitution, yet there is a consensus that it follows from the 5th, 6th, and 14th amendments.

Everyone charged with a criminal offence shall be presumed innocent until proved guilty according to law

Finally, the same principle is iterated by Article 48 of the Charter of Fundamental Rights of the European Union, which reads

Everyone who has been charged shall be presumed innocent until proved guilty according to law

In conclusion, law enforcement and administration of justice have different (although congruent) missions and follow two distinct logics; the former is ruled by the *principle of suspicion*, while the latter is governed by the *presumption of innocence*. In democracies, these two different perspectives are mutually consistent, because systematic errors made by the former are compensated by systematic errors made by the latter. In other words, the law enforcement system shows a natural tendency to minimize false negative (no criminal should get off scot-free), although this could rise false positive (an innocent could be unjustly accused); the judicial system shows the opposite behavior, say, it aims to keep false positive as lower as possible (no innocent should be wrongly condemned), although it could increase false negative (it could happen that a guilty individual is acquitted). If the two systems work properly, and each one of them respects its mission, the overall cycle functions rather well because opposite errors mutually compensate. The law enforcement tight mesh net may capture also innocents, but they will be eventually released by the judicial wider mesh net (of course, it is always possible that a criminal escapes and an innocent is condemned, I am just arguing that the two systems are theoretically complementary).

## 16.5 Probabilistic Biometrics and Presumption of Innocence

I have illustrated the probabilistic nature of biometric recognition and the way in which biometric applications deal with it, by tuning the tolerance of the system according to different user requirements. I would like now to pose a question: could law enforcement and judicial systems adopt the same degree of tolerance? Say, are user requirements the same in the two systems, which jointly form the forensic cycle? This question could seem purely technical; on the contrary, I argue that it is the main ethical question raised by forensic biometrics.

Suppose that positive identification is the critical element to determine who committed a crime (e.g., detecting and identifying terrorists who attacked a metro station). Suppose that positive identification should rely on biometrics (e.g., face recognition carried out on a videotape recorded on the crime scene) and suppose that the biometric matching system is based on a similarity score<sup>6</sup>, which range from 0.0

---

<sup>6</sup>Beyond similarity scores, there are also other mathematical tools used for comparing two biometrics, but this would not change the sense of my example.

(no match at all) to 1.0 (exact match digit by digit). As I have previously explained, a perfect match is theoretically and practically impossible, consequently the similarity score should be set by administrators at any value lower than 1.0. For instance, if one sets the system at a similarity score of 0.95, it means that only biometrics which have a score beyond this threshold will match. Now, suppose that we know (from previous performance tests) that a 0.95 score implies a very low—suppose 0.1%—risk of false recognition (that is to say, almost all recognitions are right). Yet from the same tests, we also know that the score 0.95 implies 20% failed recognition (20% guilty individuals who escape from our biometric fishnet). If we lower the score to 0.75, the failed recognition rate will decrease dramatically to 1% (only 1% guilty individuals would escape recognition). Unfortunately, such a lower score would also provoke a hike of false recognitions, which would increase to 10%, say, our fishnet will wrongly fish also 10% innocent people. What should system administrators do? In the investigation phase, they would be plenty legitimate to set the system at the lowest similarity score compatible with its correct use. This would respect the logic behind criminal investigation, which aims to avoid that a criminal gets off scot-free. This might imply that some 10% innocent people are unjustly accused of being terrorists. In an ideal world, these people would be discharged by the judicial system. What happens if probabilistic digital biometrics is admitted as an evidence in court? It happens that the 10 % innocent people of our example are burdened by a “positive identification”, notwithstanding the doubt that they could be “false positive”. The principle of presumption of innocence is de facto bypassed, because its foundation, *in dubio pro reo*, could not put up with an evidence obtained through a system whose initial aim was to detect the highest number of suspects.

I argue that if a biometric system is tuned in order to meet law enforcement requirements, its results should not be transferred, as they stand, into the judicial system, which has different, even opposite, user requirements. Also pre-digital biometrics were somehow probabilistic, at least in very rough terms, but today one can set the degree of confidence of the system. In other words, differently from the past, probabilities of recognition can be tuned. If biometrics in law enforcement were tuned to minimize the risk of false recognition, I would not see any ethical problem in being their results transferred in court as well. Yet, would law enforcement agencies ever accept to purchase (and use) a system burdened by a higher failed recognition rate when that system could be tuned in order to minimize it?<sup>7</sup> I doubt.

