The proceedings of the BIOSIG 2013 include scientific contributions of the annual conference of the Biometrics Special Interest Group (BIOSIG) of the Gesellschaft für Informatik (GI). The conference took place in Darmstadt, 04.-06. September 2013. The advances of biometrics research and new developments in the core biometric application field of security have been presented and discussed by international biometrics and security professionals.

Gesellschaft für Informatik e.V. (GI)

publishes this series in order to make available to a broad public recent findings in informatics (i.e. computer science and information systems), to document conferences that are organized in cooperation with GI and to publish the annual GI Award dissertation.

Broken down into
• seminars
• proceedings
• dissertations
• thematics

current topics are dealt with from the vantage point of research and development, teaching and further training in theory and practice. The Editorial Committee uses an intensive review process in order to ensure high quality contributions.

The volumes are published in German or English.

Information: http://www.gi.de/service/publikationen/lni/

GI-Edition

Lecture Notes in Informatics

Arslan Brömme, Christoph Busch (Eds.)

BIOSIG 2013

Proceedings of the 12th International Conference of the Biometrics Special Interest Group

04.–06. September 2013
Darmstadt, Germany

Proceedings

Arslan Brömme, Christoph Busch (Eds.) BIOSIG 2013 12th International Conference of the Biometrics Special Interest Group

ISSN 1617-5468
ISBN 978-3-88579-606-0

The proceedings of the BIOSIG 2013 include scientific contributions of the annual conference of the Biometrics Special Interest Group (BIOSIG) of the Gesellschaft für Informatik (GI). The conference took place in Darmstadt, 04.-06. September 2013. The advances of biometrics research and new developments in the core biometric application field of security have been presented and discussed by international biometrics and security professionals.
Arslan Brömme, Christoph Busch (Eds.)

BIOSIG 2013

Proceedings of the 12th International Conference of the Biometrics Special Interest Group

04.-06. September 2013 in Darmstadt, Germany

Gesellschaft für Informatik e.V. (GI)
Lecture Notes in Informatics (LNI) - Proceedings
Series of the Gesellschaft für Informatik (GI)

Volume P-212

ISBN 978-3-88579-606-0
ISSN 1617-5468

Volume Editors

Arslan Brömme
GI BIOSIG, Gesellschaft für Informatik e.V.
Ahrstraße 45, D-53175 Bonn
Email: arslan.broemme@aviomatik.de

Christoph Busch
Hochschule Darmstadt
CASED
Haardtring 100, D-64295 Darmstadt

Series Editorial Board
Heinrich C. Mayr, Alpen-Adria-Universität Klagenfurt, Austria
(Chairman, mayr@ifit.uni-klu.ac.at)
Dieter Fellner, Technische Universität Darmstadt, Germany
Ulrich Flegel, Hochschule für Technik, Stuttgart, Germany
Ulrich Frank, Universität Duisburg-Essen, Germany
Johann-Christoph Freytag, Humboldt-Universität zu Berlin, Germany
Michael Goedicke, Universität Duisburg-Essen, Germany
Ralf Hofestäd, Universität Bielefeld, Germany
Michael Koch, Universität der Bundeswehr München, Germany
Axel Lehmann, Universität der Bundeswehr München, Germany
Peter Sanders, Karlsruher Institut für Technologie (KIT), Germany
Sigrid Schubert, Universität Siegen, Germany
Ingo Timm, Universität Trier, Germany
Karin Vosseberg, Hochschule Bremerhaven, Germany
Maria Wimmer, Universität Koblenz-Landau, Germany

Dissertations
Steffen Hölldobler, Technische Universität Dresden, Germany

Seminars
Reinhard Wilhelm, Universität des Saarlandes, Germany

Thematics
Andreas Oberweis, Karlsruher Institut für Technologie (KIT), Germany

© Gesellschaft für Informatik, Bonn 2013
printed by Köllen Druck+Verlag GmbH, Bonn
Chairs’ Message

Welcome to the annual international conference of the Biometrics Special Interest Group (BIOSIG) of the Gesellschaft für Informatik (GI) e.V.

GI BIOSIG was founded in 2002 as an experts’ group for the topics of biometric person identification/authentication and electronic signatures and its applications. Over the last decade the annual conference in strong partnership with the Competence Center for Applied Security Technology (CAST) established a well known forum for biometrics and security professionals from industry, science, representatives of the national governmental bodies and European institutions who are working in these areas.

The BIOSIG 2013 international conference is jointly organized by the Biometrics Special Interest Group (BIOSIG) of the Gesellschaft für Informatik e.V., the Competence Center for Applied Security Technology e.V. (CAST), the German Federal Office for Information Security (BSI), the European Association for Biometrics (EAB), the European Commission Joint Research Centre (JRC), the TeleTrusT Deutschland e.V. (TeleTrusT), the Norwegian Biometrics Laboratory (NBL), the Center for Advanced Security Research Darmstadt (CASED), and the Fraunhofer Institute for Computer Graphics Research (IGD). This years’ international conference BIOSIG 2013 is again technically co-sponsored by the Institute of Electrical and Electronics Engineers (IEEE) and is enriched with satellite workshops by the TeleTrust Biometric Working Group and the European Association for Biometrics.

The international program committee accepted full scientific papers strongly according to the LNI guidelines (acceptance rate ~26%) within a scientific double-blinded review process of at minimum five reviews per paper. All papers were formally restricted for the printed proceedings to 12 pages for regular research contributions including an oral presentation and 8 pages for further conference contributions including a poster presentation at the conference site.

Furthermore, the program committee has created a program including selected contributions of strong interest (further conference contributions) for the outlined scope of this conference. All paper contributions for BIOSIG 2013 will be published additionally in the IEEE Xplore Digital Library.

We would like to thank all authors for their contributions and the numerous reviewers for their work in the program committee.

Darmstadt, 04th September 2013

Arslan Brömme
GI BIOSIG, GI e.V.

Christoph Busch
Hochschule Darmstadt
**Chairs**

Arslan Brömme, *GI BIOSIG, GI e.V., Bonn, Germany*

Christoph Busch, *Hochschule Darmstadt, Germany*

**Program Committee**

<table>
<thead>
<tr>
<th>Chairs</th>
<th>Program Committee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arslan Brömme, GI BIOSIG, GI e.V., Bonn, Germany</td>
<td>Emilio Mordini (CSSC, IT)</td>
</tr>
<tr>
<td>Christoph Busch, Hochschule Darmstadt, Germany</td>
<td>Axel Munde (BSI, DE)</td>
</tr>
<tr>
<td>Program Committee</td>
<td>Alexander Nouak (Fhg IGD, DE)</td>
</tr>
<tr>
<td>Harald Baier (CASED, DE)</td>
<td>Markus Nuppeney (BSI, DE)</td>
</tr>
<tr>
<td>Oliver Bausinger (BSI, DE)</td>
<td>Hisao Ogata (Hitachi, JP)</td>
</tr>
<tr>
<td>Thiriamchos Bourlai (WVU, US)</td>
<td>Martin Olsen (GUC, NO)</td>
</tr>
<tr>
<td>Patrick Bours (GUC, NO)</td>
<td>Javier Ortega-Garcia (UAM, ES)</td>
</tr>
<tr>
<td>Sebastien Brangoulo (Morpho, FR)</td>
<td>Michael Peirce (Daon, IR)</td>
</tr>
<tr>
<td>Ralph Breithaupt (BSI, DE)</td>
<td>Anika Pflug (CASED, DE)</td>
</tr>
<tr>
<td>Julien Bringer (Morpho, FR)</td>
<td>Ioannis Pitas (AUT, GR)</td>
</tr>
<tr>
<td>Arslan Brömme (GI/BIOSIG, DE)</td>
<td>Patrizio Campisi (Uni Roma, IT)</td>
</tr>
<tr>
<td>Christoph Busch (CAST-Forum, DE)</td>
<td>Fernando Podio (NIST, US)</td>
</tr>
<tr>
<td>Victor-Philipp Busch (Sybuca, DE)</td>
<td>Nathaniel Clarke (CSCAN, UK)</td>
</tr>
<tr>
<td>Patrizio Campisi (Uni Roma, IT)</td>
<td>Reinhard Posch (IAIK, AT)</td>
</tr>
<tr>
<td>Nathan Clarke (CSCAN, UK)</td>
<td>Raghu Ramachandra (GUC, NO)</td>
</tr>
<tr>
<td>Henning Daum (secunet, DE)</td>
<td>Kai Rannenberg (Uni FFM, DE)</td>
</tr>
<tr>
<td>Nicolas Delvaux (Morpho, FR)</td>
<td>Nalini Ratha (IBM, US)</td>
</tr>
<tr>
<td>Farzin Deravi (UKE, UK)</td>
<td>Christian Rathgeb (CASED, DE)</td>
</tr>
<tr>
<td>Bernadette Dorizzi (IT, FR)</td>
<td>Marek Rejman-Greene (HO, UK)</td>
</tr>
<tr>
<td>Martin Drahansky (BUT, CZ)</td>
<td>Arun Ross (WVU, US)</td>
</tr>
<tr>
<td>Julian Fierrez (UAM, ES)</td>
<td>Heiko Roßnagel (Fhg IAO, DE)</td>
</tr>
<tr>
<td>Simone Fischer-Hübner (KAU, SE)</td>
<td>Patrick Flynn (ND, US)</td>
</tr>
<tr>
<td>Patrick Flynn (ND, US)</td>
<td>Raul Sanchez-Reillo (UC3M, ES)</td>
</tr>
<tr>
<td>Lothar Fritsch (NR, NO)</td>
<td>Stephanie Schuckers (CIU, US)</td>
</tr>
<tr>
<td>Steven Furnell (CSCAN, UK)</td>
<td>Günter Schumacher (JRC, IT)</td>
</tr>
<tr>
<td>Patrick Grother (NIST, US)</td>
<td>Takashi Shinzaki (Fujitsu, JP)</td>
</tr>
<tr>
<td>Daniel Hartung (GUC, NO)</td>
<td>Max Snijder (EAB, NL)</td>
</tr>
<tr>
<td>Olaf Henniger (Fhg IGD, DE)</td>
<td>Luis Soares (IST, PT)</td>
</tr>
<tr>
<td>Detlef Hühnlein (ecsec, DE)</td>
<td>Luuk Spreeuwers (UTW, NL)</td>
</tr>
<tr>
<td>Heinrich Ihmor (BSI, DE)</td>
<td>Elham Tabassi (NIST, US)</td>
</tr>
<tr>
<td>Christiane Kaplan (softpro, DE)</td>
<td>Tieniu Tan (NLPR, CN)</td>
</tr>
</tbody>
</table>
Stefan Katzenbeisser (CASED, DE)  Cathy Tilton (Daon, US)
Tom Kevenaar (GenKey, NL)  Massimo Tistarelli (UNISS, IT)
Ulrike Korte (BSI, DE)  Carlo Trugenberger (SwissSc, CH)
Bernd Kowalski (BSI, DE)  Dimitrios Tzovaras (CiRaT, GR)
Ajay Kumar (Poly, HK)  Andreas Uhl (COSY, AT)
Herbert Leitold (a-sit, AT)  Markus Ullmann (BSI, DE)
Stan Li (CBSR, CN)  Raymond Veldhuis (UTW, NL)
Paulo Lobato Correia (IST, PT)  Anne Wang (Cogent, US)
Mark Lockie (PB, UK)  Jim Wayman (SJSU, US)
Davide Maltoni (UBO, IT)  Frans Willems (UE, NL)
Tony Mansfield (NPL, UK)  Andreas Wolf (BDR, DE)
Tsutomu Matsumoto (YNU, JP)  Haiyun Xu (UT, NL)
Johannes Merkle (secunet, DE)  Bian Yang (GUC, NO)
Didier Meuwly (NFI, NL)  Xuebing Zhou (CASED, DE)

Hosts

Biometrics Special Interest Group (BIOSIG) of the Gesellschaft für Informatik (GI) e.V.
http://www.biosig.org

Competence Center for Applied Security Technology e.V. (CAST)
http://www.cast-forum.de

Bundesamt für Sicherheit in der Informationstechnik (BSI)
http://www.bsi.bund.de

European Association for Biometrics (EAB)
http://www.eab.org

European Commission Joint Research Centre (JRC)
http://ec.europa.eu/dgs/jrc/index.cfm

TeleTrusT Deutschland e.V (TeleTrust)
http://www.teletrust.de/

Norwegian Biometrics Laboratory (NBL)
http://www.nislab.no/biometrics_lab

Center for Advanced Security Research Darmstadt (CASED)
http://www.cased.de/

Fraunhofer-Institut für Graphische Datenverarbeitung (IGD)
http://www.igf.fraunhofer.de/
Biometrics provides efficient and reliable solutions to recognize individuals. Growing interests about trustworthiness of authentication stimulate employment of biometric techniques. Nowadays, biometric applications can be found in diverse areas such as health monitoring, national ID cards, e-banking, e-commerce, etc. It rises to challenges of robustness, reliability, interoperability, scalability, system reliability, and usability.

Large-scale applications such as the European Union Visa Information System (VIS) and Unique Identification (UID) in India require high accuracy. Multimodal biometrics combined with fusion techniques can improve recognition performance for such applications. Furthermore, efficient searching and/or indexing methods can accelerate also the identification efficiency. Additionally, the quality of acquired biometric samples can strongly influence the performance.

Quality assessment methods can not only guarantee success of authentication but can also provide helpful feedback to system operators during the capturing process. Recently it was shown, that biometric recognition with low cost sensors embedded in mobile devices such as cell phones can improve deployment and acceptance of biometric systems.

Moreover, concerns about security and privacy can not be neglected. The relevant techniques in the area of presentation attack detection (liveness detection) and template protection are about to supplement biometric systems, in order to improve fake resistance, prevent potential attacks such as cross matching, identity theft etc.

BIOSIG 2013 offers you once again a platform for international experts’ discussions on biometric research and the full range of security applications.
# Table of Contents

**BIOSIG 2013 – Regular Research Papers**

Nicolas Buchmann, Roel Peeters, Harald Baier, Andreas Pashalidis  
*Security considerations on extending PACE to a biometric-based connection establishment*  
15

Ferdinand Hahmann, Gordon Böer, Hauke Schramm  
*Combination of Facial Landmarks for Robust Eye Localization Using the Discriminative Generalized Hough Transform*  
27

Andreas Uhl, Peter Wild  
*Experimental Evidence of Ageing in Hand Biometrics*  
39

Ctirad Sousedik, Ralph Breithaupt, Christoph Busch  
*Volumetric Fingerprint Data Analysis using Optical Coherence Tomography*  
51

Andreas Pashalidis  
*Simulated annealing attack on certain fingerprint authentication systems*  
63

Benjamin Tams  
*Absolute Fingerprint Pre-Alignment in Minutiae-Based Cryptosystems*  
75

Sandra Cremer, Nadege Lemperiere, Bernadette Dorizzi, Sonia Garcia-Salicetti  
*Quality driven iris recognition improvement*  
87

Rudolf Haraksim, Didier Meuwly, Gina Doekhie, Peter Vergeer, Marjan Sjerps  
*Assignment of the evidential value of a fingermark general pattern using a Bayesian Network*  
99

Chris Stein, Vincent Bouatou, Christoph Busch  
*Video-based Fingerphoto Recognition with Anti-spoofing Techniques with Smartphone Cameras*  
111

Soumik Mondal, Patrick Bours  
*Continuous Authentication using Mouse Dynamics*  
123

Asad Ali, Farzin Deravi, Sanaul Hoque  
*Spoofing Attempt Detection using Gaze Colocation*  
135
Napa Sae-Bae, Nasir Memon
A Simple and Effective Method for Online Signature Verification .............................. 147

Daria La Rocca, Patrizio Campisi, Jordi Solé-Casals
EEG Based User Recognition Using BUMP Modelling .................................................. 159

Toshiyuki Isshiki, Toshinori Araki, Kengo Mori, Satoshi Obana, Tetsushi Ohki, Shizuo Sakamoto
New Security Definitions for Biometric Authentication with Template Protection:
Toward covering more threats against authentication systems .................................... 171

BIOSIG 2013 – Further Conference Contributions ...................................................... 183

Leila Mirmohamadsadeghi, Andrzej Drygajlo
A template privacy protection scheme for fingerprint minutiae descriptors............. 185

Alina Krupp, Christian Rathgeb, Christoph Busch
Social Acceptance of Biometric Technologies in Germany: A Survey....................... 193

Su Yang, Farzin Deravi
Quality Filtering of EEG Signals for Enhanced Biometric Recognition..................... 201

Nesli Erdogmus, Sébastien Marcel
Spoofing 2D Face Recognition Systems with 3D Masks ........................................... 209

Pawel Kasprowski, Ioannis Rigas
The influence of dataset quality on the results of behavioural biometric experiments 217

Ingo Deutschmann, Johan Lindholm
Behavioral biometrics for DARPA's active authentication program ....................... 225

Georgia Koukiou, Vassilis Anastassopoulos
Eye temperature distribution in drunk persons using thermal imagery ...................... 233

Zinelabidine Boulkenafet, Messaoud Bengherabi, Omar Nouali, Mohamed Cheriet
Using the conformal embedding analysis to compensate the channel effect in the
i-vector based speaker verification system .................................................................. 241

Markus Springer
Protection of Fingerprint Data with the Glass Maze Algorithm ................................. 249
Laurentiu Acasandrei, Angel Barriga
Embedded Face Detection Implementation ................................................................. 257

Chris van Dam, Raymond Veldhuis, Luuk Spreeuwers
Landmark-based Model-free 3D Face Shape Reconstruction from Video Sequences... 265

Rafik Chaabouni
Solving Terminal Revocation in EAC by Augmenting Terminal Authentication .......... 273

Syed Zulkarnain Syed Idrus, Estelle Cherrier, Christophe Rosenberger,
Patrick Bours
Soft Biometrics Database: a Benchmark for Keystroke Dynamics Biometric Systems 281

Anastasios Drosou, Panagiotis Moschonas, Dimitrios Tzovaras
Robust 3D Face Recognition from Low Resolution Images ...................................... 289

Lei Gao, Lin Qi, Ling Guan
Selecting Discriminative Features with Discriminative Multiple Canonical
Correlation Analysis for Multi-Feature Information Fusion ................................... 297

Berkay Topcu, Hakan Erdogan, Cagatay Karabat, Berrin Yanikoglu
BioHashing with Fingerprint Spectral Minutiae ..................................................... 305

Ramon Blanco-Gonzalo, Raul Sanchez-Reillo, Oscar Miguel-Hurtado,
Judith Liu-Jimenez
Usability Analysis of Dynamic Signature Verification in Mobile Environments .......... 313

Mihails Pudzs, Rihards Fuksis, Rinalds Ruskuls, Teodors Eglitis,
Arturs Kadikis, Modris Greitans
FPGA based palmprint and palm vein biometric system ........................................ 321

Esra Vural, Steven Simske, Stephanie Schuckers
Verification of Individuals from Accelerometer Measures of Cardiac
Chest Movements ........................................................................................................ 329

Pinar Santemiz, Luuk J. Spreeuwers, Raymond N.J. Veldhuis
Automatic Landmark Detection and Face Recognition for Side-View Face Images.... 337

Martin Böckeler, Xuebing Zhou
An Efficient 3D Facial Landmark Detection Algorithm with Haar-like Features
and Anthropometric Constraints .............................................................................. 345
BIOSIG 2013

Regular Research Papers
Security considerations on extending PACE to a biometric-based connection establishment

Nicolas Buchmann*, Roel Peeters†, Harald Baier* and Andreas Pashalidis†

*da/sec Biometrics and Internet Security Research Group
Hochschule Darmstadt, Darmstadt, Germany
firstname.lastname@h-da.de

†KU LEUVEN, ESAT/COSIC & iMinds, Belgium
firstname.lastname@esat.kuleuven.be

Abstract: The regulations of the European Union (EU) Council in 2004 are the basis of the deployment of electronic passports within the EU. Since then EU member states adopt the format and the access protocols to further electronic machine readable travel documents (eMRTD) like national electronic ID cards and electronic residence permits, respectively. The security protocols to communicate with an eMRTD are based on the paradigm of strong cohesion and loose coupling, i.e., each step is designed to ensure only a particular security goal like authorisation to access a certain data group, authenticity and integrity of the data, originality of the chip, or the linkage between the eMRTD and its holder. However, recently a discussion evolved to integrate the linkage security goal within the connection establishment, which currently only aims at limiting basic access of authorised terminals to the eMRTD. For instance, the BioPACE protocol proposes to replace the knowledge-based shared ‘secret’ of PACE by a biometric-based one. The goal of the paper at hand is twofold: First, we evaluate the BioPACE protocol and propose improvements to enhance its features. Second, we analyse the expediency of integrating our BioPACE version 2 into the eMRTD domain. Our initial evaluation shows that our BioPACE version 2 is expedient if the EAC protocols and the corresponding PKI are abandoned.

1 Introduction

Since 2004 EU member states issue ePassports, which feature an embedded radio frequency (RF) chip [EU04, EU05]. This chip contains sensitive biometric data, typically including the ePassport holder’s facial image and fingerprints of two index fingers [ICA06]. In order to address the risks that arise through the electronic storage and wireless communication channel, security protocols for ePassports have been specified [ICA06, BSI10]. The privacy of ePassport holders, for example, is protected by access control mechanisms, which ensure that only trusted parties may read the fingerprints. Confidentiality of the transferred data is achieved by encrypting all communication between an inspection system and the chip. The specified protocols also ensure authenticity and integrity of the data read from the chip, as well as the originality of the chip itself.
The specified security protocols follow the paradigm of strong cohesion and loose coupling. That is, each protocol fulfils a very specific security goal and the security protocols hardly depend on each other, if there is a dependency at all. This paradigm is well established in the software engineering community [IEE90, ISO05].

Due to this principle further chip equipped cards (e.g., electronic ID cards) with similar security goals can use a subset of the ePassports’ security protocols and replace an ePassport protocol by a new one where appropriate. This does not only create a benefit for the electronic ID cards, but instead a mutual gain, because if a new improved security protocol is favoured in the electronic ID card domain it might replace the ePassport counterpart in the long term. This is currently the case for the Password Authenticated Connection Establishment (PACE, [BSI10]), which is expected to replace the Basic Access Control (BAC) protocol by the PACE-based Supplemental Access Control (SAC) in 2018 [ICA13].

Recently Deufel et al. [DMDK13] propose the BioPACE protocol as a replacement for the knowledge-based shared ‘secret’ of PACE. The BioPACE protocol uses a biometric-based secret instead.

The goal of our paper is twofold: Firstly, we evaluate the BioPACE protocol. We document weaknesses compared to PACE, especially that BioPACE enables tracking and abandons the connection of the physical document and its chip. Additionally we propose improvements to enhance its features. Secondly, we analyse the expediency of integrating our BioPACE version 2 into the eMRTD domain. We sketch the idea of replacing the expensive Extended Access Control (EAC) protocols and its related Country Verifying Public Key Infrastructure (CV PKI) by our BioPACE version 2 protocol. An initial evaluation reveals that our BioPACE version 2 actually has the potential to serve as replacement, if some of the conveniences of EAC are considered to be dispensable (e.g., fine-grained authorisation levels to different data groups).

This paper is organised as follows: Section 2 describes the security protocols, which are relevant for the later discussion of BioPACE. In Section 3 the concept and underlying idea of BioPACE is introduced. The security assessment of BioPACE is presented in Section 4. Section 5 proposes an enhanced version of BioPACE. Section 6 presents future plans to replace EAC with our BioPACE version 2, and discusses the expediency of our BioPACE version 2 in the eMRTD domain. In Section 7 conclusions are drawn and the presented improvements and the usefulness of BioPACE are discussed.

2 eMRTD protocols and their security goals

This section describes the eMRTD protocols and their security goals. Each protocol fulfils a very specific security goal. The protocols are either specified by the International Civil Aviation Organisation (ICAO) [ICA06] or the German Federal Office for Information Security (BSI) [BSI10], and are well described in [KN07].

Passive Authentication is the only protocol, which is specified as mandatory by the ICAO [ICA06]. It provides authenticity and integrity of the data stored on the chip. Passive Authentication depends on the so-called Signing PKI.
Basic Access Control (BAC) provides protection against unauthorised access to the data stored on the chip [ICA06]. Unauthorised means access to the data without the eMRTD owner handing over the document. To get access to the chip the terminal needs optical access to the data page to read the Machine Readable Zone (MRZ). The terminal authenticates itself to the chip with the data read from the MRZ, and both entities agree on session keys during BAC to establish a secure channel which provides authenticity, integrity and confidentiality of the transferred data by means of the Secure Messaging sub-protocol.

To protect the sensitive data groups BAC is not sufficient. Therefore Extended Access Control (EAC) protects data group 3 (DG3), which contains the fingerprints. EAC consists of Terminal Authentication and Chip Authentication [BSI10]. After performing EAC the terminal can read the fingerprints, capture a biometric sample from the eMRTD holder and compare the biometric data to check if the current eMRTD holder is the legitimate owner, and thereby achieves the linkage security goal.

To prevent chip cloning, two protocols exist in the eMRTD domain. Active Authentication (AA) specified by the ICAO [ICA06] and as part of EAC Chip Authentication (CA) specified by the BSI [BSI10]. Both protocols prove the authenticity of the chip (originality) to the terminal. AA achieves this goal with a challenge-response protocol and CA establishes a strong secure channel based on the Diffie-Hellman protocol to implicitly prove the originality of the chip.

Terminal Authentication (TA) is part of EAC and is a protocol by which a terminal can prove to a chip its access right to the sensitive biometric data [BSI10]. The chip forces every terminal to prove its authorisation to DG3 before granting access to the fingerprints. TA is based on a PKI for terminals called the Country Verifying PKI.

The Password Authenticated Connection Establishment (PACE) fulfils the same security goals as BAC, but provides strong session keys even in the presence of low-entropy passwords, and contrary to BAC is resistant against offline brute-force attacks [BSI10]. The shared password is denoted by $\pi$ and can either be received from the MRZ, a PIN, or the Card Access Number (CAN), which is printed on the data page of the eMRTD and consists of a six digit number. PACE is based on symmetric and asymmetric cryptography, while BAC is based solely on symmetric cryptography. PACE is depicted in Figure 1 and roughly consists of the following steps:

- First the eMRTD chip randomly chooses a nonce $s$ and encrypts it with $K_{\pi}$ which is derived from the shared password $\pi$. The chips sends the ciphertext $z = \text{Enc}_{K_{\pi}}(s)$ to the terminal.
- The terminal recovers $s$ with the shared password $\pi$ and receives $s = \text{Dec}_{K_{\pi}}(z)$.
- Chip and terminal both create ephemeral key pairs, and perform a Diffie-Hellman key agreement protocol based on these key pairs and the generated shared secret $s$. By performing Diffie-Hellman both entities agree on a new shared secret $K$.
- Based on $K$ both parties derive session keys.
- Chip and terminal exchange and verify authentication tokens based on a Message Authentication Code.
• After successfully performing PACE the Secure Messaging sub-protocol is started with the derived session keys to establish a secure channel, which provides authenticity, integrity and confidentiality.

PACE is the basic building block for the BioPACE protocol introduced in the next section.

3 BioPACE

This section presents the BioPACE security protocol and its underlying idea as introduced by Deufel et al. in [DMDK13]. Deufel et al. present BioPACE as a pre-processing step to the PACE protocol, which we describe in Section 2. We first sketch the idea of BioPACE and then describe its two phases.

The underlying idea for the pre-processing step is to make use of biometric template protection based on the ISO/IEC 24745 standard for biometric information protection [ISO11]. BioPACE does not favour a biometric modality, i.e., BioPACE may be implemented using the facial image, fingerprints, iris, etc. During personalisation of an eMRTD the biometric modality is enrolled and a feature extraction from the captured biometric sample results in a biometric reference comprising of a pseudonymous identifier $PI$ and auxiliary data $AD$. The concrete specification of $PI$ and $AD$ with respect to size and structure is neither specified by the ISO/IEC 24745 standard nor by the authors of [DMDK13]. A verification consists of a new feature extraction from a fresh biometric sample and the previously enrolled $AD$. The verification results in a new pseudonymous identifier $PI^*$, which equals $PI$ if and only if the same person provided the biometric sample and therefore a biometric match occurs.

We now explain the two phases of BioPACE in more detail. The authors of [DMDK13] call these phases the initialisation phase and the regular use phase.
During the initialisation phase the biometric enrolment is conducted, which results in $PI$ and $AD$. Additionally the eMRTD chip or a backend system creates a random CAN or PIN, which serve as input for the regular PACE protocol after the pre-processing step of BioPACE. In what follows we denote this random secret as CAN. The secret CAN is encrypted using $PI$ as encryption key resulting in $ENC_{PI}(CAN)$. Then $PI$ is discarded. The pair $(AD, ENC_{PI}(CAN))$ is then written to data group 13 (DG13) of the eMRTD logical data structure (LDS) [ICA06]. DG13 is publicly accessible without any authentication. This is justified by [DMDK13] with the consideration that the tuple $(AD, ENC_{PI}(CAN))$ is not security sensitive, because it does not disclose biometric data of the enrolled person.

After the initialisation phase BioPACE is ready for regular use. This phase is depicted in Figure 2. If an inspection system wants to perform BioPACE, it first has to read DG13 to receive $(AD, ENC_{PI}(CAN))$. The inspection system captures a biometric sample from the document holder and uses the received $AD$ from DG13 to compute $PI^*$. The inspection system then performs $DEC_{PI^*}(ENC_{PI}(CAN))$ to decrypt $ENC_{PI}(CAN)$ using $PI^*$ as decryption key to receive $CAN^*$, which will match $CAN$ if and only if $PI^*$ matches $PI$.

The secret value CAN is also known to the eMRTD chip, because it is stored in its internal memory and can therefore be used as input for the standard PACE protocol. After this pre-processing step BioPACE uses the steps of the PACE protocol, which we explain in Section 2.

### 4 Assessment of BioPACE

In this section we present our security assessment of BioPACE with respect to common security features of an eMRTD. We identified weaknesses that are introduced when replacing PACE with BioPACE. Every paragraph first presents a short assessment regarding
a specific security aspect, and then proposes possible solutions, when applicable.

**No physical to electronic linkage.** Where PACE makes a link between the printed data page of the eMRTD and the chip inside the eMRTD, BioPACE makes a link between the eMRTD owner and the chip inside the eMRTD. There is no link anymore between the printed data page of the eMRTD and the chip inside the eMRTD. As a consequence it cannot build further upon the prior established authenticity of the MRZ and CAN (by checking the optical security features on eMRTDs, such as special paper and printing techniques).

**Tracking.** While PACE guarantees the unlinkability of eMRTD occurrences on the wireless channel, BioPACE does not. The authors of BioPACE justify that data group 13 can be read freely from the chip by claiming that it does not disclose any biometric data and as such is not security-sensitive. However, the data \( (AD, ENC_{P1}(CAN)) \) provides a unique identifier for every eMRTD and can be read out by anyone within communication range of the eMRTD making tracking possible.

A possible solution would be to print \( (AD, ENC_{P1}(CAN)) \) on the data page of the eMRTD, additionally ensuring the coupling between the data page of the eMRTD and the chip. However, this would require substantial changes in the eMRTD creation and verification processes, as opposed to reading out some extra values from the chip.

**Usability degradation.** The aspect of better comfort is not proven in the paper. We doubt that reading and processing a fingerprint is faster than performing OCR on a MRZ or CAN. Implementing BioPACE instead of PACE also means that the verifier needs biometric reader equipment, even if one only wants to read the chip’s version of the holder’s name, or to verify authenticity and integrity of the chip’s data via passive authentication. At the end of the paper, it is suggested that one can always skip the biometric pre-processing step of BioPACE and fall back to the original PACE. However, if the biometric pre-processing step can be skipped, this raises questions about the benefits of BioPACE, especially towards the eMRTD owner.

**Loss of access control flexibility.** As long as the sensitive biometric fingerprints are stored on the chip BioPACE should not be considered as EU EAC replacement, because it can only provide two possible authorisation levels: read every data group or read no data group. With EAC, one can provide a more fine grained access control and the eMRTD receives an explicit authorisation from its issuing country that this terminal is indeed authorised to read certain data groups.

A possible solution is to replace the raw fingerprints by a biometric template that leaks no sensitive information.

**Double biometric linkage goal.** The basic BioPACE protocol claims to provide access control and create a link between the eMRTD holder and the chip. In the current eMRTD security protocol pool these goals are already achieved by BAC, PACE and EAC for the access control and the fingerprints stored on the chip for the biometric link. Achieving the same security goal twice has no benefit and only makes the border control check more lengthy.
Removing EAC and the raw fingerprints would justify the access control and linkage goal of BioPACE. Of course this should only be considered if the eMRTD would contain no more sensitive biometric data.

**Skimming.** BioPACE claims that no unauthorised data retrieval is possible. For eMRTDs that implement PACE, one requires access to the printed data page of the eMRTD to read the data on the chip. Handing the eMRTD over to an official for checking can be seen as an implicit authorisation from the eMRTD owner. For BioPACE to reach the same level of authorisation, eMRTD holders can only provide their fingerprint to the officials checking their eMRTDs. However, we leave our fingerprints everywhere. Anyone within wireless communication range that has access to the fingerprint of the eMRTD holder, can read out the data of the eMRTD without the owner even being aware. This makes skimming attacks easy, for example in airport bars (given that one can extract the fingerprint from a glass in a timely manner). One does not need to fool the terminal’s fingerprint reader (which is hard, since one has to make a dummy finger, possible liveness detection) but the raw image data is good enough for direct processing. As boundary condition, the attacker also needs a terminal and the attack is only justified if a name or facial image to a corresponding fingerprint is the goal of the attacker.

By making a link to the printed data page of the eMRTD this attack can be mitigated, because the printed content is not revealed in airport bars.

**Offline eMRTD owner guessing.** Because the CAN has low entropy, an offline guessing attack with respect to whom the eMRTD belongs to is possible. Assume that one wants to track a number of high profile individuals and one has access to their fingerprints (which are left behind on whatever the person in question happens to touch). From these fingerprints, together with $AD$ one can derive all possible $PI$’s. Only a subset of the corresponding $ENC_{PI}(CAN)$ will decrypt to a possible CAN (having low entropy). Of course this will not uniquely identify any one person, but it will narrow down the search space significantly.

A trivial solution would be to pad the CAN with some randomness before encryption, and discard the padding upon decryption. Note that this would not work, when using the MRZ instead of the CAN. While the MRZ has typically more entropy than the CAN, it also has more structure that is preserved regardless of the random padding.

A side note worth mentioning: If $PI$ could provide a high enough entropy it could also make BAC attractive again, because the main complaint of BAC is the low entropy of the MRZ combined with its vulnerability to offline brute-force attacks. Still PACE is resistant against offline brute-force attacks and should therefore preferred over BAC.

## 5 An improved BioPACE: BioPACE version 2

This section formalises our BioPACE version 2 protocol. It aims at fixing the flaws identified in Section 4 by changing BioPACE according to the proposals of Section 4.

Figure 3 on the next page illustrates the protocol steps of our BioPACE version 2 protocol.
The improvement consists of two main changes compared to the basic version: First, $PI^*$ is used directly as input for the PACE protocol (and not as decryption key to get the low entropy CAN). Second, $AD$ is printed on the data page of the eMRTD to link the physical document to the chip instead of storing $AD$ on the chip.

We see no reason why one should encrypt a low entropy value (CAN or MRZ) if it does not need to be transferred manually/optically. Therefore, we use $PI^*$ directly as input for the PACE protocol. This also means that the value $ENC_{PI}(CAN)$ no longer needs to be transmitted from the eMRTD to the terminal via the wireless channel. Hence there is no initial access from the terminal to DG13 of the eMRTD. By removing the initial wireless access we avoid the issue of offline guessing of the eMRTD owner, too.

$AD$ is printed on the data page of the eMRTD in form of a 2D barcode (e.g., a QR code [ISO06b] or a Data Matrix code [ISO06a]), which is shown in Figure 4 on the following page. $PI$ is not publicly available, instead it is stored in the internal memory of the eMRTD chip and therefore only available to the chip itself, but not to the terminal. Since the chip does not need to transmit $(AD, ENC_{PI}(CAN))$, there is no longer a unique identifier for the eMRTD, resolving the tracking problem.

We decided against integrating $AD$ into the present MRZ, because in our experience a 2D barcode is more reliable due to the integrated error correction code and more flexible for an $AD$ with variable size depending on the selected biometric modality. 2D barcodes become more and more popular in different areas of document security.

For instance, there is currently a discussion in the EU to integrate 2D barcodes to enhance the authenticity and integrity of non-electronic travel and ID documents (e.g., birth certificates, emergency passports, visas and driver licenses). This EU discussion is based on a new national standard, which is called the Digital Seal [BSI13].

By printing $AD$ on the data page we recreate the link between the physical eMRTD and the chip. Now a terminal needs optical access to the eMRTD to scan the 2D barcode and receive $AD$ to calculate $PI^*$. This will provide at least the same level of protection against skimming and sniffing attacks as PACE.
In the basic BioPACE protocol, matching $PI$ to $PI^*$ is done implicitly by decrypting the low-entropy $CAN$ with $PI^*$ and initialising PACE with the resulting $CAN^*$. The key space entropy of $PI$ is not specified by [DMDK13], but we consider it higher than the entropy of a numeric six digit CAN with an entropy of at most 20 bits. Therefore we directly enter $PI^*$ into PACE. Besides the higher entropy benefit, $PI^*$ is implicitly matched to $PI$ by the completion of the PACE protocol, because if $PI^*$ and $PI$ do not match the PACE protocol will fail.

Our BioPACE version 2 fixes all the mentioned security problems of Section 4. Now, it fulfils its duty as an access control mechanism and leaks no more unambiguous linkable data before the protocol successfully completes. To track a person and to read the data on the chip the attacker needs optical access to the data page. This equals the current security level of eMRTDs and does not constitute a security risk, because if an attacker gets access to the data page he can read all data in printed form anyway.

Still some problems remain and therefore we will discuss the expediency of our BioPACE version 2 in the eMRTD domain in the subsequent section.

6 Replacing EAC and raw fingerprints by BioPACE version 2

In this section we discuss our idea to replace the current infrastructure (i.e., the EAC protocols, the Country Verifying PKI, and the storage of raw fingerprints in data group 3) by our BioPACE version 2 protocol. We contrast the advantages with the disadvantages of our approach and include boundary conditions, which have to be fulfilled to make our BioPACE version 2 expedient.

Fundamental changes to an established infrastructure are a challenging task and require as a boundary condition both innovative ideas and enhanced security. We consider BioPACE version 2 to meet these demands as discussed below. In our context, for instance, a sample innovative idea is the Biocryptographic Key Infrastructure [SBB10] to replace a common
Public Key Infrastructure, yielding a higher security level. An example of enhancing an applied and proven protocol is the Biotokens [SB08] example, where biometric digital signatures and Bio-Kerberos increases security. Therefore the redundant protocols have to be dropped, and the BioPACE version 2 has to provide a significant enhancement to become a new eMRTD standard.

If BioPACE version 2 is used without a subsequent EAC accomplishment, we see the following benefits:

1. **Faster verification:** If we drop EAC and make use of a PI instead of raw fingerprints, we eliminate two bottlenecks: first, no more raw fingerprints have to be transferred from the chip to the terminal over the wireless interface. Second the lack of terminal authentication resolves the need to verify certificate chains by the eMRTD chip. This will drastically speed up the eMRTD processing times at border checks.

2. **Enhanced practical security:** According to a recent EU border control study [Com, D4.1] border control personnel does only perform an electronic check against eMRTD blacklists due to time constraints. Hence in practice the actual security level of the eMRTD chip and its infrastructure is mainly not used. A significant speed-up of the verification protocols will therefore not only make the verification more convenient for the travellers, but it will improve security, because the electronic security features will be actually used by border control personnel even under strict time schedule guidelines.

3. **Improving privacy:** Raw fingerprints are removed and replaced with a biometric template, which is stored in the eMRTD’s internal memory and therefore only accessible by the chip. Hence the privacy level is improved.

4. **Decreasing infrastructure costs:** If we abandon terminal authentication, there is no more need to maintain the complicated Country Verifying PKI. As the further expenses remain constant (e.g., the costs for the biometric personalisation of eMRTDs), the costs of the whole eMRTD infrastructure will decrease significantly.

5. **Standardised data structures:** 2D barcodes are standardised, and their integration is already discussed for non-electronic travel documents based on the Digital Seal standard [BSI13, Com, D6.1].

On the other hand BioPACE version 2 as a replacement for EAC yields the following downsides:

1. **Change of layout:** To establish the BioPACE version 2 protocol in the eMRTD domain the creation and enrolment process has to be changed, because AD needs to be printed on the data page.

2. **Coarse-grained access control:** As discussed in Section 4 BioPACE version 2 causes a loss of access control flexibility. However, if the sensitive JPEG fingerprints are removed from the chip no more sensitive data remains, which is worth protection with a flexible access control scheme.
3. **Renounce of strong cohesion paradigm:** Security protocols often follow the software engineering paradigm of strong cohesion and loose coupling. Every protocol should have a very specific goal and depend on as few as possible other protocols. Our proposal abandons this paradigm.

4. **Chip cloning:** Dropping EAC results in the loss of chip authentication and hence in giving up the current chip cloning protection. However, the physical protection through the printed AD on the document makes chip cloning useless from a practical point of view. We discuss a further electronic prevention approach of chip cloning below.

To conclude we rate the improvement with respect to run-time, practical security, and costs to be more important than the disadvantages to change the layout and the loss of fine-grained access control.

Future attention should be paid to a sample specification of the PI scheme and to the integration of a chip cloning protection into the BioPACE version 2 protocol. Bender et al. [BDFK12] present a protocol called PACE|AA, which combines PACE and Active Authentication to create a protocol, which is more efficient than the single protocols and solves a security risk of Active Authentication.

7 **Conclusion and future work**

This paper presented an assessment of the BioPACE protocol, pointed out flaws in the basic version and proposed an optimised version which fixed these flaws. The second part of this paper analysed the expediency of the BioPACE version 2 protocol and came to the conclusion that it is not expedient in its proposed form. The final section presented a drastic approach which makes our BioPACE version 2 very attractive if the EAC protocols together with the expensive CV PKI are shut down and our BioPACE version 2 gets merged with the PACE|AA protocol to also get a chip cloning protection and become a perfectly tailored monolithic security protocol for the eMRTD domains requirements.

We presented the theoretical idea of merging our BioPACE version 2 protocol with the PACE|AA protocol, therefore future work will focus on a formal security proof for this protocol based on the model proposed by Bellare et al.[BPR00].

**Acknowledgement**

This work was supported by the European Commission through the FIDELITY EU-FP7 project (Grant No. SEC-2011-284862), CASED and the Research Council KU Leuven: GOA TENSE (GOA/11/007).
References


Combination of Facial Landmarks for Robust Eye Localization Using the Discriminative Generalized Hough Transform

Ferdinand Hahmann, Gordon Böer, Hauke Schramm

Institute of Applied Computer Science
University of Applied Sciences Kiel
Grenzstraße 3, 24149 Kiel
Ferdinand.Hahmann@FH-Kiel.de

Abstract: The Discriminative Generalized Hough Transform (DGHT) is a general and robust automated object localization method, which has been shown to achieve state-of-the-art success rates in different application areas like medical image analysis and person localization. In this contribution the framework is enhanced by a novel facial landmark combination technique which is theoretically introduced and evaluated for an eye localization task on a public database. The technique applies individually trained DGHT models for the localization of different facial landmarks, combines the obtained Hough spaces into a 3D feature matrix and applies a specifically trained higher-level DGHT model for the final localization based on the given features. In addition to that, the framework is further improved by a task-specific multi-level approach which adjusts the zooming-in strategy with respect to relevant structures and confusable objects. With the new system it was possible to increase the iris localization rate from 96.6% to 97.9% on 3830 evaluation images. This result is promising, since the variation of the head pose in the database is quite large and the applied error measure considers the worst of a left and right eye localization attempt.