If a system fulfills law enforcement requirements, it cannot also fulfill judicial requirements, and vice versa. Both requirements are fully legitimate, but they should work in parallel. If one of them “takes the leadership” (notably, law enforcement requirements) this becomes a serious ethical and democratic problem.

---

<sup>7</sup>The biometric performance at different thresholds is expressed through the “Detection Error Tradeoff” (DET) curve.

## 16.6 Solutions and Trends

The ideal technical solution to this ethical challenge would be to develop more robust, accurate, and sensitive, biometric applications, with negligible false rejection and false acceptance rates. If false rejection and false acceptance rates were truly negligible, the contradiction between law enforcement and judicial system requirements would be eliminated and one could transfer results from one system to the other, without major ethical issues. This is the illusion that has ruled forensic biometrics till almost today. This illusion is no longer tenable and there are not current applications which are contemporarily—in real life (not in labs)—as specific and sensitive as to consider negligible both their false rejection and false acceptance rates. In principle, a strategy exists that could increase contemporarily both specificity and sensitivity. This strategy is based on tighten requirements on the quality of biometric input data. Unfortunately, this would imply that a larger number of people could not be enrolled in the system, because of their poor biometric features, due to various causes, including the way in which in real life conditions biometric features are presented to sensors. Notably, this solution would not be applicable to biometrics extracted from materials, which were not originally designed for collecting biometrics. This is the case of latent fingerprints, but also of other “biometric traces” (e.g., recorded images, voices, pictures, etc.) that we are increasingly able to detect and collect for forensic reasons.

This leads to the problem posed by biometrics extracted from outside the forensic context, and brought into the legal cycle only at a later stage. Recorded faces, images, and voices, found online, in the Internet, are the largest (accidental) biometric library ever created (it is enough to think of the number of freely available pictures of Facebook users). Searching these huge, dispersed, online, databases is becoming one of the main activities, not only in criminal investigations, but also in crime prevention and judicial decisions.<sup>8</sup>

Online face and voice recognition—often coupled with soft biometrics<sup>9</sup> and geolocalization—are increasingly used for searching and recognizing people. This is destined to become still more pervasive with the arrival of new and emerging biometrics. New behavioral biometrics (also including electrophysiological biometrics) and cognitive biometrics will be able not only to extract biometric features from recorded online materials, but also to elicit meaningful biometric information from online behaviors and human-machine interaction patterns. New biometrics are extremely promising for law enforcement purposes. Will these biometrics be ever

---

<sup>8</sup>To give an idea of the magnitude and pervasiveness of this phenomenon, it is enough to mention that the Internet is today the main source of evidence on marriage validity or nullity in Catholic ecclesiastic courts.

<sup>9</sup>Sex, ethnicity, age, body shape, skin color, and so.

admissible in court? I doubt if presumption of innocence still rules criminal proceedings. Yet, trends seems to go toward an opposite direction.<sup>10</sup>

In November 2013, the European Commission presented a package of proposals to strengthen procedural safeguards for citizens in criminal proceedings. *Inter alia*, these proposals included a proposal of “*Directive on the strengthening of certain aspects of the presumption of innocence and of the right to be present at trial in criminal proceedings*”,<sup>11</sup> aiming at harmonizing this principle (considered chiefly in its dimension of procedural right instead of human right) in different EU jurisdictions. When legislators feel the need to rule on great principles, it is often because they aim to mitigate them. Indeed, after a long preamble and four initial, generic, articles, Art.5 of the proposal directive focuses on *Burden of proof and standard of proof required*. Par.1 reaffirms the principle that

Member States shall ensure that the burden of proof in establishing the guilt of suspects or accused persons is on the prosecution. This is without prejudice to any ex officio fact finding powers of the trial court

But Par.2 seriously mitigates it, by allowing Member States to shift the burden of proof on the defendant for any “*sufficient important*” (!) reason

Member States shall ensure that any presumption, which shifts the burden of proof to the suspects or accused persons, is of sufficient importance to justify overriding that principle and is rebuttable.

In order to rebut such a presumption it suffices that the defense adduces enough evidence as to raise a reasonable doubt regarding the suspect or accused person’s guilt.

The Commission explained the rationale behind Art.5 Par.2 in an accompanying Communication on Making progress on the European Union Agenda on Procedural Safeguards for Suspects or Accused Persons—Strengthening the Foundation of the European Area of Criminal Justice,<sup>12</sup> which reads,

In criminal proceedings, the burden of proof should be on the prosecution and any doubt should benefit the suspect or accused person, without prejudice to the independence of the judiciary when assessing the suspect or accused’s guilt. A judgment must be based on the evidence put before it and not on allegations or assumptions. However, the ECtHR<sup>13</sup> has accepted that in specific and limited cases, the burden of proof may be shifted to the defense, and the Directive will reflect this standard, striking a balance between the public interest in effective prosecution and the rights of the defense.

<sup>10</sup>In his 2003 paper on Evaluation of Forensic Science [18], Arizona State University Professor of Law, Michael J. Saks, raised the issue of reversal of the burden of proof related to biometric evidence. His argument is different from mine, because he focuses on the fact that some courts are asking the defendant to demonstrate that biometric evidence **does not** fulfill *Daubert* criteria, instead of assessing by themselves whether it does. However, it is interesting to note that trends move in the same direction.

<sup>11</sup>COM(2013) 821 final, Brussels, 27.11.2013.

<sup>12</sup>COM(2013) 820 final, Brussels, 27.11.2013.

<sup>13</sup>European Court of Human Rights.

The initial formulation of COM (2013) 821 has raised many perplexities in the LIBE<sup>14</sup> (main) and in the JURI<sup>15</sup> (opinion) Committees of the European Parliament. More recently, the Council<sup>16</sup> has proposed to modify Art.5 Par.2 by eliminating the expression “*sufficient importance*” in its place listing cases in which it would be possible to shift the burden of proof on the defendant, say,

(...) in two situations:

- (a) in case of certain minor offences, notably traffic offences (see recital 15a);
- (b) in case of certain other types of offences, when two conditions have been complied with:
  - (i) the shifting of the burden of proof must be strictly necessary in the interest of the criminal proceedings; and
  - (ii) the shifting of the burden of proof must be justified to override the principle that the burden of proof is on the prosecution (see also recital 15b);

It is definitely out of scope of this article to discuss this important debate, and its potential consequences on the European legal system. I think that the reader has probably understood that the issue at stake is something more important than “minor traffic offences”.<sup>17</sup> This is also suggested by the web page of the *Justice Directorate of the European Commission*, which is not, of course, a Commission official statement or a policy document, but it is expected to represent the Commission’s point of view,

When designing and implementing measures in this field, it is important for the EU to get the balance right between measures that protect such rights and those that facilitate<sup>18</sup> the investigation and prosecution of crime [19]

This is why the current debate is likely to be so relevant also to ethics of forensic biometrics, because it finally concerns the delicate balance between pursue of the common good, and respect for individual liberty.

## References

1. Mordini E, Tzovaras D (2012) Second generation biometrics: the ethical and social context. Springer, Berlin
2. Mordini E, Massari S (2008) Body, biometrics, and identity. *Bioethics* 22(9):488–498
3. Mordini E (2008) Nothing to hide. biometric privacy and private sphere. In: Tistarelli M, Juul N, Drygajlo A, Schouten B (eds) BIOID 2008 Biometrics and identity management, Springer, Berlin, Heidelberg, pp 247–57

---

<sup>14</sup>European Parliament Committee on Civil Liberties, Justice and Home Affairs.

<sup>15</sup>European Parliament Committee on Legal Affairs.

<sup>16</sup>Council of the European Union, 12196/14, Brussels, 29 July 2014.

<sup>17</sup>It is worth recalling that the issue of presumption of innocence has been one of the main issues at stake in the US legal, political, and ethical debate on war on terrorism and terrorists’ detention.

<sup>18</sup>My italics.