1 Introduction

Automatic landmark localization in face images is an important first step for various computer vision applications like person recognition and tracking, gender classification or facial expression analysis. This underlines the relevance of this task, which has attracted wide scientific interest in recent years. Especially for the eyes as the most important facial landmarks, a large number of localization approaches have been proposed. Many of these techniques have in common that they were specifically developed for the given task using expert knowledge about the object’s positioning, appearance and individual adjustments. Several eye localization approaches employ the Viola & Jones face detector [VJ04] in a first step, which uses Haar-wavelets in a boosted cascade of classifiers. Although this technique is a general object localization approach it requires additional shape constraints to be successful on the task of eye localization [CC03]. Those constraints might be either manually defined [KS10] or automatically learned [CCS04]. Frequently, the method of
Viola & Jones is only used to localize a bounding box around the face in order to perform a subsequent eye localization inside the box, using a specifically developed approach. In this second localization level the given bounding box allows for a rough determination of the eye positions [KHM08] or at least an additional restriction of the search space [KS10].

A popular eye detection approach, sometimes applied in previously located bounding boxes, is to search for circular structures, representing the pupil or the iris [TB11, DLCD04, VG08, NG12]. Despite the technical differences of these approaches, they all make use of expert knowledge about the appearance of the target object and therefore cannot be directly transferred to other localization tasks.

A general and well-known object localization method is the Generalized Hough Transform (GHT) [Bal81]. This technique uses a voting procedure to transform an image into a transformation parameter space, called Hough space, in order to determine the degree of matching between a transformed shape model and the image content. An extension of this approach are Hough Forests [GL09], which learn a direct mapping between the appearance of image patches and the votes in the Hough-space. Hough Forests have already been used in different application scenarios like mouth localization for audio-visual speech recognition [FGVG09] or classification of facial expressions [FYN +12]. In both cases, however, the eyes were localized by searching for circular structures by the method of [VG08].

The idea of splitting the target object into different parts is utilized in various object localization methods [Oka09, LLS08]. Furthermore, in [CCS04] a procedure for eye localization is presented, which detects 17 different facial features using the method of Viola & Jones and learns their relative positions in a geometric model.

The success of the GHT heavily depends on the applied shape models. Therefore, the Discriminative Generalized Hough Transform (DGHT) [RBS08, Rup13] extends the GHT by a fully automated and general learning method for model generation. In [HRBS12] the DGHT was successfully used for eye localization and in [HRBS12] improved results for this task have been achieved by combining the localizations of both eyes with prior knowledge about the expected eye distance vector. In this contribution, we present a novel method for combining different landmarks in two hierarchical DGHT-based localization levels. In the first level the standard DGHT technique is used to determine the localization probabilities for different facial landmarks which are afterwards combined in a 3D feature matrix. On these features the DGHT training approach (Section 2.3) is applied to train a 3D DGHT localization model which learns the relative positioning of each of the landmarks given in level one. In addition to that, a modified multi-level approach is introduced in this work, which achieves an improved robustness by replacing the gradual reduction of the search space in [HRBS12] with a direct zooming into the eye region (Section 2.2).

Both changes have been evaluated on the public PUT Face Database (Section 3) and led to a significant improvement over the standard method (Section 3.3). The paper closes with a discussion of the experimental results (Section 4) and a conclusion (Section 5).
2 Method

2.1 Discriminative Generalized Hough Transform

The Generalized Hough Transform (GHT) is a general method for object localization. It is based on a geometric model which stores model points representing features of the searched-for object in relation to a reference point. The GHT transforms an image space into a model transformation parameter space, from which the optimal object transformation into the given image can be derived. Although, the transformation, considered in the GHT, is in general not restricted, object localization can be based on translation parameters only by determining the position with the highest degree of matching between the model and the feature image. This restriction, used throughout this contribution, allows for quick processing and works well if the model sufficiently represents the object’s variability.

A cell $c_i$ of the quantized parameter space, also called Hough space $H$, represents an image position and reflects the degree of matching between model $M$ and feature image $X_n$ by the number of corresponding feature points $e_k$ and model points $m_j$. The Hough space is generated by an efficient voting procedure and can be formalized by

$$H(c_i|X_n) = \sum_{\forall e_k \in X_n} \sum_{\forall m_j \in M} \begin{cases} 1, & \text{if } c_i = e_k - m_j \text{ and } d(e_k, m_j) \leq \vartheta \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

whereas $d(e_k, m_j)$ specifies the distance between the value of the feature and model point, which has to be lower than a threshold $\vartheta$. In case of using edge features, the values of the feature and model points are usually given by the gradient direction $\varphi_k$ and model point orientation $\varphi_j$. Thus, the distance is determined by $d(e_k, m_j) = |\varphi_k - \varphi_j|$.

Since the result of equation (1) highly depends on the quality of the model, the DGHT comprises an automatic training procedure to generate optimal models. This training procedure uses the Hough space, resulting from the explained voting procedure, to extract the model point specific contributions $f_j(c_i|X_n)$ which is the number of votes from model point $m_j$ into the Hough cell $c_i$. These model point specific votes are recombined into a Maximum-Entropy distribution [Jay57] to ensure maximum objectivity:

$$p_\Lambda(c_i|X_n) = \frac{\exp \left( \sum_j \Lambda_j \cdot f_j(c_i|X_n) \right)}{\sum_k \exp \left( \sum_j \Lambda_j \cdot f_j(c_k|X_n) \right)} \quad (2)$$

The introduced model point specific weights $\Lambda = \{\lambda_1, \lambda_2, \ldots, \lambda_J\}$, which may be negative, are optimized to reflect the importance of a model point for the correct localization and for the distinction of similar objects. For further details on the DGHT method, we refer the reader to [HRBS12].

Note that the DGHT can be used for 2D and 3D images. Although edge images are used most of the time, other features may also be utilized.
2.2 Modified Multi-Level-Approach

The Multi-Level-Approach (MLA) is a zoom-in strategy, in which the resolution is gradually increased around the suspected target point. By decreasing the considered image extract and increasing the resolution in each zoom level the visible structures range from global and coarse to local but fine structures. Since the different DGHT models, applied in the MLA, are specifically trained on the respective image extracts they learn relevant and discriminative structures in each zoom level. Therefore, the MLA is a good trade-off between keeping sufficient target object details and suppressing noise and confusing objects.

The MLA presented in prior publications \([RKL^{+}11]\) doubled the resolution and halved the size of the image extract in each zoom level, therefore keeping the number of pixel constant. For the task of eye localization on the public PUT Face Database this procedure was used with 6 zoom levels in \([HRBS12]\) (Figure 1(a)).

It could be shown in \([HRB^{+}12]\), that the standard MLA procedure is prone to a confusion of the eyes in zoom levels, where both eyes might be visible while important discriminating structures are missing. Consequently, the modified MLA uses a higher resolution in the first zoom level in order to ensure a more accurate target localization than the standard approach. This especially aims at a reliable distinction between both eyes. In the second zoom level of the modified MLA the image extract is already restricted to a region containing only a single eye which excludes a confusion with the other eye. This image extract already has the full resolution and is used for the final localization (Figure 1(b)).

![Figure 1: Comparison of the standard Multi-Level-Approach with the modified Multi-Level-Approach](image-url)
Figure 2: Illustration of the process of landmark combination: According to the standard procedure for landmark localization, the image is transformed into a feature image (edge image). Subsequently, the edge-based DGHT models $M^1_1$, $M^1_2$, $M^1_3$ are utilized for single localization of both eyes ($M^1_1$ and $M^1_2$) and the chin ($M^1_3$). The thereby generated probability distributions $X^1_{n1}, X^2_{n1}, X^3_{n1}$ are combined into a 3D feature image $X^1_n$. On this 3D feature image, a discriminatively trained 3D model $M^2$ is applied for the final localization. Hence, $M^2$ combines the information about the probable position of the individual facial landmarks related to the target landmark.

2.3 Landmark combination

The landmark combination occurs in two levels. In the first level, for each landmark $l$ special DGHT models $M^l_1$ are trained by using the standard DGHT procedure (section 2.1) and canny edge images [Can86] as features. By applying these models to new images, individual probability distributions $X^l_n$ (see Equation (2)) of target localizations are generated. Since (i) with the distribution of a landmark (e.g. left eye), the position of another landmark (e.g. right eye) can be estimated and (ii) the DGHT is neither restricted to edge images nor to 2D images, these landmark specific distributions are combined in a new 3D feature image $X_n = \{X^1_n, ..., X^L_n\}$ for the next localization level. For a given set of $N$ training images, the corresponding 3D features $X_1, ..., X_N$ are used to train a higher-level 3D DGHT model $M^2$ in the second level utilizing the standard DGHT training approach (section 2.1). This model captures the relative position of the landmarks to each other and provides the final localization result.

The feature value of a point $e_k$ in $X^l_n$ specifies the probability $p_l(e_k|I_n)$ (calculated by Equation (2)) of landmark $l$ being localized at position $e_k$ for the given image $I_n$ and model $M^l_1$. Thus, it represents the certainty of the underlying localizer in level one. This
important source of information should be directly incorporated into the GHT voting procedure of level two in order to increase the influence of areas with high localization reliability. Therefore, the standard voting procedure (Equation (1)) is adapted to directly vote with the feature value $p_l(e_k|I_n)$ instead of voting with the value 1. In addition to that, a summation over the $L$ landmarks has to be done in order to combine the results from the different landmark localizations in level one. This leads to the following modified voting procedure for the GHT in level two:

$$H(c_i|X_n) = \sum_l \sum_{e_k \in X_n^l} \sum_{m_j \in M_l^i} \begin{cases} p_l(e_k|I_n), & \text{if } c_i = e_k - m_j \\ 0, & \text{otherwise} \end{cases}$$

(3)

Note that the standard DGHT training approach (see section 2.1) is used for optimizing the models of both described localization levels.

3 Experiments

3.1 Data

The experiments were conducted using the public PUT Face Database [KFS08] in training and evaluation, which includes 9971 images from 100 subjects. The high resolution (2048 × 1536 pixels) color images were taken under controlled lighting conditions in front of a uniform background. Since 30 facial landmarks are provided for each image in this corpus it is very well suited for investigating the presented landmark combination technique. Despite of the neutral background, the corpus is challenging due to the strong variability of head positions (see Figure 3).

As in [HRB+12, HRBS12], the 100 different subjects in the corpus were divided into a training set, containing 60 subjects, and an evaluation set with the remaining 40 subjects. For better comparability the evaluation corpus is identical to [HRB+12, HRBS12] and includes 3830 images. The training was performed on 600 images which have been randomly selected from the training set.
3.2 Setup

In the modified MLA (Section 2.2), the resolution is reduced by a factor of eight in zoom level 0 (see Figure 1(b)). Around the target point, localized in this level, an image extract with original resolution and the size of one-eighth of the complete image is taken for the second and final localization step. The system works with Canny edge features [Can86] and applies a standard DGHT training procedure for generating the specific GHT models for the two localization levels. All described experiments have been performed using a 64 bit system with an Intel Xeon W3520 with 2.66 GHz and 24 GB RAM.

To further enhance the robustness of the modified MLA in zoom level 0, a combination of three landmarks (both eyes and chin) is applied by the landmark combination procedure described in Section 2.3: Using standard DGHT models, based on Canny edge features, three probability distributions for the landmark locations are generated (see Section 2.3). These distributions are combined into a 3D feature image $X_n$, ignoring values of less than 0.01 in order to decrease the processing time and to reduce noise. With a specifically trained 3D DGHT model a robust target localization in zoom level 0 is performed using the modified voting procedure (equation (3)) and the result is handed over to zoom level 1. Figure 4 gives an overview of the system with the modified MLA and landmark combination.

To determine the localization rate, the measurement explained in [JKF01] is used, in which the larger localization error of both eyes is normalized with the eye distance. An error of less than $0.1 / 0.25$ therefore corresponds to a localization result approximately located within the iris / eye. Due to slightly inaccurate annotations, provided by the PUT Face database, an error distance less than 0.1 is not meaningful since the inaccuracy would be higher than the error distance.

3.3 Results

By using the modified MLA a success rate of 97.2% for a localization within in the iris could be achieved on the evaluation corpus. This is an improvement of 0.6% compared to the previously best published result and a gain of 2.2% to the published result obtained with a standard method (Table 1). A good indicator for the localization robustness of zoom level 0 of the MLA is given by the number of target points lying outside the optimal image extract. In comparison to the standard MLA approach and a comparable image extract, this number could be reduced from 130 to 50 by applying the described modifications.
Table 1: Experimental results comparing different systems for different fault tolerances.

<table>
<thead>
<tr>
<th></th>
<th>$e &lt; 0.1$</th>
<th>$e &lt; 0.15$</th>
<th>$e &lt; 0.2$</th>
<th>$e &lt; 0.25$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kasinski et al. [KS10]</td>
<td>94.0%</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Standard MLA with 6 zoom levels [HRB+12]</td>
<td>95.0%</td>
<td>95.4%</td>
<td>96.0%</td>
<td>96.5%</td>
</tr>
<tr>
<td>Standard MLA with Model interpolation [HRBS12]</td>
<td>96.6%</td>
<td>97.1%</td>
<td>97.6%</td>
<td>98.1%</td>
</tr>
<tr>
<td>Modified MLA with 2 zoom levels</td>
<td>97.2%</td>
<td>97.6%</td>
<td>98.0%</td>
<td>98.2%</td>
</tr>
<tr>
<td>Modified MLA with landmark combination</td>
<td>97.9%</td>
<td>98.5%</td>
<td>98.9%</td>
<td>99.1%</td>
</tr>
</tbody>
</table>

A further improvement of the localization robustness in zoom level 0 of the modified MLA could be achieved by using the described landmark combination technique for three facial landmarks. This measure reduced the number of target points lying outside the optimal image extract to 20 and therefore improved the error rate to 97.9% for iris localization. Considering a less restricted fault tolerance, a localization inside the eye was achieved in 99.1% (Table 1). The generated landmark localization models $M_l$ are shown in Figure 5 (a) to (e). The model points are represented as lines to visualize their orientation while the gray value illustrates their weight. Figure 5 (f) displays the 3D DGHT model of zoom level 0. Here, the symbol of a model point indicates the corresponding landmark and the gray value represents again the individual weight as obtained by the discriminative training process.

### 4 Discussion

The significant improvement of the modified MLA can be mostly explained by a better discrimination between both eyes. This is due to an improved localization robustness in zoom level 0 which may be assigned to a better and more detailed DGHT model with a strong focus on both eyes (e.g. see Figure 5(a)). Comparing the models of the standard and the modified MLA, it is noticeable that the average number of model points has substantially increased from 357 to 1807. This rise results from the higher resolution in the modified MLA which leads to an increase of feature points and shape variation, compensated by a larger number of model points.

It is interesting to note that only a few model points of a given localization model are relevant for a single image. Therefore, the percentage of model points, voting for the best Hough cell, is only 11% on average for the standard MLA. For the modified MLA, however, this number is even smaller and amounts to only 5% which underlines the fact that the overall size of the model results from the large variation over all images.

The higher number of feature and model points also explains an increase of the processing time from about 600 ms for the standard MLA to 970 ms for the modified approach. Note, the system has not been optimized for runtime performance yet.
Figure 5: (a) to (e): DGHT models used for baseline landmark localization, where the gray value denotes the individual model point weight. (f): 3D DGHT model used for landmark combination. The symbols illustrate the corresponding landmark layer (circle: right eye layer, square: left eye layer, diamond: chin layer) and the gray value represents the model point weight. Note, that model points with negative weights, which ensure a better discrimination of similar object, are not shown for clarity since they only play a minor role in these experiments.

A clear advantage of the DGHT approach in comparison to most other state-of-the-art localization techniques is the visual interpretability of the models, which reveal the shape of the most discriminative structures as well as the importance of each individual model point. In the localization models of zoom level 0 (Figure 5 (a) to (c)), for example, it can be seen that the localization heavily relies on both eyes and the mouth. The nose, is hardly represented by model points since it is a facial structure which is rarely visible in the feature images and, in addition to that, highly variable (Figure 6(b)). Another interesting aspect, which can be seen in the model images, is that they represent different head positions at the same time to cope with the strong head pose variation contained in the PUT database. For demonstrating this aspect, Figure 6 shows (a) some original images with overlaid model, (b) the corresponding edge feature images, and (c) the model points which voted for the best localization hypotheses.

In zoom level 1 (Figure 5 (d) und (e)), the eye localization models clearly display two concentric circles, representing the iris and the pupil respectively. This search structure has also been integrated in many other systems by using expert knowledge [TB11, DLCD04, VG08, NG12] which demonstrates that the DGHT may learn and incorporate this kind of knowledge fully automatically without the need for a detailed insight into the localization problem. Other model points, contained in the localization model, represent the eyebrows and eyelids, which have different positions depending on the viewing direction, and reflec-
Figure 6: (a) Original images with overlayed model, (b) corresponding feature images, (c) model points which voted for the best localization hypotheses in the respective image. The used model is identical to Figure 5(b).

Figure 7: Examples of image extracts in zoom level 1 with corresponding feature images

tions of the flash on the eyeball (see Figure 7). This also contradicts the common modeling assumption that the sclera is always brighter than the iris, which in turn is brighter than the pupil.

When studying the model for the landmark combination (Figure 5 (f)), it is apparent that model points of the chin have a large scattering and very similar weights while the important points, representing the eye, are much more focussed. This is because of the lower reliability of the chin localizer, which has a mean error of 49 pixels in comparison to 21 and 23 pixels for left and right eye, respectively. It is also worth mentioning that the increased robustness of the landmark combination goes together with a loss in accuracy since the model is more blurred. The increase of the eye localization mean error to 29 and 31 pixels for the left and right eye in zoom level 0 after the landmark combination is compensated by the more precise edge based localization model applied in zoom level 1. In zoom level 1, the mean error was reduced to 12 and 10 pixel.
5 Conclusion

In this contribution two novel techniques for an improved eye localization in portrait images based on the Discriminative Generalized Hough Transform have been presented. By using a task-specific multi-level strategy and a novel facial landmark combination technique it was possible to increase the iris localization rate from 96.6 to 97.9%. This result is promising, since the variation of the head pose in the used public PUT face database is quite large and the applied error measure considers the worst of a left and right eye localization attempt.

The general standard MLA, which gradually zooms into the target object by halving the search space in each level, could be shown to be suboptimal. A more task-specific approach, adjusting the zooming strategy with respect to the relevant structures and confusable objects, may significantly improve the success rate. For the given task of eye localization, with two very confusable objects, a good strategy is an early limitation of the search space to a region, covering only a single eye.

The novel approach for facial landmark detection, which has been introduced in this paper, could be combined with the modified MLA and further increases the robustness of the system in the first zoom level. With this framework, it could be shown for the first time that the DGHT is applicable for both, the individual localization of various landmarks and combined usage in a higher-level localization model. This comes together with the possibility to visually interpret the generated DGHT models in the different stages unveiling discriminative structures and important model parts.

Although in this contribution only three facial landmarks, both eyes and the chin, have been combined with the novel method, the approach may theoretically incorporate an unlimited number. Since the applied discriminative training procedure identifies and penalizes model points of weak landmarks, not supporting the localization, it is possible to select the most discriminative ones from a large set of candidates. A systematic evaluation of this idea, selecting optimal landmarks in an iterative training procedure as well as evaluation on other databases and comparision with other methods will be done by our group in the next future.

Acknowledgments. This work is partly funded by the Innovation Foundation Schleswig-Holstein under the grant 2010-90H. Additionally, the authors are grateful to the anonymous reviewers for their valuable comments.

References


[CCS04] D. Cristinacce, T. Cootes, and I. Scott. A multi-stage approach to facial feature detec-


Experimental Evidence of Ageing in Hand Biometrics

A. Uhl and P. Wild

Multimedia Signal Processing and Security Lab (WaveLab)
Department of Computer Sciences, University of Salzburg, Austria
Email: {uhl, pwild}@cosy.sbg.ac.at

Abstract: Biometric systems build upon the critical property of measuring behavioral, physiological or chemical human properties remaining stable over time. But both, the age of users and ageing of the user’s template may affect performance due to the accumulation of personal changes and indirect behavioral effects like less accurate ability to present the biometric to the sensor. This paper compares short-timespan versus long-timespan effects on different hand-based features presenting the first high-resolution hand-ageing database and identifying features resistant and prone to ageing. Ageing goats, i.e. users responsible for low matching scores across features, are investigated and single-sensor multibiometrics is highlighted to target the ageing problem.

1 Introduction

From the prerequisites for biometric modalities proposed by Jain et. al. in 2004 [JRP04]: universality (availability of the property across populations), distinctiveness (high interpersonal variability), permanence (invariance over periods of time) and collectability (the characteristic can well be measured), it is the accumulation of changes in a person over time - ageing - affecting permanence which constitutes a major challenge to keep high performance of biometric systems. Especially for skin-related modalities, like fingerprints, hand geometry and palmprint, which dominate over 68% of the biometrics market [Int09] ageing is an important issue, since due to loss of collagen skin becomes dryer [MEWK07] and biometric signals may be extracted less accurately. While in the past, at many points claimed permanence and uniqueness properties of fingerprints, e.g. the claimed stability from the 24th week of gestation [GHL+11] and uniqueness of fingerprints in 1892 [Gal92], have been questioned, like in the Brandon Mayfield fingerprint misidentification case in 2004, ageing effects have been superficially treated so far.

This work extends existing ageing studies focusing on the question How does ageing affect different hand-based biometric features? by providing a quantification of the impact on recognition identifying ageing-sensitive and ageing-invariant hand biometric features. The collected dataset is currently the only existing fingerprint and multibiometric full hand ageing dataset. It provides with 5 years timespan a 5 times larger time-lapse between recordings than comparable fingerprint-only datasets, like KFRIA [RJK07] (1 year, 100 subjects). In order to cope with the intrinsic problem of little test data in ageing (e.g. in [ABI05] “Np=30 people took part in the test”, in [BG04] “experiments have been ac-
complished on a data set made of 30 subjects”, or in [GHL+11] “identification test was conducted [...] of all 48 subjects”), we employ significance tests. The novelty of the paper lies not only in setting ageing effects of different hand-based features into context, but to highlight the observation, that degradation is much higher than could be expected and second, that multibiometric fusion can target the ageing problem.

The paper is organized as follows: Section 2 presents related work. In Section 3 we introduce the multibiometric hand recognition system under test. Section 4 presents all experimental evaluations as well as employed ageing datasets and discusses ageing impact in detail. The paper concludes with a summary and outline on future research directions in hand biometric ageing in Section 5.

2 Related Work

Existing work in ageing mainly concentrates on the fingerprint modality. The impact of adult age groups (18-25, 26-39, 40-62 and 62+) on fingerprints recognition performance is highlighted by Modi et. al. [MEWK07] and in Uhl and Wild [UW09] for youth fingerprints (age groups 3-10, 11-18, 19+) observing overall a decreased performance for the boundary age groups, i.e. old and young fingerprints. Reasons for this behavior are given in Modi and Elliott [ME06] observing, that for age classes 18-25 vs. 62+ ageing causes global quality to decrease while false minutiae and the total average number of minutiae drastically increases (90 vs. 55). Sickler et. al. [SE05] show, that young fingerprints exhibit more moisture. As a solution to the problem of fingerprint age prediction Bevilacqua and Gherardi [BG04] watershed segment capacitance images and derive the sum of cells smaller than a given area (observing 13 pixels for 1 year age to 51 pixels for 91 years age). Uchida et. al. [UKM+96] quantify skin ageing by analyzing the 3D profile of subjects aged 20-60 using 2D DFT features (assessing skin ridges) resulting in less high frequency components for elder people - but also wide scattering.

Besides cross-ageing tests, also the ageing of fingerprint features has been subject to investigations: Arnold et. al. [ABI05] quantify the degradation in fingerprint matching performance, using a long-term database from German federal criminal police office, as an FRR increase of factor 2 for 10 years time lapse. Ryu et. al. [RJK07] confirm the degradation observing an EER increase of a factor even greater than 2 over three sensors and a time delta of 1 year employing the KFRIA ageing database. In order to cope with age group effects, Gottschlich et. al. [GHL+11] propose an isotropic rescaling method improving EERs from 11-14% to 5-6% for three different extraction and matching methods. Alternatively, template update techniques are common approaches in current biometric systems, like Kekre and Bharadi’s [KB09] adaptive feature set algorithm, to account for intra-personal variability. Marcialis et. al. [MDP+12] investigate self update algorithms providing a conceptual explanation using a path-based clustering view highlighting the critical task of selecting initial templates and the need for threshold relaxation in case of high environmental variability.

Related to other hand-based modalities, only few ageing studies exist. Uhl and Wild [UW09] extended their comparison of verification performance of kids and adults for fingerprints to also palmprint, hand-geometry and digitprint biometrics resulting in sim-
Figure 1: Sample full-hand images of (a) 24-year-old male user #6 in session 1 (b) 54-year-old female user #32 in session 2 and (c) Architecture of the tested multibiometric system.

Similar results - i.e. age degradation for the youngest age group with only few exceptions - like an inverse behavior for the hand geometry feature. Lanitis et. al. [LT11] investigate template-ageing for face, fingerprint, and palm modalities and formulate a generic AI factor metric assessing the impact of ageing. According to their experiments on FG-NET face ageing and POLYBIO2 multibiometric (face, fingerprints, speech, palm) databases, features derived from faces tended to cause highest AI factors, followed by palm-based features and fingerprints - yet no direct impact on recognition accuracy is provided. Zheng et. al. [ZjWB07] investigate biometric features in hand geometry that are both distinctive and invariant to projective transformations, therefore can be expected to tolerate more variation - caused by, e.g., ageing effects.

Latest trends in fingerprint-based biometrics propose the extraction of fingerprints from full-hand images, employing either high-resolution scans [UW09] or even extracted from video [QYRL10]. In favor of traditional single finger optical, capacitive, ultrasound and thermal single sensors, multi-biometric hand-based acquisition from a single sensor has the advantage, that all hand-based modalities can be extracted at the same time - even by using commodity hardware. This way, accuracy can be further increased by employing fusion techniques. Especially for ageing societies ageing effects causing failures to enrol or false accepts and rejects may decrease the acceptability. Therefore, the approach involving multiple biometric traits is also favored by Rebera and Guihen [RG12] assessing the social impact of biometric ageing.

3 System

We assess the age impact on three textural and three geometrical hand-based features related to the modalities fingerprint, digitprint, palmprint and hand-geometry, described in the following subsections. All these modalities are extracted from a single high-resolution (500 dpi) scan of the palmar surface of the hand using commodity flatbed scanner hardware (HP Scanjet 3500c), see Fig. 1. We extract: (1) minutiae from regions of interest
(ROI) at finger tips; (2) digitprints of individual fingers; (3) a palmprint ROI using again a fixed size region centered in the middle-ring finger valley, and (4) the entire hand image. Preprocessing segments skin from background using Otsu’s thresholding [Ots79], the largest connected object is rotationally aligned using moment-based ellipse fitting. Contour extraction is based on the center-of-mass, detected salient points are used to map the image onto palm coordinates [UW09]. We employ the valley between middle and ring finger as origin and use an approximation of the outer palm line for orientational alignment. Fingers separated by valley positions are fitted with ellipses. Geometrical and textural features operate on either contour data or ROIs.

3.1 Fingerprint Extraction and Comparison

We employ local (level-2) fingerprint features [MMJP09], i.e. minutiae tracking position \((x, y)\) and orientation \((\theta)\) of bifurcations and terminations of ridge lines. For this task, NIST’s mindct extractor [NIS06] has been employed on the normalized (CLAHE) finger-axis-aligned fingerprint patches using a fixed fraction of the finger’s length (one third of its height, one half for the thumb) as ROI. For comparison, we employ NIST’s bozorth3 comparator and combine results of each finger using sum rule fusion.

3.2 Digitprint Extraction and Comparison

As digitprint textural feature we employ the classical Turk-Pentland [TP91] Eigenspace-based feature extraction method projecting each finger (and the palmprint) onto the space spanned by the 25 most significant principal components, trained from a separate dataset. Since this method exploits similarities at low resolution employing a compact representation minimizing the reconstruction error tracking the projection coefficients as feature vectors, it may be more robust to ageing changes than other features. Comparison is executed in the domain of Eigenspace coefficients employing the \(L_1\) norm on the template vectors and product rule fusion to combine all fingers and palm.

3.3 Palmprint Extraction and Comparison

For the palmprint region we extract both, an Eigenspace-based feature (see Sect. 3.2) as well as a local block-based variance feature. The first feature is fused with digitprints to contribute to a common Eigenhand feature. The latter follows Kumar et. al.’s approach [KWSJ03] tracking 144 variances of overlapping 24 times 24 pixel sized blocks and is applied to a mean-variance normalized square (using the average finger length as unit length) palmprint edge region (using Prewitt filtering), centered in the finger valley between ring and middle finger at an offset of 0.2 times the finger length. Two feature vectors are compared based on the L2 norm.
3.4 Hand-Geometry Extraction and Comparison

Kukula and Elliott [KE06] investigate hand geometry as one of the oldest (since 1960s) commercialized modalities and report high acceptability (93%), high universality (0% FTE) and high accuracy (0.98% FRR) for a test set of 129 persons. Common measures relate to widths and height of fingers, palm, segments, etc. in the order of 10-30 features [JRP99, SRSAGM00, KWSJ03]. Geometric features are known to be less suited for identification, but can effectively be employed for fast screening and additional plausibility checks in order to strengthen security (it is generally believed that a multibiometric system is more difficult to circumvent than unimodal systems, although the attacker may specifically target the weakest biometric). Especially for ageing, it is interesting to verify claimed fragility of geometry-related measurements. The employed hand geometry feature tracks all 5 finger lengths (including segment lengths of proximal, intermediate and distal phalanx), finger silhouettes (contours as distances with respect to the finger’s centroid and enclosed area) and a shape-based feature. The latter feature takes local finger widths (scans of the y-monotone finger contour building for each slice the average width of in-object pixels in a total of 15 components) into consideration. As comparators the L1 norm for finger shape, dynamic time warp matching [MR81] of the silhouette and L2 norm for the finger length, respectively, are employed.

4 Experiments

For ageing experiments, we collected a database of high-resolution human handprints from 28 members in our labs with 127 hand images in the first session (Old) captured in November 2007 and 135 hands in the second session (New) captured in October/November 2012, i.e. exhibiting a time lapse of approximately 5 years between recordings adhering to the same strict recording protocol (users were allowed to wear rings or watches and obtain an arbitrary position on the scanner as long as fingers did not touch each other)\(^1\). The system is initialized (parameters for preprocessing, training of feature space) using data from a separate dataset involving a distinct person set (i.e. for the calculation of Eigenfingers and Eigenpalms). Performance is evaluated in terms of pairs of Genuine Acceptance Rate (GAR) at certain False Acceptance Rate (FAR) in form of receiver operating characteristics (ROC) to compare different features under ageing effects. We chose the Equal Error Rate (EER), i.e. the operating point with equal FAR and FRR=1-GAR as main comparison criterion. For the estimation of statistical significance of results, McNemar tests [Yat84] are conducted. In the following subsections several claims related to ageing are examined.

\(^1\)see http://www.wavelab.at/sources/Uhl13b for available material related to this study.
Figure 2: ROC curves for template-ageing effects on hand-based features.
Table 1: Ageing-effects on EER

<table>
<thead>
<tr>
<th>Feature</th>
<th>EER Short</th>
<th>EER Long</th>
<th>X²</th>
<th>McNemar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minutiae</td>
<td>0</td>
<td>0.63 %</td>
<td>12.02</td>
<td></td>
</tr>
<tr>
<td>Eigenhand</td>
<td>1.22 %</td>
<td>1.22 %</td>
<td>1.35</td>
<td></td>
</tr>
<tr>
<td>Palmprint</td>
<td>3.95 %</td>
<td>11.40 %</td>
<td>35.40</td>
<td></td>
</tr>
<tr>
<td>Silhouette</td>
<td>7.81 %</td>
<td>10.12 %</td>
<td>8.38</td>
<td></td>
</tr>
<tr>
<td>Shape</td>
<td>2.55 %</td>
<td>7.15 %</td>
<td>39.75</td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>7.67 %</td>
<td>8.10 %</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Short-Timespan ageing effects

<table>
<thead>
<tr>
<th>Feature</th>
<th>EER 2007</th>
<th>EER 2012</th>
<th>X²</th>
<th>McNemar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minutiae</td>
<td>0</td>
<td>0 %</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Eigenhand</td>
<td>1.3 %</td>
<td>0.8 %</td>
<td>2.04</td>
<td></td>
</tr>
<tr>
<td>Palmprint</td>
<td>3.4 %</td>
<td>4.5 %</td>
<td>1.26</td>
<td></td>
</tr>
<tr>
<td>Silhouette</td>
<td>8.5 %</td>
<td>7.3 %</td>
<td>3.32</td>
<td></td>
</tr>
<tr>
<td>Shape</td>
<td>3.7 %</td>
<td>1.3 %</td>
<td>9.19</td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td>7.2 %</td>
<td>7.5 %</td>
<td>2.56</td>
<td></td>
</tr>
</tbody>
</table>

4.1 Claim 1: “Hand-ageing has no statistical significant impact on recognition”

In order to disprove claim 1, we compared short-timespan (1 day) and long-timespan (5 years) performance of individual features. We therefore examined comparisons of 491 (605) genuine and 16288 (16286) imposter pairs of handprints (numbers in brackets refer to long-timespan pairs). From the ROC curves in Fig. 2 and error rates in Table 1 we can see, that for all of the employed features, an increased time-span of 5 years between recordings decreases recognition accuracy over almost the entire operating range for almost all features. Still, in the interesting EER range ageing impact on some of the features was less expressive. Claim1 disproved.

4.2 Claim 2: “Within a modality, ageing has comparable impact on features”

Comparing classifiers based on thresholds set to achieve close-to-EER performance and using a 95% confidence level (3.84), Minutiae, Palmprint, Silhouette, and Shape classifiers are significantly affected by ageing, while the longer timespan showed no significant effect on Eigenhand and Length features. In order to be able to conduct McNemar significance testing we relate short-timespan and long-timespan comparisons by exchanging the probe template \( p_i \) in the gallery-probe pair \( (g_j, p_i) \) by the corresponding i-th sample of the parallel session. Furthermore, for balancing reasons, we restrict imposter comparisons to the same amount of genuine comparisons and test the classifier in close-to-EER setup based on training data.

Especially for the highly accurate Minutiae feature providing perfect separation for short-timespan data, the long-timespan comparison yields errors due to ageing effects: 0.63% EER. Interestingly, system error increase is observed to be least significant for the Length feature (7.67% vs. 8.10% EER, respectively) tracking finger lengths – a feature which is known to be rather varying during bone growth. However, it has to be considered, that original matching rates are already quite high. For similar reasons, the Silhouette feature is only slightly degraded from 7.81% to 10.12%. Besides for high-security configurations, Eigenhand provides both, high accuracy and no significant changes (stable 1.22 % EER.
Table 3: Users responsible for 5% lowest match scores (numbers in brackets indicate repeated appearance)

<table>
<thead>
<tr>
<th></th>
<th>Minutiae</th>
<th>Eigenhand</th>
<th>Palmprint</th>
<th>Silhouette</th>
<th>Shape</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2007 vs. 2012</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minutiae #32 (20), #30 (7), #15, #12, #24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eigenhand #21 (9), #14 (6), #6 (4), #30 (3), #15 (3), #13 (3), #12, #16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palmprint #31 (6), #7 (5), #12 (5), #30 (3), #13 (3), #10 (2), #28 (2), #4, #1, #9, #21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silhouette #30 (5), #31 (5), #10 (3), #4 (3), #16 (3), #19 (3), #14 (2), #6 (2), #28 (2), #12, #9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shape #21 (15), #30 (6), #12 (4), #14 (4), #8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length #6 (13), #18 (7), #30 (4), #14 (3), #31, #13, #16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Minutiae</th>
<th>Eigenhand</th>
<th>Palmprint</th>
<th>Silhouette</th>
<th>Shape</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2007</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minutiae #24 (4), #15 (3), #12 (3), #30 (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eigenhand #6 (4), #14 (3), #10 (3), #21, #16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palmprint #7 (3), #21 (2), #18 (2), #14, #12, #31, #1, #16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silhouette #19 (3), #6 (2), #31 (2), #16 (2), #10, #9, #30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shape #30 (4), #14 (3), #19 (2), #9, #31, #12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length #6 (6), #18 (4), #10, #7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Minutiae</th>
<th>Eigenhand</th>
<th>Palmprint</th>
<th>Silhouette</th>
<th>Shape</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2012</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Minutiae #32 (9), #24 (3), #30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eigenhand #13 (5), #6 (3), #31 (2), #30 (2), #11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Palmprint #26 (4), #31 (3), #13 (2), #1, #21, #7, #12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Silhouette #6 (4), #4 (3), #31 (2), #11 (2), #29, #13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shape #30 (6), #8 (2), #31 (2), #32, #22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length #6 (6), #18 (3), #13 (2), #31, #32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

and $\chi^2 = 1.35$ in the McNemar test). Due to fusing results from 5 fingers and palm this feature benefits of using multiple geometric and textural properties from various different parts of the hand. Palmprint with 11.4% EER and Shape with 7.15% turned out to be most affected (degradation greater than factor 2). Claim 2 disproved.

4.3 Claim 3: “Short-time intra-personal variability increases with age”

Since [UW09] report different age group performance of youth handprints, we examined intra-short-timespan performance of the two session recordings in 2007 and 2012. Results are given in Table 2. While some features even exhibited higher accuracy for 2012 data compared to 2007, performance differences were not pronounced and McNemar significance tests yielded no significant performance differences for all but the Shape feature. Furthermore, in this configuration the number of genuine comparisons is halved compared to the short vs long-timespan experiments. Therefore, there may be intra-session ageing effects, but Claim 3 could not be confirmed using the given datasets. On the other hand, results confirm validity of the experimental configuration, i.e. set 2012 is not “more challenging”.

46
Figure 3: Long-Timespan 2007 (a-e) and 2012 (f-j) Minutiae genuine pair illustrating age variation due to false minutiae detected near changed creases and worn fingerprints.

Figure 4: Lowest scored Eigenfinger Long-Timespan 2007 (a-e) and 2012 (f-j) genuine pair still closely resembles original data.

Figure 5: Low-scored Palmprint genuine pair (more expressive wrinkles).

Figure 6: Low-scored Shape and Silhouette Long-Timespan genuine pair (thicker fingers).
4.4 Claim 4: “Short-term goats extend to long-term goats”

Related to Doddington’s [DLM98] classification of users, we investigated by analyzing users responsible for the 5% of lowest matching scores (see Tab. 3), whether short time goats (users exhibiting problems in being accepted) extend to long-time goats. For this experiment we separately analyzed 236 and 255 genuine as well as 7765 and 8523 imposter comparisons, respectively for each of the sessions 2007 and 2012 (short-timespan) and 605 genuine and 16286 imposter comparisons for the cross-session (2007 vs. 2012) experiment. We found, that: (1) there are users with low matching scores across all features, e.g. #30; (2) many short-term goats are also long-term goats (100% of lowest-5%-scored Minutiae users, 78% for Eigenhand, 60% for Palmprint, 73% for Silhouette, 44% for Shape and 57% for Length) confirming Claim 4, and; (3) between features rather different users cause problems suggesting for efficient fusion.

4.5 Discussion

From examinations in previous sections, we see, that increased time variation has significant impact on recognition accuracy. Still, one might object, that by measuring the impact of time variation also factors other than age - like dirt, bruising or growing nails may affect recognition accuracy. By adhering to the same strict attended recording protocol and assessing also intra-session effects, we tried to minimize other influencing factors. However, in order to highlight ageing factors, this section illustrates some of the observed visual effects. In Fig. 3 showing a long-timespan genuine tenprint we see (1) the problem of additional/different creases leading to false minutiae, and (2) worn fingerprints (especially Middle and Index). Fig. 4 illustrates the lowest-score Eigenfinger pair highlighting that low resolution does not track skin wrinkles and dryness, in contrast to e.g. the higher-resolved Palmprint feature illustrated in Fig. 5. Also changes in weight may affect recognition accuracy, e.g. thicker fingers for Shape, see Fig. 6.

Finally, when combining all modalities based on weighted sum score fusion (using weights inverse proportional to corresponding feature’s EERs ceiled to 0.5%) perfect separation can be achieved for also aged templates.

5 Conclusion

For hand-based features ageing has been shown to decrease accuracy with least impact on Eigenhand, Length and Silhouette. Ageing caused the Minutiae feature with perfect separation for both intra-session experiments to exhibit an EER of 0.63%. While the latter two features exhibit rather high error rates, Eigenhand with 1.22% EER turned out to be rather stable under ageing effects. Fusion has been shown to be able to target ageing issues, with features exhibiting different ageing goats. In the future, age-adaptive fusion techniques should be investigated. Furthermore, the authors plan to relate ageing effects
to biometric quality issues to isolate long-term ageing-effects from short-term variation.

References


Volumetric Fingerprint Data Analysis using Optical Coherence Tomography

Ctirad Sousedik\(^1\), Ralph Breithaupt\(^2\), Christoph Busch\(^1\)

\(^1\)Norwegian Biometrics Laboratory  
Gjøvik University College  
Teknologiveien 22  
2815 Gjøvik, Norway  
crirad.sousedik@hig.no  
christoph.busch@hig.no

\(^2\)Federal Office for Information Security  
53133 Bonn, Germany  
ralph.breithaupt@bsi.bund.de

Abstract: The increasing usage of fingerprint biometrics as a security technology requires the biometric systems to be resistant against moderate or even high attack potential. To date, state-of-the-art fingerprint sensors can be deceived by an accurate imitation of the ridge/valley pattern of an enrolled fingerprint that can be produced utilizing low-cost, commonly available materials and techniques. The structure of high-resolution 3D volumetric scans of fingertips, acquired by using the Optical Coherence Tomography (OCT), has been analyzed by this work so that a Presentation Attack Detection method could be developed that would render the artefact production process extremely difficult or even practically impossible.

1 Introduction

The fingerprint biometrics is increasingly being applied as a means of access control to security sensitive facilities or sensitive data. The economic constraints further motivate cost saving decisions and thus unsupervised operation of fingerprint based access control has been considered. Recently, the wide availability of the fingerprint sensors on the market has enabled their application also to rather common purposes or facilities such as recreation areas, fitness centers, etc.

The increasing usage of fingerprint biometrics as a security technology requires the biometric systems to be resistant against moderate or even high attack potential [Mun12]. The fingerprint biometrics, as well as any other biometric mode, can be deceived by an accurate artificial representation of the original characteristic and spoofing of fingerprint systems, as well as the development of efficient countermeasures, is one of the most relevant research areas at present. The majority of the state-of-the-art fingerprint sensors can be deceived by an accurate imitation of the fingerprint created by using low-cost and widely available materials and techniques [GFAFMD11].