4. Mordini E, Rebera AP (2013). The biometric fetish. In: About I, Brown J, Lonergan G (eds) *People, papers, and practices: identification and registration in: transnational perspective*, 1500–2010 London, Palgrave, pp 98–111
5. Mordini E (2009) Ethics and policy of biometrics. In: Tistarelli M, Stan ZL, Chellappa R (eds) *Handbook of remote biometrics for surveillance and security*. Springer, Berlin Heidelberg, pp 293–309
6. Mordini E, Rebera AP (2011) No identification without representation: constraints on the use of biometric identification systems. *Rev Policy Res* 29(1):5–20
7. Moenssens AA (1963) Admissibility of fingerprint evidence and constitutional objections to fingerprinting raised in criminal and civil cases. *Chicago-Kent Law Rev* 40(2):85–124
8. Kaye DH (2003) Questioning a courtroom proof of the uniqueness of fingerprints. *Int Stat Rev* 71:521–533
9. Haber L, Haber RN (2004) Error rates for human latent print examiners. In: Bolle R, Natalini R (eds) *Advances in automatic fingerprint recognition*. Springer, New York, pp 339–360
10. Stacey R (2004) Report on the erroneous fingerprint individualization in the madrid train bombing case. *J Forensic Ident.* 6(54):706–718
11. Champod, C. (2000). Standards of proof. In: Siegel J(ed) *Encyclopaedia of forensic sciences*, Academic Press p 890
12. Egli NM, Champod C, Margot P (2007) Evidence evaluation in fingerprint comparison and automated fingerprint identification system—Modelling within finger variability. *Forensic Sci Int* 167:189–195
13. Champod C, Lennard C, Margot P, Stoilovic M (2004) Fingerprints and other ridge skin impressions. CRC Press—Taylor & Francis, London
14. Cole S (2008) Comment on ‘scientific validation of fingerprint evidence’ under Daubert’. *Law Probab Risk* 7:120–132
15. Hirsch Ballin MF (2012) *Anticipative criminal investigation*. Springer, Berlin
16. Pollock F (1922) *Essays in the law*. Oxford University Press, Oxford
17. Summers RS (1999) Formal legal truth and substantive truth in judicial fact-finding—their justified divergence in some particular cases. Cornell Law Faculty Publications, Paper 1186
18. Saks MJ (2003) The legal and scientific evaluation of forensic science (especially fingerprint expert testimony). *Seton Hall Law Rev* 1167–87
19. European Commission (2014) Rights of suspects and accused. [http://ec.europa.eu/justice/criminal/criminal-rights/index\\_en.htm](http://ec.europa.eu/justice/criminal/criminal-rights/index_en.htm). Accessed 2 June 2015

# Index

## Numerical

3D face recognition, 9

## A

Accidental biometric library, 362

Acquisition techniques, 20, 22, 24, 26, 32

Administration of justice, 354, 358, 360

Advanced technology, 4

Age, 200–204

Atered fingerprint detection, 86, 87, 90, 120

Analysis tool, 330, 344, 348

Anticipation of crime, 358

Antispoofing, 63, 65, 67, 69, 82

Automated Fingerprint Identification System (AFIS), 37–40, 49, 51, 56, 58

## B

Bayesian interpretation framework, 222, 228, 235, 236

Behavioral biometrics, 362

Benchmark, 208–210, 212, 213, 215

Biometric evidence, 222–224, 236, 237

Biometrics, 5–7, 9, 10, 177, 178, 188, 330

Blood oxygenation, 70

Burden of proof, 359, 363, 364

## C

Calibrated LR, 233, 235, 236

Calibration, 307, 316, 323–325

Conclusion scale, 205, 206

Confidence interval, 356

Contact-free measurement, 291, 293

Court cases, 1, 3, 10

Covariance discriminative learning, 157

Covert operation, 154

Criminal investigation, 354, 358, 361, 362

## D

Data-driven, 127, 128, 130–132, 145

Daubert criteria, 357

Deep convolutional neural networks, 160

Digital biometrics, 355–357, 361

Direct method, 225, 229, 230, 236

Dynamic model-based approach, 149, 150, 170, 172

Dynamic signature recognition, 336

## E

Electrical properties, 69

Environmental effects, 63

Equal proportion probability (EPP), 231, 232, 234–236

Evidence, 354, 357, 359, 361, 363

Expert, 196, 197, 199, 203–206, 208, 209, 215

## F

Face Inversion, 201, 208

Face recognition, 5, 8, 149–151, 154, 160, 163–167, 170–172, 177–190

Face recognition by computer, 196, 201, 214

Face recognition from videos, 177

Face recognition systems, 197, 201, 209

Face sketch recognition, 144

Face sketch synthesis, 128, 130, 132, 133, 138, 139, 144, 145

Facial expression, 200, 213

Facial image comparison, 205, 206, 208, 209

Facial images, 290, 291, 293, 296

False acceptance, 362

False negative, 206

False positive, 206

False rejection, 356, 362

Familiar face, 197–200, 202, 213, 214

Fingermarks, 24, 25, 27, 37, 44, 45, 50, 51, 54, 56, 57

Fingerprint, 20–23, 26, 30, 32, 37, 38, 40, 41, 43–51, 53, 54, 56, 57, 63–67, 69, 71–74, 79–82, 310, 313, 317, 319–322, 354, 362

- Fingerprint mutilation, 87  
 Fingerprint recognition, 99, 102  
 Forensic, 306–309, 313, 315–320, 322, 325, 330–332, 335, 338, 339  
 Forensic automatic speaker recognition (FASR), 222–224, 231, 234, 236, 237  
 Forensic evaluation, 51, 53, 56, 57  
 Forensic expert, 357  
 Forensic face recognition, 178  
 Forensic gait analysis, 247, 252  
 Forensic identification, 202, 206, 207, 215  
 Forensic image analysis, 242  
 Forensic intelligence, 50  
 Forensic investigation, 39, 51, 52, 56, 57, 241, 242  
 Forensic science, 2–9  
 Forensic signature analysis, 329, 336, 339  
 Forensic speaker recognition (FSR), 221  
 Forgery detection, 21  
 Fusion, 207, 212, 213
- G**  
 Geometrical model-based methods, 150, 168, 172
- H**  
 Handwriting, 330  
 Heartbeat, 289–293, 296–299  
 Hidden Markov models, 166  
 Human, 196, 197, 199–202, 207–215  
 Human identification, 177
- I**  
 Identification accuracy, 39  
 Identity management, 49, 50  
 Image quality, 204–206, 215  
 Image resolution, 201
- J**  
 Jurisdiction, 354, 357, 363
- L**  
 Latent fingerprints, 20–22, 25–30, 33  
 Law enforcement, 354, 355, 358, 360–362  
 Legal requirements, 33  
 Legal truth, 359  
 Likelihood ratio (LR), 222, 223, 225–237, 306, 307, 309, 311, 313, 314, 318, 320–322, 325  
 Liveness detection, 65, 81  
 Logistic regression, 314, 322–324
- M**  
 Manifold-based methods, 150, 154, 172  
 Manifold discriminant analysis, 156
- Minutiae cylinder code (MCC), 42, 44, 45  
 Minutiae, 22–24, 29, 30  
 Minutiae analysis, 96  
 Minutiae points, 96, 114, 115  
 Missing persons, 49, 50  
 Model-based approach, 149, 150, 168, 172
- N**  
 National Institute of Standards and Technology (NIST), 209, 211–213  
 Novel biometrics, 289, 290
- O**  
 On-line signature recognition, 330  
 Orientation field, 89–92, 94, 95, 103, 105, 107, 111, 120  
 Other race effect, 200, 202, 211
- P**  
 Passport, 197, 204, 205  
 Performance, 198–201, 204–214  
 Performance characteristics, 231, 235, 236  
 Performance evaluation, 222, 236  
 Performance metrics, 231, 232, 236  
 Person search, 269, 270  
 Photogrammetry, 242, 247  
 Photo sketches, 5  
 Poincaré index, 103, 105, 107  
 Pool adjacent violators (PAV), 315, 316, 322  
 Pores, 22, 24  
 Pre-crime, 358  
 Presentation attack detection, 86  
 Presumption of innocence, 359–361, 363  
 Probabilistic graphical model, 136  
 Probabilistic methods, 150, 163, 172  
 Probability of misleading evidence, 231–233, 235  
 Professional, 196, 203, 205  
 Projective metric learning, 159, 160, 163  
 Psychology, 201, 215
- R**  
 Rare minutiae, 318–320, 322, 325  
 Representation, 197, 199, 201, 211, 214
- S**  
 Scalability, 56  
 Score, 306, 311–318, 320–325  
 Scoring method, 225, 227–230, 236  
 Security camera, 150, 279  
 Semantic description, 268  
 Sensor, 356, 362  
 Signal detection analysis, 197  
 Similarity score, 360, 361

- Singular points, 89, 95, 99, 103, 105, 107  
Skin disease, 63, 72, 81  
Skin model, 66, 81  
Soft-biometric, 267, 268, 290, 298, 299  
Software platform, 241, 242, 247, 248, 264  
Source, 199, 205, 214  
Sparse-coding, 149, 150, 172  
Sparse representation, 130, 132, 134, 135  
Spoofing, 63  
Statistical analysis, 53, 54, 57  
Strength of evidence, 222, 226, 227, 231, 234, 236  
Surveillance, 267–270  
Suspect identification, 128
- T**  
Temperature properties, 67, 79  
Tenprints, 28  
Trace age estimation, 20
- Training, 196, 197, 204–208, 210, 212, 214, 215  
Training set, 210
- U**  
Unconstrained face recognition, 188–190  
Uncontrolled environment, 182  
Unfamiliar face, 198–203, 209, 214
- V**  
Video analysis, 149  
Video-based face recognition, 149–151, 153, 163–168, 170, 172, 189  
Video traces, 177, 178, 182
- W**  
Within-identity variation, 198