Numerous approaches to solve the problem of fingerprint liveness detection have been suggested in literature [SMB10]. In order to avoid the costs related to an update of the
sensing technology hardware, researchers and manufacturers attempted to develop liveness detection methods that make use of the 2D fingerprint representations provided by the classical sensors. The methods analyze patterns in the 2D scans that occur due to differences between the genuine fingers and their artificial counterparts, such as inaccurate reproduction of a genuine finger, different elasticity of the artefact material, absence of the skin perspiration etc. However, the limited performance of the state-of-the-art 2D based liveness detection methods rather points to the fact that 2D scans might not be sufficient to be able to distinguish between genuine and artefact or cut-off fingers [YGD+12].

Another group of approaches suggest an update of the hardware parts of the sensing technology by using supplementary sensors that would provide for additional information usable for the liveness detection. Due to large variety of possible artefact material and production techniques, a single aspect dedicated liveness detection sensor can usually be deceived if an appropriate new combination of materials and techniques has been proposed. In order to increase the difficulty of producing an artefact, some manufacturers try to include a larger number of such supplementary sensors that would capture information about multiple aspects of the scanned characteristic. This approach requires application of machine learning techniques to process information from all sensors and take the final decision whether a genuine characteristic has been presented. Since performance of a machine learning based classifier depends on the training data, the sensor can still be vulnerable if an entirely new material and production technique has been used to produce the artefact characteristic.

The above mentioned challenges have motivated a search for a single sensing technology that would provide for a sufficiently detailed representation of the biometric characteristic and could be used both for the presentation attack detection[ISO], as well as biometric recognition. This paper proposes a method for analyzing the structure of fingerprint scans provided by the Optical Coherence Tomography (OCT). The OCT allows to obtain a 3D volumetric representation of the scanned fingerprint, capturing not only the outer fingerprint but also the inner skin structures up to depth of 2-3 mm. Specifically we are interested in the boundary between the inner dermis and the outer epidermis, which contains the master-template of the fingerprint pattern. This inner fingerprint is responsible for regeneration and stability of the outer fingerprint pattern and it can be captured and visualized by the OCT. Small internal structures, such as sweat glands, are also represented in an OCT scan and can be used for liveness detection purposes. A reliable automated liveness detection analysis of an OCT scan would render the production of an appropriate fake representation (i.e. artifact) extremely difficult or even impossible in practice.

2 Optical Coherence Tomography for Fingerprints

The Optical Coherence Tomography is a 3D scanning technology that has originally been developed for medical purposes. It is capable of capturing volumetric representations of the scanned object up to depth of 2-3mm under the surface and the captured signal is a function of the scattering properties of the material. In the classical Time-Domain OCT (TD-OCT) design, a beam of light of low coherence length and wavelength of about 830nm
is split into two separate beams. The first beam is targeted at the object and reflects back. The second beam is targeted by means of a reference arm in such manner that it interferes with the beam reflected from the object. Due to low coherence length of the light source, the interference takes place only with the light reflected from a certain depth under the object surface and the amount of the light reflected can be measured [CSF06].

The classical TD-OCT design requires the reference arm to be positioned in all three axes \(x, y, z\), in order to capture a single voxel, which makes it rather slow, if a full volumetric representation is to be acquired. The OCT based scanning devices exist in multiple adaptations of the original design, in order to achieve higher scanning speeds. Full-Field OCT (FF-OCT) makes use of a 2D array of detectors that can capture the entire \((x, y)\) slice in a single measurement. The Fourier-Domain OCT (FD-OCT) and Swept-Source OCT (SS-OCT) utilize the properties of Fourier Transform to acquire full depths scans without movement of the reference arm along the axis, \(z\) [CSF06].

The potential of the OCT technology for the fingerprint presentation attack detection scenario was researched by the project OCT-Finger [OCT10]. The test group individuals were able to achieve almost 100\% manual recognition rate on a large-scale database of OCT scans of genuine and artefact fingerprints. The study has demonstrated that, apart of the outer fingerprint pattern, the OCT technology has also the potential to capture the structure of the inner fingerprint tissues that are responsible for stability and regeneration of the outer fingerprint, as well as small internal skin structures such as sweat glands (Fig. 1).

The project has demonstrated that the OCT technology allows for analysis of multiple aspects that differentiate genuine and artefact fingers. The 3D representation of the outer fingerprint provides for more information than a classical 2D fingerprint scan. The sweat glands underneath could be detected and their helix-like structure, as well as other properties, verified. In genuine scans, the inner fingerprint pattern copies the pattern of the outer fingerprint and therefore successful comparison of the two patterns could provide for a strong evidence of a genuine finger scan. Other possibilities include verification of the scattering properties of the fingertip skin that potentially differs from the scattering properties of an artefact material. It can be assumed that replication of the internal structure and the scattering properties represented in an OCT scan by means of an artefact would be extremely difficult.

Nevertheless, an efficient liveness detection method based on volumetric scans of fingertips has to cope with a number of challenges. A volumetric OCT scan represents a significant amount of data that increases very quickly with its resolution. A scan of resolution \(200 \times 200 \times 512\) voxels \((width \times height \times depth)\) represents 19,5 MB of data and a scan of \(1024 \times 1024 \times 512\) voxels 512 MB of data if a voxel is represented by a single byte. The acquisition time typically increases with the scanning resolution, which limits practically achievable resolutions. For practical applicability, all the data need to be captured and processed in a matter of a few seconds. As depends on the quality of the OCT scanner and the scattering properties of the fingertip skin, the resulting scan is subject to some amount of speckle noise along with a number of faulty measurements. The structure of the noise can vary slightly among the scanning devices, and the method should deal with the noise in device-independent fashion. The intra-class variability of properties of
genuine human fingers is relatively large and an efficient classification method has to be able to reliably distinguish between the genuine structures and the structures achievable in artefact fingers. In order to reliably classify previously unknown artefacts, the method should validate properties of genuine fingers rather than try to detect anomalies caused by artefact fingers.

Based on the data collected in the framework of the OCT-Finger project, Menrath and Breithaupt [Men11] have proposed an automatic method for analysis and classification of the OCT scans into genuine and artefact ones. They preprocess the OCT volumetric scans by using 3D median filters to suppress the influence of the speckle noise and faulty measurements in the analysed data. The filtered volume is then searched for the inner and outer fingerprint and sweat glands are attempted to be extracted. Due to 3D filtering of a large amount of data, the approach turned out to be an order of magnitude slower than required for practical application. Even though an attempt to perform massive parallelization of their method successfully resulted in a speed-up of an order of a magnitude, the method remains very computationally intensive, which limits its applicability to scans of larger resolutions.

3 Fingerprint Skin Layer Separation

3.1 Database

The OCT scan database used in this work has been composed within the framework of the OCT-Finger [OCT10]. The OCT scanning device was based on the Fourier-Domain OCT (FD-OCT) technology with an acquisition time of 2.24s per scan and operating on a wavelength of 1300 ± 55nm. A scan represents 4 × 4 × 2.5mm-large volume (width × height × depth) of the fingertip at a resolution of 200 × 200 × 512 voxels.

The scans of genuine living fingers represent fingers of 226 subjects of which 96 (42%) were males and 130 (58%) females. The age structure has been as follows: 18 subjects (8%), 5-20 years; 172 subjects (76%), 20-60 years; 36 subjects (16%), 60-80 years. For each subject, at least the right thumb, the right index finger and the right little finger was scanned. Each finger instance was scanned 11 times, which provided for the minimum total amount of 7458 samples of genuine living finger scans.

The fake fingerprint scans represent 30 different classes of artefact fingerprints. The variation artefacts is caused by the mold material composition (gelatin, silicone, latex, window paint, wood glue etc.) and the artefact material composition (glycerol, graphite, window paint etc.). For each class, at least 9 artefact fingerprints were produced. Each of the artefact fingerprints was used to acquire at least 11 scans, yielding for minimum number of 2970 artefact fingerprint scans.

In addition to that, fingers of 5 male and 5 female dead bodies were scanned. For each body, 3 fingers were scanned, 11 times each, providing for 330 dead finger samples.
3.2 Scan Structure

As illustrated by Fig. 1, the OCT technology is able to capture the layered structure of the skin of human fingers. The scan captures a strong reflection from the boundary between the air and the outer fingerprint (Fig. 1b). The outer fingerprint layer is followed by a layer containing the sweat glands. Thickness of this layer as well as its scattering properties vary in the skin of different subjects. The amount of sweat glands also varies strongly among subjects. Some subjects can provide ten times more sweat glands than other subjects and for some individuals the OCT scan is completely lacking any sweat glands. The layer containing sweat glands is followed by a layer that contains the inner fingerprint. The inner fingerprint (i.e. dermis boundary) is a cell structure responsible for stability and regeneration of the outer fingerprint. Unless the inner fingerprint is damaged, the outer fingerprint can fully regenerate to the original pattern. The structure of the outer fingerprint is typically a copy of the inner fingerprint (however, the thickness, clarity and strength of its reflection widely varies among subjects).

![Figure 1](image1.png)

Figure 1: (a) OCT scan of a genuine finger (b) OCT scan of a genuine finger, thresholded 200+, heat-map scale change allows for clearer visualization of fine structures such as sweat glands

A typical structure of a scan of an artefact fingerprint differs in various aspects (Fig. 2) from the genuine counter piece. A thin layered artefact typically yields a strong reflection layer from its outer surface. The structure of the next layer depends on scattering properties and inner structure of the artefact fingerprint (Fig. 2a). Unless the artefact fingerprint is too thick or impenetrable for the OCT scanner, a second strong reflection of the boundary between the artefact and the genuine fingerprint appears. The following structure then copies the genuine fingerprint structure. If the artefact fingerprint is too thick or its material too difficult to penetrate for the OCT scanner, only an outer fingerprint reflection appears (Fig. 2b). The following layer of data represents scattering properties and inner structure of the artefact material (bubbles etc.). A second layer of the inner fingerprint does not appear, nor does it copy the pattern of the outer fingerprint.
3.3 Automatic Method

In order to cope with the large amount of volumetric data in an OCT scan, the initial detection of positions of the specific layers in an OCT fingerprint scan should ideally avoid computationally expensive preprocessing of the data by 3D filtering. We propose a layer detection method robust to a certain degree of speckle noise and faulty measurements that can perform the initial layer detection directly. The volumetric scan, \( V(x, y, z) \), of size \( w \times h \times d \), is analyzed as follows. A grid of size \( w_g \times h_g \) is positioned over the \((x, y)\) plane of the volume \( V(x, y, z) \) (Fig. 3). Each grid cell yields a column volume, \( C_{m,n}(x, y, z) \). From each of the columns, \( C_{m,n}(x, y, z) \), a column accumulation function, \( f_{m,n}(z) \), is computed as follows (Fig. 3):

\[
 f_{m,n}(z) = \sum_{x=0}^{\lfloor w/w_g \rfloor - 1} \sum_{y=0}^{\lfloor h/h_g \rfloor - 1} C_{m,n}(x, y, z)
\]  

(1)
For genuine living fingers, the shape of the column accumulation function \( f_{m,n}(z) \) typically contains two intuitively observable peaks, one for the outer fingerprint and one for the inner fingerprint layer (Fig. 3). However, the two peaks are not always associated with the two main global maxima of the function. Due to noise in the volumetric data, one peak typically consists of multiple smaller peaks, while the actual center of the layer lies in the center of mass of the smaller peaks, rather than on one of them. A particular peak width range should be taken into consideration while searching for the peaks. In addition, the peaks themselves do not often form a maximum even in a smaller window of the function. They rather represent a sudden change on the shape of a function with an otherwise constantly growing or decreasing trend, typically a peak-like fluctuation that does not represent a maximum. Therefore, the peak localization method should look for the fluctuations taking the width range of the peaks into consideration and it should not be affected by the general slope of the function \( f_{m,n}(z) \).

For an otherwise constant discrete function, \( f(n) \), defined on interval \([0, l-1]\), where \( l \) is a peak width parameter, a position \( p \) and an energy \( e \) of a single peak on its shape can be detected as the phase and amplitude of function \( f_p \) as follows (Fig. 4):

\[
\begin{align*}
\sin(n) &= \sin\left(\frac{1}{2} \frac{2\pi}{l} + n \frac{2\pi}{l}\right) \quad (2) \\
\cos(n) &= \cos\left(\frac{1}{2} \frac{2\pi}{l} + n \frac{2\pi}{l}\right) \quad (3) \\
a &= -\sum_{n=0}^{l-1} f(n) \sin(n) \quad (4) \\
b &= -\sum_{n=0}^{l-1} f(n) \cos(n) \quad (5) \\
f_p &= a \cdot \sin(n) + b \cdot \cos(n) \quad (6) \\
&\quad \quad \text{ or } \\
e &= \sqrt{a^2 + b^2} \quad (7) \\
p &= \begin{cases} 
\frac{\arctan2(a,b)}{\pi} \frac{1}{2} l & \text{ if } \arctan2(a,b) \geq 0 \\
\frac{l + \arctan2(a,b)}{\pi} \frac{1}{2} l & \text{ if } \arctan2(a,b) < 0 
\end{cases} \quad (8)
\end{align*}
\]

![Figure 4: Peak detection for an otherwise constant function](image)

However, if the discrete function, apart from the peak range, exhibits a constantly increasing or decreasing trend, the above mentioned method fails. Instead of pointing to the peak position, the method rather analyzes the strength of the global slope of the function \( f(n) \).

For an otherwise constantly increasing or decreasing function \( g(n) \) defined on interval \([0, N-1]\) and \( N = \frac{3}{2} l \), the following method can be used to detect a peak of energy \( e \) and position, \( p \), as the phase and amplitude of function \( f_p \) (Fig. 5):
\[ sn(n) = \sin\left(\frac{1}{2} \frac{2\pi}{l} + n \frac{2\pi}{l}\right) \quad (9) \]
\[ cs(n) = \cos\left(\frac{1}{2} \frac{2\pi}{l} + n \frac{2\pi}{l}\right) \quad (10) \]
\[ a = \sum_{n=0}^{l-1} g(n)sn(n) \quad (11) \]
\[ b = \sum_{n=0}^{l-1} g(n)cs(n) \quad (12) \]
\[ c = \sum_{n=0}^{l-1} g(n + \frac{1}{3} N)sn(n) \quad (13) \]
\[ d = \sum_{n=0}^{l-1} g(n + \frac{1}{3} N)cs(n) \quad (14) \]
\[ f_p = (c - a)sn(n - \frac{1}{3} N) + (d - b)cs(n - \frac{1}{3} N) \quad (15) \]
\[ e = \sqrt{(c - a)^2 + (d - b)^2} \quad (16) \]
\[ p = \begin{cases} \tan^{-2}(c - a, d - b) \frac{1}{2} l + \frac{1}{2} l & \tan^{-2}(c - a, d - b) \geq 0 \\ l + \tan^{-2}(c - a, d - b) \frac{1}{2} l + \frac{1}{2} l & \tan^{-2}(c - a, d - b) < 0 \end{cases} \quad (17) \]

Figure 5: Peak detection an otherwise constantly increasing or decreasing function

The actual detection of positive fluctuations on the shape of a column accumulation function \( f(n) \) defined on interval \([0, N_f - 1] \) is done by means of the following approach. The interval \([0, N_f - 1] \) is divided into overlapping windows \( W_m = [m \frac{1}{4} l, m \frac{1}{4} l + \frac{3}{2} l - 1] \), where \( m \geq 0 \) (Fig. 6). For each window \( W_m \) the peak position \( p_m \) and energy \( e_m \) is computed using equations (9)-(17). Since \( \sin(\alpha + \frac{\pi}{2}) = \cos(\alpha) \) and \( \cos(\alpha + \frac{\pi}{2}) = -\sin(\alpha) \), the following set of equations can be used to efficiently compute \( a_m, b_m, c_m, d_m \) (eq. (11)-(14)) for a window \( W_m \):

\[ sn(n) = \sum_{n=0}^{o-l/4+l/4-1} f(n) \sin\left(\frac{1}{2} \frac{2\pi}{l} + n \frac{2\pi}{l}\right) \quad (18) \]
\[ cs(n) = \sum_{n=0}^{o-l/4+l/4-1} f(n) \cos\left(\frac{1}{2} \frac{2\pi}{l} + n \frac{2\pi}{l}\right) \quad (19) \]
if $m$ is even:

$$a_m = \sum_{n=0}^{3} sns(m + n)$$  \hspace{1cm} (20)  

$$b_m = \sum_{n=0}^{3} css(m + n)$$  \hspace{1cm} (21)  

$$c_m = \sum_{n=0}^{3} sns(m + n + 2)$$  \hspace{1cm} (22)  

$$d_m = \sum_{n=0}^{3} css(m + n + 2)$$  \hspace{1cm} (23)  

if $m$ is odd:

$$a_m = \sum_{n=0}^{3} -css(m + n)$$  \hspace{1cm} (24)  

$$b_m = \sum_{n=0}^{3} sns(m + n)$$  \hspace{1cm} (25)  

$$c_m = \sum_{n=0}^{3} -css(m + n + 2)$$  \hspace{1cm} (26)  

$$d_m = \sum_{n=0}^{3} sns(m + n + 2)$$  \hspace{1cm} (27)  

Figure 6: Overlapping windows for peak detection

Values of functions $sns(o)$ and $css(o)$ can be precomputed by a single multiplication of the analyzed function $f(n)$ with the function $sn(n)$ (eq. (9)), and the function, $cs(n)$ (eq. (10)), while storing the intermediate results. Each of the weights, $a_m, b_m, c_m, d_m$, can then be computed by adding 4 of these precomputed values. The method does not perform any convolution of any core function with the original function $f(n)$, which greatly reduces computational complexity as compared to a convolution based method. In addition the values of functions $sns(n)$ and $css(n)$ can be computed in parallel with no data overlap.

The windows $W_m$ of length $\frac{3}{2}l$ are overlapping each other with an interval of $\frac{1}{4}l$. Consequently each of the windows $W_m$ provides a reliable peak position $p_m$ and peak energy $e_m$ around its center in an interval of length, $\frac{1}{4}l$. If the detected peak position $p$ falls outside of the central interval, it is discarded. The peaks that are detected inside of the central interval of their detection window $W_m$ are considered in the further processing. Depending on the settings, $P$ peaks ($P = 2$ for detection of two layers) with highest energy $e$ are stored.

By means of the above described method $P$ peaks are detected for each column accumu-
lation function $f_{m,n}(z)$ yielding for a sparse approximation of the position of the $P$ most apparent layers in the scan volume $V(x, y, z)$.

4 Results

Manual inspection of the genuine fingerprint scans has identified about 11% of the scans as being of insufficient quality due to non-compliant behavior of the capture subjects. Clearly visible anomalies such as sudden missing parts of the scans and discontinuities in the structure of the scans are related to the fact that the finger was not held still during the scanning period, it was put too close to the scanning surface, removed before the scanning was finished, etc.

An example of the results of the method applied to genuine fingerprint scans ($P = 2$) is shown by Fig. 7. For the genuine fingerprint scans without the anomalies related to the non-compliant subject behavior, the method successfully detected the outer and inner fingerprint layer in 90% of the scans (less than 5% outliers, sparsely distributed) with an average processing time of 0.02s in a single thread on a 3.6GHz 64bit CPU, as opposed to 56s in a single thread on a 2.5GHz CPU as reported by Menrath and Breithaupt [Men11]. The unsuccessful cases were caused by a very weak or completely missing representation of an inner fingerprint in the OCT scan. If the inner fingerprint was present to some extent, the method still did succeed to represent its position in most of the columns, generating more than 5% of outliers. In the scan, the inner fingerprint is represented as a scattered point cloud, rather than a continuous surface. The clarity and density of the point cloud varies among different subjects, from a very clear almost continuous representation to a complete lack of observable inner fingerprint, which causes the method to generate the largest number of outliers for the inner fingerprint layer.

![Figure 7](image)

Figure 7: (a) OCT scan of a genuine finger, thresholded (b) Segmentation into 2 layers ($P = 2, l = 36$)

Fig. 8 illustrates detection of 3 layers ($P = 3$) in a fake fingerprint scan.

In addition, an experiment was carried out regarding usability of the overall energy of the layers for artefact detection. The layer detection was run with $P = 3$, in order to detect three layers in an OCT scan. For each layer detected, the energy, $e$, of the peaks was summed, providing for a vector $(S_{e1}, S_{e2}, S_{e3})$. The vector was used for classification by
means of an SVM classifier with a Gaussian kernel. The Table 1 lists the classification results, cross-validated using 5 different randomly chosen configurations of 50% training and 50% testing data. The results are presented in terms of the ISO/IEC 30107 False Non-Live Detection Rate (FNLD) and False Live Detection Rate (FLD) metrics [ISO]. The FNLD metric represents the proportion of genuine fingerprint scans incorrectly classified as being artefact ones, while the FLD metric represents the proportion of artefact fingerprint scans incorrectly classified as being genuine. FLD train and FNLD train represent results on the training data, while FLD test and FNLD test demonstrate results on the testing data.

Table 1: Results of the SVM classification using the overall energy of three layers ($P = 3$)

<table>
<thead>
<tr>
<th>our approach</th>
<th>FLD train</th>
<th>FNLD train</th>
<th>FLD test</th>
<th>FNLD test</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>10.99%</td>
<td>3.86%</td>
<td>11.01%</td>
<td>3.36%</td>
</tr>
<tr>
<td>#2</td>
<td>11.73%</td>
<td>3.42%</td>
<td>11.37%</td>
<td>3.88%</td>
</tr>
<tr>
<td>#3</td>
<td>10.60%</td>
<td>3.75%</td>
<td>12.12%</td>
<td>3.34%</td>
</tr>
<tr>
<td>#4</td>
<td>11.45%</td>
<td>3.44%</td>
<td>11.11%</td>
<td>3.55%</td>
</tr>
<tr>
<td>#5</td>
<td>11.31%</td>
<td>3.77%</td>
<td>10.97%</td>
<td>3.47%</td>
</tr>
<tr>
<td>average performance</td>
<td>11.22%</td>
<td>3.65%</td>
<td>11.32%</td>
<td>3.52%</td>
</tr>
<tr>
<td>[Men11]</td>
<td></td>
<td></td>
<td>25.37%</td>
<td>6.17%</td>
</tr>
</tbody>
</table>

5 Conclusions and Future Work

A method for an efficient detection of boundaries between layers in an OCT fingerprint scan has been developed. Our method is able to estimate positions of the boundaries at least two orders of magnitude faster than the method by Menrath and Breithaupt [Men11]. Our classification results demonstrate that even by simply analyzing the energies of the detected layers, our approach outperforms the method by Menrath and Breithaupt [Men11], which confirms robustness of the layer detection method and its potential for further de-
Future work would concern further automated analysis of the detected outer and inner fingerprint layers. Development of a new database of OCT fingerprint scans is planned in the framework of the project OCT-Finger II. A new OCT scanning technology with better signal to noise ratios, faster acquisition times and higher resolutions [KCGK12] is expected to provide for data that would allow for reliable analysis of fine structures in an OCT scan. It can be attempted to compare the outer and the inner fingerprint structure, since the patterns are identical if the fingerprint OCT scan is genuine. In case of the artefact fingerprints, a falsely detected inner fingerprint layer would yield different fingerprint pattern than the outer layer. Structure and density of the layers could be also analyzed for liveness detection. Genuine fingerprints typically produce the second inner fingerprint layer in a form of a scattered point cloud, in contrast to a rather continuous thin second layer in artefact fingerprint scans. The presence of three distinct layers or only a single layer provide for a strong evidence that the fingerprint is an artefact. In addition, the structure of the inner layer between the first and the second detected layer could be analyzed. In genuine OCT fingerprint scans, this layer has a specific inner structure and typically contains a number of sweat glands. Verification of the appropriate inner structure would reject the artefact fingerprint scans, since their inner structure between the first and the second layer is determined by the properties of the artefact material.

References


Simulated annealing attack on certain fingerprint authentication systems

Andreas Pashalidis

KU Leuven
ESAT/SCD-COSIC & iMinds
Kasteelpark Arenberg 10
3001 Heverlee, Belgium
andreas.pashalidis@esat.kuleuven.be

Abstract: This paper describes a simple and generic attack against minutiae-based fingerprint authentication systems. The aim of the attack is to construct a fingerprint minutiae template, compliant to ISO/IEC standards, that matches a fixed but unknown target fingerprint. Our attack is expected to be most effective against systems that employ vicinity-based matching algorithms, i.e. systems that divide fingerprints into multiple regions and then compute similarity over these regions. The effectiveness of our attack is experimentally demonstrated against the recently proposed ‘Protected Minutiae Cylinder Code’ (PMCC) scheme.

1 Introduction

Biometric authentication systems work as follows. Initially, users are enrolled in the system. A user’s enrolment involves capturing a biometric trait, such as his fingerprint, and storing the captured data in the form of a ‘reference’ template. An enrolled user who wishes to authenticate himself does so by providing a fresh biometric sample. This is again captured by the system, and transformed into a ‘sample’ template. This sample template is then compared against the stored reference template. If the two templates ‘match’, then the user is deemed authenticated.

In order to decide whether or not two templates match, every biometric authentication system employs an algorithm called the ‘matcher’.

1 The matcher takes as input two biometric templates and outputs a similarity score. During authentication, this score is compared against a fixed threshold value; if a given comparison yields a similarity above the threshold, then the two templates are deemed to originate from the same user. The threshold value determines the accuracy of the biometric authentication system: if the threshold is too high, then too many biometric samples that should be accepted by the system will actually be rejected. If, on the other hand, the threshold is too low, then too many

1The terms ‘comparison’, ‘(dis)similarity measurement’, and ‘(dis)similarity estimation’ are sometimes used in the biometrics literature instead of ‘matching’; see [ISO05b] for a recent attempt to harmonize the use of these and related terms.
samples that should be rejected will actually be accepted. The False Match Rate (FMR) and False Non-Match Rate (FNMR) are two well-known measures, which we use in this paper, that depict the dependency of system accuracy on the threshold value. These measures enable an informed choice for this value, also having in mind the context in which the system is or will be deployed.

Unfortunately, access to the matcher’s output enables an adversary to launch ‘hill climbing’ attacks against the system. In such an attack, the adversary first provides a synthesized biometric sample to the system and observes the matcher’s output. It then iteratively modifies this sample depending on how the matcher’s output reacts to the modifications: modifications which yield worse similarity are discarded while modifications leading to better similarity score are further modified. Using this approach, the adversary may end up with an artificially constructed biometric sample that passes the threshold and can therefore impersonate a particular user. Using template inversion techniques (see, for example, [FJ11] and Chapter 2 of [Nag12]) the adversary may be even able to reconstruct a ‘complete’ matching biometric imitation such as a ‘gummy finger’ [MMYH02]. In this paper, we describe a ‘simulated annealing’ attack, which uses a slightly more sophisticated technique than hill climbing. Our attack focuses on minutiae-based fingerprint systems that use vicinity-based matchers.

Since biometric data is considered to be sensitive personal data, it is important to safeguard reference templates. To this end, special biometric template protection schemes have been developed (see, for example, [Sim12, Nag12]). These schemes encode biometric templates in a way that is irreversible. Among other things, they aim to ensure that, even if an adversary gets access to stored protected biometric templates, this data cannot be abused to reconstruct original biometric data. An adversary, however, with access to protected biometric templates, can launch hill climbing and similar attacks even without access to the matcher’s output; assuming that the matcher logic is not secret, the adversary simply uses its own implementation of the matcher. In this case, the attack can be launched in an offline fashion. Such offline attacks are particularly undesirable because their occurrence cannot be detected. In this paper, we demonstrate the effectiveness of our attack against the ‘Protected Minutiae Cylinder Code’ (PMCC) scheme [FMC12], which is a template protection scheme for fingerprints. That is, we demonstrate our attack in a setting where it can be launched in an offline manner.

We expect our attack to be successful only against fingerprint systems that employ vicinity-based matchers. A vicinity-based matcher roughly works as follows. It first divides all templates into multiple regions. Each (or some) region of a sample template is then ‘paired’ with some region of the reference template on the basis of how similar they are. Subsequently, the matcher computes similarities over the resulting region pairs, and, finally, merges the different similarity scores into a single, consolidated score.² The PMCC scheme employs such a matcher.

The remainder of this paper is organised as follows. The next section surveys some related work, Section 3 describes our attack, and Section 4 describes our experiments. Section 5

---

²The reason we call such matchers ‘vicinity-based’, rather than ‘region-based’, is because we are only aware of schemes where region boundaries are defined with respect to detected minutiae points; that is, a region is defined to be a central minutia’s vicinity.
reports our results and, finally, Section 6 concludes.

2 Related work

Ratha, Conell and Bolle describe a generic attack model for biometric authentication systems and examine the entropy of minutiae templates with respect to matching algorithms that require a minimum number of minutia points to ‘match up’ [BS01, RCB01]. They explain that, if sufficiently many minutiae points are present in the template, then synthesizing a matching template by exhaustively searching through the space of all possible templates, becomes increasingly infeasible. The search strategy employed in this paper ignores large parts of the search space; instead, it divides the template area into a number of smaller regions, and navigates through the space by operating on these these regions individually. Moreover, it is not our goal to reconstruct the original minutiae; the goal is to synthesize a minutiae template that contains a sufficient number of small regions, each with a configuration of minutiae that is ‘good enough’ for the matching algorithm. As a result, the probabilistic analysis of [BS01] does not apply to our approach, and to vicinity-based matchers in general.

Hill climbing techniques have been used to attack various biometric systems. Uludag and Jain, for example, present such an attack [UJ04]. Their attack works against a minutiae-based fingerprint authentication system where matching scores are based on the number of minutiae with a similar location and orientation [JHPB97]. While conceptually similar, the attack in this paper differs in the three following aspects. Firstly, it uses simulated annealing [KGV83], which is a form of hill climbing that is able to escape from local optima. Secondly, it explores the search space in a way that is optimised for vicinity-based matchers; moreover it does not explore the entire search space and its goal is not reconstruct the original minutiae template. Thirdly, the attack is applied against PMCC, i.e. a template protection technique where the adversary is explicitly allowed to access protected templates. This means that the attack can be launched in an offline manner and all safeguards proposed in [UJ04] are not applicable.

We also note that, in order to discourage hill climbing attacks in general, Soutar proposed to use only coarse similarity scores, for example by using large quantization intervals [Sou]. While this countermeasure would make our attack less efficient, it only works if the adversary does not have access to reference templates (or in the case where the matching algorithm is secret).

3 Attack description

The logic of our attack is depicted in the algorithms shown in Figures 1 and 2. Table 1 explains the meaning of the different input parameters. The algorithm of Figure 1 outputs a synthesized ‘gummy’ template compliant to [ISO05a] that is meant to imitate a fixed ‘target’ fingerprint; the attack is successful if this gummy template passes the matcher’s
threshold in an authentication attempt for the user to whom the target fingerprint belongs. How exactly the target template is fixed is outside our scope. The algorithm in Figure 1 simply invokes \( \text{match}(T) \), which is assumed to return the system matcher’s output, where \( T \) is a fingerprint sample compliant to [ISO05a] as generated by our algorithm, and where the reference biometric is the target fingerprint. Note that the target/reference fingerprint does not need to be available in a format compliant to [ISO05a]; our attack merely requires that the matcher handles fresh samples in this or an equivalent format.

The attack algorithm starts by composing an initial ‘current’ template \( (T) \) which consists of multiple, \( ha \times va \) to be precise, vicinities. Each of these vicinities is populated with a number of minutiae that have, within the vicinity, random locations and orientations. The number of minutiae per vicinity may differ per vicinity and is randomly chosen between \( \nu_{\text{p}_{\text{min}}} \) and \( \nu_{\text{p}_{\text{max}}} \). The algorithm then constructs a candidate template \( (T') \) by replacing one of the vicinities of the initial template by a fresh, randomly generated, vicinity. If the candidate template yields a better matching score than the current template, then the current template is replaced by the candidate template, and the procedure is repeated for \( REP \) times. Sometimes, the algorithm also replaces the current template with a candidate that yields a worse similarity score. This is done according to a ‘cooling schedule’ which influences the way in which the search avoids getting trapped in local optima. Since we expect that no single cooling schedule will work efficiently against all systems, we do not define this schedule in detail here; we describe the particular schedule we used in our experiments in Section 4. In the end, the algorithm outputs the template which yielded the best similarity throughout the entire search \( (T_{\text{best}}) \).

| \( w \) | Template width, in pixels |
| \( h \) | Template height, in pixels |
| \( ha \) | Number of vicinities, horizontal |
| \( va \) | Number of vicinities, vertical |
| \( \nu_{\text{p}_{\text{min}}} \) | Minimum number of minutiae per vicinity |
| \( \nu_{\text{p}_{\text{max}}} \) | Maximum number of minutiae per vicinity |
| \( r_{\text{min}} \) | Minimum vicinity radius, in pixels |
| \( r_{\text{max}} \) | Maximum vicinity radius, in pixels |
| \( REP \) | Number of search steps until gummy template is output |

### 4 Experimental setup

In order to demonstrate the effectiveness of our approach, we conducted a number of experiments which were driven by the FVC2006 fingerprint database DB2_A [JFGR07]. This database is a collection of 1680 bitmap images, each of size \( 400 \times 560 \) pixels, that depict the fingerprints of 140 fingers, with 12 different impressions per finger. In or-
**ConstructGummyTemplate** (input: $w, h, ha, va, vp_{\min}, vp_{\max}, r_{\min}, r_{\max}, REP$)

1. Generate an initial template $T$ by dividing the area of $w \times h$ pixels into $ha \times va$ non-overlapping rectangles, each of size $(w/ha) \times (h/va)$ pixels. For each rectangle:
   (a) Compute $(c_x, c_y)$ as the coordinates of the center of the rectangle.
   (b) Run **RandomVicinity**($c_x, c_y, vp_{\min}, vp_{\max}, r_{\min}, r_{\max}$) and add the resulting vicinity to $T$.

2. Set $sc = \text{match}(T)$.

3. Set $sc_{\text{best}} = sc$ and $T_{\text{best}} \leftarrow T$.

4. Repeat $REP$ times:
   (a) Set $T' \leftarrow T$.
   (b) Pick a random vicinity $V$ in $T'$, and replace it with a fresh random vicinity as output by **RandomVicinity**($c_x, c_y, vp_{\min}, vp_{\max}, r_{\min}, r_{\max}$), where $(c_x, c_y)$ are the coordinates of $V$'s center.
   (c) Set $sc' = \text{match}(T')$.
   (d) If $sc' > sc_{\text{best}}$, then set $sc_{\text{best}} \leftarrow sc'$ and $T_{\text{best}} \leftarrow T'$.
   (e) If $sc' > sc$, then set $sc \leftarrow sc'$ and $T \leftarrow T'$; otherwise check whether or not the cooling schedule permits moving to worse solution candidates. If it does, then set $sc \leftarrow sc'$ and $T \leftarrow T'$.

5. Output the minutiae in the vicinities of $T_{\text{best}}$.

Figure 1: Simulated annealing attack

**RandomVicinity** (input: $c_x, c_y, vp_{\min}, vp_{\max}, r_{\min}, r_{\max}$)

1. Pick a random radius $r \in \{r_{\min}, r_{\min} + 1, \ldots, r_{\max}\}$ and a random vicinity population $vp \in \{vp_{\min}, vp_{\min} + 1, \ldots, vp_{\max}\}$.

2. Start with an empty vicinity $V$.

3. For all values of $i \in \{1, 2, \ldots, vp\}$:
   (a) Pick two random offsets $x', y' \in \{1, 2, \ldots, r\}$, a random angle $\theta \in [0, 2\pi]$, and two random signs $s_x, s_y \in \{-1, 1\}$.
   (b) Generate minutiæ $m_i = (c_x + s_x x', c_y + s_y y', \theta)$ and add it to $V$.

4. Output $V = \{m_1, m_2, \ldots, m_{vp}\}$.

Figure 2: Generating a random vicinity
order to prepare our experimental environment, we extracted minutiae templates compliant to [ISO05a] from these images using the fjfxSample command line utility. Since this utility only accepts ‘portable gray map images’, we first converted the images into this format using the bmp2pnm utility. Unfortunately, fjfxSample was unable to extract minutiae from one of the images (fifth finger, first impression); hence we ended up with 1679 templates.

Then we derived protected fingerprint templates for each of these 1679 templates using the PMCC scheme [FMC12]. For this, we used the PMCC implementation made available by University of Bologna’s BioLab, in particular version 1.3 of the Minutia Cylinder-Code Software Development Kit that is available on their website. In fact, we derived two datasets, each consisting of 1679 PMCC templates. For the first dataset we used parameter value $K = 64$ and for the second dataset we used parameter value $K = 128$. Hence, we call the two datasets ‘K64’ and ‘K128’.

The parameter $K$ influences the length of PMCC templates; higher values lead to lengthier PMCC templates and better system accuracy [FMC12]. The reason we chose these two values is because (a) we expect the templates derived under these parametrisations to be more resistant to our generic attack than templates derived with smaller values for $K$, and (b) the results in [FMC12] show that these values yield better accuracy than any other examined values.

After completing the above preparations, we started our main experiments in which we aimed to separately attack each of the 1679 templates in both the K64 and the K128 dataset, using the algorithm from Figure 1. We aimed to conduct three experiment sets for each of the K64 and the K128 dataset, namely one experiment set for each of the repetition values REP = 300, 500, 1000. That is, we aimed to conduct six experiment sets in total, with each experiment set generating a collection of 1679 ‘gummy’ fingerprint templates. Unfortunately, due to the non-optimised nature of PMCC code and its unstable integration of our attack code, we were unable to complete our experiments. However, for all target templates processed so far, our algorithm has generated a gummy template, and the results shown in the next section are based on the gummy templates that were generated at the time of writing; more precisely, for $K = 64$, and for REP = (300, 500, 1000), (501, 508, 217) gummy templates were generated respectively while, for $K = 128$, the corresponding numbers are (416, 476, 0).

The other attack parameter values we used in our experiments were fixed: $w = 400$, $h = 560$, $ha = va = 4$, $vp_{\text{min}} = vp_{\text{max}} = 4$, and $r_{\text{min}} = r_{\text{max}} = 40$. The cooling schedule we employed caused the algorithm to move to worse candidates with probability $4r/(3\text{REP})$, where $r$ is the number of the current repetition, for the first REP/2 repetitions. For the remaining REP/2 repetitions, the cooling schedule caused the algorithm to behave like a standard hill climbing algorithm, i.e. no longer moving to worse candidates. The above parameter values and cooling schedule were chosen based on manual observations during early experimentation. We did not systematically optimise over the parameter space in an automated fashion.

---

3http://www.digitalpersona.com/fingerjetfx/
4http://biolab.csr.unibo.it
It should be noted that, when enrolling and matching PMCC templates, a number of further parameters, specific to the PMCC scheme, have to be fixed as well. In our experiments, we used the default parameter values, as documented in [FMC12] and provided in the configuration files with which the PMCC software is shipped.

5 Results

Figures 3 and 4 show our main results. The figures show the FNMR and the FMR rates we obtained when executing the ‘FVC protocol’ over the K64 and K128 datasets using the PMCC matcher. More precisely, the FNMR curves were obtained from 9229 ‘genuine’ comparisons: each PMCC template in the dataset was compared to all other templates of the same finger, but without comparing any pair of templates more than once. The FMR curve, on the other hand, was obtained from 9730 ‘impostor’ comparisons: the first PMCC template of each finger was compared to the first sample of all remaining fingers in the dataset, but, again, without comparing any pair of templates more than once.

![Figure 3: FNMR and FMR curves for the K64 dataset](image)

The figures also show three more FMR curves; each of these curves was obtained by comparing each of the ‘gummy’ fingerprint minutiae templates generated by our attack to the corresponding enrolled PMCC template. The three curves show how the amount of
computational effort, i.e. for REP = 300, 500, 1000, affects the effectiveness of the attack. Tables 2 and 3 show the Equal Error Rate (EER), the FMR\textsubscript{1000} and the Zero\textsubscript{FMR} rates as determined from the curves in Figures 3 and 4, as well as the corresponding values reported in [FMC12]. Our slightly higher FVC protocol rates are probably due to the fact that the minutiae extractor used in [FMC12], both for the tuning of PMCC-specific parameters and the subsequent construction of PMCC templates, differs from the extractor used in our experiments.

<table>
<thead>
<tr>
<th>Impostor Templates</th>
<th>EER</th>
<th>FMR\textsubscript{1000}</th>
<th>Zero\textsubscript{FMR}</th>
</tr>
</thead>
<tbody>
<tr>
<td>FVC Protocol on DB2_A reported in [FMC12]</td>
<td>0.32</td>
<td>0.47</td>
<td>1.07</td>
</tr>
<tr>
<td>FVC Protocol on DB2_A</td>
<td>1.7</td>
<td>3.6</td>
<td>4.9</td>
</tr>
<tr>
<td>Gummy templates, REP = 300</td>
<td>13.2</td>
<td>21.2</td>
<td>21.5</td>
</tr>
<tr>
<td>Gummy template, REP = 500</td>
<td>15.2</td>
<td>25.0</td>
<td>25.4</td>
</tr>
<tr>
<td>Gummy template, REP = 1000</td>
<td>17.9</td>
<td>29.8</td>
<td>30.3</td>
</tr>
</tbody>
</table>

Table 2: EER, FMR\textsubscript{1000} and Zero\textsubscript{FMR} rates over the K64 dataset (percentage values)

From our results it is evident that our attack is efficient and effective; even after a number of iterations as small as 300, a significant proportion of impostor matches is achieved. We
also observe that, as the number of iterations grows, so does the success degree of the attack. However, the relationship between the increase of computational effort and the attack success does not seem to be linear. Due to various technical constraints related to implementation details, we were unable to perform tests with more realistic amounts of iterations (e.g. REP = 10^7). Experiments of this order of magnitude therefore remain a topic of future research. While there is certainly a point beyond which additional efforts will only yield diminishing returns, our results suggest that, with only moderate additional computational efforts, the success degree of the attack can still be increased significantly.

### 6 Concluding remarks

One of the limitations of our attack is that it does not take into account constraints imposed by human nature. For example, the attack does not take into account the issue of inter-ridge distance. As a result, some gummy templates may contain minutiae points that are closer to each other than would be possible on a natural fingerprint. This shortcoming can be used to automatically detect an attack template generated by our algorithm. Therefore, a future improvement of our attack would be the incorporation of constraints regarding the issue of inter-ridge distance, and statistical properties of minutiae locations more generally [CM06].

Another interesting future research question is whether or not it is possible to turn our attack, which merely aims to find minutiae templates that are ‘equivalent’ to a given target template, into an attack that actually reconstructs the original target template. Our intuition on this idea is simple: after running the attack on the same target multiple times, each time with a different random initial template, the intersection of all ‘common’ minutiae points of the resulting gummy templates may resemble the original minutiae points.

We envision our attack to become one of the evaluation tools for vicinity-based fingerprint authentication systems. There is certainly a lot of further evaluation work to be done, both with respect to different parametrisations of PMCC, but also with respect to other schemes (e.g. [YBBG10, CFM10, BD10]). Finally, we would like to stress that, if a scheme appears robust against our attack, this does not imply that it is secure in any general sense. We envision our attack to be merely useful as an evaluation tool in the first stages of the development of minutiae-based fingerprint authentication systems; naturally, attacks that are specific to any particular scheme will perform better than the generic one described in this paper.
Acknowledgements

The author would like to thank Matteo Ferrara for his extensive help with setting up and integrating the PMCC toolkit for the experiments, as well as his feedback on earlier versions of this paper. The author would also like to thank Roel Peeters for insightful discussions. This work was supported by the Flemish Government, IWT SBO SPION, FWO G.0360.11N Location Privacy, and by the Research Council KU Leuven: GOA TENSE; and by the European Commission through the FIDELITY project (contract number 284862).

References


Absolute Fingerprint Pre-Alignment in Minutiae-Based Cryptosystems

Benjamin Tams*

Institute for Mathematical Stochastics
University of Goettingen
Goldschmidtstr. 7
D-37077, Goettingen
btams@math.uni-goettingen.de

Abstract: Most biometric cryptosystems that have been proposed to protect fingerprint minutiae make use of public alignment helper data. This, however, has the inadvertent effect of information leakage about the protected templates. A countermeasure to avoid auxiliary alignment data is to protect absolutely pre-aligned fingerprints. As a proof of concept, we run performance evaluations of a minutiae fuzzy vault with an automatic method for absolute pre-alignment. Therefore, we propose a new method for estimating a fingerprint’s directed reference point by modeling the local orientation around the core as a tented arch.  

1 Introduction

In 2002, Juels and Sudan [JS02] proposed the fuzzy vault scheme. It quickly has drawn the attention of biometric researchers as a promising tool for storing fingerprint minutiae templates protected as a part of a security application. While there are other biometric cryptosystems (see [JW99, DORS08, Tru11, JA07]) being considered as a potential tool for protecting fingerprint templates, the fuzzy vault scheme is one of the most dominant ones because it well conceptualizes differences of and partial overlap between fingerprints as errors and erasures, respectively, by means of Reed-Solomon codes. A common approach to protect a fingerprint template via fuzzy vault is to hide its genuine minutiae within a large number of chaff minutiae. The fuzzy fingerprint vault draws its security from the problem in distinguishing genuine from chaff minutiae (without knowing a matching minutiae template). On authentication, a second minutiae template from the alleged same user is used to distinguish genuine from the chaff. If the candidate minutiae have a significant overlap with genuine minutiae, the protected minutiae template can be recovered using techniques

*The support of the Felix Bernstein Institute for Mathematical Statistics in the Biosciences and the Volkswagen Foundation is gratefully acknowledged.

1A program implementing our method for estimating a fingerprint’s directed reference point will be provided for download from http://www.stochastik.math.uni-goettingen.de/biometrics by the date of the conference.
from the discipline of error-correcting codes.

For first implementations of the fuzzy fingerprint vault (e.g., [UJ06, NJP07]) it quickly turned out that brute-force attacks are not only possible but rather easy to perform [SB07, MMT09]. To improve security, the incorporation of minutiae descriptors in addition to mere minutiae has been proposed [NNJ10] which has the potential of providing reasonable brute-force security at a comparably usable authentication performance. However, an intruder having intercepted different application’s databases may link genuine vault correspondences across them via correlation, i.e. cross-matching. Even worse, given two genuine vault correspondences it can be possible to recover the protected templates at quite a high rate via the correlation attack [KY08]. It is known that such attacks via record multiplicity are enabled due to the different vault’s chaff being generated randomly [SB07] and these attacks can obviously be circumvented by rounding genuine minutiae to a rigid system of which unoccupied elements encode the chaff (e.g., see [MMT09] for a discussion of using non-random chaff). An implementation as well as a performance evaluation (assuming a well-solved alignment framework) of this approach can be found in [Tam13].

There remains the risk of attacks that are yielded by a cryptographically non-negligible false acceptance rate, i.e. false-accept attacks. These attacks have the potential of breaking current fingerprint cryptosystems very easily—even if brute-force attacks are impractical to perform. And yet another question remains concerning the information that is leaked from public helper data for assisting in fingerprint alignment on genuine authentication.

On authentication, query fingers have to be pre-aligned such that two vault minutiae matching with query minutiae are of sufficient similarity. Current implementations achieve a reasonably accurate relative pre-alignment by fitting the query fingerprints to auxiliary data publicly stored along with the vault [NJP07, JA07, LYT+08]. This, however, has the inadvertent effect of information leakage about the protected finger from the vault: An adversary having intercepted a vault record may use the additional data to improve off-line attacks. Formal analyses for quantifying the amount of leaked information are missing and may be hard to achieve. Unless analyses yielding negligible information leakage of auxiliary alignment data (given the vault) are available, we should not propagate the use of public alignment helper data.

In [MIK+11] the use of additional data is circumvented. On enrollment, a coarse absolute pre-alignment of the fingerprint is determined before its minutiae are protected. On authentication, as on enrollment, a coarse absolute pre-alignment of the query fingerprint is estimated and then the pre-aligned minutiae are adjusted to well match vault minutiae. The construction, however, does not ensure resistance against cross-matching or the correlation attack. Furthermore, if it is adopted for constructions protecting rounded minutiae to achieve cross-matching resistance (see [JW99, DORS08, KBK+11, Tam13]), refining the alignment of coarsely pre-aligned minutiae will not work anymore. Therefore, absolute fingerprint pre-alignment should not be only coarse but even reasonably robust. Alternatively, constructions using alignment-free features can be considered as in [LYC+10]. However, to date no resistance against cross-matching has been ensured for such constructions. In this paper, we focus on the option of absolute fingerprint pre-alignment to protect minutiae templates in order to avoid the use of additional alignment data.

Minutiae of a fingerprint can be represented w.r.t. a Cartesian coordinate system using
well-known techniques from linear algebra. There is an evident one-to-one correspondence between Cartesian coordinate systems and points constituted with a direction (also see Figure 1). To implement a robust absolute fingerprint pre-alignment, it is essential to robustly estimate directed reference points or, equivalently, an intrinsic Cartesian coordinate system. There have been proposals for estimating stable reference points such as the core [NK98] and the focal point [RA00]. For an intrinsic coordinate system, however, the estimation of a robust intrinsic direction is required in addition.

[Hot09] proposed to estimate a fingerprint’s intrinsic coordinate system using the global quadratic differential model fitted to the fingerprint’s orientation field [HHM08]. The fitted model’s origin was used as the reference point’s coordinate and the direction of the longitudinal axis as its direction. The method requires an estimation of all singular points of the finger, i.e. cores and deltas. This can, however, not be ensured in general—in particular, if some of the singular points are not visible on a fingerprint. Bazen and Gerez [BG01] partitioned a fingerprint into regions containing no singular points and defined hypothetical lines between encompassing regions as the axes of an intrinsic (non-Cartesian) coordinate system. The proposed construction, however, still requires solutions for reliable partitioning the fingerprint into regular regions. We argue with [MMJP09] that there have been no definite solutions enabling reliable absolute fingerprint pre-alignment.\footnote{To the best of the author’s knowledge, no significant advancements have been made in estimating robust absolute pre-alignments of fingerprints since the publication of [MMJP09].}

The paper is outlined as follows. In Section 2 we describe an implementation based on the fuzzy vault scheme adopted from [Tam13] with minor modifications to protect the minutiae of absolutely pre-aligned fingerprints. In Section 3 we propose a new method for estimating a directed reference point in a fingerprint, which yields an intrinsic Cartesian coordinate system to which minutiae can be absolutely pre-aligned. In Section 4, we incorporate our proposed method into the minutiae fuzzy vault of Section 2. Final discussions are given in Section 5.
2 Minutiae Cryptosystem Used for Testing

This paper focuses on the potential of biometric cryptosystems to protect absolutely pre-aligned fingerprints. For testing, we used the cross-matching resistant implementation described in [Tam13] of which source code is publicly available. In this section, we briefly describe its functioning in the language of absolutely pre-aligned minutiae.

Given an absolutely pre-aligned minutiae template, each of its minutia is coarsely quantized first. Let \( m = (a, b, \theta) \) be a minutia at \((a, b)\) and of angle \( \theta \in [0, 2\pi) \). Let \( p_j \) be a point of a hexagonal grid \( \{p_0, \ldots, p_{r-1}\} \) laying within the region in which pre-aligned minutiae can occur. Let \( p_j \) be a hexagonal grid point that best approximates \((a, b)\). Furthermore, let \( j' = \lfloor \theta/(2\pi) \cdot s \rfloor \) where \( s \) denotes a parameter controlling the number of values into which minutiae angles are quantized. Now, the integer \( j + r \cdot j' \) encodes the quantization of \( m \). Let \( x_j, j' \in \mathbb{F} \) denote the finite field element encoding \( j + r \cdot j' \) by a fixed convention. The quantization of the minutia \( m \) is encoded by \( x_j, j' \).

On enrollment, we assume that the minutiae of an absolutely pre-aligned fingerprint are provided and that they are sorted decreasingly with respect to their quality. The feature set \( A \) is defined as to contain the (at most the first \( t_{\text{max}} \)) quantizations of the minutiae. The next step is to bind the template quantized as \( A \) to a secret polynomial \( f \in \mathbb{F}[X] \) of degree \(< k \). This is done as usual by letting the genuine set \( G = \{(x, f(x)) \mid x \in A\} \). To achieve resistance against cross-matching via correlation, every element in \( E \) not contained in \( A \) is used to encode a chaff point. More precisely, \( C = \{(x, y) \mid x \in E \setminus A\} \) where the \( y's \) are chosen uniformly at random from \( \mathbb{F} \) with \( y \neq f(x) \). The vault consists of the union of genuine and chaff points. Furthermore, a cryptographic hash value \( \text{SHA}(f) \) of the secret polynomial is stored along with the vault. Thus the public vault is the tuple \((V, \text{SHA}(f))\) where \( V = G \cup C \).

An intruder having intercepted \((V, \text{SHA}(f))\) can recover \( f \) as well as the template \( A \) by running off-line attacks. From the difficulty in running such attacks the implementation draws its security.

On authentication, an absolutely pre-aligned query minutiae template is provided. The query feature set \( B \) is extracted from the query template in the same way as \( A \) was extracted from the enrolled template. Using \( B \), the unlocking set is built \( U = \{(x, y) \in V \mid x \in B\} \). \( U \) contains exactly \(|A \cap B|\) genuine points. Thus, if \(|A \cap B| \geq k\), the secret polynomial \( f \) can potentially be obtained from \( U \) using a systematic decoder. In [Tam13] the implementation uses a randomized decoding procedure. In particular, the higher the overlap \(|A \cap B|\) the higher is the probability that the correct polynomial \( f \) can be recovered.
3 Directed Reference Point Estimation

To date, there have been no definite solutions enabling absolute pre-alignment of fingerprint minutiae for biometric template protection. In this section, we describe a new method for estimating a reference point with a direction from a fingerprint, thus yielding an intrinsic Cartesian coordinate system to which the minutiae can be absolutely pre-aligned. Like in [Hot09], we use the quadratic differential model [HHM08] as an elementary ingredient to estimate a directed reference point of a fingerprint but our method only requires the presence of the core on the fingerprint and not the presence of all singular points. Our method is based on the assumption that a fingerprint’s orientation field can be modeled as a tented arch in a local neighborhood of its core. Using the quadratic differential model, we fit the orientations of a fixed tented arch to an estimation of the fingerprint’s orientation field. Therein, the orientations being close to a predefined distance from the tented arch’s core are taken into account with a higher weight which reflects our basic assumption. The position of the fitted tented arch’s core serves as the estimated reference point’s location and the direction of the fitted arch’s longitudinal axis as its direction.

3.1 (Tented) Arch Model

\begin{equation}
\psi(z) = \lambda^2 \cdot (z^2 - R^2)^2
\end{equation}

where \( \text{Im}(z) > 0 \) and \( \lambda, R \) are real parameters. \( z \) is the function’s complex variable. The (undirected) orientation \( \varphi \in [0, \pi) \) at the location \((x, y)\) with \( y > 0 \) is given by \( \varphi = 0.5 \cdot \arg(\psi(x + i \cdot y)) \). If \( y \leq 0 \), the orientation is defined to be 0 which can be ensured by letting \( \psi(x + i \cdot y) = 1 \). A tented arch model essentially is an arch of which orientation field is influenced by a core and a delta on its longitudinal axis. We control the positions of the core and the delta by their distance \( d_{\text{core}} \) and \( d_{\text{delta}} \), respectively, to the origin where \( 0 \leq d_{\text{delta}} \leq d_{\text{core}} \). The orientation field of a tented arch can be modeled as
\[ \tau(z) = \psi(z) \cdot \frac{(z-i \cdot d_{\text{core}})^2}{(z-i \cdot d_{\text{delta}})^2} \] (see [HHM08]).

The function \( \tau(z) \) models the orientation field of a tented arch w.r.t. the origin \( 0 + i \cdot 0 \) and abscissa’s direction \( 1 + i \cdot 0 \). To fit \( \tau(z) \) to a fingerprint’s orientation field, we may plug an isometry \( \alpha \cdot z + \beta \) with \( |\alpha| = 1 \). Furthermore, the orientations of the fitted model must be rotated in accordance with the isometry’s rotation part \( \alpha \), i.e., a multiplication with \( \alpha^{-2} \). Thus, a tented arch model w.r.t. the origin \( \omega = -\alpha^{-1} \cdot \beta \) and abscissa axis’ direction \( \alpha^{-1} \) is given by \( \tau_{\alpha,\beta}(z) = \alpha^{-2} \cdot \tau(\alpha \cdot z + \beta) \).

### 3.2 Description of the Method

In this section, we describe how we can fit a tented arch model to an orientation field estimation of a fingerprint. Using a Gaussian function, the orientations being close to a predefined distance from the core are taken into account with a higher weight. Finally, the core of the adjusted tented arch is used as the reference point’s location; its direction is given by the longitudinal axis of the fitted tented arch.

In the following we assume that the parameters of a tented arch (barring translation and rotation) are fixed. These are \( R \), \( \lambda \), \( d_{\text{core}} \), and \( d_{\text{delta}} \). \( R \) controls the abscissa pole’s distance to the origin; \( \lambda \) controls the stretching of the tented arch; \( d_{\text{core}} \) and \( d_{\text{delta}} \) control the distance of the core and delta, respectively, from tented arch’s origin along the longitudinal axis.

For an isometry \( \alpha \cdot z + \beta \), \( |\alpha| = 1 \), we valuate the goodness of \( \tau_{\alpha,\beta}(z) \) with the help of a cost function measuring the agreement of \( \tau_{\alpha,\beta}(z) \) to a fingerprint’s estimated orientation field \( \{(z_j, v_j)\} \): Here \( v_j \) encodes the estimated orientation at \( z_j = x_j + i \cdot y_j \); if \( \theta_j \in [0, \pi) \) is the orientation at \( (x_j, y_j) \) then \( v_j = \cos(2\theta_j) + i \cdot \sin(2\theta_j) \). We can valuate the goodness of \( \tau_{\alpha,\beta}(z) \) as \( \kappa(\alpha, \beta) = \sum_j w_{\alpha,\beta}(z_j) \cdot |\tau_{\alpha,\beta}(z_j) - v_j|^2 \) where \( w_{\alpha,\beta}(z_j) \) denotes the weight of which the orientation at \( z_j \) is taken into account. We design the weight function \( w_{\alpha,\beta}(z) \) as follows. Let \( \gamma_{\alpha,\beta} \) be the core on the ordinate axis of \( \tau_{\alpha,\beta}(z) \), i.e., \( \gamma_{\alpha,\beta} = \omega_{\alpha,\beta} + d_{\text{core}} \cdot i \cdot \alpha^{-1} \) where \( \omega_{\alpha,\beta} = -\alpha^{-1} \cdot \beta \) is the origin of the coordinate system. Now, the weight function the form \( w_{\alpha,\beta}(z) = \exp \left( -\frac{|z - \gamma_{\alpha,\beta} - \rho|^2}{2 \sigma^2} \right) \) where \( \rho \geq 0 \) and \( \sigma > 0 \).

We can fit a tented arch to \( \{(z_j, v_j)\} \) by minimizing the cost function \( \kappa(\alpha, \beta) \) where \( |\alpha| = 1 \). The minimization process that we used, consists of three atomic steps: 1. Perform a global search for an initial model: For example, for \( \alpha = 1 \), we search a complex \( \beta \) such that \( \kappa(1, \beta) \) is small, assuming that the tented arch’s core \( \gamma_1, \beta \) is on the fingerprint’s foreground; 2. rotate the model around the core such that the cost function \( \kappa(\cdot, \gamma_{\alpha,\beta}) \) is minimized; 3. update the translation part \( \beta \) minimizing the function \( \kappa(\alpha, \cdot) \) for fixed \( \alpha \).

The second and third step can be repeated until convergence of \( \alpha \) and \( \beta \). If the core \( \gamma_{\alpha,\beta} \) lays on the fingerprint’s foreground, it is output as the reference point and \( \theta \in [0, 2\pi) \) with \( \exp(i \cdot \theta) = i \cdot \alpha^{-1} \) as its direction. Otherwise, if \( \gamma_{\alpha,\beta} \) is outside the fingerprint’s foreground, we repeat the procedure using another initial \( \beta \). It is possible, that no initial
model yields a core laying on the fingerprint’s foreground. Therefore, we should only try a few (20, say) initial models and report a corresponding error message if none yielded a valid reference point. The first step can be realized by an iteration over a grid laid on the fingerprint’s foreground. The second and third steps can be solved using a steepest descent method for finding local minimums. In the following, we discuss details concerning the three fitting steps.

3.2.1 Initial Model

Let $I = [0, N] + i \cdot [0, M)$ be the region of the fingerprint image. Furthermore, let $S \subset I$ be an estimation of the fingerprint’s foreground. To find an initial model $\tau_{1,\beta}(z)$, we can iterate its core $\gamma_{1,\beta}$ over a finite subset of $S$, e.g., a grid. Therefore, the core $\gamma_{1,\beta}$ and the delta $\delta_{1,\beta}$ must be distinct from the $z_j$. Otherwise, attempting to evaluate the cost function $\kappa(\alpha, \beta)$ results in a division by zero: $\tau_{\alpha,\beta}(\gamma_{\alpha,\beta}) = 0$ and $\tau_{\alpha,\beta}(\delta_{\alpha,\beta}) = 0$.

To ensure that $\gamma_{1,\beta}$ is distinct from the $z_j$, we can choose them to lay on a grid $z_j \in ((h, h + g, h + 2g, \ldots) + i \cdot (h, h + g, h + 2g, \ldots)) \cap I$ with $g > 0$ and $h \geq 0$. Then $\gamma_{1,\beta}$ can be iterated over a grid arranged concurrently to the $z_j$, e.g., $\gamma_{1,\beta} \in ((h + g/2, h + g/2 + g, h + g/2 + 2g, \ldots) + i \cdot (h + g/2, h + g/2 + g, h + g/2 + 2g, \ldots)) \cap S$. By choosing $d_{\Delta}$ carefully, we can ensure that the delta $\delta_{1,\beta}$ is disjoint from the $z_j$. The model $\tau_{1,\beta}(z) \in ((h + g/2, h + g/2 + g, h + g/2 + 2g, \ldots) + i \cdot (h + g/2, h + g/2 + g, h + g/2 + 2g, \ldots)) \cap S$ is chosen as the initial model for which $\kappa(1, \beta)$ is minimal.

3.2.2 Fitting the Direction

Given $\tau_{\alpha,\beta}(z)$, we may rotate it around the core for adjustment to $\{(z_j, v_j)\}$. This will change both $\alpha$ and $\beta$. Let $\xi = \cos(\theta) + i \cdot \sin(\theta)$ be the complex number describing the rotation by the angle $\theta$. More precisely, the rotated model $\tau_{\alpha',\beta'}(z)$ is given by $\alpha' = \xi \cdot \alpha$ and $\beta' = \gamma_{\alpha,\beta} - i \cdot d_{\text{core}} \cdot \xi \cdot \alpha^{-1}$. To find the adjusting rotation angle $\theta$, we may minimize the function $\theta \mapsto \kappa(\alpha', \beta')$ for $\theta \in (-\pi, \pi)$ where $\alpha'$ and $\beta'$ define the rotated model $\tau_{\alpha',\beta'}(z)$ as above. A (local) minimum of the function can be found using a steepest descent method starting with $\theta = 0$.

3.2.3 Fitting the Translation Part

For further refinement, we may update the translation part $\beta$ of $\tau_{\alpha,\beta}(z)$. Therefore, we can minimize the function $(x_1, x_2) \mapsto \kappa(\alpha, x_1 + i \cdot x_2)$. Let $(x'_1, x'_2)$ be a (local) minimum which can be found using a steepest descent method starting with $(\text{Re}(\beta), \text{Im}(\beta))$. Then the updated model is $\tau_{\alpha,\beta'}(z)$ where $\beta' = x'_1 + i \cdot x'_2$. 
4 Training and Evaluation

This paper discusses the potential of biometric cryptosystems to protect absolutely pre-aligned fingerprint minutiae. We used the fuzzy vault implementation described in Section 2 to test our proposed method for automatically estimating a directed reference point from a fingerprint and thus for absolutely fingerprint pre-alignment. In this section, we describe how we determined a good configuration for our proposed method. Furthermore, we describe the result of an evaluation. Throughout, we used minutiae templates corresponding to the FVC 2002 DB2 [MMC+02] database (DB2-B for training and DB2-A for evaluation). The minutiae templates have been extracted using a commercial extractor.\textsuperscript{3} The orientation fields were estimated using the well-known gradient method [KW87] following the description of [MMJP09]. Furthermore, we estimated a fingerprint’s foreground by selecting the largest connected region after Otsu thresholding and then computing its convex hull via Graham scan.

4.1 Training

In [Tam13], the parameters for the vault implementation have been determined during a training in where a good alignment was achieved manually on genuine authentication. We resumed the training to find a good configuration for our tented arch model to implement an automatic absolute pre-alignment. A best configuration observed was $\lambda = 1.81$, $R = 175$, $d_{\text{delta}} = 48$, $d_{\text{core}} = 186$, $\sigma = 45$, and $\rho = 12$ which resulted in 262 (among 280) genuine correspondences with at least $k = 7$ common elements.\textsuperscript{4}

4.2 Evaluation

Table 1: Authentication performances of minutiae-based cryptosystem with absolutely pre-aligned fingerprint with our method on a 3.2 GHz desktop computer using a single processor core. For authentication performances achievable with relative pre-alignment, we refer to [Tam13].

<table>
<thead>
<tr>
<th>$k$</th>
<th>sub-GAR (GAR)</th>
<th>FAR</th>
<th>GDT</th>
<th>IDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>= 7</td>
<td>= 91% (≈ 74.12%)</td>
<td>≈ 0.91%</td>
<td>≈ 0.081 sec</td>
<td>≈ 0.270 sec</td>
</tr>
<tr>
<td>= 8</td>
<td>= 88% (≈ 65.93%)</td>
<td>≈ 0.16%</td>
<td>≈ 0.130 sec</td>
<td>≈ 0.329 sec</td>
</tr>
<tr>
<td>= 9</td>
<td>= 85% (≈ 56.91%)</td>
<td>≈ 0.06%</td>
<td>≈ 0.195 sec</td>
<td>≈ 0.405 sec</td>
</tr>
<tr>
<td>= 10</td>
<td>= 80% (≈ 48.67%)</td>
<td>= 0%</td>
<td>≈ 0.263 sec</td>
<td>≈ 0.462 sec</td>
</tr>
<tr>
<td>= 11</td>
<td>= 73% (≈ 40.20%)</td>
<td>= 0%</td>
<td>≈ 0.351 sec</td>
<td>≈ 0.539 sec</td>
</tr>
<tr>
<td>= 12</td>
<td>= 68% (≈ 32.40%)</td>
<td>= 0%</td>
<td>≈ 0.447 sec</td>
<td>≈ 0.624 sec</td>
</tr>
</tbody>
</table>

\textsuperscript{3}Neurotechnology Ltd. Verifinger SDK 5.0, \url{http://www.neurotechnology.com}.
\textsuperscript{4}The minutiae templates were absolutely pre-aligned using the directed reference points estimated by our method. Each pre-aligned minutiae template was quantized as a subset of a finite field with at most $t_{\text{max}} = 44$ elements using a hexagonal grid of distance $\lambda = 29$ centered in $[-559, 559] \times [-559, 559]$, which covers the region in where minutiae locations can occur. The minutiae angles were quantized into $s = 6$ quanta.
Figure 3: Excerpt from the training — The core (green diamond) and the direction of the longitudinal axis (bold yellow line) give the location and the direction, respectively, of the estimated directed reference point. The blue lines correspond to the orientation of the tented arch. Their transparencies indicate the weight of which the orientations around the core were taken into account.

Using the parameter configuration determined during training, we ran performance evaluations with absolutely pre-aligned minutiae templates following the FVC protocol. We measured the genuine acceptance rate (GAR), false acceptance rate (FAR), avg. time on genuine decoding (GDT), and avg. time on impostor decoding (IDT). In order to allow comparison with other implementations of fingerprint cryptosystems, e.g., [NJP07,NNJ10, LYC+10], we also kept track of the genuine acceptance rate (sub-GAR) accounting only for the first two impressions of each finger. The results can be found in Table 1. We furthermore observed, that the avg. time for estimating a fingerprint’s directed reference point with our implementation (including orientation field estimation and segmentation) was \( \approx 3.05 \) sec per fingerprint. If for a fingerprint the estimation of a valid reference point failed, it was not taken into account, neither for enrollment nor for authentication. For 2.5% of the finger a pre-alignment failure were reported while for the sub-database no pre-alignment failure was reported. As a consequence, the GARs and sub-GARs have been measured from 2,562 and 100 observed genuine authentication attempts, respectively. The FARs were measured from 4,950 impostor authentication attempts.

4.3 Security

In this section, we give a brief security discussion for the above cryptosystem with absolute pre-alignment. For more details, we refer to [Tam13]. First, we note that the implementation is resistant against cross-matching via correlation and the correlation attack. Second, our implementation’s resistance against brute-force attacks is at least as in [Tam13]. Thus, for simplicity, we assume the same number of chaff points \( n = 1,452 \) if we analyze our construction against brute-force attacks. However, analyses of false-accept attacks correspond to a more realistic estimate of the system’s overall security. We used the method in [Tam13] to estimate false-accept security. The results can be found in Table 2.
Figure 4: Excerpt from the evaluation — The estimated directed reference points are quite robust (a)-(c) and our method has the potential to work even for arches (d)-(f).

Figure 5: (a)–(c) Examples where the estimation of the directed reference point was evidently unstable because of the fingerprint ridge flow being not well tented. (d)–(f) Examples where failures were reported due to the fingers being not well provided by the users.

Table 2: Security evaluation — The times correspond to a 3.2 GHz desktop computer using a single processor core.

<table>
<thead>
<tr>
<th>secret size $k$</th>
<th>brute-force security $(\frac{44}{k})/(\frac{1}{k^{452}})$</th>
<th>false-accept security FAR (for $D = 1$)</th>
<th>expected time for a successful brute-force-attack</th>
<th>expected time for a successful false-accept attack</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>$\approx 2^{-36}$</td>
<td>$\approx 2^{-22}$</td>
<td>$\approx 2$ days</td>
<td>$\approx 9$ sec</td>
</tr>
<tr>
<td>8</td>
<td>$\approx 2^{-41}$</td>
<td>$\approx 2^{-24}$</td>
<td>3–4 months</td>
<td>$\approx 1$ min</td>
</tr>
<tr>
<td>9</td>
<td>$\approx 2^{-47}$</td>
<td>$\approx 2^{-27}$</td>
<td>$\approx 14$ years</td>
<td>$\approx 11$ min</td>
</tr>
<tr>
<td>10</td>
<td>$\approx 2^{-52}$</td>
<td>$\approx 2^{-30}$</td>
<td>$\approx 703$ years</td>
<td>1–2 hours</td>
</tr>
<tr>
<td>11</td>
<td>$\approx 2^{-57}$</td>
<td>$\approx 2^{-33}$</td>
<td>$\approx 35,198$ years</td>
<td>13–14 hours</td>
</tr>
<tr>
<td>12</td>
<td>$\approx 2^{-63}$</td>
<td>$\approx 2^{-36}$</td>
<td>$\approx 1,78 \cdot 10^6$ years</td>
<td>5–6 days</td>
</tr>
</tbody>
</table>

5 Discussion

In this paper, we discussed a countermeasure to information leakage from auxiliary alignment data in minutiae-based cryptosystems. As a proof of concept, we performed test
with a cross-matching resistant cryptosystem protecting absolutely pre-aligned minutiae templates. Therefore, we proposed a new method for automatically estimating a directed reference point from a fingerprint which yields a method for absolutely pre-aligning fingerprints. Another nearby approach is to protect alignment-free features as in [LYC+10] but it is not yet clear how cross-matching resistance can be ensured for such implementations and how this would affect the authentication performances. Currently, the authentication rates that we achieved with absolute pre-alignment are clearly inferior to authentication rates achievable in a well-solved alignment framework (compare Table 1 with Table 5 in [Tam13]). However, the method for estimating a directed reference point and the biometric template protection scheme are building blocks which can be replaced by possibly more robust implementations. The search for methods improving the authentication rates of biometric cryptosystems protecting absolutely pre-aligned minutiae will be part of our future research. Finally, we stress that single-finger cryptosystems are currently not sufficient for providing a reasonable amount of security at a usable genuine acceptance rate (e.g., see Table 2). Our research has the purpose to yield to and to improve multi-finger cryptosystems—or even systems for protecting multiple biometric modalities. First steps have already been made, e.g., see [MIK+11, NNJ12].

References


Quality driven iris recognition improvement

Sandra Cremer, Nadege Lemperiere, Bernadette Dorizzi, Sonia Garcia-Salicetti

Thales Communications & Security
20-22 rue Grange Dame Rose, 78141 Velizy Cedex
Institut Mines-Telecom, Telecom SudParis
9 rue Charles Fourier, 91011 Evry Cedex

sandra.cremer@thalesgroup.com, nadege.lemperiere@thalesgroup.com
bernadette.dorizzi@telecom-sudparis.eu, sonia.garcia@telecom-sudparis.eu

Abstract: The purpose of the work presented in this paper is to adapt the feature extraction and matching steps of iris recognition to the quality of the input images. To this end we define a GMM-based global quality metric associated to a pair of normalized iris images. It quantifies the amount of artifact in these images as well as the amount of texture in artifact-free regions. First we use this metric to adjust, for each pair of irises, the proportion of the normalized image selected on a local quality criteria for feature extraction. This approach is tested with two matching techniques: one performs a bit to bit comparison of binary feature vectors and the other one computes local cross-correlations between real valued vectors. We show that our approach is effective with both techniques. Then we use our metric to choose the matching technique that is best adapted to each image pair in order to make a good compromise between accuracy and speed.

1 Introduction

Iris possess a very rich pattern that is believed to be different between persons, therefore iris recognition has become one of the most reliable and accurate biometric identification systems available. A detailed survey on iris recognition was published by Bowyer et al. [BHF08].

The first successful algorithm for iris recognition was proposed by Daugman [Dau93]. After a preprocessing of the iris that includes a segmentation and a normalization step, this algorithm uses convolution with two dimensional Gabor filters to extract the texture from the polar iris image. Each filter’s phase response is then quantized in a pair of bits, so the information from the iris image is represented by a binary code. Following this, different images of irises can be compared through their binary codes using bitwise operations.

Daugman’s algorithm is used in most of today’s commercial iris recognition systems for it is very efficient computationally and its evaluations on very large databases have proven its high recognition accuracy [Dau04]. However it has been shown that its performance considerably decreases on images with degraded quality, for instance images affected by defocus [KDS+02], occlusion [KDS+02] or pupil dilation [HBF09].
A lot of research has been done to quantify iris image properties that are influential on recognition performance. The National Institute of Standards and Technology has conducted the Iris Quality Calibration and Evaluation (IQCE) [TGS] to this end. A number of global quality metrics were defined to qualify iris images. They can be used to reject low quality images in the recognition process and thus improve performance. Some work has also been done on defining local quality metrics, namely by Chen et al. [CDJ05], Li and Savvides [hLS09] or Cremer et al. [CDGSLpt]. They are used to diminish the impact of low quality regions on the computation of the Hamming distance when using techniques for feature extraction and matching that are similar to Daugman’s.

In [CDGSLpt], we showed that our local quality metric enables to localize regions of the polar iris images that contain artifacts, as well as to quantify the amount of texture in artifact-free regions. In this paper we exploit this local quality metric to define a global quality metric. Unlike most global quality metrics defined in literature, it is computed from the polar iris images. It qualifies these images in terms of amount of artifacts and of iris texture which enables in particular to detect segmentation inaccuracies. Our aim is to use it to improve the feature extraction and matching steps of the recognition process when dealing with different levels of quality image degradations.

More specifically, the purpose of this work is to combine the exploitation of the local quality metric presented in [CDGSLpt] and of our global quality metric in order to adapt the feature extraction and matching steps to each pair of iris images to compare. In [CDGSLpt] local image quality had been incorporated in Daugman’s algorithm to select the best regions in the polar iris images in terms of occlusions and amount of texture. It had been shown that the proportion of the initial image selected is a parameter that is very influential on performance. Here we go further and show that the optimal proportion of the initial image to select is directly linked to the global quality of the image pair to compare. We therefore chose an approach in which we use a local quality metric to localize the regions to select for feature extraction and matching and a global quality metric to decide how many regions should be selected. We test this approach with two different techniques for feature extraction and matching: Daugman’s [Dau93] and Krichen’s [KGSD09]. We chose Krichen’s technique because it is presented as being more robust to degraded input iris images. Krichen et al. keep Daugman’s idea of using Gabor phases to describe the iris image, however they use a real-valued feature vector to represent the iris and perform local correlation between these vectors to compute a similarity score.

Moreover we exploit our global quality metric to decide which feature extraction and matching technique out of Daugman’s or Krichen’s is the best adapted to each iris pair in order to make a good compromise between recognition accuracy and speed. The main novelty of this paper is to use a global quality metric to adapt the recognition process to each pair of images instead of using it to eliminate low quality images, as is done in most systems.

This paper is organized as follows. Section 2 describes briefly the local quality measure defined in [CDGSLpt] as well as the region selection strategy for feature extraction and matching. Section 3 shows how the proportion of the image that is selected can be adapted to the global quality of images, both when it is used with Daugman’s algorithm and Krichen’s. In Section 4, we show that one or the other algorithm performs best de-
pending on the value of this global quality measure. Finally, in Section 5, we propose an iris recognition process in which we combine the local and global quality measures. We work on two databases ND-IRIS-0405 [PSO+10] and CASIA-IrisV3-Lamp [CI] for they are large databases that contain images with a large range of qualities. The ND-IRIS-0405 database contains 64,980 images from 356 unique subjects and the CASIA-IRISV3-Lamp database contains 16,212 irises from 411 subjects.

2 Using local quality to select the right regions for feature extraction

2.1 Definition of the exploited local quality measure

In the same way as in [CDGSLpt], we used a Gaussian Mixture Model (GMM) to characterize high quality sub-images. The sub-images that we have chosen to train the GMM are free from occlusion, well-focused and highly textured. Four local observations, measured in a neighborhood of each pixel form the input vector of our GMM: the local mean, local variance and local contrast in addition to the values of the pixel grey-level.

The quality measure associated to a sub-image \( w \) is obtained by applying the formula 1, where \( a.b \) is the size of the sub-image \( w \) and \( x_i \) is the input vector of our GMM associated to a pixel indexed \( i \). \( P(x_i|\lambda) \) is the likelihood given by the model \( \lambda \) to the input vector \( x_i \). \( LL_{train} \) is the mean log-likelihood on the training set, computed over the training set of sub-images and the spatial region of each sub-image.

\[
Q(w) = \exp\left(-\frac{1}{a \cdot b} \sum_{i=1}^{a \cdot b} \left| \log(P(x_i|\lambda)) - LL_{train}\right| \right)
\]  

(1)

\( Q(w) \) will be comprised between 0 and 1. It is shown in [CDGSLpt] that it allocates the highest values to sub-images that are artifact-free and highly textured and the lowest values to sub-images containing artifacts. Intermediate values are given to artifact-free and lowly textured sub-images.

2.2 Selection of the best regions for feature extraction

It has often been demonstrated as in [KDS+02] that iris recognition performs best when the features have been extracted in high quality regions of the polar iris image, namely non-occluded and highly textured zones. As in [CDGSLpt], we use the local quality measure described here above to find those regions. More specifically we proceed in 4 steps for each pair of gallery and probe iris images to compare:

Step 1: In each polar iris image sized 64x512 pixels, we compute the local quality of \( M \) sub-images sized 11x51 pixels. \( M \) is the same for all the images in the database.

Step 2: For each pair of gallery and probe polar images, we fuse the gallery local qualities
with the probe ones. This is done by selecting, for each one of the $M$ pairs of sub-images, the minimum between the gallery and the probe quality value.

**Step 3:** We sort the pairs of (gallery, probe) sub-images according to their local quality.

**Step 4:** We keep the $N$ best sub-images for feature extraction and matching.

This process is done for different translations of the probe polar image in order to allow in-plane rotations of the raw iris in the image.

In [CDGSLpt], this strategy has been set-up with Daugman’s feature extraction and matching algorithm on three public databases: ND-IRIS-0405 [PSO$^{+10}$], CASIA-IrisV3-Twins [CI] and CASIA-IrisV3-Interval [CI]. It was shown that, this strategy performs better than other strategies that select the regions used for recognition by computing a segmentation mask, for it allows to keep control on the quantity of information exploited through the choice of the value $N$, for a set value of $M$. Moreover, different values of the ratio $N/M$ have been tested. It has been demonstrated that when using the same value of $N/M$ for all (gallery, probe) pairs of images, there is an optimal value of $N/M$ equal to 1/3 whichever database was used and whichever was the value of $M$. This ratio represents the proportion of the polar iris image that is exploited for recognition. Its optimal value compromises between minimizing the amount of occluded regions taken into account for feature extraction and maximizing the amount of information available for matching. Indeed, when this ratio is too high, regions of bad quality are taken into account for recognition. On the other hand, when it is too low, the amount of information exploited for matching is not high enough and the impostor distribution is biased.

## 3 Using global quality to decide what proportion of the polar iris should be kept

In Section 2, the proportion of the polar iris image exploited for feature extraction and matching was the same for all (gallery, probe) pairs of images to compare. In this section, we wish to demonstrate that the optimal value of this proportion, represented by the ratio $N/M$, can change for each (gallery, probe) pair depending on the pair’s quality. This is one of the novelties of the present work.

It appears obvious that the less occluded an image is, the more regions of the image can be used for matching without taking into account occlusions. Therefore the higher the image quality is in terms of occlusion, the higher the optimal value of the ratio $N/M$ should be. Moreover, non-occluded regions that contain very few texture won’t necessarily bring discriminating information for matching. So the amount of texture can also have an influence on the optimal value for $N/M$.

To qualify a pair of (probe, gallery) images globally in terms of occlusion and amount of texture we have exploited the local quality described in Section 2. Since we want to define the quality of each pair of (probe, gallery) images, we fuse the $M$ local qualities of the probe image with the $M$ local qualities of the gallery image. This is done by selecting the minimum value between the gallery and the probe local quality value. Indeed,
when matching the features corresponding to a gallery and a probe sub-image, one of both sub-images containing an artifact is enough for the chances of false rejection to increase drastically, so the selection of the minimum quality seems logical.

As a result, we define the global quality measure associated to a pair of images as the mean of the $M$ fused local measures corresponding to the $M$ pairs of (gallery, probe) sub-images as in equation 2

$$GlobQ = \frac{1}{M} \sum_{i=1}^{M} \min(Q_{gal}^{w}(w_i), Q_{probe}^{w}(w_i))$$

(2)

$GlobQ$ is directly linked to the local measure used for feature selection. We therefore expect the ratio $N/M$ to depend on $GlobQ$.

3.1 Experiments using Daugman’s algorithm

As in [CDGSLpt], we have used the OSIRIS-V2 version of Daugman’s algorithm. However we have replaced the segmentation module by the one described in [LDGS+13], in order to better model the iris’ inner and outer boundaries. The segmented images are then normalized using Daugman’s rubber-sheet model [Dau93] to perform a polar transformation of the iris. After this, we apply the strategy described in Section 2.2 to select the $N$ best sub-images out of the total number $M$, for each (probe, gallery) pair of polar images. $M$ is set to a constant value throughout this experiment. The $N$ selected sub-images will then be convolved with two dimensional Gabor filters and each filter’s phase response will be quantized in a pair of bits. So the information of each iris image is summarized in a binary code. The probe and gallery binary codes will be compared by the computation of their Hamming distance.

We have applied this recognition process on images from the ND-IRIS-0405 and the CASIA-IrisV3-Lamp databases. More specifically, in order to demonstrate the link between the global quality defined in formula 2 and the optimal value of $N/M$, we have generated matching lists for these databases and then divided them into five categories according to the value of the global quality associated to each (gallery, probe) pair. Then we have applied our recognition process on each category for different values of $N$ and compared the achieved performance. The values of the global quality measure that define the boundaries of each of the five categories were determined by analyzing the range of the global quality on the entire list and dividing the resulting quality interval into five equal-sized intervals. Table 1 presents the percentage of the full matching list in each category for the ND-IRIS-0405 and CASIA-IrisV3-Lamp databases and Figure 1 gives examples of pairs of (gallery, probe) polar iris images belonging to categories 1, 3 and 5.

We have started by applying our recognition process to the first category of matches ($0 < GlobQ < 0.13$) for different values of $N$. The values tested for $N/M$ were 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.9. Figure 2 presents the ROC curves, i.e the False Rejection Rate (FRR) as a function of the False Acceptance rate (FAR), obtained for 4 values of $N/M$.
<table>
<thead>
<tr>
<th>GlobQ</th>
<th>[0; 0.13]</th>
<th>[0.13; 0.26]</th>
<th>[0.26; 0.39]</th>
<th>[0.39; 0.52]</th>
<th>[0.52; 0.65]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND-IRIS</td>
<td>11%</td>
<td>16%</td>
<td>34%</td>
<td>33%</td>
<td>6%</td>
</tr>
<tr>
<td>CASIA</td>
<td>13%</td>
<td>21%</td>
<td>39%</td>
<td>24%</td>
<td>3%</td>
</tr>
</tbody>
</table>

Table 1: Percentage of the full matching list for each category of GlobQ.

![Figure 1: Examples of (gallery,probe) polar iris image pairs for different global qualities.](image)

on this category corresponding to the lowest-quality images of the CASIA-IrisV3-Lamp database.

As we can see, the best performance is achieved for $N/M = 0.4$. We have carried out the same experiment on the four other categories of matches. Table 2 presents the values of $N/M$ that lead to the best performance for each category for the CASIA-IrisV3-Lamp and the ND-IRIS-0405 databases.

As we can see, the optimal value of $N/M$ is the same for the two first categories and increases with GlobQ for the three other ones. This can be explained by the fact that a higher value of GlobQ implies that the quality of the images is better in terms of occlusion and amount of texture, so taking into account more regions for feature extraction and matching adds useful information for recognition. It is interesting to note, that the optimal value of $N/M$ is the same for categories 1 and 2, which means that when GlobQ goes beneath a certain threshold, $N/M$ stays stable and does not decrease anymore. One reason for this is that this value of $N/M$ is the value underneath which too few information is taken into account for the matching to be significant. Consequently, under a certain value of GlobQ, taking into account regions of low quality has a less negative impact on recognition performance than taking into account too few regions.

Table 2 shows that the optimal value of $N/M$ for $(0.39 < \text{GlobQ} < 0.52)$ is 0.6. On the

<table>
<thead>
<tr>
<th>GlobQ</th>
<th>[0; 0.13]</th>
<th>[0.13; 0.26]</th>
<th>[0.26; 0.52]</th>
<th>[0.39; 0.52]</th>
<th>[0.52; 0.65]</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND-IRIS</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>CASIA</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Table 2: Optimal values for $N/M$ for different databases and different values of GlobQ when using Daugman’s algorithm.
other hand, Figure 2 shows that using $N/M = 0.6$ for $0 < GlobQ < 0.13$, instead of $N/M = 0.4$ leads to a 30% increase of the FRR for an FAR=$10^{-4}$. This demonstrates that setting $N/M$ appropriately with the knowledge of $GlobQ$ can considerably improve performance compared to the use of an inadequate value of $N/M$.

Moreover, for a given range of $GlobQ$, the optimal value for $N/M$ is the same whether the exploited database is ND-IRIS-0405 or CASIA-IrisV3-Lamp. So this value only depends on the image quality.

### 3.2 Experiments using Krichen’s algorithm

We have conducted the same type of experiment with the recognition algorithm proposed by Krichen et al. [KGSD09]. More specifically we have segmented and normalized the iris images in the same way as is section 3.1. and applied our strategy to select the $N$ best sub-images in polar iris images. We have then performed local normalized cross-correlations on the Gabor phases of these sub-images to compare the iris images. So the recognition process we have used here for a (gallery, probe) pair can be cut into the following steps after segmentation and normalization of the gallery and probe images:

**Step 1:** computation of the local qualities of $M$ sub-images from the galley and probe images

**Step 2:** fusion of the local qualities of the gallery and probe sub-images

**Step 3:** selection of the $N$ best sub-images for convolution with 2D Gabor filters

**Step 4:** local normalized cross-correlations for the $N$ pairs of (gallery, probe) sub-images

**Step 5:** similarity score computation

The 3 first steps are identical to the ones done in section 2.2 with $M$ set to a constant
value. The last steps comply with Krichen’s algorithm [KGSD09]: instead of quantizing the Gabor phases we perform local normalized cross-correlation between the phase sub-images allowing a possible translation between gallery and probe sub-images. Each local correlation leads to a Peak to Side lobe Ratio (PSR) value and a Peak Position (PP). The similarity score (SS) is computed with the following formula, where $std$ represents the standard deviation.

$$SS = \frac{\text{mean}_{i \in [1;N]}(\text{PSR}(w_i))}{\text{std}_{i \in [1;N]}(\text{PP}(w_i))}$$  \hspace{1cm} (3)$$

This process for iris recognition has been applied on the same five categories of matching lists as in 3.1, generated from ND-IRIS-0405 and CASIA-IrisV3-Lamp. Different values have been tested for $N$. The optimal value of $N/M$ obtained for the different categories for each database are exactly the same as the ones obtained when using Daugman’s feature extraction and matching technique that are presented in Table 2. So once again a relation appears between $N/M$ and $GlobQ$ and it is interesting to note that this relation does not seem to depend on the database or on the technique used for feature extraction and matching.

### Table 3: FRR for FAR=$10^{-3}$ on the ND-IRIS database for different values of GlobQ and $N/M$ when using the recognition process on Daugman’s and Krichen’s techniques.

<table>
<thead>
<tr>
<th>GlobQ</th>
<th>$[0; 0.13]$</th>
<th>$[0.13; 0.26]$</th>
<th>$[0.26; 0.52]$</th>
<th>$[0.39; 0.52]$</th>
<th>$[0.52; 0.65]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N/M$</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Daugman</td>
<td>33%</td>
<td>11%</td>
<td>3%</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>Krichen</td>
<td>19%</td>
<td>6%</td>
<td>2%</td>
<td>1%</td>
<td>0%</td>
</tr>
</tbody>
</table>

4 Using global quality to choose between two classifiers

We have demonstrated that the global quality we have defined in section 3 can be used to improve performance by adjusting the proportion $N/M$ of the images used for feature extraction and matching, whether the technique used is Daugman’s or Krichen’s. In this section we present another useful way of exploiting this global quality measure.

We compare the performance of Daugman’s and Krichen’s algorithms on the five categories of image lists described in section 3.1 when they are used with the local-quality-based region selection strategy for feature extraction as in sections 3.1 and 3.2. We use the results from Table 2, to set the value of $N/M$ for each category as the value that maximizes performance. The comparative performance of the Daugman and Krichen based processes is presented in Table 3 and Table 4 for the ND-IRIS-0405 database and CASIA-IrisV3-Lamp database respectively.

It is interesting to notice that the Daugman based algorithm performs less well than the one based on Krichen’s technique on the categories of lower global quality. However the
Table 4: FRR for FAR=10^{-3} on the CASIA-IrisV3-Lamp database for different values of GlobQ and N/M when using the recognition process on Daugman’s and Krichen’s techniques.

<table>
<thead>
<tr>
<th>GlobQ</th>
<th>[0; 0.13]</th>
<th>[0.13; 0.26]</th>
<th>[0.26; 0.52]</th>
<th>[0.39; 0.52]</th>
<th>[0.52; 0.65]</th>
</tr>
</thead>
<tbody>
<tr>
<td>N/M</td>
<td>0.4</td>
<td>0.4</td>
<td>0.5</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Daugman</td>
<td>17%</td>
<td>11%</td>
<td>3%</td>
<td>1%</td>
<td>0%</td>
</tr>
<tr>
<td>Krichen</td>
<td>13%</td>
<td>7%</td>
<td>2%</td>
<td>1%</td>
<td>0%</td>
</tr>
</tbody>
</table>

performance is equivalent on the category of best global quality.

This can be explained by the fact that Krichen’s algorithm does local correlations between real feature vectors, whereas Daugman’s algorithm does a bit to bit comparison allowing only global radial translations. Therefore Krichen’s algorithm is more robust to local distortions as well as a reduced amount of texture available than is Daugman’s. Both of these elements are linked to our global quality measure. Indeed local distortions mostly occur in the presence of occlusions which are detected by our quality measure. Moreover the amount of texture is also directly measured by our quality measure. We can therefore understand why Krichen’s algorithm performs better than Daugman’s on pair of images that have a low global quality according to our measure.

Moreover, it is important to note that the algorithm based on Krichen’s technique is much slower. The best compromise between speed and precision would therefore be to apply Krichen’s technique only on the image pairs for which it will perform best. In practice this can be done thank to our global quality measure, since the benefit of Krichen’s algorithm towards Daugman’s depends on GlobQ.

5 Final strategy

Based on the results presented in Section 3 and 4, we propose a complete iris recognition strategy that combines Krichen’s and Daugman’s algorithms as well as the local and global use of a GMM-based quality measure.

More specifically, for each pair of iris images to compare, we compute the associated local and global quality measures. We then set the proportion of the polar images to keep for feature extraction depending on the value of the global quality value, based on the results presented in Table 2. The regions to keep are chosen as the ones with the highest local quality values. Following this we use the global quality value once more to decide which feature extraction and matching technique to apply. If GlobQ < GlobQ_{th} we perform the correlation-based method described in section 3.2. Otherwise we perform the bit-to-bit comparison method described in section 3.1.

Based on the results of Section 4, we have set the value of the threshold GlobQ_{th} to 0.4. Indeed, since Krichen’s technique is more than one hundred times slower than Daugman’s, we only wish to use it when it performs significantly better than Daugman’s in terms of accuracy. Table 3 shows us that this is the case when the global quality is under 0.4.
Figure 3: ROC curves for iris image comparisons of the ND-IRS-0405 database for which $GlobQ < 0.4$ (upper graph) or $GlobQ \geq 0.4$ (lower graph).

Figure 3 presents the ROC curves obtained when applying the iris recognition we propose to image pairs from the ND-IRIS-0405 database. Note that the images used for these comparisons are different ones than those used in sections 3 and 4. To plot these ROC curves we have separated the comparisons in two categories: those for which $GlobQ < GlobQ_{th}$ and those for which $GlobQ \geq GlobQ_{th}$. We have compared our strategy to two other algorithms:

**Krichen04**: using the local quality-based region selection strategy with the same value of the $N/M$ parameter for all images: $N/M = 0.4$ and Krichen’s feature extraction and matching technique

**Daugman06**: using the local quality-based region selection strategy with the same value of the $N/M$ parameter for all images: $N/M = 0.6$ and Daugman’s feature extraction and matching technique

For each of these two algorithms, the constant value of $N/M$ for all comparisons has been chosen as the one that maximizes performance for the range of global qualities considered.

We can see that the best performance is achieved with the method we propose, whatever
the value of $GlobQ$. When $GlobQ < GlobQ_{th}$, the performance is very close to the one achieved with Krichen04, but it outperforms Krichen04 when $GlobQ \geq GlobQ_{th}$. On the contrary, the performance of our method is very close to the one of Daugman06 when $GlobQ \geq GlobQ_{th}$ but it exceeds Daugman06 performance when $GlobQ < GlobQ_{th}$.

6 Conclusion

In this article we associated a global GMM-based quality measure to each pair of (gallery, probe) polar iris images we wish to match in order to quantify the amount of occlusions present in the images as well as the amount of texture in the non-occluded areas. We demonstrated that this global quality measure can be used for two purposes. First it can be exploited to adjust, for each pair of irises, the proportion of the polar image selected for feature extraction on a local quality criteria. The value of this parameter has a major influence on recognition performance whether the selection strategy is applied to Daugman’s feature extraction and matching technique or Krichen’s. Setting it too high can imply taking into account regions with artifacts for matching, but setting it too low can lead to missing out on useful information for matching. Its optimal value decreases when the global quality of the image pair is reduced until a threshold value under which matching becomes insignificant being based on too few information. Moreover, the relationship between the global quality of an image pair and the optimal value of the proportion of the polar image to use for matching is the same whether the matching technique used is Daugman’s or Krichen’s. Therefore, the knowledge of the global quality of an image pair allows us to set this parameter so as to maximize performance whichever technique is used for matching.

Secondly, the global quality measure we define can be used to choose between using Daugman’s or Krichen’s feature extraction and matching technique for each pair. Indeed, we showed that Krichen’s technique performs better than Daugman’s on low quality images according to our measure, but is much slower.

In consequence we proposed a new strategy for iris recognition that combines Krichen and Daugman’s algorithms as well as the local and global use of a GMM-based quality measure. We use the local quality measure to select the regions in the probe and gallery polar iris images that are of highest quality in terms of non-occlusion and amount of texture. These regions will then be used for feature extraction and matching using either Krichen’s or Daugman’s techniques. The global quality measure we have defined is associated to a pair of polar images and is used for two purposes: to decide how many regions are selected for feature extraction and matching, and to choose which technique to use for this extraction and matching. More precisely, these two decisions are taken automatically for each pair of images given the value of the associated global quality measure.

The experiments we carried out showed that our strategy leads to a higher recognition accuracy than a strategy for which we would exploit the same proportion of the iris polar image for all images and the same technique for feature extraction and matching.
References


Assignment of the evidential value of a fingermark general pattern using a Bayesian Network

Rudolf Haraksim, Didier Meuwly, Gina Doekhie, Peter Vergeer and Marjan Sjerps

WISK
Netherlands Forensic Institute
Laan van Ypenburg 6
2497 GB, The Hague
r.haraksim@nfi.minvenj.nl

Abstract: When visible on a fingermark, the general pattern maintains its importance in the fingerprint examination procedure, since the difference between the general pattern of a fingermark and a fingerprint is sufficient for exclusion. In the current work, the importance of the general pattern is extended by evaluating the strength of evidence of a match given corresponding general pattern. In current practice (due to the lack of statistical support for the general pattern evidence) the fingerprint examiners assign personal probabilities to the general pattern evidence based on their knowledge and experience, while in this work the probabilities are calculated using a Bayesian Network which is fed by empirical data.

1. Introduction

In this article, we aim to assign a value to the correspondence of the general patterns (GP) in terms of descriptive and inferential statistics. We have developed two Bayesian Networks (BN) – one at the level of finger and one at the level of person – to assist the fingerprint examiners in statistical quantification of probabilities they assign to the general pattern evidence. The main motivation for using BNs is their ability to model the dependencies between different types of evidence in a logically correct framework.

When a fingerprint examiner compares a fingermark retrieved from a crime-scene to a reference fingerprint of a suspected person, (s)he exploits all the available information to assign its evidential value: properties of the ridge flow (level 1), of the minutiae (level 2) and of the ridges themselves (level 3). Recently tools producing Likelihood Ratios (LR) have been developed, allowing the fingerprint examiners to quantify the evidential value of spatial configurations of minutiae [Ne11, ECM07, AJR13, NCJ12, FSS07]. According to [Ne11], the evidential value assigned to the spatial configuration of the minutiae present in a fingermark can be expressed using a likelihood ratio (LR) and a set of propositions at the level of the finger\(^1\):

\[
\begin{align*}
H_p: & \text{ the fingermark was left by a specific finger} \\
H_d: & \text{ the fingermark was left by an unknown finger}
\end{align*}
\]

\(^1\) The factfinders phrase their questions at the level of the person, which is then investigated at the level of the finger.
or at the level of the person:

Hp: the fingermark was made by the person who made the set of fingerprints
Hd: the fingermark was made by some unknown person

In absence of realistic data, the numerator of the LR has been reduced by a factor of 10 in [Ne11] when the propositions are considered at the level of the person, to account for the uncertainty in relation to which of the ten fingers of a donor the fingermark originates. The aim of this article is to complement these approaches, using real forensic fingerprint and fingerprint data as well as a BN to account for the probability from which of the 10 fingers of a donor the fingermarks retrieved from crime-scene originated and to quantify the evidential value of the shape of the ridge flow classified as a GP.

In the following sections of this article we will provide firstly an insight in the datasets used for constructing the networks, and secondly present each of the BNs proposed paired with a case example. Finally we will assign the evidential value in form of likelihood ratios. Such likelihood ratios can be combined with the evidential value assigned to the other corresponding features of the fingerprint and fingerprint, for example the minutiae configuration.

## 2. Data used and descriptive statistics

### 2.1 Data labeling

By convention, the fingers are numbered from 1 to 10, starting from the right thumb (labeled finger 1) and ending to the left little finger (labeled finger 10). Numerous systems exist to assign GP to the shape of the ridge flow. In this work, the data are labeled according to the GP classification codes of the ANSI/NIST-ITL 1-2000 format [NIST11]: plain arch, tented arch, left loop, right loop and whorl. A 6th class labeled “unknown” merges the ANSI-NIST codes “unable to print” and “unable to classify”.

![Pattern Classification](image)

Figure 1 – General pattern classification

In 1975, A. J. Brooks conducted a study on the fingermarks identified in Chicago during the period from 1969 to 1973, to determine from which of the 10 fingers of a donor the
fingermarks retrieved from crime-scenes originated [Br75]. Since this time, too little attention has been paid to the study of datasets of identified fingermarks [RJM12]. More attention has been dedicated recently to the study of the distribution of the GP on the 10 fingers [Sw05, NBMM09, GARG08]. These studies use various GP classification codes and only the results presented in [GARG08] classify the shape of the ridge flows, with codes similar enough to the ANSI-NIST codes to be compared to the results of the present study.

Due to their age, rarity, diversity, or origin, we have replicated these studies independently in 2012 in our country using the most recent operational data, to ensure the applicability of the results in this country and at the present time.

2.2 Identified fingermarks – finger number

A total of 11555 identified fingermarks\(^2\) from the years 2010 (4032 identifications) and 2011 (7523 identifications) was used to determine from which of the 10 fingers of a donor the fingermarks retrieved from crime-scenes originate. These data reflect the operational activity as processed by the national police force in the field of fingerprint examination in these two years. For each identified fingermark, the finger number, the GP and the gender of the donor of the (corresponding) reference fingerprint general were provided. The results summarizing the distribution of fingers identified in the police investigations are presented in the Table 1.

<table>
<thead>
<tr>
<th>Finger Number</th>
<th>Brooks Identified Fingermarks</th>
<th>Police Identified Fingermarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15.06</td>
<td>15.59</td>
</tr>
<tr>
<td>2</td>
<td>11.69</td>
<td>16.97</td>
</tr>
<tr>
<td>3</td>
<td>13.57</td>
<td>10.64</td>
</tr>
<tr>
<td>4</td>
<td>10</td>
<td>6.9</td>
</tr>
<tr>
<td>5</td>
<td>2.22</td>
<td>2.1</td>
</tr>
<tr>
<td>6</td>
<td>14.05</td>
<td>15.26</td>
</tr>
<tr>
<td>7</td>
<td>10.17</td>
<td>9.62</td>
</tr>
<tr>
<td>8</td>
<td>13.2</td>
<td>11.67</td>
</tr>
<tr>
<td>9</td>
<td>7.81</td>
<td>8.07</td>
</tr>
<tr>
<td>10</td>
<td>2.22</td>
<td>3.18</td>
</tr>
</tbody>
</table>

The proportions of identified fingermarks will be integrated into the “FingerNumber” node of both BNs (described in the following section). The results have also been compared to the results of the Brooks study.

\(^2\) We are aware that no ground truth exists for a decision regarding identification of a crime-scene fingerprint and a corresponding reference fingerprint of a suspect. Due to the fact that 12-minutiae numerical standard is adopted in many countries (including ours) we consider the identifications carried out by fingerprint examiners based on this standard as an acceptable ground truth by proxy.
Despite the 35 years separating the two studies, the diversity of the populations studied and the fact that the quantity of data of the present study supersedes almost 4 times the dataset of Brooks, we observe similar results. The descriptive statistics presented indicate that differences smaller than 2% are observed between the two datasets. Our interpretation is that inferences made using these results are valid on the long term and are not sensitive to the diversity of the populations. We also observe fact that both hands are similarly represented in the criminal activity (47% left hand vs. 53% right hand), despite the fact that the majority of the human population is right-handed.

![Graph showing distribution of identified fingers in %](image1)

![Bar chart showing distribution of identified fingers by hand](image2)

**Figure 2 – Results of the comparison of the results of the present study with the results of the Brooks study [Br75]**

### 2.3 Reference fingerprints – General pattern

The dataset consists of inked, digitized and encoded 10-print cards of the police fingerprint database. The GP of these prints has been assigned manually by fingerprint examiners. For each print, additional information regarding the finger number and the gender of the donor is available. 10-print cards from 312,484 individuals have been randomly selected from the original dataset to study the distribution of the GP over the 10 fingers. 72.5% of the data originates from male donors and 26.8% from female donors. For 0.7% of the data the gender was unknown.

**Table 2 – GP distribution (%) on different fingers of the right hand (females and males)**

<table>
<thead>
<tr>
<th>Finger No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GP \ Gender</strong></td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>Plain arch</td>
<td>2.2</td>
<td>4.1</td>
<td>5.3</td>
<td>7.8</td>
<td>3.7</td>
</tr>
<tr>
<td>Tented arch</td>
<td>1.1</td>
<td>1.5</td>
<td>11.8</td>
<td>10.3</td>
<td>7.3</td>
</tr>
<tr>
<td>Right Loop</td>
<td>47.3</td>
<td>54.9</td>
<td>29.6</td>
<td>36.6</td>
<td>65.6</td>
</tr>
<tr>
<td>Left Loop</td>
<td>0.4</td>
<td>0.4</td>
<td>16.4</td>
<td>12</td>
<td>1.4</td>
</tr>
<tr>
<td>Whorl</td>
<td>48.7</td>
<td>39</td>
<td>36.3</td>
<td>32.9</td>
<td>21.5</td>
</tr>
<tr>
<td>Unknown</td>
<td>0.3</td>
<td>0.2</td>
<td>0.7</td>
<td>0.4</td>
<td>0.5</td>
</tr>
</tbody>
</table>
The information related to the GP, to the finger number, and to the gender have been exploited in combination in order to study the distribution of the GP on the 10 fingers. The results for the female and male donors are presented in the Tables 2 and 3. They will be integrated into the node variable “GeneralPattern” of the two BNs described in the next section.

Table 3 – GP distribution (%) on different fingers of the left hand (females and males)

<table>
<thead>
<tr>
<th>Finger No.</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>GP \ Gender</td>
<td>M</td>
<td>F</td>
<td>M</td>
<td>F</td>
<td>M</td>
</tr>
<tr>
<td>Plain arch</td>
<td>3.9</td>
<td>6.5</td>
<td>5.3</td>
<td>8.1</td>
<td>4.6</td>
</tr>
<tr>
<td>Tented arch</td>
<td>1.7</td>
<td>2</td>
<td>12</td>
<td>11.5</td>
<td>7.7</td>
</tr>
<tr>
<td>Right Loop</td>
<td>0.5</td>
<td>0.9</td>
<td>14.3</td>
<td>15.8</td>
<td>1.1</td>
</tr>
<tr>
<td>Left Loop</td>
<td>55.3</td>
<td>55</td>
<td>34.1</td>
<td>32.9</td>
<td>64.8</td>
</tr>
<tr>
<td>Whorl</td>
<td>38.3</td>
<td>35.4</td>
<td>33.8</td>
<td>31.3</td>
<td>21.4</td>
</tr>
<tr>
<td>Unknown</td>
<td>0.3</td>
<td>0.2</td>
<td>0.6</td>
<td>0.4</td>
<td>0.5</td>
</tr>
</tbody>
</table>

In the Figure 3 the results are compared to the results of the Gutierrez [GARG08] study. As the entries in the Tables 2 and 3 indicate minor differences of the order of 2% between the relative frequencies of GPs for females and males. The prints labeled as Plain and Tented Arch of our study have been merged into one class labeled Arch to fit the classification codes used in [GARG08].

Figure 3 – Distribution of the GPs on the 10 fingers (Police vs Gutierrez datasets)
Despite the fact that 312,484 individuals were used in our study and only 200 individuals in the study of Gutierrez, we observe similar proportions of right and left loops distributed over different fingers. However, proportions of arches and whorls appear to be quite different for some fingers. Unfortunately, the difference in encoding standards used by Nithin [Sw05] and Swofford [NBMM09] prevent a direct comparison.

3. Inferential statistics using Bayesian Networks

From practice and experience the fingerprint examiners acquire an implicit knowledge of the distribution of the GP over the 10 fingers and of the relative contribution of the 10 fingers to the fingermarks retrieved from crime scenes. They make use of this knowledge when assigning evidential value to the correspondences and differences observed between a fingermark and a fingerprint.

Two BNs integrating the descriptive statistics described in Section 2 have been built to quantify the evidential value resulting from the combination of the GP and the relative contribution of the 10 fingers. The utility of these networks is to assist the fingerprint examiners to refine the numerator of the LR when they consider propositions at the finger and person level. In other words, the use of BNs allows the examiners to support their personal probabilities with statistical data. Concretely, we propose two BNs to assist the examiner, the first one for the finger level (3.1) and the second one for the person level (3.2). The BN models are „built for purpose“ and their implicit validation and justification is subject to further research.

3.1 Finger level (Distinctiveness of the GP)

At the finger level the BN informs about the rarity of a GP observed on each finger number of a random person (based on the population). The node “Finger Number” contains the distribution from which of the 10 fingers of a random donor the fingerprint originated; the node “Hand” encapsulates the proportion of right / left handed in the identified fingermarks; the node general “General Pattern” contains the distribution of the GP over the 10 fingers and the node “Gender” contains the proportions of male / female / unknown donors of identified fingermarks. We express the dependency of the GP node on the finger number and the gender by $P(GP|FN,G)$.

---

3 The gender dependency has been made explicit, despite the minor differences in the GP distribution between the male and female population (see Section 2.2 of this article)
A fingermark containing a GP labeled as a whorl is recovered from the surface of a ceramic mug. The BN calculates the probability (posterior odds) for this whorl to have been left by each finger of a randomly selected donor. In this case the BN indicates that this mark has the highest probability to have been left by the finger number 1 and the lowest probability to have been left by the finger number 5. This result is useful for 2 purposes. Firstly, it allows for searching the database per finger number, starting from the most common finger. Secondly, the posterior odds indicate that the evidential value expected strongly depends on which finger of a donor it can be paired to.

The propositions to be tested are: the mark originates from donor’s finger 1 vs. the mark originates from any other finger (2-10) of a the same donor. The posterior odds provided by the BN allow to calculate the posterior odds ratio $P(F_1|GP) / P(F_{2-10}|GP)$. The uninformed prior odds of 1/10 in absence of data are updated using the descriptive statistics of the Table 4. The evidential value for a whorl observed on a fingermark paired to the finger number 1 (vs. on any other finger) as calculated in the table 4 is 1.46. In other words, it is 1.46 times more likely to observe a whorl if it originates from the finger number 1 than if it originates from any other finger number of a donor randomly selected. The calculation for the highest and lowest evidential value has been added for illustration purposes.

**Table 4 – LR values for the most rare, case example and most common GP**

<table>
<thead>
<tr>
<th>Evidence</th>
<th>Evidential Value</th>
<th>Prior Odds (in %)</th>
<th>Posterior Odds (in %)</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whorl on Finger 1</td>
<td>Example given</td>
<td>15.59/84.51</td>
<td>21.26/78.74</td>
<td>1.46</td>
</tr>
<tr>
<td>Right loop on</td>
<td>Highest</td>
<td>3.18/96.82</td>
<td>0.02/99.98</td>
<td>0.000609</td>
</tr>
<tr>
<td>Finger 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right loop on</td>
<td>Lowest</td>
<td>2.10/97.90</td>
<td>6.22/93.78</td>
<td>3.09</td>
</tr>
<tr>
<td>Finger 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
\[
\frac{P(F_1 \mid GP)}{P(F_{2-10} \mid GP)} = LR \times \frac{P(F_1)}{P(F_{2-10})}
\]
\[
\frac{21.26}{78.74} = LR \times \frac{15.89}{84.51}
\]

Equation 1 – LR calculation from the prior and posterior odds

### 3.2 Person level

A few extra nodes need to be introduced in the previous BN to address the propositions at the person level. The node “Suspect Print GPs” contains the classification code of the 10 GPs for the donor of the 10-print card paired to the fingerprint (GP code). The node “Source of the Mark” contains the pair of alternative propositions to be tested: the mark originates from the donor of the 10-print card vs. the mark originates from a donor randomly selected. For a practical reason the prior odds ratio for these 2 propositions is set to \( \frac{1}{2} \) (prior odds = 1). The choice for the prior of 1 is a conscious choice to force the posterior odds to be equal to LR. We do not mean to imply that equal prior odds are a good choice for any other purpose than extracting the LR from the BN.

The probabilities of the GP of the mark (“Mark General Pattern” node) directly depend on the finger number, the gender and GP code of the donor of the 10-print card. In the case of correspondence between the GP code and finger number of the fingerprint of the donor of the 10-print card, the numerator of the likelihood ratio is equal to 1; it is equal to 0 in the case of a difference. For the denominator of the LR, the probability of correspondence between the GP code and finger number of the fingerprint and the fingerprint of another person is determined by the data of the Tables 1, 2 and 3.

![Diagram](image_url)

**Figure 5 – the person level BN**

**Case example**

At the person level we use the same fingerprint as in the previous example: a whorl found on a ceramic mug on the crime-scene. Based on eyewitness testimony the police
arrest a person, from which a 10-print card is produced. The Table 5 summarizes the GP
classification codes of this donor.

Table 5 – Description of the general pattern code the donor of the 10-print card
(A – Arch, W – Whorl, R – Right loop, L – Left loop)

<table>
<thead>
<tr>
<th>Finger Numer</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Pattern</td>
<td>R</td>
<td>A</td>
<td>R</td>
<td>W</td>
<td>R</td>
<td>L</td>
<td>L</td>
<td>W</td>
<td>L</td>
<td>L</td>
</tr>
</tbody>
</table>

In the BN shown in Figure 5, the GP of the fingerprint mark is given in the node “Mark
General Pattern” and the GP code of the donor is given in the node “Suspect Print GPs”.
The probability that the fingerprint mark originates from the donor of the GP code given
the finger number and the correspondence of the GP divided by the probability that the
fingerprint mark originates from another person given the same evidence: \( P(Hp|GP) / P(Hd|GP) \) is calculated in the node “Source of the Mark”. This posterior odds ratio is
equivalent to the LR, since the prior odds for the 2 propositions have been set to 0.5
(odds 1/1).

The BN at the level of the person uses the general pattern code of the donor (distribution
of the GPs over all 10 fingers) together with the whorl found on the crime-scene
information to calculate the LR at the person level directly. This information is
quantified for the two sets of propositions in the node “Source of the Mark”: \( P(Hp|GP) = 30.73 \) and \( P(Hd|GP) = 69.27 \) (as shown in Figure 5). The value calculated as
presented in Table 6 is 0.44 (30.73/69.27). It means that it is slightly less probable to
observe a whorl if it originates from the donor of the 10-print card than if originates from
a donor randomly selected. For illustration purposes, the LR has also been calculated for
the other available GPs: arch, right loop and left loop.

Table 6 – LR values for the most rare, case example and most common GP

| Data from the BN | P(GP|Hp) in % | P(GP|Hd) in % | LR |
|-----------------|--------------|--------------|----|
| Arch (given suspects GP) | 80.73 | 19.27 | 4.19 |
| Whorl given suspects GP (case example) | 30.73 | 69.27 | 0.44 |
| Right loop (given suspects GP) | 51.43 | 48.57 | 1.06 |
| Left Loop (given suspects GP) | 57.64 | 42.36 | 1.36 |

Tables 2 and 3 show that the arch is the most rare classification code for a GP. Similarly,
as in the previous example, we can attempt to evaluate the smallest and largest LR.
Unlike in previous example however, we now operate at the level of the person, hence
the LR depends not only on the GP found on the crime scene but also on distribution of
the GPs in the population. It is directly dependent on the general pattern of the suspect as
well. Given the general patterns of the suspect in this case the smallest LR corresponds
to the whorl found on the crime-scene and the biggest LR corresponds to the arch. These
LR values remain modest, but the strategy consists in measuring and combining the evidential value of each characteristic available in the fingermark that can be paired with a reference fingerprint. Clearly, the LRs obtained for the first level information are calculated with the intention to combine them with the LRs calculated for the second level of information, based on the spatial arrangement of the minutiae.

4. Conclusions

When no prior information is available it is reasonable to assume the refinement of 1/10 when moving from the finger to person level as proposed by Neumann et al [Ne11].

The two BNs developed in this article combine the statistical information regarding the GP distribution over different fingers contained within the fingermarks (police identifications) and fingerprints (police database). The main motivation for using the BNs is their ability to model the dependencies between different types of evidence. They also provide a practical solution when quantifying the rarity of the GP found on the crime-scene fingermark and a finger of a random donor (level of the finger) or when quantifying the weight of the GP evidence found on the crime-scene fingermark and GP code of a donor (level of the person).

The choice between the two BNs proposed depends on the needs of the fingerprint examiner and/or operational conditions of the systems used to assign the evidential value of the second level details. Our aim in both cases was to quantify the evidential value contained in the first level detail fingermark/fingerprint comparison in meaningful LR values, which could be further combined with LR values obtained from the second level detail fingerprint evidence evaluation process or any other case related evidence.

5. Future work

Future work will include validation of the BN models developed in terms of enhanced application scenarios, sensitivity analysis and further improvement of the BN to support any finger combination. Also, further investigation is needed when utilizing the tools developed in combination with other fingerprint/case related evidence.

Acknowledgements

This research was motivated by the work carried out at the University of Lausanne presented by prof. Christophe Champod at the IPES 2010 in Florida entitled “The Use of Probabilistic Networks in the Area of Fingerprints”. Our work was carried out in cooperation with the National Police Services Agency of the Netherlands (KLPD), the University of Amsterdam, department of interdisciplinary research and statistics (WISK) of the Netherlands Forensic Institute, and the European Union MC-ITN FP7 BBfor2 project. More detailed summary of the work presented can be found in [Do13].

108
References


Video-based Fingerphoto Recognition with Anti-spoofing Techniques with Smartphone Cameras

Chris Stein¹, Vincent Bouatou², Christoph Busch¹

Hochschule Darmstadt - CASED¹ Safran Morpho - Identification Division²
Mornewegstraße 32 Boulevard Gallieni 11
D-64295 Darmstadt F-92130 Issy les Moulineaux

chris.stein@cased.de, vincent.bouatou@morpho.com, christoph.busch@h-da.de

Abstract: This work is concerned with the acquisition of fingerprints samples on smartphones with the built-in smartphone camera. A novel approach to capture multiple fingerphotos in a videostream with a smartphone camera and the processing of the photos for the finger recognition is discussed in this paper. The proposed technique offers a convenient and efficient way to capture multiple samples of a biometric instance in a short time frame. Due the fact that fingerphotos can be easily replicated with low effort (e.g. print outs with an ordinary printer) and thus are vulnerable to presentation attacks, anti-spoofing algorithms were developed to detect such spoof attempts. The algorithms for the detection and segmentation of the finger as well the preprocessing of the photo with graphical operations and anti-spoofing were implemented in a prototype as application for the Android operating system. User tests are performed to evaluate the usability and to create a database of biometric samples for offline evaluation of the recognition performance. Further tests are done with diverse artefacts such as printed finger images, fake fingers of gelatin, gummy and silicon as well finger replay videos to measure the resistance of the developed solution against presentation attacks.

1 Motivation

1.1 Field of application

Smartphones can be secured with a biometric authentication system based on fingerprints that use the fingerphoto recognition. The built-in camera of the smartphones is used to capture the biometric characteristics of the finger. The latest smartphones have at least one integrated camera to capture the finger in sufficient quality and have enough computational capacities to process the photos and execute algorithms for the fingerphoto recognition. Hence, there are no extra devices needed to perform the solution proposed in this work. Biometrics offers an authentication factor that is more reliable since knowledge-based authentication schemes since observed biometric characteristics cannot be delegated, forgotten or copied like e.g. passwords.
1.2 Advantages of the capture method

The proposed capture method with a camera has advantages over the widely spread used touch-based solutions for capturing fingerprints: Body contact is avoided while capturing the fingerprint sample and thereby there is no risk of leaving a latent fingerprint on a sensor. There are no deformations of the finger potentially caused by high pressure of the finger on a touch-based sensor and thus no risk of decreased quality of the captured sample due inadequate pressure. The video stream input enables the possibility to capture multiple samples from a biometric instance in a short time frame and with minimal user interaction. Such video frame sequences can be used to improve the quality of the biometric templates by consolidating the biometric information from multiple frames.

2 Related Work

This work is related to the work by Lee et al. [LD-2008], Derawi et al. [DM-2011] and the work of Stein et al. [SC-2012]. In the related work of Stein et al. a first complete authentication system for the Android operating system based on single fingerprint was developed and evaluated. The prototype was written as a module for MBASSy [WH-2010]. MBASSy is a framework which allows the user to utilize various biometric authentication methods. The observed Equal Error Rate (EER) was in the range of 20% [SC-2012]. In that work a minutia extractor and comparator with low complexity was used to be executed directly on the smartphone. These components seem to be the main reason for the weak performance. In this work we have adopted the finger detection and segmentation algorithms from Stein et al. and have integrated an industrial solution for minutia extraction and comparison. Thus in this work, we are able to evaluate the recognition rate of a single photo capture and to benchmark it with our proposed video-based capture method that is based on an industrial minutia extractor and template comparator [Morp].

3 Objectives and Approach

The intention of our work was to improve the proposed fingerprint authentication system for the Android smartphones [SC-2012] in terms of enhanced usability, recognition rate and anti-spoofing resistance. The capture method with single fingerprint was replaced by a video-based approach. The algorithms for the finger detection and quality assurance are adapted and optimized for continuously use on the video stream. An anti-spoofing technique is implemented that requires performing a challenge response of the user. The position and distance of the finger as well the edge density (metric for sharpness on fingerprint photos [SC-2012]) and the light reflection on the finger caused from the LED of the camera is measured to detect spoof attempts. The developed solution is evaluated in user tests to determine the usability and to create a biometric database of fingerprint photos that is used in an offline evaluation to determine the recognition rates. A commercial minutia extractor and template comparator namely the MorphoLite SDK from
Morpho [Morp] is applied to extract and compare minutiae in an offline technology test. The algorithms for spoof detection were tested with genuine presentations and different attack presentations with diverse artefacts such as printed finger images, fake fingers of gelatin, gummy and silicon as well finger replay videos.

4 Hardware Requirements and Camera Settings

In order to capture usable finger photos that contain the friction ridge pattern of the finger, the built-in smartphone camera must be able to focus on very close objects (<10cm) in front of the camera. The camera must also have a built-in LED that is used for the implemented anti-spoofing technique. The continuous capture and processing of the frames of the video stream demand a smartphone with at least a powerful dual-core CPU with more than 1GHz clock frequency in order to process the frames fast enough. Otherwise the finger and anti-spoofing detection rates can be affected negatively because too many frames cannot be processed in time and must be discarded. Suitable smartphones those fulfill these requirements were the Galaxy Nexus and Galaxy S3 from Samsung.

The “macro” mode of the camera is used, such that the camera uses the closest possible focus. The LED is switched on during the capture process. The LED spotlights the finger such that it appears brighter than the background. This simplifies the detection and segmentation of the finger against the background. Another advantage is the reduced camera noise and risk of blurring caused from hand-motion due the high brightness from the LED. Further advantages using the LED are stabilized lighting conditions and a more homogeneous illumination. The usage of the LED is also important for the implemented anti-spoofing technique (see Section 8).

5 Capture Process

The user simply positions his finger close in front of the camera (see Figure 1) in order to capture a defined amount of biometric samples from the video stream. The orientation of the finger can be random. During the capture the user has the option to rotate his finger slightly in x- and/or y-axis to capture the finger from different perspectives. The usage of multiple perspectives of the finger can improve the recognition rates when a consolidated template from several video frames is generated. The constant input stream from the camera is processed by the finger detection and quality assurance algorithms that is adapted for video stream input of the camera from the prior work of Stein et al. [SC-2012] to detect the Region of Interest (ROI) and determine the quality of the sample. The amount of the processed frames (finger photos) per second is limited by the processing power of the CPU. Frames that pass the quality assurance will be segmented, preprocessed and stored for the offline evaluation.
6 Segmentation and Preprocessing of the Photos

The ROI of the captured photos that have passed the quality assurance will be further processed to prepare it for the minutiae extractor. These steps are applied on the ROI with the functions of the OpenCV framework [OSCV] in the following order:

1. Segmentation of foreground and background area
   The foreground area (the finger) is segmented from the background to remove the pixels at the background that are not relevant for the fingerphoto recognition. This can be achieved when all values of the ROI below a defined value are set to black. All other values remain unchanged. This results in the segmented finger foreground area. Only the red channel is evaluated for the segmentation. A threshold of 100 (red value range 0…255) has been proven as an optimal value for the segmentation.

2. Transformation of the image from RGB to gray-scale
   The color information is not used anymore after the ROI was detected from the finger detection algorithm. The computation of only one channel reduces the computation time for the following preprocessing steps significantly.

3. Median filter
   A simple median filter with a kernel size of 3 is applied to reduce the camera noise.

4. Adaptive threshold
   The ROI is binarized after this operation. The calculation is done by analyzing the gray values of the neighborhood pixels of a certain block size to determine the average value. A pixel is set to “white” if this average is above the threshold; otherwise it is set to “black”. The Morpho minutiae extractor can handle with regular and inversed binarized images. Thus, an inversion of the binarized data (valleys are “white” and ridges are “black”) would also work properly as input for the minutiae extractor. The binarization step is required for reliable detection of the minutiae of the finger with a minutiae extractor. The best results were achieved with a block size of 19 in combination with the used input resolution of 1280x720 pixels from the camera of the test device and the Morpho minutiae extractor.

5. Scaling to a fixed width
   The dimensions of the images must be normalized because the capture method allows different distances of the finger to the camera those results in different dimensions of the image. The ROIs width is scaled to a fix value and the height is changed according the calculated scale factor to keep the aspect ratio. This operation ensures to generate from finger images always a geometrically normalized template that can be processed with the template comparator.

6. Cropping of the height
   A very long image indicates that the border of the first finger segment was not properly detected. In this case the lower part of the image that does not contain any essential information for the fingerphoto recognition is removed, such that it does not exceed the defined maximal height.
Fig. 2: Preprocessing steps with the detected ROI from the finger detection algorithm: 1. Rotated ROI, 2. Segmentation and RGB-to-gray-scale transformation, 3. Median-filter and adaptive threshold, 4. Scaling to a fixed width and cropping height

7 Implementation

The application is written in Java for the Android operating system. Common middle to high end smartphones had at least a dual-core processor. The program workflow is optimized for dual-core processors to maximize the performance on such devices. The preprocessing of the finger photos requires more computing power than the finger detection and quality assurance algorithm together. Thus, the preprocessing is done in a separate (asynchronous) worker thread, so it does not block the main thread due heavy work load. This also allows the parallel preprocessing of photos and the capture of frames. The open source framework OpenCV [OSCV] is available for the Android operating system and used to perform the graphical operations on the images. For performance optimization, the preprocessing code is called over the JNI (Java Native Interface). The anti-spoofing algorithms are running also in separate thread to guarantee a high performance during the evaluation of the challenge response.

8 Presentation Attack Detection

8.1 Principle

After the probe photos are taken in authentication mode, the application enters the challenge response mode. In this mode, the user is prompted to move his fingertip slowly towards the camera. The shape and the consistency of the finger and in combination with the slow movement of the fingertip towards the camera lead to a characteristic strong reflection at the fingertip from the camera’s LED. Other materials like 2D print outs and (unprocessed) fake fingers do not possess such reflection properties and thus do not pass the challenge response. The reflection must be detected near the fingertip and must be strong enough in order to exceed the defined threshold for a positive challenge response. Figure 3 shows the reflection characteristics of a genuine finger and other typical fake fingers. The calculation of the light reflection is described in Section 8.3.
Additional checks regarding the edge density and the position of the finger as well the distance of the finger to the camera are performed to detect unusual sharpness values and keep the link with the shown finger from authentication mode (see next Sub-Section 8.2).

![Image of genuine and fake fingers with challenge response photos]

Fig. 3: The genuine finger (with moved fingertip towards the camera) reflects enough light from the LED during the challenge response to pass the challenge response (upper row). The fake fingers reflect (even with shiny and glossy materials) much lower light. However, the fake finger of gelatin with glycerin treatment on its surface (lower row) can also generate a very high reflection like the moved genuine finger.

8.2 Challenge Response

During the challenge response, the position of the finger, the distance of the finger to the camera, the edge density and the light reflection caused from the LED is continuously measured in the video stream:

1. Position of the finger
The measured position of the finger must not differ significantly from the last captured photo of the video stream: The position of the finger is determined with the finger detection algorithm. The position of the left and right boundaries from the last captured ROI and the ROI from the challenge response is checked against the set value (in pixels) for the movement tolerance.
The finger in the photo from the challenge response must also cover the whole area of the quality assurance and must not exceed the image border towards the direction of the fingertip. Otherwise the check fails (even when the measured movement of the finger is lower than the set movement tolerance). The check of the position of the finger keeps the link of the presented finger from authentication mode so it cannot be exchanged by a fake finger. Figure 4 illustrates a valid example of a position check on a (rotated) fingerprint: The shadowed area in the center must be covered from the finger but the outer shadowed area must not; the non-shadowed area indicates the allowed movement tolerance.

2. Distance of the finger to the camera
The measured width of the finger must not differ significantly from the last captured photo of the video stream: The width of the last captured ROI and the ROI from the challenge response is checked against the set tolerance value. Lowering the distance of the finger to the camera can produce an overexposed image because more light of the LED is captured on the photo on closer distances. The camera compensates too much light incidence by closing the shutter but the correction is delayed. Therefore, fast distance changes of the finger can produce overexposed images before the shutter correction is applied (see Figure 5). Those images would achieve higher reflection values and could pass the challenge response falsely. To avoid this issue, distance changes are limited and detected by the change of the fingers width.

3. Edge density of the finger
The edge density on the ROI of the challenge response is calculated and must not exceed a defined maximum threshold. Print outs from a printer have a typical raster pattern (see Figure 6). The raster pattern causes a very high edge density value of 10+. This check detects the usage of a print out during the challenge response.

4. Light reflection in the inner area of the ROI
The measured light reflection in the core area of the finger (see next Sub-Section 8.3) must exceed a threshold: A strong reflection near the fingertip appears due the movement of the fingertip towards the camera and must exceed the threshold for a positive challenge response.
5. Light reflection in the outer area of the ROI

The measured light reflection outside the core area (outer area) of the finger (see next Sub-Section 8.3) must not exceed a threshold: Artificial light reflections can be produced from high reflecting materials (in the background) and can be "guided" near the fingertip to achieve a positive challenge response. This measure detects presentation attacks.

For a positive challenge response, all mentioned criteria 1 to 5 must be fulfilled. The challenge response is unresolved and will be continued when all criteria are fulfilled except criterion 4. If one criterion (except 4) is not fulfilled, then the challenge response is aborted with a negative result. In this case the user must restart the authorization process for a new attempt.

8.3 Light Reflection Measurement

The ROIs of the captured frames of the video stream during the challenge response were calculated and converted into gray-scale images. The half width in the center of the ROI and the upper half height of the ROI will be defined as core area and is used to detect the reflections in the central part of the fingerprint for positive authentication of the challenge response. The outer part is the rest of the ROI with all values set to black of the core area and is used to detect reflections at the edges of the finger for negative authentication of the challenge response (see Figure 7). Light reflections in this area do not occur from a genuine finger but from spoof attempts.

Only pixels with a maximum value (full white = 255) are kept to detect the light reflections. All other pixels are set to black. An additional morphological operation “erosion” is done to filter small areas of white pixels those are not large enough. The remaining white pixels are count and summed up for the core and outer area of the finger separately and represent the strength of the light reflection for each area. A higher value represents a stronger measured light reflection.

9 Evaluation and Results

A biometric database was created with the single photo capture technique and video-based photo capture technique in user tests to evaluate the EER with the minutia extractor and template comparator from Morpho [Morp]. The usability of the presentation attack detection method was tested in a separate user test to determine the genuine acceptance rate. The presentation attack detection rate (PADR) [ISO-2012] was determined with several fake fingers and methods.
9.1 Recognition Rates

Capture Environment and Data Set
Finger photos from 37 data subjects were captured with the single finger photo capture method. Six to eight photos from the left and the right index finger with the smartphones “Nexus S” and “Galaxy Nexus” from Samsung were captured in two sessions. The resulting test data set consists of 569 unique finger photos from the “Nexus S” and 541 photos form the “Galaxy Nexus”. 11 data subjects have participated on capturing finger photos with the video-based finger photo method. Those photos were captured with the “Galaxy Nexus”. Six capture sessions have been performed for capturing the left and right index finger (three sessions each). 15 photos are captured on each session. The resulting test data set consists of 990 unique finger photos. All captured photos with both capture technologies were only accepted when they have passed the quality assurance. The finger photos were captured indoor in a standard office environment. The rooms were well lightened from natural daylight.

Evaluation Procedure
The algorithms from the MorphoLite SDK were not available for the Android operating system [Andr] at the time of writing this paper. Thus, the found and preprocessed ROIs from the input finger photos were stored on the smartphones file system and then transferred to the PC for the minutiae extraction and template creation. The created templates of the video-based photo capture on each session will be consolidated into one template with the algorithms from the MorphoLite SDK. With all templates from the database genuine and imposter comparisons were computed with MorphoLite SDK in order to determine the error rates.

Results

![DET Curve](image)

Fig. 8: DET curves of the single photo and video-based capture method on the tested devices
The integration of the minutiae extractor and template comparator from the MorphoLite SDK lead to significant better recognition rates over the results reported in previous work [SC-2012]. The achieved error rates are shown in Figure 8 in a DET-diagram and in a table in Figure 9.

**Computation Time / Frame Rate of the Video-based Approach**

The measured performance during authentication and enrolment on the “Galaxy Nexus” is about 2.27 frames per second (440ms computing time per frame). 4.55 frames per second (220ms computing time per frame) are achieved during the challenge response. The determined values are the average from the collected values during the user tests.

**9.2 Performance of the Presentation Attack Detection**

**Genuine Tests**

26 subjects have participated on a voluntarily basis to test the challenge response of the application. In order to determine the usability and the false detection rate of the challenge response, each user tries to pass the challenge response with his index finger. Each user has repeated the procedure 10 times after a short instruction and demonstration to the application from the operator. The amount of successful and failed attempts was count. The needed amount of moving the fingertip to the camera was also count in case of a successful challenge response.

**Results**

201 of 260 performed challenge responses were successful. The result is a recognition rate of 77.3% (see Figure 11). 80 of them were recognized at the first time as the finger was moved to the camera. In 69 cases the fingertip movement must repeated twice to the camera in order to achieve a successful challenge response. In 44 cases three attempts was needed for a successful challenge response. Four or more attempts were needed in 8 cases. Subjects that have held the smartphone in a brighter area e.g. near a window, had more difficulties to pass the challenge response. Light conditions decrease
the effect of the light reflection of the LED and make it harder to pass the challenge response.

**Artefact Detection Tests**

The following fake tests are done:

- Several fake 2D print outs with original unprocessed and binarized fingers in different sizes printed with a laser printer on ordinary paper (*see Figure 12*).
- Several fake fingers of gummy, silicon and gelatin those differ in color and shape (*see Figure 13*).
- A replay attack with a video captured and injected with another “Galaxy Nexus” smartphone (*see Figure 14*).

**Results**

Many repeated fake attempts with the above mentioned fake attacks were performed. The determined true PADR is 0.83 based on the tests of the six different attack presentation characteristics: 2D print out, fake finger of gummy, silicon and gelatin, fake finger with post treatment and replay attack. The results give first impressions about the potential of the developed anti spoofing technique. Extended tests are needed to make more meaningful statements about the spoofing resistance.

**2D print outs**

The fakes possess a higher edge density value due the raster pattern effect (*see Figure 6*) and are normally detected by the edge density check. However, the fake can pass this check when the fake is not properly in focus. But such a fake is not able to pass the reflection check because the material and the raster pattern effect (only the pigments of the toner are reflected) do not provide such high reflection strength as a genuine finger.

**Fake fingers of gummy, silicon and gelatin**

None of the tested (unprocessed) fake fingers was able to produce the necessary light reflections to pass the challenge response. However, a half-transparent fake finger of gelatin with a special treatment of the surface with glycerin can simulate a similar reflection behavior of a real finger and was able to pass the challenge response.

**Video replay attack**

The reflection strength from the shown finger on the display in the video is far lower...
than a natural reflection. The surface of the (glossy) screen itself reflects much more light and makes a proper capture respectively an injection almost impossible resulting in a failure of the challenge response due various finger detection errors like too small ROI or too high finger movement. After all, the measured reflection of the screen is still lower and do not pass the challenge response.

10 Conclusions and Future Work

A new approach to capture fingerphotos over a video stream with a smartphone camera has been implemented and evaluated. An EER of about 3% was achieved. The existing prototype has been improved in aspects of recognition rate, usability and anti-spoofing resistance. However, a smartphone with at least a fast dual-core processor is needed to achieve a usable frame rate. Otherwise the CPU is not able to process the frames in time and a lot frames must be discarded due the lack of available CPU resources which results in a decreased capture rate and decreased anti-spoofing detection rate. The developed anti-spoofing technique can detect spoof attempts with fake print outs and fake fingers of gummy, silicon and gelatin as well video replay attacks. However, advanced techniques with special treatments of the surfaces of finger fakes can simulate similar light reflections of a real finger those cannot detected reliably with the current implementation of this technique.

Further development can be the modification of the finger-detection algorithm to detect multiple fingers per frame in the video stream. This will decrease the effective needed time per capture further and enables a convenient way to capture multiple fingerprints from different fingers from one subject.

Acknowledgement

This work was partly sponsored by Safran Morpho group. The authors also like to thank the volunteers that supported the technology and usability testing in this work.

References

[DM-2011] Derawi, Mohammad Omar; Yang, Bian; Busch, Christoph: Fingerprint Recognition with Embedded Cameras on Mobile Phones. In: MobiSec 2011, Denmark, 2011
Continuous Authentication using Mouse Dynamics

Soumik Mondal and Patrick Bours
NISlab, Gjøvik University College
Postboks 191, 2820, Gjøvik, Norway
soumik.mondal@hig.no
patrick.bours@hig.no

Abstract: In this paper, we demonstrate a new way to perform continuous authentication using Mouse Dynamics as the behavioural biometric modality. In the proposed scheme, the user will be authenticated per mouse event performed on his/her system. We have used a publicly available mouse dynamics dataset and extracted per event features suitable for the proposed scheme. In this research, we have used the mouse dynamics data of 49 users and evaluated the system performance with 6 machine learning algorithms. In this approach, the genuine user has never been classified as an impostor throughout a full session whereas the average number of mouse actions an impostor could perform before detection is 94 from the best classification algorithm with a person based threshold.

1 Introduction

For most existing computer systems, once the user identity is verified at login, the system resources are available to that user until he/she exits the system or locks the session. In fact, the system resources are available to any user during that period. This may be appropriate for low-security environments, but can lead to session hijacking, in which an attacker targets an open session. In high risk environments or where the cost of unauthorized use of a computer is high, continuous verification/authentication of the user’s identity is extremely important. Continuous Biometric Authentication Schemes (CBAS) has built around the biometrics supplied by the user behavioural characteristics and continuously check the identity of the user throughout a session [Bou12].

In this research, we are going to evaluate the Mouse Dynamics Biometric modality for Continuous Authentication. Mouse Dynamics is the way a user is interacting with his/her system through the mouse. Similar to Keystroke Dynamics [add references] does Mouse Dynamics (MD) not require any special hardware to capture the biometric data. From 2003, has MD become an interesting research topic in the area of behavioural biometrics due to its non-intrusiveness and convenience [EM03, GF03] for the user.

We found that the current research on continuous authentication reports the results in terms of EER or FAR and FRR over either the whole test set or over chunks of a large, fixed number of events. This is then in fact no longer continuous authentication, but at best periodic authentication. In this paper we focus on actual continuous authentication that
reacts on every single mouse action from a user. The contribution in this paper is the following:

- New scheme for continuous authentication;
- Verifying the genuineness of the user for every mouse event;
- Use 6 different classification algorithms to measure the system performance; and
- Modification of the trust model from [Bou12].

The remainder of this paper will be as follows. Section 2 will describe some of the research done in this area. In Section 3, we will discuss some basic knowledge about the machine learning algorithms which we have used in this research. We will discuss the dataset and the feature extraction process in Section 4. In Section 5 the trust model will be described. Section 6 will analyse the results obtained in this research, while Section 7 presents the results of our analysis method on the used dataset. Finally, in Section 8 we will conclude this research with future work.

2 State of the Art

Gamboa et al. [GF03, GF04] considered mouse strokes for continuous authentication. Each stroke was characterized by a 63-dimensional feature vector including spatial parameters such as angle and curvature, and temporal parameters such as velocity and acceleration. They used the data of 50 users and report that the EER is the function of iteration time and the number of strokes per user. They have reported results on 1 stroke (EER of 48.9%), 50 strokes (EER of 2%) and 100 strokes (EER of 0.7%).

Zheng et al. [ZPW11] used angle-based metrics of mouse movements for user verification. They have used 30 users (different ages, educational backgrounds, and occupations) and SVM as a classification tool. They have reported an EER of 1.3% with the requirement of 20 mouse clicks.

Pusara et al. [PB04] build a continuous authentication system using mouse movements and mouse events as features. They performed an experiment with 18 users (on average 2 hours of data per user) and they used Decision Tree Classifier with smoothing filters for classification. They reported an FRR of 0.43% and a FAR of 1.75% with the verification time ranging from 1 minute to 15 minutes.

Ahmed et al. [AT07, NTA10] build a continuous authentication mechanism with mouse dynamics. They collected data of 49 users and achieved a FAR of 0% and an FRR of 0.36%. Their dataset is publicly available and was also used in this research. They used fuzzy classification based on the learning algorithm for Multivariate Data Analysis and used a score-level fusion scheme to merge corresponding biometric scores.

Feher et al. [FEM++12] used individual mouse actions (contrary to using a histogram over a number of mouse actions) as a feature for continuous authentication. They have used 25
volunteers (21 male and 4 female) to collect data in their experiment and used Random Forest Classifier for data analysis. They have achieved an EER of 1.01% (for 30 actions) with the authentication time of less than 2 minutes.

Lin et al. [LCL12] build a continuous authentication system by using everyday mouse interaction data on a windows computer. They have used data of 11 volunteers and created 3 sample sets. Set A contained the feature vectors of the mouse movements for the complete file-related operations in Windows Explorer. For comparison, set B contained the feature vectors of the mouse movements for operating Windows Explorer, and set C contained the feature vectors of the mouse movements for operating the computer. Their best results were obtained with data set A as was to be expected.

Schulz [Sch06] has claimed a continuous authentication system using mouse dynamics. They collected in his experiment, data of 72 users. He has used three features (1) Length and number of a movement sample of the mouse curve; (2) Curvature and inflection; and (3) Curve straightness characteristics. Furthermore did they use Euclidean distance for classification. They achieved an EER of 24.3% when using a sample size of 60 mouse curves. This performance improved to 11.2% when using a sample size of 3600 mouse curves.

Shen et al. [SCG12] designed a mouse interaction based continuous authentication system. In their system, there is no need for impostor training data. They build their system based on the data of 28 users focusing on different mouse events, mouse operations and mouse behaviour patterns and used different classifier for classification. The best result was a FAR of 0.37% and an FRR of 1.12%, obtained using a One Class SVM detector. The authors also showed the effect of the length of a session. In particular for a session of 5 minutes they reported a FAR of 7.78% and an FRR of 9.45% while for a sessions of 10 minutes these values dropped to a FAR of 2.75% and an FRR of 3.39%.

3 Background Knowledge

For our analysis we tested various different machine learning algorithms. These were Naive Bayes Classifier, $k$-Nearest Neighbour ($k$-NN), Decision Tree Learning, Multilayer Perceptron, Radial Basis Function Network, and Support Vector Machine (SVM). We used the WEKA [HFH+09] and the LibSVM [CL11] software tools for the analysis of our data. Initial testing showed that SVM was the only method that always detected the impostors. For the other methods the probability that impostors were not detected ranged from 12% for $k$-NN to 76% for Decision Tree Learning. For continuous authentication it is obviously of the highest importance that impostors are detected and for this reason we present in the remainder of this paper only the results obtained with SVM. Some details of SVM are presented below.
3.1 Support Vector Machine

Support Vector Machine (SVM) is a very well-known supervised learning model which can be used for classification and regression analysis. This model is capable of creating a decision margin that is as wide as possible, depending on the Support Vectors (SV). The SV are those data points from the different classes that are closest to the decision line. In this research, we have used the LibSVM software distribution for the SVM classifier [CL11]. Initially we tried SVM with a linear kernel, but found that the classifier did not perform well due to the small feature set (see Section 4.2). We decided to use Gaussian kernel as a similarity measure function in this research.

4 Data Description and Feature Extraction

In this section, we are going to describe the dataset that we have used in this research. In Section 4.2 we describe the features that are extracted from the raw data for the analysis we perform.

4.1 Raw Data Description

We have used a publicly available mouse dynamics dataset [AT07, NTA10]. This dataset contains the mouse dynamics data collected from 49 volunteers. The volunteers were asked to use their computer and mouse in a normal, everyday fashion, without any restrictions on the tasks they had to perform. The data collection software stored the following 4 features for each mouse action from a volunteer:

- Type of action (1: Mouse Move; 2: Silence; 3: Point and Click; or 4: Drag and drop);
- Travelled distance in pixels;
- Elapsed time (with a 0.25 second accuracy);
- Direction of movement (a value between 1 and 8 according to the movement of the mouse. See Figure 1 for which direction corresponded to which value).

4.2 Feature Extraction

In [AT07, NTA10], as well as in many other works on continuous authentication, are statistical features extracted from the raw data. In our scheme we want to verify the identity of the user from every single mouse action. Therefore we cannot use statistical features
derived from the raw data, but are we looking for single event based features. We have extracted the following 5 features from the raw data:

**Type of action** We have explicitly removed the "Silence event" actions from the raw data because we want to focus on the behaviour of the user. We therefore only used the other actions that were recorded, i.e. Mouse Move (MM), Point and Click (PC), and Drag and Drop (DD).

**Direction of movement** Taken directly from the raw data.

**Speed of the mouse movement** This equals the Travelled distance in pixels / the Elapsed time.

**Reciprocal Acceleration of the mouse movement** Equal to the Elapsed time / Speed of the mouse movement. We first tried to use the Acceleration of the mouse movement, but we got much better results when using the reciprocal of the acceleration.

**Travelled distance in bins** We did not use the travelled distance in its raw form, but decided to use a limited number of bins for the travelled distance range. The first 20 bins contained a 50 pixel range each, for example if the travelled distance was between 1 and 50 pixels then we assigned bin 1, if the travelled distance was within 51-100 pixels we have assigned bin 2 and so on. After that the bins grew in range, in particular we used 38 bins according to the following schedule:

- From 1 to 1000 pixels: Bin size is 50 pixels, so 20 bins in total;
- From 1001 to 2000 pixels: Bin size is 100 pixels, so 10 bins in total;
- From 2001 to 3000 pixels: Bin size is 200 pixels, so 5 bins in total;
- From 3001 to 4000 pixels: Bin size is 500 pixels, so 2 bins in total;
- More than 4001 pixels: Treated as a separate bin.
4.3 Data Separation

For each of the 49 users has the data set been split into a training set and a test set. The total training set was a combination of the genuine user’s training set and some impostor user’s training sets. To avoid classifier biasness is the number of samples from the impostor users in the total training set is equal to the number of samples from the genuine user. The genuine user’s part of the total training set was randomly chosen from the genuine user data. The amount of training data from the genuine user was 50% of his total data set. The imposer user’s part was chosen at random from the other 48 users’ data.

The total test set consisted of users’ data samples that were not used in the training process. The amount of data used from the genuine user for testing equals 50% of his total data set.

5 Trust Model

Bours [Bou12] has described a trust model concept for continuous authentication using keystroke dynamics. He demonstrated that the trust level will increase or decrease according to the distance between the template and the current typing. In this research, we have used the classifier score (that is the probability of the genuineness of that event) to increase or decrease the trust value (denoted by $C$). We have used the distribution of the classifier score for the impostor users and the genuine user to adjust the parameter of the trust model.

Figure 2 shows an example of the box plots for the distribution of the classifier scores for a randomly chosen genuine user (left) and a randomly chosen impostor user (right) when compared to the given genuine user. We can see that the median score for the genuine user lies above 0.5. On the other hand, for the impostor user we can see that the median value is below 0.45. We have adjusted the trust model from [Bou12] by having 1 type of reward, and 2 different levels of penalty. We have tested our system with various parameter values in the trust model. The values of the parameters represented in this paper were the ones giving the results from Section 7. The used trust model algorithm is as follows:

1. Build the classifier model with the training data of both the genuine and the impostor users (see Section 3).

2. Start at 100% trust (so initially we have $C = 100$).

3. For each event in the test set of a specific user:

(a) Determine from the classifier model what the probability $P_g$ of the genuineness of this event is.

(b) Increase or decrease the trust value $C$ according to the probability $P_g$ of the genuineness of the event:

$$C = \begin{cases} \min(C + P_g, 100) & \text{if } P_g \geq 0.5 \\ \max(C - (1 - P_g), 0) & \text{if } 0.3 \leq P_g < 0.5 \\ \max(C - 1, 0) & \text{if } P_g < 0.3 \end{cases}$$

(1)
4. Compare the trust value $C$ with the lock out threshold $Tr$:

(a) If $C \geq Tr$ then continue with step 3.

(b) If $C < Tr$ then lock the system. During analysis this means a simulation of a new log on and continuation at step 2.

In our analysis we used both global thresholds and person specific thresholds. When testing the system we found that a global threshold was not giving good results, meaning that impostors were not detected. A personal threshold, based on the stability of the typing of the genuine person was also used in our research. The main goal was to have a personal threshold where the genuine user was never locked out of the system. This was achieved by selecting the threshold slightly below the least attained trust value when evaluating the test data of the genuine user. In a real system this cannot be done in this way of course. In that case will the system, after the training of the classifier model, first use some data to determine the personal threshold level for the specific user. Only after also the personal threshold has been set will the system of the user be protected with continuous authentication via mouse dynamics.
6 Analysis method

In this section, we are going to discuss the methodology of our system. The system was divided into two basic phases (see Figure 3).

I. Training Phase: In the training phase, the training data (see Section 4.3) was used to build the classifier model and store the model in a database for use in testing phase. Each genuine user has his/her own classifier model. This means that we have build 49 different classifier models.

II. Testing Phase: In the testing phase, we are going to use test data which was separated from the training data for comparison. In the comparison, we will use the model stored in the database and obtain the classifier score (probability) on each sample of the test data. This score will then be used to update the trust value \( C \) in the trust model (see Section 5). Finally, the trust value \( C \) was used in the decision module, to determine if the user will be locked out or can continue using the PC. This decision was made based on the current trust value and the lockout threshold.

In the testing phase will the performance of the system be measured in terms of Average Number of Genuine Actions (ANGA) and Average Number of Impostor Actions (ANIA). In this case an action of the user can be anything done with the mouse, for example moving the mouse, clicking, or drag-and-drop. This is done by counting the number of test data samples of the genuine or an impostor user that can be used in the model from Section 5 before a user is locked out. Any user will always start at the trust value \( C = 100 \) and in
case of a trusted test data sample, i.e. the probability outputted by the classifier model is above 50%, this trust value will go up (with a maximum of 100). In case the probability from the classifier model is below 50% then the trust value $C$ will decrease. This means that if the user’s typing is in accordance with the classifier model, then the trust in the genuineness increases, otherwise it will decrease.

The data samples from the genuine user will in most cases get a high probability from the classifier and sometimes a low probability. This means that most often the trust will increase and sometimes it will decrease. This then results in a trust value that will remain at a high level. For impostor users this situation is the opposite. Often the trust value will decrease and sometimes, when the behaviour of the genuine user is mimicked correctly, the trust value will go up. The general trend for the trust value will however be downwards, and once the trust value reaches below the threshold, then the user will be locked out.

The average number of actions an impostor can do before being locked out will be the ANIA value, while the average number of actions for the genuine user will be denoted by ANGA. The goal is obviously to have ANGA as high as possible (in fact we try to never lock out a genuine user), while at the same time the ANIA value must be as low as possible. The last is obviously to assure that an impostor user can do as little as possible, hence he/she is detected as quick as possible. Given a fixed model of a user, then changing the parameters in the trust model of Section 5 will either increase both values of decrease both values.

7 Performance

In this section we are going to discuss the result we have achieved from this research and the discussion related to the result.

Although we used 6 different ML algorithms (see Section 3) to evaluate the system, we will report here only the results of the SVM classifier. As mentioned before were some impostors not detected for each of the other 5 classifiers, even when using a personal threshold. For the other 5 classifiers the probability that an impostor user was not detected when using a personal threshold ranged from 12% (for $k$-NN) to 76% (for Decision tree learning). Therefore we will focus only on the results obtained with SVM in this paper.

We created 49 SVM classifier models, i.e. one for each user. For the performance analysis we calculated the ANGA and ANIA values both for fixed lockout thresholds and for personal thresholds. We first tested our 49 models with 5 fixed lockout thresholds of 40, 60, 80, 85, and 90. We observed that the genuine users were never locked out because those users are very stable in their way of using the mouse with respect to the used model. In particular does this mean that the trust value of a genuine user in his own classifier model never drops below 90. In Table 1 the results are shown for these fixed thresholds. This table shows the average ANIA but also the number of impostors that are not detected by the system, given the particular settings.

From Table 1 we see that a lower fixed threshold means that an impostor can do more actions and more impostors will not be detected. Although this is a logical conclusion,
<table>
<thead>
<tr>
<th>Threshold</th>
<th>ANIA</th>
<th>undetected</th>
</tr>
</thead>
<tbody>
<tr>
<td>40</td>
<td>1363</td>
<td>190</td>
</tr>
<tr>
<td>60</td>
<td>999</td>
<td>135</td>
</tr>
<tr>
<td>80</td>
<td>583</td>
<td>45</td>
</tr>
<tr>
<td>85</td>
<td>385</td>
<td>28</td>
</tr>
<tr>
<td>90</td>
<td>192</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 1: ANIA results for fixed lockout threshold

the table is intended to give an indication of how much more an attacker can do when the threshold is lowered. In particular we can see that lowering the threshold from 90 to 85 means more or less a doubling of the average number of actions an impostor can do, i.e. from 192 to 385. On top of that is the number of impostors that are not detected more than 4 times as high for threshold 85 compared to threshold 90.

Based on the results for the fixed threshold did we decide to use a user specific threshold. This user specific threshold was chosen in such a way that the genuine user was never locked out. More specifically did the personal thresholds range between 90.8 and 96.7. A user with a higher personal threshold will be more consistent in his own mouse usage and will have fewer (consecutive) penalties that decrease the trust value. The average ANIA value was in this case equal to 96 (with a standard deviation of 79), meaning that an impostor was, on average, detected after 96 mouse actions.

The value of ANIA depended highly on the specific genuine user. One of the users had an ANIA of only 9, i.e. on average an impostor was detected after only 9 mouse actions. On the other hand, in the given dataset, the worst performance was from a user where, based on his classification model, the value of ANIA was 344. It is obvious that there is a wide spread of ANIA values, meaning that not only the average ANIA value of 96 is important, but also the standard deviation of the ANIA values should be considered.

Figure 4 shows the trust value of a genuine user when tested against almost 5000 test data samples. Although it can be seen that the trust value goes down slightly, will this be corrected automatically because the genuine user uses the mouse in the correct manner, according to the classifier model. Figure 5 shows the trust value of an impostor user, and we see that generally the trust value will go down. After each lockout the trust value is reset to 100 again.

8 Conclusion and Future work

In this research, we have shown a new way of building a continuous authentication system where the system will decide the trust in the genuineness of the user in each and every event. We have tested this idea with 6 different ML algorithms in a publicly available mouse dynamics dataset. The genuine user has never been locked out throughout the session whereas the Average Number of Impostor Actions (ANIA) is 96 for the SVM clas-
sification algorithm with the person based threshold. We have found that the performance of the user is very much dependent upon the stability of user’s own mouse dynamics.

We stress here that the results are obtained given the specific trust change function from Equation 1. The boundaries for $P_g$ or the changes in the trust value $C$ used in this equation are taken as fixed, and we have not attempted to optimize these values. Even better results could have been reached, if these values are optimized, or even if these are not taken as global values, but also as person dependent. In fact we have done some tweaking of the trust function from Equation 1, where the value of 0.3 was replaced with 0.4 and in that case the average ANIA went down to 94 (with standard deviation of 78).

The results we have obtained from this research are satisfactory to validate this approach of continuous authentication using mouse dynamics, but we need more per event features to create a real world continuous authentication system which is out of scope for this dataset. Other examples of per event feature attributes can be the size of the active region where the mouse event occurs, the centroid of the active region and the relative position of the

Figure 4: Trust value for genuine user tested with the genuine test set.

Figure 5: Trust value for genuine user tested with the impostor test set.
mouse in that region, curvature, or duration between mouse button pressure and release. Future work will be to build our own dataset to prove our concept.

References


Spoofing Attempt Detection using Gaze Colocation

Asad Ali, Farzin Deravi and Sanaul Hoque

School of Engineering and Digital Arts
University of Kent, Canterbury, Kent, CT2 7NT, United Kingdom
E-mail: {aa623, f.deravi, s.hoque}@kent.ac.uk

Abstract: Spoofing attacks on biometric systems are one of the major impediments to their use for secure unattended applications. This paper presents a novel method for face liveness detection by tracking the gaze of the user with an ordinary webcam. In the proposed system, an object appears randomly on the display screen which the user is required to look at while their gaze is measured. The visual stimulus appears in such a way that it repeatedly directs the gaze of the user to specific points on the screen. Features extracted from images captured at these sets of colocated points are used to estimate the liveness of the user. A scenario is investigated where genuine users track the challenge with head/eye movements whereas the impostors hold a photograph of the target user and attempt to follow the stimulus during simulated spoofing attacks. The results from the experiments indicate the effectiveness of the gaze colocation feature in detecting spoofing attack.

1 Introduction

Despite the successes in biometric recognition systems in recent decades, they still remain vulnerable to increasingly sophisticated spoofing attacks with the use of fake artifacts. These artifacts may be created from the biometric information of genuine users and presented at the system sensor(s). An impostor can present a fake biometric sample of a genuine user to a biometric recognition system to gain access to unauthorised data or premises. This type of spoofing is a direct attack on the sensor (also known as presentation attack); the impostor does not require any a priori knowledge about the internal operation of the biometric system. To prevent such sensor-level attacks, biometric systems need to establish the liveness of the source of an acquired sample.

Amongst biometric modalities, face recognition has emerged as being socially acceptable, accurate and convenient and is therefore used for a variety of security applications. But face recognition systems may be considered to be more vulnerable to abuse compared to other biometric modalities, because a simple photograph or video of a genuine user can be used to deceive such systems [Tr11]. Therefore, by introducing a liveness detection mechanism, the security of biometric systems can be substantially improved.
Photographs, masks, and videos are the spoofing artifacts that may be used for attacks at sensor level. Photo spoofing can be prevented by detecting motion, smile, eye blinks, etc. However, such techniques can be deceived by presenting a video of the genuine user to the face recognition system. The subtle differences between a photograph (or video) of an individual and the live person needs to be used to establish liveness of the presentation at the sensor.

An important source of liveness information is the direct user interactions with the system that are captured and assessed in real time. In this paper we present a novel challenge/response mechanism for face-recognition systems, using a standard webcam, by tracking the gaze of the user moving in response to a visual stimulus. The stimulus is designed to facilitate the acquisition of distinguishing features based on the colocation of sets of points along the gaze trajectory.

The paper is organized as follows. In Section 2 a brief overview of the state of the art is presented. Section 3 describes the proposed techniques while Section 4 reports on its experimental evaluation. Finally Section 5 provides conclusions and offers suggestions for further work.

2 Related Work

Various approaches have been presented in the literature to establish liveness and to detect presentation attacks. Liveness detection approaches can be grouped into two broad categories: active and passive. Active approaches require user engagement to enable the facial recognition system to establish the liveness of the source through the sample captured at the sensor. Passive approaches do not require user co-operation or even user awareness but exploit involuntary physical movements, such as spontaneous eye blinks, and 3D properties of the image.

Passive anti-spoofing techniques are usually based on the detection of signs of life, e.g. eye blink, facial expression, etc. For example Pan et al [PWL07] proposed a liveness detection method by extracting the temporal information from the process of the eye blink. They used Conditional Random Fields to model and detect eye-blinks over a sequence of images. Jee et al’s method [JJY06] uses a single ordinary camera and analyses the sequence of the images captured. They locate the centre of both eyes in the facial image. If the variance of each eye region is larger than a preset threshold, the image is considered as a live facial image; otherwise the image is classified as a photograph. Wang et al [WDF09] presented a liveness detection method in which physiological motion is detected by estimating the eye blink with an eye contour extraction algorithm. They use active shape models with a random forest classifier trained to recognize the local appearance around each landmark. They also showed that if any motion in the face region is detected the sample is considered to be captured from an impostor. Kollreider et al [Ko09, Ko08, Ko05] combined facial components (nose, ears, etc.) detection and optical flow estimation to determine a liveness score. They assumed that a 3D face produces a special 2D motion. This motion is higher at central face parts (e.g. nose) compared to the outer face regions (e.g. ears). Parts nearer to the
camera move differently to parts which are further away in a live face. A translated photograph, by contrast, generates constant motion at various face regions. They also proposed a method which uses lip-motion (without audio information) to assess liveness [Ko05].

Some anti-spoofing techniques are based on the analysis of skin reflectance, texture, noise signature etc. Li et al [Li04] explored a technique based on the analysis of 2-D Fourier spectra of the face image. Their work is based on two principles. They proposed the principle that as the size of a photograph is smaller than the real image and the photograph is flat, it therefore has fewer high frequency components than real face images. Kim et al [Ki12] proposed a method for detecting a single fake image based on frequency and texture analyses. They exploited frequency and texture information using power spectrum. They also used Local Binary Pattern (LBP) features for analyzing the textures. They fused information of the decision values from the frequency-based classifier and the texture-based classifier for detecting the fake faces. Pinto et al [Pi12] used the noise signatures generated by the recaptured video to discriminate between live and fake attempts. They suggested noise was the artifact generated from video captured from other video (and not from real scenes). They used the Fourier spectrum, computation of the visual rhythm and extraction of the grey level co-occurrence matrices as feature descriptors.

Systems based on the challenge-response approach belong to the active category, where the user is asked to perform specific activities to ascertain liveness such as uttering digits or changing his or her head pose. For instance Frischholz et al [FW03] investigated a challenge-response approach to enhance the security of the face recognition system. The users were required to look in certain directions, which were chosen by the system randomly. The system estimated the head pose and compared the real time movement (response) to the instructions asked by the system (challenge) to verify the user authenticity. Ali et al [ADH12] presented a method of liveness detection based on gaze tracking. Users are required to follow a moving object with their head/gaze while a camera captures images of the user’s face. The path of the object is designed in such a way that a number of collinear points are visited. Work has also been reported on using the gaze trajectory as a source of biometric information [DG11].

The work presented here explores a new feature set, hereby referred to as the gaze colocation feature set, for the detection of presentation attacks. Although a similar setup to the one in [ADH12] has been used the novel features proposed here establish the ability of the natural gaze to return to the same location consistently. Here the users gaze is directed to some pre-selected random positions on the display and features are extracted from sets of gazes at these colocated targets. The underlying hypothesis is that the variance in gazes for colocated positions should be small in genuine user attempts. This phenomenon is then exploited to differentiate between a photo spoof attack and a genuine user input. Video spoofing presents an even greater challenge. A video camera is required and, as reported, sophisticated methods such as video background control, 3D masks, 3D facial images and placing fiducial points in the background are all being employed to prevent video spoofing [Tr11, Pa11]. In this paper, however, we do not report results of tests on the proposed system under video spoofing attacks.
3 Liveness Detection through Gaze Tracking

The scenario considered in this paper is that of a face verification system using an ordinary camera (webcam). A block diagram of the proposed system is shown in Figure 1. An object appears on the display and the camera (sensor) captures the frames as the position of the object on the display changes. The gaze colocation features are extracted from the pupil centres in the captured frames which are then classified as genuine or fake.

![Figure 1: System block diagram](image)

3.1 Visual Stimulus

A small object appears at random locations on the screen and the user is required to find and follow it with head/gaze movement. It is not necessary to space these targets uniformly but ideally these should not be too close to one another and each should be visited multiple times. At each appearance of the stimulus, the camera captures an image of the user’s face. The presentation of the challenge takes approximately 130 seconds to complete, capturing 90 still images at each location of the challenge. The object appears in a random sequence to prevent predictive video attacks. The object visits each position at least three times. In this way a number of colocated sets of gaze can be identified. In Figure 2(a) a genuine user is seen tracking the challenge to establish liveness, while in Figure 2(b) the impostor is responding to the challenge by carefully shifting a high quality printed photo to gain access to the system.

![Figure 2: Example of (a) Genuine attempt, and (b) Spoof attempt](image)
3.2 Facial Landmark Detection

The images captured during the challenge-response were analysed using STASM [MN08] to extract facial landmark points. STASM returns 68 different landmarks on the face region using an active shape model technique. The coordinates of the center of the pupils were used for feature extraction in the proposed scheme.

3.3 Gaze Colocation Features

The gaze colocation features are extracted from images when the stimulus is at a given location. The ‘x’ and ‘y’ coordinates of the object on the display are same when they reappear at a given place at different times during this exercise. It can therefore be assumed that the ‘x’ and ‘y’ coordinates of the pupil centres in the corresponding frames should also be very close. This should result in a very small variance in the observed x- and y-coordinates of the pupil centres in genuine attempts. A feature vector is thus formed from the variances of pupil centre coordinates for all the occasions where the stimulus is colocated.

Similar features can be extracted from other facial landmarks, but were not used in the results reported here.

4 Experiments

The system setup was similar to the one shown in Figure 3. The setup consists of a webcam, a PC and a display monitor. The camera used is a Logitech Quick Cam Pro 5000, and is centrally mounted on the top of a 21.5” LCD screen, a commonly used monitor type, having a resolution of 1920×1080 pixels and 5ms response time. The distance between the camera and the user was approximately 750 mm. This distance was not a tight constraint but had to be such that the facial features could be clearly acquired by the camera.

![Figure 3: System Setup](image-url)
Data was collected from 8 subjects in 3 sessions. Each subject provided data for both genuine and impostor attempts, creating 26 sets of each. During spoofing attacks a high quality colour photo of a genuine user was held in front of the camera while attempting to follow the stimulus. Each attempt acquired 90 image frames of resolution 352×288 pixels. This resolution provided a good enough picture quality to recognize the facial landmarks. In total, 30 sets of x-y coordinates of the pupil centres from collocated gaze targets were extracted resulting in a feature vector of size 60 for each eye. There were a small number frames where the pupil centres were not detected by STASM and such frames (and associated colocation points) were excluded from the feature extraction process.

For this data, the (x,y) coordinates of the pupil centres from frames captured while users are looking at the central stimulus location are plotted in Figure 4 displaying deviations from their mean for all the genuine and fake attempts respectively. It can be observed that the range of the points in genuine attempts is much smaller compared to that of the spoof attempts. This is because the impostor, relying on hand-eye coordination, is unable to align the photo back to the same spot as accurately as a genuine user.

![Figure 4: Pupil centre deviations capture during (a) genuine attempt and (b) spoof attempt for the central location of the target](image)

### 4.1 Evaluation Framework

Face liveness detection is a two-class problem and there are four possible outcomes of the classification process: true positive, true negative, false negative and false positive. When a genuine (live/non-spoof) attempt is classified as genuine and a false (fake/spoof) attempt is classified as genuine, these are termed true positive (TP) and false positive (FP) classifications respectively. Similarly, when a genuine attempt is classified as a fake and fake attempt is classified as fake these are called true negative (TN) and false negative (FN) respectively. FP and FN are the erroneous outcomes of the process and the rates of their occurrence is reported as False Positive Rate (FPR) and False Negative Rate (FNR) in this report in order to facilitate the assessment and comparison of system performance. The term True Positive Rate (TPR) is also used and is equal to 1-FNR. Total Error Rate (TER) can be defined as the proportion of misclassified attempts out of all the attempts, including both genuine and fake.
For the experiments reported here, the database was divided into two disjoint sets for training and testing purposes. Of the 52 samples, 12 were chosen for testing and the remaining 40 for training the classifier. For training the classifier, 20 random samples from fake and 20 from genuine attempts were chosen. The experiments were repeated 50 times, and on each occasion the system used randomly selected samples for testing and training. The mean error rates are reported here.

4.2 Experimental Results

Error rates were calculated for a range of system parameters and are reported in this section. True Positive Rates at a set of predefined FPR values were obtained and used for comparison. The ROC curve of the proposed scheme using features from the single eye is presented in Figure 7. It is apparent that the system did not perform very accurately when the entire feature vectors are used. However, the performance improved significantly when subsets of the available features were used for training and testing. The forward feature selection method was used to rank the features [BL97]. The best results were achieved when a subset of the best 15 features was used (as shown in Figure 7). At 10% FPR, the TPR was above 90% which was only around 40% when using the entire feature set.

![Figure 7: Performance with features extracted from a single eye](image)

4.3 Feature Combination Schemes

While the colocation features from each eye may be used in isolation it is interesting to explore if there is complementarity in these feature sets and if a greater accuracy can be achieved by their combination. Therefore, both feature and score fusion schemes were explored to find if there can be gain in accuracy by combining information from features extracts from the two eyes. The following sub-sections will cover each of these fusion schemes in turn.
4.3.1 Feature Fusion

The features extracted from both the eyes were concatenated to form a larger feature vector which was then used for training and testing. All 60 features from the left eye and the 60 features from the right eye were combined in a feature-level fusion scheme. The scheme is illustrated in Figure 8. A feature selection method was incorporated to find the optimum feature subsets for this scheme.

Figure 8: Feature fusion using left and right eye

Figure 9 shows the ROC curves for different feature dimensions. The TPR of the system was found to be lower than the instances when only one eye was used. Reducing the number of features improved the performance but the best TPR (at 10% FPR) of the system was about 80% while for the single-eye system it was above 90%.

Figure 9: Feature fusion performance

4.3.2 Score fusion

An alternative to the feature fusion strategy, a score fusion scheme is often implemented. In the score fusion scheme, these features were extracted from the right and left eyes and independent classifiers were used to obtain classification scores for each eye. In this multi-classifier system two k-NN classifiers were used for each eye. The a posteriori probabilities from the two classifiers were combined using the ‘product rule’ for liveness
detection [Ki98]. Figure 10 illustrates the scheme and Figure 11 shows the corresponding ROC curves. The scheme achieved a TPR of 99% at FPR of 10%.

![Score fusion scheme](image)

Figure 10: Score fusion scheme

![ROC curves](image)

Figure 11: Score fusion performance

In order to establish the tradeoff between the feature dimensionality and liveness detection accuracy experiments were performed to establish the performance of the system as the number of dimensions was steadily reduced. Figure 12 illustrates total error rates for different feature dimensions selected using the forward feature selection method. It can be seen that the lowest total error rate was observed when the feature dimension was reduced to around 15. The total error rate started increasing when the feature set was further reduced. The system produced higher total error rates when the feature dimension was large. The reason for this might be that only a small amount of data was available for training given the size of the feature set.
Table 1 presents a comparative performance of the proposed methods at various levels of FPR. The feature fusion scheme gave the highest error rates in all cases. While using features from only one eye, the system TPR was up to 91%. This improved vastly when the score fusion approach was implemented. At 1% FPR, a TPR of 93% was achieved using score fusion. At 10% FPR, this rose to 99%.

Table 1: Performance comparison of the three schemes

<table>
<thead>
<tr>
<th></th>
<th>TPR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>@FPR =0.01</td>
</tr>
<tr>
<td>Single Eye</td>
<td>84%</td>
</tr>
<tr>
<td>Feature Fusion</td>
<td>47%</td>
</tr>
<tr>
<td>Score Fusion</td>
<td>93%</td>
</tr>
</tbody>
</table>

Table 2 shows a comparative analysis of our experimental observations with the performances reported for similar spoof attacks published in the literature. Although the results are from different databases they suggest possible comparative ranking of these various methods and indicate that the proposed method compares favourably with these schemes.
5 Conclusion

This paper presents a novel feature set for liveness detection in the presence of photo spoofing for face verification systems. A challenge-response approach is described which uses a visual stimulus to direct the gaze. The test scenario did not constrain the users to move either their head or eyes exclusively. However, the proposed gaze colocation features provided a robust measure for discriminating between live and fake attempts.

Initial experiments prove the potential viability of this approach, however, more data is required to establish the performance of the proposed approach with confidence. Although video attacks are excluded in the tests it is expected that within the proposed challenge response framework they would be difficult to mount due to the need for synchronisation with the challenge sequence. Future work will expand the experiments to include a larger database of users and will also explore incorporation of additional features for improving the anti-spoofing capabilities of the system in response to more sophisticated attacks. In particular the relative position of eye centres within the face will be a subject of further study.

References


A Simple and Effective Method for Online Signature Verification

Napa Sae-Bae and Nasir Memon
Computer Science Department, NYU-Poly
Six Metrotech Center, Brooklyn, New York, 11201
nsae-b01@students.poly.edu, memon@poly.edu

Abstract: This paper presents a simple and efficient method for online signature verification. The technique is based on a feature set comprising of several histograms that can be computed efficiently given a raw data sequence of an online signature. The features which are represented by a fixed-length vector can not be used to reconstruct the original signature, thereby providing privacy to the user’s biometric trait in case the stored template is compromised. To test the verification performance of the proposed technique, several experiments were conducted on the well known MCYT-100 and SUSIG datasets including both skilled forgeries and random forgeries. Experimental results demonstrate that the performance of the proposed technique is comparable to state-of-art algorithms despite its simplicity and efficiency.

1 Introduction
A handwritten signature is a socially and legally accepted biometric trait for authenticating a human. Typically, there are two types of handwritten signature verification systems: off-line and online systems. In an off-line system, just an image of the user’s signature is acquired without any additional attributes, whereas, in an online system, a sequence of x-y coordinates of the user’s signature along with associated attributes like pressure, time, etc., are also acquired. As a result, an online signature verification system usually achieves better accuracy than an off-line system [FJS11].

The increasing number of personal computing devices that come equipped with a touch sensitive interface and the difficulty of entering a password on such devices [FWW11] have led to an increased interest in developing alternative user authentication mechanisms on such devices [SBAIM12]. In this context, an online signature verification system is a plausible candidate given the familiarity users have with the concept of a signature for the purpose of authentication. This paper presents a simple online signature verification system that is suitable for use on personal computing devices. It has high accuracy with low computation, and space complexity as well as it requires a small number of enrollment samples. In addition, the stored template in the proposed system does not reveal the user’s signature thereby providing privacy protection to an original biometric trait.

1.1 Previous Work
There have been two approaches proposed in the literature for online signature verification, namely, function-based and feature-based approaches [Pla89]. The former refers to a system where the matching process is done using, directly or indirectly, the original time sequence of the signature. The latter refers to a system where the matching process is done using descriptive features of the signature. Examples of well-known function-based approaches include Dynamic Time Warping Algorithm(DTW) [KY08,FZ07, FW03], and Hidden Markov Model(HMM) [OGFAS+03].

The major advantage of a function-based approach is that it generally yields better verification accuracy than a feature-based system [FOGRGR07]. However, a user’s biometric information is not protected, since, during the matching process, a dynamic construction of an original signature is revealed. Furthermore, the system is generally more complex and slower than feature-based systems [FW03]. Even worse, when a template protection approach is applied in order to provide biometric privacy and/or security, then
the verification performance can get significantly degraded. For instance, Maiorana et al [MMDC+08] have proposed a convolution scheme to protect the original signature sequence of the user that can be directly applied to a function based approach in general. The idea is to split the original input sequence into \( W \) subsequences. Each subsequence may have a different selected length based on a random parameter. This technique has been applied with HMM and DTW based verification systems as reported in [MCN09, MMDC+08, NMLC10]. They also reported that, with this convolved version of a signature, verification rates were lower when compared to the original version of the signature.

With a feature-based system, a clear template of a user’s signature does not have to be stored. This results in increased biometric privacy and security. Further, there are many known algorithms to derive a cryptographic key [LTT12, CZC] from feature sets which are typically fixed length. However, the major difficulty for a feature-based approach for online signature verification is to derive a good set of descriptive features that can be used to effectively and efficiently verify an online signature [Pla89, FZ07, FW03].

There have been many proposals to derive a feature set from an online signature. In 2005, Fierrez-Aguilar et al [FANLP+05] proposed a set of 100-features, such as total duration of signature, number of pen ups, sign changes of \( (dx/dt) \) and \( (dy/dt) \), etc. to represent an online signature and applied a feature selection method to rank the proposed features. Based on this 100-feature set, Nanni [Nan06] proposed a multi-matcher method to verify an online signature. The system achieved outstanding performance when two factor authentication is applied, namely a signature sample, and a user-specific token. In addition, Guru and Prakash [GP09] derived a symbolic representation of an online signature and introduced the concept of writer independent threshold in order to improve verification accuracy. Regardless of these efforts, however, the system performance has not been promising when only a feature set is used without a second factor.

Recently, Argones et al [ARMACC12] have proposed a set of HMM model features from a universal background model. The best reported verification performance obtained by their system is promising. However, the system extracts 4800 features from tuning 16 different HMM models, which is a computationally expensive task. Moreover, the universal background model is trained from a pool of 2500 genuine and forged signatures from 50 users on the same device specification, and, in addition, a user-specific classifier is trained from 10 signatures. These make it less feasible to be employed in a mobile device application scenario, where the embedded sensors are different from model to model and only a very few signatures can be taken from a user during enrollment.

1.2 Contributions

This paper presents a method to extract a model-free non-invertible feature set from an online signature. Specifically, the proposed feature set comprises of sets of histograms that capture distributions of attributes generated from several raw signature data sequences and their combinations. Benefits of the proposed method are as follows.

1. The feature set can be computed efficiently, i.e. in linear time proportional to the length of an online signature.
2. The features stored in the system for verification are irreversible. In other words, the original dynamic construction of an online signature is not revealed even when the features are revealed. This is a desirable property from a biometric privacy point of view.
3. Verification performance of the proposed system is superior to several state of the art algorithms on common data sets.
4. There is no large and extensive training set required by the system to train global model parameters. A classifier is derived using only a set of enrolled samples from a specific user. Therefore the verification performance does not depend on the representativeness of the training set which may differ between sensors, native languages of users, and population distributions of training subjects.
5. Features employed in the system are derived from global statistical characteristics of a signature and hence are more robust to fluctuation in local extreme points. This results in competitive verification accuracy even when a small number of samples from a specific user are used to enroll.

The rest of the paper is organized as follows. Section 2 presents a process of deriving a set of histograms from a given online signature, gives details of the proposed online signature verification system, and analyzes its complexity. Experimental results are given in section 3. Section 4 provides conclusions and discussion on future work.

2 The Proposed Online Signature Verification System

This section presents a histogram feature based online signature verification system that comprises of a feature extractor, a template generator and matcher. First, the input online signature is processed by the feature extractor module to extract a set of histograms from which a feature vector is computed. Then, the system constructs a user-specific template from the feature sets derived from multiple enrollment signatures. This template is later used by the matching process to compare against a query signature to verify whether it has been input by the genuine user. The details of each of these components are described in this section.

2.1 Histogram Features

This subsection describes how a set of histograms are computed from an online signature. These histograms are designed to capture essential information of an online signature attributes as well as the relationships between these attributes. Hence they can be used as a succinct representation of an online signature.

Histograms are widely used as feature sets to capture attribute value distribution statistics in many recognition tasks, for instance, in object recognition and off-line signature verification [QLT07]. Using histograms for online signature verification was first suggested by Nelson et al [NTH94]. They have also been used as part of the feature set in [FJS11, FANLP+05]. However, in [FJS11, FANLP+05], they limit the use of histograms only to the angles derived from data points of an online signature. In fact, much more information can be used to construct histograms as a discriminative feature set. These include x-y trajectories, and the corresponding angles, pressure, speed, as well as their derivatives. This paper shows that, when such information is included, the verification performance of the system is significantly improved and it outperforms many of the other state of the art techniques while retaining the inherent simplicity in a histogram based approach.

The feature extraction process of the proposed method begins by decomposing time-series data of a signature to a sequence of cartesian vectors and other attributes as well as deriving their derivatives. Then, each cartesian vector is also converted to a vector in polar coordinate system. Finally, histograms from these vector sequences are derived. Details of the feature extraction process are given below.

Let $X = \{x_1, x_2, \ldots, x_n\}, Y = \{y_1, y_2, \ldots, y_n\}$, and $P = \{p_1, p_2, \ldots, p_n\}$ be the sequences of position in x-axis and y-axis, and pressure, respectively, of an original online signature time-series with length $n$ sampled at times $T = \{t_1, t_2, \ldots, t_n\}$. For the rest of this paper, it is assumed that the time interval between these samples is constant, and hence the time information is implicit and is ignored. It should be noted that if the time interval is not a constant, a normalization process using information from $T$ can be applied to the sequences $X, Y$, and $P$ prior to being processed by the system. We leave this investigation for future work. To construct a set of histogram features, first the descriptive feature vectors $X^k, Y^k$ and
$P^k$ are computed as follows:

\[
X^1 = \{x_i^1 | x_i^1 = x_{i+1} - x_i\},
\]

\[
Y^1 = \{y_i^1 | y_i^1 = y_{i+1} - y_i\},
\]

\[
P^1 = \{p_i\}\]

, where \(i = 1, 2, ..., n - 1\)  \hspace{1cm} (1a)

and

\[
X^k = \{x_i^k | x_i^k = x_{i+1} - x_i \}
\]

\[
Y^k = \{y_i^k | y_i^k = y_{i+1} - y_i \}
\]

\[
P^k = \{p_i^k | p_i^k = p_{i+1} - p_i \}
\]

, where \(k > 1\) and \(i = 1, 2, ..., n - k\)  \hspace{1cm} (2a)

Noting that, by computing a sequence of differences between each pair of successive points as $X^1$ and $Y^1$, the above features serve to eliminate the effect of the first drawing position of a signature (in principle, the system should always accept the signature from the same and honest user regardless of its beginning position.) And by repeating this process of taking differences, $X^k$ and $Y^k$ yields the \(k^{th}\) order derivative of the original $X$ and $Y$ sequences respectively.

Then, a sequence of vectors $V = \{v_i^k\}$, is constructed where each of the vector element, $v_i^k = [v_{x_i}^k, ..., v_{y_i}^k]$ is the concatenation of $v_i^k$ which is a five-tuple consisting of the \(k^{th}\) order derivative of the cartesian and polar coordinates and pressure attributes as follows:

\[
v_i^k = \langle x_i^k, y_i^k, r_i^k, \theta_i^k, p_i^k \rangle
\]

\[r_i^k = \sqrt{(x_i^k)^2 + (y_i^k)^2}
\]

\[\theta_i^k = \tan^{-1}\left(\frac{y_i^k}{x_i^k}\right)
\]

\[i = 1, 2, ..., n - k\]  \hspace{1cm} (3)

A set of histograms from the feature vectors above is then computed from their attribute value distribution (figure 1 illustrates the process of deriving $\theta$ distribution.) The details of these uniform width histograms are given in Table 1. Specifically, they consist of two types of histograms:

1. One dimensional histograms – these capture the distribution of a single attribute. For example, the histogram $\Phi^1$ captures the angle distribution of an online signature since this can be used to broadly reflects the similarity between two signature shapes. Similarly, $\Phi^2$ is used to capture the...
distribution of the angles of the first derivative since it provides more information about how these vectors are aligned, an aspect that is completely ignored in the histogram \( \Phi^1 \). \( R^1 \) is used to capture the speed distribution of an online signature which is one of the distinctive features that is unique among users and is especially useful in combating skilled forgeries.

2. Two dimensional histograms – these capture the distribution of relationship between pairs of attributes, for example, \( < \Phi^1, R^1 >_{(1)} \) and \( < \Phi^1, R^1 >_{(2)} \) capture the distribution of the dependence between speed and angle of the first and the second halves of an online signature. \( < \Phi^1, \Phi^1_{d(1,2)} > \) are used to capture the distribution of the relationship between three consecutive angles of an online signature sequence as well as to provide warping flexibility when comparing two different signatures from the same user.

Table 1: Descriptions of histograms that are used in the proposed technique

<table>
<thead>
<tr>
<th>Histogram</th>
<th>Input Attributes</th>
<th>Min</th>
<th>Max</th>
<th>Number of bins</th>
<th>Output Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Phi^1 )</td>
<td>( {\theta^1_1, \ldots, \theta^1_n} )</td>
<td>(-\pi)</td>
<td>(\pi)</td>
<td>24</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( \Phi^2 )</td>
<td>( {\theta^2_1, \ldots, \theta^2_n} )</td>
<td>(-\pi)</td>
<td>(\pi)</td>
<td>24</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( \Phi^1_{d(1,2)} )</td>
<td>( {\theta^1_{1,2}, \ldots, \theta^1_{n-1,2}, \theta^1_1, \theta^1_n} )</td>
<td>(-\pi)</td>
<td>(\pi)</td>
<td>8</td>
<td>Actual frequency</td>
</tr>
<tr>
<td>( R^1 )</td>
<td>( {r^1_1, \ldots, r^1_n} )</td>
<td>0</td>
<td>(\mu + 3\sigma)</td>
<td>16</td>
<td>Actual frequency</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>( {r^2_1, \ldots, r^2_n} )</td>
<td>0</td>
<td>(\mu + 3\sigma)</td>
<td>16</td>
<td>Actual frequency</td>
</tr>
<tr>
<td>( X^1 )</td>
<td>( {x^1_1, \ldots, x^1_n} )</td>
<td>(\mu - 3\sigma)</td>
<td>(\mu + 3\sigma)</td>
<td>8</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( Y^1 )</td>
<td>( {y^1_1, \ldots, y^1_n} )</td>
<td>(\mu - 3\sigma)</td>
<td>(\mu + 3\sigma)</td>
<td>8</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( X^2 )</td>
<td>( {x^2_1, \ldots, x^2_n} )</td>
<td>(\mu - 3\sigma)</td>
<td>(\mu + 3\sigma)</td>
<td>8</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( Y^2 )</td>
<td>( {y^2_1, \ldots, y^2_n} )</td>
<td>(\mu - 3\sigma)</td>
<td>(\mu + 3\sigma)</td>
<td>8</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( X^1, X^2 )</td>
<td>( {x^1_1, x^2_1} )</td>
<td>(\mu - 3\sigma)</td>
<td>(\mu + 3\sigma)</td>
<td>6</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( Y^1, Y^2 )</td>
<td>( {y^1_1, y^2_1} )</td>
<td>(\mu - 3\sigma)</td>
<td>(\mu + 3\sigma)</td>
<td>6</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( &lt; \Phi^1, R^1 &gt;_{(1)} )</td>
<td>( {\theta^1_1, \ldots, \theta^1_{n/2}} )</td>
<td>(-\pi)</td>
<td>(\pi)</td>
<td>8</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( &lt; \Phi^1, R^1 &gt;_{(2)} )</td>
<td>( {\theta^1_{n/2}, \ldots, \theta^1_n} )</td>
<td>0</td>
<td>(\mu + 3\sigma)</td>
<td>4</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( &lt; \Phi^2, R^1 &gt;_{(1)} )</td>
<td>( {\theta^2_1, \ldots, \theta^2_{n/2}} )</td>
<td>(-\pi)</td>
<td>(\pi)</td>
<td>8</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( &lt; \Phi^2, R^1 &gt;_{(2)} )</td>
<td>( {\theta^2_{n/2}, \ldots, \theta^2_n} )</td>
<td>0</td>
<td>(\mu + 3\sigma)</td>
<td>4</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( &lt; \Phi^1, R^2 &gt;_{(1)} )</td>
<td>( {\theta^1_1, \ldots, \theta^1_{n/2}} )</td>
<td>(-\pi)</td>
<td>(\pi)</td>
<td>8</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( &lt; \Phi^1, R^2 &gt;_{(2)} )</td>
<td>( {\theta^1_{n/2}, \ldots, \theta^1_n} )</td>
<td>0</td>
<td>(\mu + 3\sigma)</td>
<td>4</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( P^1_{(1)} )</td>
<td>( {P^1_1, \ldots, P^1_n} )</td>
<td>0</td>
<td>(\mu + 3\sigma)</td>
<td>8</td>
<td>Actual frequency</td>
</tr>
<tr>
<td>( P^1_{(2)} )</td>
<td>( {P^1_{n/2}, \ldots, P^1_n} )</td>
<td>0</td>
<td>(\mu + 3\sigma)</td>
<td>8</td>
<td>Actual frequency</td>
</tr>
<tr>
<td>( P^2_{(1)} )</td>
<td>( {P^2_1, \ldots, P^2_n} )</td>
<td>(\mu - 3\sigma)</td>
<td>(\mu + 3\sigma)</td>
<td>8</td>
<td>Relative frequency</td>
</tr>
<tr>
<td>( P^2_{(2)} )</td>
<td>( {P^2_{n/2}, \ldots, P^2_n} )</td>
<td>(\mu - 3\sigma)</td>
<td>(\mu + 3\sigma)</td>
<td>8</td>
<td>Relative frequency</td>
</tr>
</tbody>
</table>

The histograms above are computed by splitting the range of the feature (specified by Min and Max columns in Table 1), into a number of equal width bin intervals (also given in Table 1), and counting the number of elements that fall into each particular bin. For an angle attribute and its derivative, the range of its histogram is defined as \([-\pi, \pi]\). For an input attribute which has no explicit boundary, an outlier process with cutoff at three standard deviations from its mean is applied prior to computing the mean and standard deviation of the attribute in order to derive its implicit range described in Table 1. For example, histogram \( \Phi^1 \) is derived from a sequence \( \{\theta^1_i; i = 1, \ldots, n\} \) by forming a 24 bin histogram with equal
width bin intervals beginning from \(-\pi\) to \(\pi\) and counting the number of elements, \(\{\theta^i\}\), that fall into each of the 24 bins. It then results in a vector of 24 bin frequencies.

It should be noted that histogram’s frequencies are divided into two types: absolute frequency – the actual count of elements that fall into a particular bin, and relative frequency – frequencies that are normalized by the total number of elements in the histogram which is essentially the factor of a signature’s length, \(n\). Using absolute frequency results in more implicit importance given to the length of the signature whereas using relative frequency ignore the length of the signature. In this paper, we choose to use relative frequency counts more often than absolute frequency counts. Out of the 21 histograms listed in Table 1, only 5 are absolute frequencies. These are the speed and its first derivative histograms, the pressure histograms of the first and second half of a signature, and the \(<\Phi^1,\Phi^1_d(1,2)>\) histogram as they are derived from the lowest order derivative of considered sequences, which are the most consistent ones in our empirical experiment. In future work, the effect of this choice will be investigated in more detail.

Once, all the histogram vectors are computed, they are concatenated and used as an online signature’s feature vector as follows. Let \(B_i\) be a vector of bin frequencies of \(i^{th}\) histogram. An online signature’s feature vector \(F\) is defined as \(F = \{B_1||B_2||...||B_j\}\), where \(j\) is the total number of histograms, and || is the concatenation operator. Once the feature vector \(F\) is constructed, each of the elements is used independently as a feature component of an online signature. So for the rest of the paper we treat \(F\) as a feature vector and do not distinguish which histogram each feature belongs to. Hence we say we have a feature vector \(F = \{f_i; i = 1,...,M\}\) where \(M\) is the total number of histogram bins from all \(j\) histograms.

2.2 User Template Formation and Verification

This subsection describes the proposed verification system. Generally, an online signature verification system comprises of two stages:

1. Enrollment stage – where a user enrolls in the system by giving multiple online signatures which will later be used to verify a user,
2. Verification stage – where a user claims an identity by inputting a signature on the system’s sensor and the system accepts the signature if the distance between the enrolled template corresponding to that identity and the newly input one is less than a pre-defined threshold.

In the proposed system, an online signature is represented by the set of features derived by the feature extraction module described in the previous sub-section. During the enrollment process, multiple signatures are acquired from a user and features are computed for each sample. The set of feature vectors are then used to identify feature variations of each feature component for the specific user. A user-specific uniform quantizer is constructed for each feature component and the resulting user specific vector of quantization step sizes, \(Q^u\), which we call that quantization step size vector is created. \(Q^u\) helps recover a reliable quantized feature vector from noisy biometric data during the verification process. Each feature vector from the enrollment samples is then quantized according to \(Q^u\) and a feature vector template \(\tilde{F}^u\) for the user is obtained by averaging the quantized feature vectors.

In the verification stage, the claimed user is asked to produce a single signature which is again represented by the set of features derived by the feature extraction module. The system then derives a signature’s quantized feature vector from a given signature using the stored feature quantization step size vector and compares it against the stored user-specific quantized feature vector template. The signature is accepted if the Manhattan distance between these two quantized vectors is less than a predefined threshold, otherwise it is rejected. A summary of the system is shown in Figure 2. The details on how to derive the quantization step size vector \(Q^u\) and the template feature vector are given below.

Let \(S\) be the total number of enrolled samples, and let \(F^u = \{f^u_i; i = 1,...,M\}\) be the feature vector of the enrolled sample \(s\) of the user \(u\) where \(1 \leq s \leq S\), and \(M\) is the total number of online signature features.
The quantization step size vector of the user $u$, $Q^u = \{q_i^u \mid i = 1, \ldots, M\}$, is obtained by computing the standard deviations over all the enrolled samples for each feature and using a multiple of this as the quantization step size. That is,

$$q_i^u = \beta \sqrt{\frac{1}{S} \sum_{s=1}^{S} (f_i^s - \mu_{f_i^u})^2}, \; i = 1, \ldots, M$$  \hspace{1cm} (4)

where $\mu_{f_i^u} = \frac{1}{S} \sum_{s=1}^{S} f_i^s$; $\beta$ is experimentally fixed at 1.5. Then, a quantized feature vector, $\hat{F}^{(s|u)} = \{\hat{f}_i^{(s|u)} \mid i = 1, \ldots, M\}$ is derived from each sample $s$ using the quantization step sizes $q_i^u$ in $Q^u$ (adding a small $\epsilon$ to prevent division by zero) as follows:

$$\hat{f}_i^{(s|u)} = \left[ \frac{f_i^s}{q_i^u + \epsilon} \right], \; i = 1, \ldots, M$$  \hspace{1cm} (5)

where $\epsilon$ is at 0.002 and 0.8 for histograms with absolute and relative frequencies, respectively. Lastly, the user-specific feature vector template, $\bar{F}^u = \{\bar{f}_i^u \mid i = 1, \ldots, M\}$, is derived by averaging the quantized feature vectors of all the enrolled online signature samples from the user $u$.

$$\bar{f}_i^u = \left[ \frac{\sum_{s=1}^{S} \hat{f}_i^{(s|u)}}{S} \right], \; i = 1, \ldots, M$$  \hspace{1cm} (6)

A pair $\left(Q^u, \bar{F}^u\right)$ comprising of the quantization step size vector and its associated feature vector template is then stored as the user $u$’s template and used to verify a claimed signature of the user $u$. During the verification, given that $t$ is claimed to be a sample from user $u$, $\hat{F}^{(t|u)}$ is calculated using $Q^u$. Then the system derives a dissimilarity score using manhattan distance between $\hat{F}^u$ and $\hat{F}^{(t|u)}$ as,

$$Score = \sum_{i=1}^{M} |\bar{f}_i^u - \hat{f}_i^{(t|u)}|$$  \hspace{1cm} (7)

The system then accepts the sample $t$ if the dissimilarity score is less than a predefined threshold, otherwise it rejects.

2.3 Complexity

Given $n$ as the length of an online signature’s sequence, $X^k$, $Y^k$, $R^k$, $\Phi^k$, and $P^k$ can be computed in time $O(n)$. Then, they are used to derive $h$ histograms which yields $O(h \cdot n)$ or $O(n)$ time complexity in deriving a feature vector as $h$ is a constant. For the classification process, a feature vector is first quantized and then used to construct or to compare against the feature vector template. Since the number of features is a constant, the time complexity is $O(1)$. The space required to store a template is clearly a constant as it consists of two fixed-length vectors. As a result, the proposed method requires constant space to store a user’s template and achieves linear time complexity for enrolling and verifying a signature.
3 Experiments

In this section we provide experimental results for the proposed technique and compare its performance to others published in the literature.

Experiments were performed with the well known MCYT dataset [OGFAS+03], which consists of signatures from 100 individuals with 25 genuine samples and 25 skilled forgery samples, and the SUSIG dataset [KY08], which consists of signatures from 94 individuals with 20 genuine samples from two separate sessions and 10 skilled forgery samples. Additional details of these two datasets can be found in [OGFAS+03, KY08].

In terms of training samples, there have been two approaches taken in the literature. Some papers [GP09, YK09] randomly select $k$ samples as the training set and then they average over multiple such random selections to arrive at the final performance result. The reasoning is that such a strategy better captures within-user variation. Other papers [MCN09, ARMACC12, KY08, OGFAS+03, MMDC+08, MCN09] choose the first $k$ samples, according to the original order in which the data was acquired, as the training set. For online signature verification most papers have chosen this second approach to perform experiments since it provides a more realistic result in the sense of mimicking what an application will actually do. In any application, at first enrollment samples will be acquired which would then be repeatedly used to test against query samples. The results from such an approach also captures the timing effect [YK09, FOGRGR07], which is, in fact, one of the major causes for degradation of verification performance called the template aging problem [FOGRGR07, YK09, URJ04] which a system designer often needs to take into consideration.

In our experiments, the first $k$ samples of the set from a specific user were used to enroll a template, and the rest were used to evaluate the False Rejection Rate (FRR) at different threshold levels. In the random forgery scenario or zero knowledge attack, i.e., an attacker simply attacks the system using his own signature, all samples from all other individuals were used to evaluate the False Acceptance Rate, namely FAR-RF. On the other hand, for the skilled forgery scenario, 25 skilled forgery samples from MCYT dataset and 10 skilled forgery samples from SUSIG dataset for each user were used to evaluate the False Acceptance Rate, namely FAR-SF. The Equal Error Rate (EER), the rate at which FAR and FRR are equal, was also used to compare the verification performance of different approaches.

Noting that, in this work, first and second derivative sequences of vectors $v_1^* = [v_1^1, v_2^1], v_2^* = [v_2^1, v_2^2]$ were derived from an online signature. Then a set of 448 features, which consisted of the histograms described in Table 1 were extracted.

3.1 Effectiveness of 1-D versus 2-D histogram features through verification performance

As mentioned in the previous section, the proposed histogram features are derived from two types of histograms namely, one dimensional histograms and two dimensional histograms. These histograms capture the distributions of a single attribute and relationship between pairs of attributes, respectively. The former is commonly used in many feature based systems [NTH94, FANLP+05, Nan06] whereas the latter one, to the best of our knowledge, has never been explored in the research literature. In this subsection, we investigate the effectiveness of using these two types of histograms by evaluating their verification performance against both skilled and random forgery. The plot of the receiver operator characteristic (ROC) curve obtained from MCYT-100 dataset when each of these histograms is used as well as when they are combined, using 10 enrollment samples, is depicted in Figure 3.

The results show that 2-D histograms are indeed a more effective feature set in terms of discrimination power against both skilled and random forgeries compared to 1-D histograms provided that they are employed with larger bin widths. They also appear to work well with less information since pressure information was not included in the 2-D histograms we computed. The results also demonstrate that 1-D histogram features provide complementary information since the best result is observed when the two sets are combined.
histogram features provide complementary information since the best result is observed when the two information was not included in the 2-D histograms we computed. The results also demonstrate that 1-D employed with larger bin widths. They also appear to work well with less information since pressure. The results show that 2-D histograms are indeed a more effective feature set in terms of discrimination they are combined, using 10 enrollment samples, is depicted in Figure 3. (ROC) curve obtained from MCYT-100 dataset when each of these histograms is used as well as when performance against both skilled and random forgery. The plot of the receiver operator characteristic we investigate the effectiveness of using these two types of histograms by evaluating their verification one, to the best of our knowledge, has never been explored in the research literature. In this subsection, selections to arrive at the final performance result. The reasoning is that such a strategy better captures others published in the literature.

In this section we provide experimental results for the proposed technique and compare its performance 3 Experiments

As mentioned in the previous section, the proposed histogram features are derived from two types of histograms namely, one dimensional histograms and two dimensional histograms. These histograms, v

3.1 Effectiveness of 1-D versus 2-D histogram features through verification performance Section 2.1, were derived from an online signature. Then a set of 448 features, which consisted of the v

Table 1 were extracted.

3.2 Verification performance of the proposed system

The performance of the proposed verification system for different number of training samples per user derived from MCYT-100 dataset is reported in Table 2. The results demonstrate that the proposed system can effectively verify a user’s online signature even if only three signatures are supplied by a user during enrollment. However, verification performance at every operating point slightly improves as the number of training samples grows. Using the same dataset, the plot of receiver operator characteristic (ROC) when 5 and 10 training samples are supplied are shown in Figure 4. In Figure 5, we show the the distributions of dissimilarity scores for a pool of 1,500 genuine samples, 2,500 skilled forgery samples and 247,500 random forgery samples drawn from MCYT-100 dataset when 10 training samples were provided.

Table 2: EER of the proposed system derived from MCYT-100 dataset when different number of samples are used for training

<table>
<thead>
<tr>
<th>Number of Training Samples</th>
<th>EER-SF</th>
<th>EER-RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5.74</td>
<td>1.43</td>
</tr>
<tr>
<td>5</td>
<td>4.02</td>
<td>1.15</td>
</tr>
<tr>
<td>7</td>
<td>3.43</td>
<td>0.87</td>
</tr>
<tr>
<td>10</td>
<td>2.72</td>
<td>0.44</td>
</tr>
<tr>
<td>20</td>
<td>2.72</td>
<td>0.35</td>
</tr>
</tbody>
</table>

3.3 Comparison with previous work

The online signature verification systems that have been proposed in the literature can be broadly classified into two types based on the signature input that is presented to a matcher: feature based and function based. The first type of systems is generally more preferable due to the computation and space complexity requirement of the system. However, it is claimed that the latter one usually yields better verification performance [FOGRGR07]. In this subsection, a comparison of performance between the systems that are considered as the state of the art for these two approaches and the proposed approach, which is considered as the feature based system, is provided. Results reported on the proposed system as well as the other system are on the same dataset. The function based approach considered here includes Dynamic Time Warping technique (DTW), Hidden Markov Model (HMM), and their template protection approach. The feature based approach compared includes one utilizing Fourier descriptor features, and a 100-feature system in conjunction with three different classifiers.

Table 3 lists the verification performance of these different techniques on MCYT-100 dataset. As seen from the table, the proposed system outperforms other systems especially when a few training samples are supplied. These results emphasize the competitiveness of the proposed system considering that it is a...
Table 4 reports verification performance for the previous techniques listed above on the SUSIG dataset. As mentioned in [YK09, KY08], they chose to more heavily weigh the signing duration feature as they observed that a skilled forgery signature typically takes twice as long as a genuine one on the average. Hence their FAR-SF is lower than FAR-RF. However, as reported in [YK09, KY08], the EER for the skilled forgery case in this dataset is greatly induced by the significance that a classifier gives to the length disparity between two given signatures. On the other hand, in the proposed system, less weight is given to this length disparity since most of histogram features in the proposed set are attributed by their relative frequency in which the actual length of the signature is ignored. Only 112 histogram features or 25% in the set are attributed by absolute frequency, where the actual length of the signature gets reflected in the feature values. This results in lower FAR-RF but higher FAR-SF than the system in [YK09,KY08]. However, when more weight is given to the histograms with frequency attributes, the EER of skill forgery is reduced from 5.86 to 4.59 whereas that of random forgery remains unchanged. This implies that the verification performance could potentially improve if there are more type of histograms created using absolute frequency and not relative frequency. It also demonstrates that the forger’s skill of these two datasets are very different. However, we acknowledge that reported verification rate is not always the best justification for a system’s effectiveness, since each system might have been trained and tested differently as well as the difference in employing skilled forgery model. In addition, the system might apply different set of features to employ their classifiers. Nevertheless, the results show that the proposed system, at the very least, comparable to others in terms of the verification performance.
The proposed method

The proposed method*

Table 3: EER of different verification approaches on MCYT-100 dataset

<table>
<thead>
<tr>
<th>Matching Types</th>
<th>Approaches</th>
<th>EER-SF</th>
<th>EER-RF</th>
<th>EER-SF</th>
<th>EER-RF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Function-based</td>
<td>DTW [MCN09]</td>
<td>5.53</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DTW [KY05]</td>
<td>9.81</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>DTW-Protected [MCN09]</td>
<td>8.13</td>
<td>-</td>
<td>5.22</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>HMM [MMDC ’08]</td>
<td>10.29</td>
<td>-</td>
<td>6.33</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>HMM-Protected [MMDC ’08]</td>
<td>13.30</td>
<td>-</td>
<td>7.95</td>
<td>-</td>
</tr>
<tr>
<td>Feature-based</td>
<td>Fourier Descriptors in [YK09]</td>
<td>14.53</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>100 global features in [FANLP’05]</td>
<td>6.89</td>
<td>2.2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>100 global features in [Nan06]</td>
<td>7.1</td>
<td>1.6</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>100 global features in [GP09]</td>
<td>6.12</td>
<td>2.05</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>UBM-HMM [ARMACC12]</td>
<td>-</td>
<td>-</td>
<td>2.785</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>The proposed method</td>
<td>4.02</td>
<td>1.15</td>
<td>2.72</td>
<td>0.44</td>
</tr>
</tbody>
</table>

Table 4: EER of different verification approaches on SUSIG dataset

<table>
<thead>
<tr>
<th>Matching Types</th>
<th>Approaches</th>
<th>EER-SF</th>
<th>EER-RF</th>
<th>EER-SF</th>
<th>EER-RF</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Function-based</td>
<td>DTW [KY05, YK09, KY08]</td>
<td>3.30</td>
<td>4.08</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Feature-based</td>
<td>Fourier Descriptors in [YK09]</td>
<td>6.20</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The proposed method</td>
<td>6.08</td>
<td>2.94</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The proposed method*</td>
<td>4.37</td>
<td>2.91</td>
<td>-</td>
<td>-</td>
<td>* when the weight of R and Φ = Φ[1:2] histogram attributes is given 3 times of others</td>
</tr>
</tbody>
</table>

4 Conclusion and Future Work

This paper proposes a simple and effective online signature verification system. First, a histogram based feature set that can be efficiently computed is introduced. Second, a model-free Manhattan distance classifier based on a quantized feature vector is used to verify an online signature sample. This implies that a user-specific classifier can be trained using only a few enrollment samples without requiring a training set with a large number of samples. Therefore the technique is suitable for employment in a mobile device application where sensors may differ from one device to another. More importantly, since the feature set employed in the proposed system represents only statistics derived from the original sequence, the transformation is non-invertible where the privacy of the original biometric data is protected. Testing with MCYT dataset, the proposed system achieves competitive performance when compared to other proposed systems.

The limitations of the current work are as follows. First, it is currently possible to match different signature templates generated from the same online signature samples and thereby learn that two leaked biometric templates belong to the same user. Work needs to be done on how one can inject some randomness in the original template generation process so that different templates can be generated from the same set of signatures without compromising system performance. In addition, we plan to further investigate the use of other biometric key binding approaches like fuzzy commitment that could be applied on the proposed feature set in order to strengthen security of the system. Secondly, it is also important to evaluate performance on the dataset that collected in mobile authentication context, i.e., users may sit or stand while signing on the current handset devices that may have non-uniform sampling rate. Then the proposed algorithm can be modified accordingly.

Acknowledgments

The authors were supported by NSF grant 1228842. We also would like to thank the anonymous reviewers for their comments and feedbacks towards this work.
References


EEG Based User Recognition Using BUMP Modelling

Daria La Rocca*, Patrizio Campisi*, Jordi Solé-Casals**

*Section of Applied Electronics, Department of Engineering, University of Roma Tre,
Via Vito Volterra 62, 00146, Roma, Italy
daria.larocca@uniroma3.it, patrizio.campisi@uniroma3.it
** Escola Politècnica Superior, Universitat de Vic
C/ de la Laura, 13, 08500 Vic, Catalunya
jordi.sole@uvic.cat

Abstract: In this paper the use of electroencephalogram (EEG) as biometric identifier is investigated. The use of EEG within the biometric framework has already been introduced in the recent past although it has not been extensively analyzed. In this contribution we apply the “bump” modelling analysis for the feature extraction stage within an identification framework, in order to reduce the huge amount of data recorded through EEG. For the purpose of this study we rely on the “resting state with eyes closed” protocol. The employed database is composed of 36 healthy subjects whose EEG signals have been acquired in an ad hoc laboratory. Different electrodes configurations pertinent with the employed protocol have been considered. A classifier based on Mahalanobis distance have been tested for the enrollment of the subjects and their identification. An information fusion performed at the score level has shown to improve correct classification performance. The obtained results show that an identification accuracy of 99.69% can be achieved. It represents an high degree of accuracy, given the current state of research on EEG biometrics.

1 Introduction

EEG waves have been largely studied in the past with the aim of analyzing brain functioning, reflected in rhythmic activity of local or widespread neurons’ networks during specific mental states. This activity can be measured extracting from the EEG signals the main brain rhythms involved in the integration of different processes, which show specific frequency content: δ([0.5 − 4]Hz), θ([4 − 8]Hz), α([8 − 14]Hz), β([14 − 30]Hz) and γ (over 30Hz). The description of the brain functioning represented by EEG oscillations, recorded from the different scalp regions, is supposed to reflect individual-specific anatomical and psycho-physiological traits. In the last decade EEG signals have been proposed to be used in biometric based recognition systems. The aim of such frameworks is to detect and quantitatively evaluate inter-subject variability in EEG features, expected to show stability in an individual over time, as discussed in [CLRS12]. Some promising results have been obtained employing different EEG acquisition protocols, involving both resting conditions with closed or opened eyes, and response to specific internal or external events. EEG signals present some peculiarities, which are not shared by the most commonly used biometrics, like face, iris, and fingerprints. Specifically, brain signals, as a result of the electrical activity of the cortex, are not exposed like face, iris, and fingerprints. Therefore,
they are more privacy compliant than other biometrics since they are "secret" by their nature, being impossible to capture them at a distance. This property makes EEG biometrics also robust against the spoofing attack at the sensor, since an attacker would not be able to collect and feed the EEG signals, which are the result of ionic current flows within the neurons of the brain. Moreover, being brain signals the result of a cognitive process, they cannot be synthetically generated and fed to a sensor, which also addresses the problem of liveness detection. Also, the level of universality of brain signals is very high. In fact people with some physical disabilities, preventing the use of biometrics like fingerprint or iris, would be able to get access to the required service using EEG biometrics. Due to a not practical acquisition process, also very sensitive to external and physiological noise, applications of EEG biometrics could be designed within high security contexts, given the actual technology. In fact, applications for everyday life access to personal utilities would result very impractical. Some interesting evidences have already been obtained in the recent literature, see for example [PRCE99], [BK10], [RSFC+], [MdRM07], and [CSB+11] where a review on the state of the art of EEG biometrics is also given. In the aforementioned papers, different acquisition protocols have been employed, such as closed or open eyes resting conditions, where the user is asked to rest and not to perform any specific task, or response to specific stimuli. Despite the promising results a systematic analysis of EEG biometric traits aiming at identifying individual specific features is missing.

The purpose of this work is to provide an analysis of parametric time-frequency maps of the acquired EEG signals, in order to identify distinctive measures of the map parameters for each individual. Therefore, in the proposed EEG-based biometric framework we focus on the features extraction stage, evaluating the significance of each considered variable in catching the differences between individuals. We rely on the resting state acquisition protocol to acquire data from 36 healthy subjects. Different configurations for the spatial placement of the electrodes are tested. Specifically, sets of three acquisition channels are considered selecting inter-hemispheric symmetrical configurations and combinations of mid-line electrodes. The analysis of each frequency band is also carried out in order to evaluate features of different brain rhythms, functionally involved in the integration of different kind of brain activities. The so acquired signals, after proper preprocessing, are then modeled employing wavelet decomposition to extract time-frequency maps. The wavelet analysis of EEG signals allows to obtain a dynamic representation of the frequency content able to catch even transient phenomena. The multi-scale approach allows to locate accurately time-frequency oscillations with different resolutions in the time-frequency space. Bump modelling of wavelet maps is subsequently performed in order to reduce the data dimensionality, and to provide relevant features. Fusion at the matching score level is also performed. The proposed system uses the combination of scores obtained considering different electrodes configurations and different frequency bands. A Mahalanobis distance based classifier is tested for identification purpose.

The paper is organized as follows. The employed modelling is detailed in Section 2, and in Section 3 the acquisition protocol, the template extraction stage and the classification procedure are described. Experimental results are given in Section 4, where the recognition performance is reported for all the tests carried out. Finally conclusions are drawn in Section 5.
2 Bump modelling

Given raw EEG signals collected through a multi-channel EEG device, we seek to obtain a set of parameters allowing characterizing subjects from their brain electrical activity. In order to achieve this goal, we use bump modelling [VMD+07], which is a technique for modelling a time-frequency map, with the aim of representing the map with a limited number of elementary functions. The purpose is to reduce the huge quantity of parameters that describe a time-frequency map, tens to hundreds of thousands, to a sum of parametric functions, a few functions with some tens of parameters. These parameters will be used in order to characterize each subject.

The main idea of this method is to approximate a time-frequency map with a set of predefined elementary parameterized functions called bumps; therefore, the map is represented by the set of parameters of the bumps, which is a very sparse encoding of the map, resulting in information compression rates that range from one hundred to one thousand (rationales for this procedure, proofs and technical details are explained in [VMD+07]).

The algorithm performs the following steps on the time-frequency maps (after appropriate normalization):

i window the map in order to define the zones to be modelled (those windows form a set of overlapping sub-areas of the map);

ii find the window that contains the maximum amount of energy;

iii adapt a bump $\beta$ to the selected zone, and withdraw it from the original map. The parameters of the bumps are computed using the BFGS algorithm [PFTV92] in order to minimize the cost function $C$ defined by:

$$C = \frac{1}{2} \sum_{t,f \in W} (z_{f,t} - \beta(f,t))^2$$  \hspace{1cm} (1)

where the summation runs on all pixels within the window $W$, $z_{f,t}$ are time-frequency coefficients at time $t$ and frequency $f$, and $\beta(f,t)$ is the value of the bump function at time $t$ and frequency $f$;

iv if the amount of information modelled by the bumps reaches a threshold, stop; else return to (iii).

EEG signals are transformed to time-frequency maps using Complex Morlet wavelets, as they are appropriate for time-frequency analysis of electroencephalographic signals because of its symmetrical and smooth Gaussian shape both in time and frequency domains [KMR87].

$$w(t) = \exp\left(-t^2/2\sigma_t^2\right)\exp(2i\pi rt)$$  \hspace{1cm} (2)

Bump functions used are half ellipsoids. Half ellipsoids (see Figure 1) are defined by:

$$\beta(f,t) = a\sqrt{1-v} \hspace{0.5cm} for \hspace{0.5cm} 0 \leq v \leq 1$$

$$\beta(f,t) = 0 \hspace{0.5cm} for \hspace{0.5cm} v > 1$$  \hspace{1cm} (3)
where $v = (e_f^2 + e_t^2)$ with $e_f = \frac{f - \mu_f}{l_f}$ and $e_t = \frac{t - \mu_t}{l_t}$. $\mu_f$ and $\mu_t$ are the coordinates of the center of the ellipsoid, $l_f$ and $l_t$ are the half-lengths of the principal axes, $\alpha$ is the amplitude of the function, $t$ is the time and $f$ the frequency.

Figure 2 shows a typical example of bump modelling of the time-frequency map of an EEG recording. Each bump is described by 5 parameters: its coordinates on the map (2 parameters), its amplitude (one parameter) and the lengths of its axes (2 parameters). All the experiments performed in this work have been done using the BUTIF Toolbox [VSCD+09]¹

### 3 Data Analysis

In the data collection stage brain activity was recorded using a BrainAmp recording system, from Brain Products², operating at a sampling rate $S_r = 200Hz$. The EEG recordings of $N_C = 36$ healthy volunteers have been acquired. Informed consent was obtained from each subject after the explanation of the study, which was approved by the local institutional ethical committee. During the experiment, the participants were comfortably

²details on the amplifier device in http://www.brainproducts.com/index.php
seated in a reclining chair with both arms resting on a pillow in a dimly lit room, properly
designed in order to minimize external sounds and noise, not interfering with the attention
and the relaxed state of subjects. The EEG was continuously recorded from $C_T = 56$
sites on the scalp, positioned according to the 10 $-$ 20 international system as shown in
Figure 3, and potentials were referenced to the average signal from the ear lobes. Before
starting the recording session, the electrical impedance of each electrode was kept lower
than $10k\Omega$ through a dedicated gel maximizing the skin contact and allowing for a low-
resistance recording through the skin. A set of $N_C = 36$ EEG digital recordings from
$C_T = 56$ channels $V_{i,ch}[n]$, for $i = 1, \cdots, N_C$, $n = 1, \cdots, N_T$, and $ch = 1, \cdots, C_T$ has
been obtained. The recorded signals have been consequently preprocessed as described in
3.2.

3.1 EEG acquisition protocol

Since the earliest applications of the EEG signals, particular interest has been shown in the
study of cerebral activity during resting sate, due to the amount of information conveyed in
it with respect to brain functioning and organization. Specific features of the brain activity
during resting, as described by EEG signals, have given some indications about their capa-
bility to distinguish among people. In fact the resting state protocol has been employed in
the biometric framework [ASLA10], [LRCS12] for recognition purpose. In the presented
work we refer to EEG signals acquired during resting conditions. Specifically, the subjects
were asked to perform one minute of “resting state with closed eyes”. Therefore in each
session an $N_T = 200 \times 60$ samples long record was provided for each acquisition channel.
3.2 Preprocessing

In the preprocessing stage raw signals are first downsampled applying a decimation factor to the collected data. A sampling rate of $S_r = 100\text{Hz}$ and its anti-aliasing FIR filter are selected according to the Nyquist theorem. A further band-pass filtering stage is applied after down-sampling in order to retain spectral information in the band $[0.5, 40]\text{Hz}$, containing most of the frequency components of interest referring to the resting state condition. The so obtained signals are then segmented into $M$ overlapped frames of length $T$. A frame length of $T = 15\text{s}$ is selected, in order to achieve a trade-off between the sample size and the quality of the time-frequency decomposition at lower frequency range. Specifically, a time interval of one second is considered between the beginnings of consecutive frames. Such a considerable overlap resulting from the described segmentation is needed, due to the small sample size, in order to obtain an adequate number of instances which generate class distributions required when training the proposed classifier. $M = 45$ EEG overlapped frames are obtained for each subject and each channel. Subsequently the DC component is removed from each EEG segment, and a $z$-score normalization of each signal is performed in order to make signals referring to different acquisition sessions (users) reliably comparable, removing differences due to scale factors. The so obtained datasets, \{ $Z_{i,ch}^{m}$ \}, with $i = 1, \cdots, N_C$, $ch = 1, \cdots, C_T$, and $m = 1, \cdots, M$ is further processed to extract time-frequency maps from each user brain signal, as described in the following section.

3.3 Modelling and features extraction

In the herein proposed study, particular interest was directed to the data modelling in the time-frequency domain as described in Section 2. In more detail, for each subject $i$ and each channel $ch$, the wavelet decomposition of EEG signals was computed employing Complex Morlet wavelet, so that each frame $m$ was represented by the extracted wavelet coefficients in the time-frequency space. The frequency range from $3\text{Hz}$ up to $40\text{Hz}$ was investigated, while lower frequencies were removed from the analysis, since the selected frame length limited the estimation accuracy of the related wavelet coefficients. Maps of coefficients were obtained analyzing the different brain rhythms $\theta, \alpha, \beta$ and $\gamma [30 - 40]\text{Hz}$ individually, as well as altogether. Therefore for each of the $M$ EEG frames, five maps were provided $z_{f,t}^{ch}$, with $b \in \{ [3 - 40]\text{Hz}, \theta, \alpha, \beta, \gamma [30 - 40]\text{Hz} \}$. Bump modelling was then employed to obtain a parametric description of most energy contained in the wavelet maps, as described in Section 2. The publicly available BUTIF toolbox was used to extract the 5 parameters for each bump fitting the energy spots in the time-frequency representation of the EEG signals. Subsequently the analysis of the obtained parametric maps was performed in order to extract discriminant features, as reported in the following. A statistical analysis was carried out considering the median values for the amplitude ($A$), the central frequency ($f_c$), the volume ($v$), the surface of the meridian ellipse ($a$), the height (extension in frequency $h$) and the width (extension in time $w$) of the extracted bumps, modelling the time-frequency maps. Also we included in the analysis of features the total number of bumps for each map, the sum and max values of $A$, $f_c$, $v$, $a$, $h$ and $w$, the number of bumps within different ranges of the investigated time-frequency domain, the amplitudes and central frequencies of bumps showing max values of parameters $a$, $v$, $w$.
and \( h \). Moreover for each bump map we obtained the spectral centroid \( c(t) \), that is the “center of mass” of the bumps in the frequency domain defined for each time step \( t \), computed as the weighted mean of the frequencies present in the map for each instant, with the corresponding map amplitudes as the weights. The average value over time of \( c(t) \) was considered as a further feature. Furthermore, each of the listed parameters were also computed for filtered maps where a threshold value \( \tau = 0.4 \) was applied to the bump amplitude \( A \). Therefore a feature vector of 39 components \( \zeta_{tot} = [\zeta, \zeta_\tau] \) was obtained for each EEG frame, where parameters referred respectively to the unfiltered and the filtered maps were concatenated. The one-way Analysis of Variance (ANOVA) [Sch99] was performed in order to explore the within-subject and between-subjects variance of each feature extracted from the parametric time-frequency maps. Within the assumptions of normal distributions of features, homogeneity of variance of the different group distributions, and independence of observations, the null hypothesis to test was that data extracted from all groups (corresponding to subjects) show the same stochastic distribution, and that observed differences of values assumed by a certain feature between subjects is due to the case. According to the outcome of the test, the features which showed to vary significantly between subjects were selected to train the classifier for the recognition purpose. The so selected features were further tested performing the same ANOVA analysis for each pair of subjects, in order to verify the disjunction of all subjects from each-others. A final combination of features was selected, such that all pairs of subjects showed significant difference for most of them (at least two thirds). As a result of the statistical analysis, two thirds of the computed features were discarded and the remaining 13 discriminant parameters were considered to form the feature vector used to train and test the classifier. The features which individually showed to significantly discriminate group (subject) distributions with a confidence level of 99% are: the median value of \( A \) for both the unfiltered and filtered maps, the median value of \( v \), the number of bumps within different adjacent ranges of the investigated time-frequency domain (6 elements), the number of bumps with \( A > \tau \), the sum of the values of \( A \) for both the unfiltered and filtered maps, and the mean of \( c(t) \) over time. Different spatial distributions of sets of three electrodes were considered, placed according to both symmetrical inter-hemispheric and mid-line configurations, as shown in Figure 3 with the marked electrodes. Features related to channels of each set under analysis were concatenated in a unique vector. Therefore, the resulting feature vector was obtained for each subject \( i \), each set of channels \( Ch \), each rhythm \( b \) and each frame \( m \), providing a dataset \( Ch^b_\hat{\zeta}(m) \) to evaluate the recognition performance as discussed below.

3.4 Classification

A linear discriminant analysis was performed for the identification purpose, in order to predict the class, namely the user identity, to which the observed feature vector \( \hat{\zeta} \) belongs to. The model used for the discriminant analysis assumes that the vector \( \hat{\zeta} \) has a Gaussian mixture distribution, the same covariance matrix for each class, and that only the means vary. The extracted feature vectors were divided into a training dataset \( \hat{\zeta}_{train} \) used to enroll users, and a test dataset \( \hat{\zeta}_{test} \) used to test the classification performance in terms of correct recognition. The correct recognition rate (CRR) we computed to evaluate the system’s accuracy is defined as the average over the diagonal of the resulting misclassification matrix. It represents the percentage of test trials which led to a correct identification of each user.
within the cross-validation framework provided, averaged over the subjects. The classifier we implemented is based on the assumption of Gaussian mixture distribution drawn by the feature vectors of the training dataset. Mahalanobis distances, in squared units, are computed between each observation in the test dataset, that is feature vectors averaged over frames, and the mean of each of the $N_C$ class distributions representing the training dataset. For each tested condition, the classification of the extracted feature vectors was carried out within a cross-validation framework providing 45 runs. In each cross-validation run a different partition of the entire dataset into training and test subsequent frames was provided, consecutively selecting one of the $M = 45$ frames to start the training. Due to the small sample dimension when considering 60 seconds EEG recordings, initial tests were carried out including overlapping frames between the train and test datasets in the analysis. In this initial experiments 35 frames were used to train the classifier and the remaining 10 frames to test the recognition accuracy. The best performing conditions were further examined removing the overlap between the training and test datasets, and the resulting performance shift was evaluated. In this latter case 20 consecutive frames were employed in the training stage, while only 5 consecutive frames form the test dataset. For each considered frequency band and each set of channels, the training stage consists of the evaluation of the class distributions of the feature vectors $\hat{\zeta}$, which were supposed belonging to the Gaussian mixture. In the test stage for each user $j$ to be identified, the Mahalanobis distances $d(j, i)$ between the mean of the feature vectors $\hat{\zeta}^{\text{test}}_j$ and the means of the observed $N_C$ class distributions $\mu_i$ were evaluated. For each considered vector a transformation of the classifier score $d(j, i)$ was obtained performing the multiplicative inverse $s(i, j) = 1/d(j, i)$. An information fusion integrating multiple sensors distributions and brain rhythms was then performed at the match score level, which is the most common approach in multibiometric systems [RNJ06]. The aim was to determine the best sets of channels configurations and frequency bands that could optimally combine the decisions rendered individually by each of them. The score fusion was obtained through the sum $\sum_{b \in N_B} \sum_{Ch \in S} s_b^{Ch}(i, j)$ of scores related to specific bands $b \in N_B$ and selected sets $Ch \in S$ composed of three electrodes. All tests performed and obtained results are reported in the next Section.

4 Results and Discussion

For the purpose of user identification EEG signals have been acquired from 36 subjects in resting conditions, and modelled in the time-frequency domain, in order to extract discriminant features. Bump modelling has been employed to reduce the dimensionality of the time-frequency representation of each EEG segment, while retaining most of the energy content, as discussed in Section 2. Within the feature extraction stage of the proposed biometric framework an extensive analysis of the parameters which describe the bump maps led to obtain a set of discriminant traits. Different tests have been carried out in order to infer about the best performant electrodes configuration and to analyze the distinctive contribution of each brain rhythm. More in detail, given the “resting state with eyes closed” acquisition protocol here investigated and the 56 employed channels shown in Figure 3, we considered different subsets of acquisition channels in order to find

---

3 The authors thank prof. F. Babiloni Fondazione Santa Lucia, Rome, for having provided the dataset
the best performing spatial arrangements of the electrodes while minimizing their number. To achieve this goal we selected sets of three symmetrical inter-hemispheric and sets of mid-line electrodes. Template extraction has been performed as described in Section 3.3, by first preprocessing the EEG signals, which includes decimation with sampling rate \( S_r = 100 \) Hz, band-pass filtering to remove very low and high frequency noise, z-score normalization and segmentation into overlapping frames of \( T = 15 \) s. Then the so obtained frames are modeled through wavelet decomposition and parametric functions are used to obtain compact representations of the extracted wavelet coefficients in the time-frequency domain, employing the BUTIF toolbox. Different features have been extracted from the bump maps, and the analysis of variance was employed to evaluate the inter-subject variability of each of them, so that 13 features have been selected as discussed in Section 3.3. The subbands related to the different brain rhythms \( \theta, \alpha, \beta, \gamma([30 - 40]\) Hz), which are the ones interested by the “resting state with eyes closed” protocol, and their combination \([3 - 40]\) Hz are individually modelled and analyzed. For each identity the template is subsequently obtained by concatenating the selected features related to the different electrodes in the set under analysis, thus generating feature vectors of length \( 13 \times 3 \) for the sets of three electrodes listed in Table 2, first column.

In Table 1 the results obtained from the analysis of different subbands using the classifier based on the Mahalanobis distance, described in Section 3.4, are given for some of the tested electrodes configurations. The correct recognition percentages reported in the table refers to overlapped training and test datasets, and are obtained within a cross-validation framework as described in Section 3.4. In Table 2 the same analysis for a larger set of scalp electrodes configurations is provided considering disjoint datasets, which are obtained by removing overlapping frames. It is worth pointing out that the classification performance varies considerably for the different scalp regions and rhythms under analysis. Moreover the performance significantly decreases for the disjoint datasets than for the overlapped frames. On the other hand, the match score fusion obtained as discussed in Section 3.4 has led to a dramatic increase in recognition accuracy, especially for the otherwise poorly performing case of disjoint training and test datasets, as observed in Figure 4. For the selection of the frequency bands to combine, the best performing set of three channels \( PO_3 - PO_z - PO_4 \) (see Table 2) was considered, and subsequent score fusions were performed. To this aim the brain rhythms were sorted in descending order of performance achieved individually, and sequentially combined within a forward-backward stepwise approach, retaining in the information fusion only those bands which improved the correct classification. Results reported in Figure 4 showed that a significant improvement was obtained combining \( \beta, \alpha \) and \( \theta \) rhythms. The same approach was considered to select the combination of electrodes configurations, while considering the best performing band fusion \( \beta \cup \alpha \cup \theta \). For the case of disjoint training and test datasets a correct recognition percentage of 99.69% could be achieved when the sets of three inter-hemispheric and midline channels \( PO_3 - PO_z - PO_4, O_1 - PO_z - O_2, CP_5 - CP_z - CP_6, TP_7 - CP_z - TP_8, PO_4 - PO_z - PO_2, PO_2 - P_z - CP_2, T_7 - C_z - T_8, P_5 - P_2 - P_6, CP_z - C_z - FC_z \), and the three rhythms \( \theta, \alpha, \beta \) containing most information were combined into the match score fusion. It should be noticed that the selected channels result located in the posterior region of the head, where the considered rhythms are mainly detected. Figure 4 reports the improvements obtained in the subsequent steps of the information fusion. Moreover the re-
Figure 4: Improvement of the correct recognition rate obtained performing subsequent score fusions (see Section 4). Curves refer to the combination of different brain rhythms (top x-axis) and different electrodes sets (bottom x-axis). Labels in the x-axes refers to the score added at the related step.

Resulting misclassification matrix showed that a 100% of correct identification was obtained for all users but one, who presented correct recognition rate of about 89%. The result obtained did not differ significantly from the perfect recognition performance (100%) obtained for the case of overlapped datasets, within the same information fusion approach and cross-validation framework.

5 Conclusions

In this paper the feature extraction from parametric representations of EEG signals, within the framework of EEG based biometric recognition, has been addressed. The simple “resting state with eyes closed” protocol has been employed to acquire a database of 36 people. Electrodes configurations have been selected in order to find the brain region containing the most significant information for the user identification purpose. Extensive simulations have been performed by considering different sets of 3 electrodes with respect to their positioning. Bump modeling has been employed for the parametric representation of the time-frequency maps, and the extraction of different features. Different subbands have been tested. In summary our analysis has shown that a very high degree of correct recognition can be achieved with a multi-biometric approach consisting in an information fusion at the match score level, combining specific sets of electrodes configurations and brain rhythms.

A small set of features of parametric time-frequency maps obtained from EEG segments showed a significant variability between users, as observed from recognition accuracy obtained employing a simple Mahalanobis based classifier. In future works other classifiers could be considered in order to overcome the assumptions of the linear discriminant analysis, and to best fit the separation surfaces between the decision regions in the classification problem, since such surfaces are generally not hyper-planes. Different task performances
<table>
<thead>
<tr>
<th>Channels</th>
<th>3 – 40Hz</th>
<th>$\theta$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 Cz C2</td>
<td>62.22</td>
<td>76.60</td>
<td>83.02</td>
<td>83.21</td>
<td>73.27</td>
</tr>
<tr>
<td>C3 Cz C4</td>
<td>73.27</td>
<td>84.26</td>
<td>87.10</td>
<td>87.90</td>
<td>76.67</td>
</tr>
<tr>
<td>C5 Cz C6</td>
<td>75.56</td>
<td>86.54</td>
<td>89.94</td>
<td>91.67</td>
<td>83.64</td>
</tr>
<tr>
<td>T7 Cz T8</td>
<td>78.83</td>
<td>85.43</td>
<td>93.09</td>
<td>92.96</td>
<td>85.19</td>
</tr>
<tr>
<td>PO1 POz PO2</td>
<td>79.69</td>
<td>86.23</td>
<td>90.43</td>
<td>93.77</td>
<td>79.01</td>
</tr>
<tr>
<td>TP7 CPz TP8</td>
<td>80.68</td>
<td>87.90</td>
<td>92.65</td>
<td>93.70</td>
<td>83.09</td>
</tr>
<tr>
<td>P7 Pz P8</td>
<td>81.60</td>
<td>89.44</td>
<td>94.57</td>
<td>92.90</td>
<td>83.33</td>
</tr>
<tr>
<td>O1 POz O2</td>
<td>82.59</td>
<td>86.30</td>
<td>94.14</td>
<td>94.38</td>
<td>83.70</td>
</tr>
<tr>
<td>PO3 POz PO4</td>
<td>84.69</td>
<td>89.69</td>
<td>95.74</td>
<td>95.19</td>
<td>83.21</td>
</tr>
</tbody>
</table>

Table 1: Classification performance in terms of correct recognition rate. Results refer to overlapped training and test datasets. The analysis of individual subbands is reposted in subsequent columns, and the set of 3 channels considered is listed in the first column.

<table>
<thead>
<tr>
<th>Channels</th>
<th>3 – 40Hz</th>
<th>$\theta$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1 Cz C2</td>
<td>24.81</td>
<td>38.61</td>
<td>43.70</td>
<td>45.65</td>
<td>35.56</td>
</tr>
<tr>
<td>CP1 CPz CP2</td>
<td>28.98</td>
<td>42.69</td>
<td>44.35</td>
<td>48.61</td>
<td>33.89</td>
</tr>
<tr>
<td>FCz Fz AFz</td>
<td>29.54</td>
<td>42.13</td>
<td>47.96</td>
<td>46.57</td>
<td>34.44</td>
</tr>
<tr>
<td>F3 Fz F4</td>
<td>29.63</td>
<td>41.02</td>
<td>54.35</td>
<td>51.39</td>
<td>38.98</td>
</tr>
<tr>
<td>Cz FCz Fz</td>
<td>30.19</td>
<td>39.17</td>
<td>49.07</td>
<td>46.85</td>
<td>39.44</td>
</tr>
<tr>
<td>CPz Cz FCz</td>
<td>30.46</td>
<td>44.81</td>
<td>49.72</td>
<td>53.70</td>
<td>41.02</td>
</tr>
<tr>
<td>C3 Cz C4</td>
<td>31.76</td>
<td>43.80</td>
<td>48.43</td>
<td>52.22</td>
<td>40.00</td>
</tr>
<tr>
<td>Pz CPz Cz</td>
<td>32.59</td>
<td>42.96</td>
<td>57.78</td>
<td>55.83</td>
<td>39.72</td>
</tr>
<tr>
<td>Fz AFz Fpz</td>
<td>34.17</td>
<td>38.43</td>
<td>44.63</td>
<td>51.20</td>
<td>31.30</td>
</tr>
<tr>
<td>CP5 CPz CP6</td>
<td>37.04</td>
<td>50.28</td>
<td>53.61</td>
<td>53.43</td>
<td>36.20</td>
</tr>
<tr>
<td>CP3 CPz CP4</td>
<td>37.59</td>
<td>45.09</td>
<td>49.35</td>
<td>55.28</td>
<td>36.76</td>
</tr>
<tr>
<td>FC3 FCz FC4</td>
<td>37.87</td>
<td>43.06</td>
<td>52.69</td>
<td>56.67</td>
<td>47.31</td>
</tr>
<tr>
<td>C5 Cz C6</td>
<td>40.00</td>
<td>46.11</td>
<td>57.78</td>
<td>58.70</td>
<td>48.98</td>
</tr>
<tr>
<td>P5 Pz P6</td>
<td>41.57</td>
<td>49.63</td>
<td>69.17</td>
<td>63.06</td>
<td>40.65</td>
</tr>
<tr>
<td>T7 Cz T8</td>
<td>42.41</td>
<td>42.41</td>
<td>62.41</td>
<td>62.69</td>
<td>50.09</td>
</tr>
<tr>
<td>POz Pz CPz</td>
<td>42.87</td>
<td>49.17</td>
<td>59.54</td>
<td>61.02</td>
<td>37.41</td>
</tr>
<tr>
<td>PO1 POz PO2</td>
<td>46.57</td>
<td>52.78</td>
<td>62.13</td>
<td>66.39</td>
<td>45.56</td>
</tr>
<tr>
<td>P7 Pz P8</td>
<td>46.94</td>
<td>53.98</td>
<td>71.11</td>
<td>62.31</td>
<td>47.50</td>
</tr>
<tr>
<td>TP7 CPz TP8</td>
<td>49.07</td>
<td>48.61</td>
<td>64.44</td>
<td>69.26</td>
<td>48.52</td>
</tr>
<tr>
<td>P3 Pz P4</td>
<td>49.91</td>
<td>45.28</td>
<td>64.91</td>
<td>62.04</td>
<td>43.98</td>
</tr>
<tr>
<td>O1 POz O2</td>
<td>52.59</td>
<td>51.57</td>
<td>64.91</td>
<td>69.91</td>
<td>52.04</td>
</tr>
<tr>
<td>PO3 POz PO4</td>
<td>52.78</td>
<td>55.56</td>
<td>67.31</td>
<td>73.98</td>
<td>49.81</td>
</tr>
</tbody>
</table>

Table 2: Classification performance in terms of correct recognition rate. Results refer to disjoint training and test datasets. The analysis of individual subbands is reposted in subsequent columns, and the set of 3 channels considered is listed in the first column.
could be also proposed for the acquisition of EEG signals to be represented through the bump modelling approach, in order to investigate the significance of the information related to the event timing for user recognition.

References


New Security Definitions for Biometric Authentication with Template Protection: Toward covering more threats against authentication systems

Toshiyuki Isshiki¹, Toshinori Araki¹, Kengo Mori¹, Satoshi Obana², Tetsushi Ohki³, Shizuo Sakamoto¹

¹ NEC Corporation ² Hosei University ³ National Institute of Advanced Industrial Science and Technology (AIST)

¹ {t-issiki@bx, t-araki@ek, ke-mori@bx, s-sakamoto@bu}.jp.nec.com ² obana@hosei.ac.jp ³ tetsushi.ohki@aist.go.jp

Abstract: Existing studies on the security of biometric authentication with template protection have considered the adversaries who obtain only protected templates. Since biometric authentication systems transmit data other than the protected templates, we need to consider how to secure biometric authentication systems against adversaries with those data. In this paper, we propose a classification of adversaries in biometric authentication with template protection into the following three types in accordance with their knowledge: (1) protected template data, (2) data transmitted during authentication, and (3) both types of data. We also propose a new security metric unforgeability, which provides authentication security against attacks by adversaries impersonating someone else on authentication systems even when they cannot obtain the biometric information of a claimant. We then give security definitions against each type of adversary we classified. We also propose a biometric authentication scheme with template protection that is irreversible against all types of adversaries.

1 Introduction

Biometric authentication provides advantages in terms of better usability in systems with person identification functions by freeing users from having to remember something or carry around a token. However, biometric authentication may have some vulnerabilities, which appear in various system elements including users, environmental conditions, operating conditions, biological data, and biometric equipment. Among these vulnerabilities, template leakage is the most critical. Leaked templates can be abused by adversaries for replay attacks. Even though replay attacks are prevented, biometric information can be forged by using the leaked templates [Ma03, Ca07]. Hence, the biometric templates must be protected for security.

Many techniques such as cancelable biometrics [NNJ10] and biometric cryptosystems [Tu05] have already been proposed to improve biometric template security. The performance of template protection techniques can be evaluated not only by recognition performance but also from security and privacy aspects. The latter metrics have not been established yet,
and many researchers are still seeking them. Previous works mainly focus on security and privacy against stored templates. Simoens et al. proposed and evaluated irreversibility and unlinkability of biometric cryptosystems [STP09]. Zhou also defined a systematic evaluation framework to assess irreversibility and unlinkability [Zh11]. Recently, Inuma et al. proposed alternative definitions of irreversibility and unlinkability [IO12]. Their work also focuses on security and privacy against stored auxiliary data. Wang et al. proposed revocability and reusability of biometric cryptosystems [Wa11], and Nagar et al. also proposed similar properties for cancelable biometrics and bio-hashing [NNJ10]. ISO/IEC 24745 [ISO11] defined reference architecture for template protection, and Simoens et al. [Si12] proposed criteria and several metrics that comply with ISO/IEC 24745.

The above metrics only consider the adversaries with stored data. However, since biometric authentication systems transmit data other than stored data, we need to consider how to secure the biometric authentication systems against adversaries who obtain those data. It is especially important to consider how to secure biometric authentication systems against adversaries who obtain query data, because query data transmitted during authentication are usually generated from biometric information of claimants. For this purpose, in this paper, we classify adversaries into the following three types in accordance with their knowledge. The first obtains protected templates. These adversaries have been considered in existing studies. The second obtains query data transmitted during authentication. The third obtains both protected templates and query data, whose importance will show in Section 4.1. As discussed above, no security properties against the second and third types of adversaries have been defined yet. We thus present the security definitions against each type of adversary. We also propose a new security metric for biometric authentication with template protection: unforgeability. Unforgeability provides authentication security against attacks by adversaries impersonating someone else on authentication systems even when they cannot obtain the biometric information of a claimant. We show that this property is not weaker than the irreversibility in [Si12]. Furthermore, we evaluate the security of the scheme in [Tu05] under our security definitions. We show that though the scheme [Tu05] is irreversible against the first and second types of adversaries respectively, the third type can break its irreversibility. Then, we propose a biometric authentication scheme with template protection that is irreversible against all types of adversaries. We cannot have provided a proof that the proposed scheme is unforgeable in this paper yet. Constructing a scheme possessing the unforgeability is a future work.

2 Biometric Authentication with Template Protection

This section introduces the model, components, and schemes of biometric authentication with template protection.
2.1 Model

In this paper, we deal with the model in which the biometric references and the identity references are stored on a server. In the target model, the query data extracted from the biometric information of a claimant are transferred to the server. The model can be applied to verification (i.e., one-to-one matching) and to identification (one-to-many matching). The merit of such model is that a claimant requires only his/her biometric characteristic for verification/identification and does not need to memorize any secret data nor bring any physical devices such as smart cards. Figure 1 shows our model, which is slightly modified from Model A [ISO11]. More specifically, we introduce session specific data (SSD for short). The motivation of introducing such data is to secure the system against replay attacks. Details are described in Section 3.

![Figure 1: The model referenced in this paper.](image)

2.2 Components

The biometric authentication schemes consist of the following five components.

- **Feature Extraction**: given a biometric sample \( b \) as input, extracts a feature set \( fs \) of the sample \( b \). We denote \( fs \leftarrow FE(b) \).

- **Pseudonymous Identifier Encoder**: takes a feature set \( fs \) as input and generates a **protected template** \( PT \). We denote \( PT = (PI, AD) \leftarrow PIE(fs) \), where \( PI \) is a pseudonymous identifier and \( AD \) is auxiliary data.

- **Pseudonymous Identifier Recoder**: takes a feature set \( fs' \), \( AD \), and session specific data \( SSD \) as input and generates query data \( QD \), where \( AD \) is a set of \( ADs \). We denote \( QD \leftarrow PIR(fs', AD, SSD) \).
• **Pseudonymous Identifier Comparator**: takes query data \( QD \), a set \( PI \) of PIs, and session specific data \( SSD \) as input and outputs a comparison result \( cr \), where \( cr \subset PI \) or \( cr = \bot \). We denote \( cr \leftarrow \text{PIC}(QD, PI, SSD) \).

• **Decision Subsystem**: takes a comparison result \( cr \) and an identity reference set \( IR \) as input, and outputs a result \( result \), which is a set of accepted IDs or is \( \text{reject} \). We denote \( result \leftarrow \text{DS}(cr, IR) \).

### 2.3 Biometric Authentication Schemes with Template Protection

We use the notation \( (y_c, y_s) \leftarrow P[C(x_c), S(x_s)](x) \) to denote an interactive protocol \( P \) between client \( C \) with private input \( x_c \) and server \( S \) with private input \( x_s \) to run with common input \( x \). At the end of the protocol \( P \), \( C \)’s output is \( y_c \) and \( S \)’s output is \( y_s \). If a party has no input or output, we use the placeholder “-”.

One of the biometric authentication schemes \( \langle \text{ENROLL}, \text{IDENTIFY} \rangle \) is for identification with template protection that consists of two protocols: \( \text{ENROLL} \) and \( \text{IDENTIFY} \). We assume that all entities know the system parameters \( \text{param} \).

The enrollment protocol \( \text{ENROLL} \) is an interactive protocol between \( C \) and \( S \). \( C \) which privately takes a biometric sample \( b_c \) as input and \( S \) which maintains an enrollment storage, collaboratively execute the enrollment protocol. In the enrollment protocol, \( C \) extracts the feature set \( fs \) by the feature extraction component \( \text{FE} \) with input \( b_c \). Then, \( C \) generates the protected template \( PT \) by the pseudonymous identifier encoder component \( \text{PIE} \) with input \( fs \). The \( PT \) consists of a pseudonymous identifier (\( PI \)) and auxiliary data (\( AD \)). The \( PI \) represents the individual and is used as reference for identification. The \( AD \) helps to generate query data \( QD \) in the identification phase. \( C \) sends \( PT = (PI, AD) \) to \( S \) and \( S \) stores them with original user identity reference \( ID \) in the enrollment storage. We denote this protocol by \( (-, ID) \leftarrow \text{ENROLL}[C(b_c), S(-)](\text{param}) \).

The identification protocol \( \text{IDENTIFY} \) is an interactive protocol for identifying an individual from the enrollment storage after sharing the session specific data \( SSD \) between \( C \) and \( S \) if needed. \( C \) with a biometric sample \( b_c' \) and \( S \) with a set \( PI \) of PIs and a set \( IR \) of IDs as private inputs collaboratively execute an identification protocol with common input \( AD \) and \( SSD \). \( AD \) is a set of auxiliary data \( AD \) generated in \( \text{ENROLL} \). In the identification protocol, \( C \) extracts his/her feature set \( fs' \) by the feature extraction component \( \text{FE} \) with a biometric sample \( b_c' \) as private input. \( C \) then generates query data \( QD \) by the pseudonymous identifier encoder component \( \text{PIE} \) with input \( fs' \), \( AD \), and the session specific data \( SSD \) and sends the query data \( QD \) to \( S \). \( S \) executes the pseudonymous identifier comparator component \( \text{PIC} \) with input \( PI, SSD, \) and \( QD \), and obtains a comparison result \( cr \). \( cr \) is a subset of \( PI \) or is \( \bot \). \( S \) executes the decision subsystem \( \text{DS} \) with input \( cr \) and \( IR \) and obtains a result \( result \). \( result \) is a subset of \( IR \) or is \( \text{reject} \). We denote this protocol by \( (-, result) \leftarrow \text{IDENTIFY}[C(b_c'), S(PI, IR)](\text{param}, AD, SSD) \).

Note that we can introduce a one-to-one verification scheme by adding \( ID \) to the common input in the identification protocol and replacing \( AD \) with \( AD \) corresponding to \( ID \). \( S \) executes \( \text{PIC} \) with input \( PI_{ID}, QD, \) and \( SSD \), where \( PI_{ID} \) is the protected template of \( ID \).
We denote the verification protocol by \((-\rightarrow, \text{result}) \leftarrow \text{VERIFY} [C(b'_c), S(PI, IR)](\text{param}, AD, ID, SSD)\).

3 New Security Definitions

3.1 Proposed Classification of Adversaries

Existing studies on the security of biometric authentication with template protection have considered only adversaries who obtain PIs. However, since biometric authentication systems transmit data other than stored data, we need to consider how to secure biometric authentication systems against adversaries who obtain those data. Thus, we classify adversaries into the following three types in accordance with their knowledge, based on PIs as well as QDs.

- **The adversary with PIs** obtain leaked protected templates.
- **The adversary with QDs** obtain QDs transmitted in authentication.
- **The adversary with PIs and QDs** obtain leaked protected templates and transmitted QDs.

The first type of adversary is captured in the metrics by Simoens et al. [Si12]. However no properties have been defined against the second and the third. The second covers the replay attack in which the adversary who observed the QD sent by a genuine user tries to impersonate him/her by re-transmitting the QD as is. To prevent such attacks, QD should not be independent of session (i.e., QD accepted in a session should not be accepted in the other sessions). To make QD session dependent, we introduce SSD into our model, where (possibly random) SSD can be used to compute QD. If we employ a secure channel, we may obtain a scheme secure against the adversaries with QDs. However, the secure channel technique is not sufficient against the adversaries with compromised servers, classified into the third.

3.2 Proposed Security Definitions

Irreversibility [Si12] focuses on the adversaries who want to obtain biometric information of a claimant. However, from the view-point of authentication systems, they may be considered to be broken even if the adversaries successfully impersonate someone else. Therefore, we define a new metric, unforgeability, which captures such adversaries, in the same manner as the security of digital signatures.

In the following, we propose three types of unforgeability in accordance with Section 3.1. Note that we only discuss the one-to-many identification scheme, but it is easy to expand our discussion to the one-to-one verification scheme in a similar way.
3.2.1 Unforgeability against Attacks with PIs

Unforgeability against attacks with PIs is defined via the following game involving an adversary $A$ and a challenger $C$. Let $n$ be the number of enrolled claimants.

**Setup.** $C$ generates system parameters $\text{param}$ and randomly $n$ samples $\{b_1, b_2, \ldots, b_n\}$. Then, for $i = 1, 2, \ldots, n$, $C$ executes $f_{s_i} \leftarrow \text{FE}(b_i)$ and $(P_{I_i}, A_{D_i}) \leftarrow \text{PIE}(f_{s_i})$. We assume that for $i = 1, 2, \ldots, n$, both $P_{I_i}$ and $A_{D_i}$ are indexed as $I_{D_i}$ and denote the set of $I_{D_i}$ by $I_{R}$. $C$ sends $\text{param}$ and $A_{D}$ to $A$.

**Phase 1.** $A$ is permitted to make queries to $O_{\text{Corrupt}}, O_{\text{Enroll}}, O_{\text{PI}},$ and $O_{\text{PIC}}$:

- $O_{\text{Corrupt}}$ takes $I_{D_i}$ and returns a biometric sample $b_i$. This oracle captures corruption of a claimant. By querying to this oracle, the adversary can corrupt at most $n - 1$ claimants.
- $O_{\text{Enroll}}$ takes $(\hat{P}_{I}, \hat{A}_{D})$. $O_{\text{Enroll}}$ updates $P_{I}$ to $P_{I} \cup \{\hat{P}_{I}\}$, $A_{D}$ to $A_{D} \cup \{\hat{A}_{D}\}$, and $I_{R}$ to $I_{R} \cup \{\hat{I}_{D}\}$. Then $O_{\text{Enroll}}$ returns an index $\hat{I}_{D}$ of $\hat{P}_{I}$. This oracle captures addition of a new claimant.
- $O_{\text{PI}}$ takes $I_{D_i}$ and returns a protected identifier $P_{I_i}$.
- $O_{\text{PIC}}$ takes $Q_{D}$ and $S_{SD}$ and returns a result $\text{result}$. If $c_{r}$ such that $c_{r} \leftarrow \text{PIC}(Q_{D}, P_{I}, S_{SD})$ is not $\bot$, then $\text{result}$ is a subset of $I_{R}$ corresponding to $c_{r}$. Otherwise, $\text{result}$ is $\bot$. By querying this oracle, the adversary can check whether $Q_{D}$, which is made by himself is acceptable or not.

**Challenge.** $C$ generates and sends session specific data $S_{SD}^{*}$ to $A$.

**Phase 2.** $A$ is permitted to make queries to $O_{\text{Corrupt}}, O_{\text{Enroll}}, O_{\text{PI}},$ and $O_{\text{PIC}}$ as same as Phase 1. Finally, $A$ generates $Q_{D}^{*}$ and outputs it with $S_{SD}^{*}$.

We define a set of identities not queried to $O_{\text{Corrupt}}$ and not returned by $O_{\text{Enroll}}$ by $U_{NC}$. Then we define the unforgeability against the attacks with PIs as follows.

**Definition 1 (Unforgeability against Attacks with PIs)** Let $c_{r}^{*} \leftarrow \text{PIC}(Q_{D}^{*}, P_{I}, S_{SD}^{*})$. In the above game, the adversary $A$ wins if there exists $I_{D_i} \in \text{result} \cap U_{NC}$, where $\text{result} \leftarrow DS(I_{R}, c_{r}^{*})$. A biometric authentication system is said to be unforgeable against the attacks with PIs if for an arbitrary polynomial time adversary $A$ in the security parameter, the probability $\Pr[ A \text{ wins}]$ is negligible.

We show that if a biometric authentication system satisfies unforgeability, then the system also satisfies authorized-leakage irreversibility [Si12]. Let $A$ be an adversary who can break authorized-leakage irreversibility. That is, $A$ can compute $f_{s}$, which matches the unprotected template in the unprotected system from $P_{I}$. Then, by invoking $A$, an adversary $B$ can obtain $f_{s}$ and generate new $Q_{D}$ by using $f_{s}$. This means we can construct an adversary $B$ who can break unforgeability by invoking $A$. Therefore, unforgeability

---

1Hereafter, $S_{SD}^{*} = \emptyset$ for the system without $S_{SD}$.
is not weaker than authorized-leakage irreversibility. Similarly, we can also prove that unforgeability is not weaker than pseudo-authorized-leakage irreversibility [Si12].

We assume that if a sample is accepted in the unprotected system, then it is also accepted in the protected system. Clearly, if a biometric authentication system meets the metric of authorized-leakage irreversibility, then this system also meets the metric of full-leakage irreversibility [Si12]. Therefore, unforgeability is not weaker than any type of irreversibility.

### 3.2.2 Unforgeability against Attacks with $QDs$

Unforgeability against attacks with $QDs$ is defined via the following game involving an adversary $A$ and a challenger $C$.

**Setup.** Same as the setup phase described in Section 3.2.1.

**Phase 1.** $A$ is permitted to make queries to $O_{\text{Corrupt}}$, $O_{\text{Enroll}}$, $O_{\text{QD}}$, and $O_{\text{PIC}}$, where $O_{\text{Corrupt}}$, $O_{\text{Enroll}}$, and $O_{\text{PIC}}$ are described in Section 3.2.1. $O_{\text{QD}}$ takes $ID$ and returns query data $QD$ such that for some $SSD$, $PI_{ID} \in cr$, where $cr \leftarrow \text{PIC}(PI, QD, SSD)$.

**Challenge.** $C$ generates and sends session specific data $SSD^*$ to $A$.

**Phase 2.** $A$ is permitted to make queries to $O_{\text{Corrupt}}$, $O_{\text{Enroll}}$, $O_{\text{QD}}$, and $O_{\text{PIC}}$. Finally, $A$ generates $QD^*$ and outputs it with $SSD^*$.

We define the unforgeability against attacks with $QDs$ as follows.

**Definition 2 (Unforgeability against Attacks with $QDs$)** Let $cr^* \leftarrow \text{PIC}(PI, QD^*, SSD^*)$. $U_{NC}$ is defined in Section 3.2.1. In the above game, the adversary $A$ wins if there exists $ID_i \in \text{result} \cap U_{NC}$, where $\text{result} \leftarrow DS(IR, cr^*)$. A biometric authentication system is said to be unforgeable against attacks with $QDs$ if for an arbitrary polynomial time adversary $A$ in the security parameter, the probability $Pr[A \text{ wins}]$ is negligible.

We can also define three types of irreversibility (i.e., full-leakage, authorized-leakage, and pseudo-authorized-leakage irreversibility) against attacks with $QDs$ in the similar manner as Simoens et al. [Si12]. We can show that unforgeability is not weaker than any of irreversibility in the same manner as Section 3.2.1.

### 3.2.3 Unforgeability against Attacks with $PIs$ and $QDs$

Unforgeability against attacks with $PIs$ and $QDs$ is defined via the following game involving an adversary $A$ and a challenger $C$.

**Setup.** Same as the setup phase described in Section 3.2.1.

**Phase 1.** $A$ is permitted to make queries to $O_{\text{Corrupt}}$, $O_{\text{Enroll}}$, $O_{\text{PI}}$, $O_{\text{QD}}$, $O_{\text{SSD}}$, and $O_{\text{PIC}}$, $O_{\text{Corrupt}}$, $O_{\text{Enroll}}$, $O_{\text{PI}}$, $O_{\text{QD}}$, and $O_{\text{PIC}}$ are described in Section 3.2.1 and 3.2.2. $O_{\text{SSD}}$ takes $QD$ as input and outputs $SSD$, which is used for generating $QD$. By querying this oracle, the adversary can obtain $SSD$, which is secretly shared between the (non-corrupted) client and the server.
**Challenge.** C generates session specific data SSD* and sends it to A.

**Phase 2.** A is permitted to make queries to $O_{\text{Corrupt}}, O_{\text{Enroll}}, O_{\text{PI}}, O_{\text{QD}}, O_{\text{SSD}},$ and $O_{\text{PIC}}$. Finally, A generates QD* and outputs it with SSD*.

We define the unforgeability against attacks with PIs and QDs as follows.

**Definition 3 (Unforgeability against Attacks with PIs and QDs)** Let $cr^* \leftarrow \text{PIC}(PI, QD^*, SSD^*)$. $U_{NC}$ is as defined in Section 3.2.1. In the above game, the adversary $A$ wins if there exists $ID_i \in \text{result} \cap U_{NC}$, where $\text{result} \leftarrow DS(IR, cr^*)$. A biometric authentication system is said to be unforgeable against the attacks with PIs and QDs if for an arbitrary polynomial time adversary $A$ in the security parameter, the probability $Pr[A \text{ wins}]$ is negligible.

We can also define three types of irreversibility (i.e., full-leakage, authorized-leakage, and pseudo-authorized-leakage irreversibility) against attacks with PIs and QDs by the similar manner as [Si12]. We can show that unforgeability is not weaker than any type of irreversibility in the same manner as in Section 3.2.1.

### 4 Proposed Scheme

In this section, we evaluate the security of the scheme in [Tu05] under classification of adversaries. We show that their scheme is not irreversible against the adversaries with PIs and QDs. Then we propose a scheme that is irreversible against all types of adversaries.

We introduce some notations to describe the proposed scheme. Let $x = (x_{n-1}x_{n-2}\cdots x_0)_2$ and $y = (y_{n-1}y_{n-2}\cdots y_0)_2$ be two $n$ bit integers. Then, we denote the hamming distance between $x$ and $y$ by $d_H(x, y) = \sum_{i=0}^{n-1} |x_i - y_i|$.

A linear binary error correcting code ECC with parameters $(K, s, d)$ consists of two algorithms: ENCODE and DECODE. ENCODE takes $s$-bit data $x$ as input and outputs its $K$ bits codeword $CW$, written as $CW \leftarrow \text{ENCODE}(x)$. DECODE takes a $K$ bits codeword $CW'$ as input and outputs $s$-bit $x'$ or ⊥, written as $\{x', \bot\} \leftarrow \text{DECODE}(CW')$. If $d_H(CW, CW') \leq d$, then $x' = x$. We write $CW \leftarrow \text{ENCODE}(x)$ and $x' \leftarrow \text{DECODE}(CW')$, respectively. A BCH code is employed, which is a linear binary ECC. Note that a BCH code possesses the following property: for all information symbols $x_1$ and $x_2$, $\text{ENCODE}(x_1) \oplus \text{ENCODE}(x_2) = \text{ENCODE}(x_1 \oplus x_2)$.

We denote an inner product operation of two vectors $v$ and $w$ by $\langle v, w \rangle$ where the length of the result $\langle v, w \rangle$ is $s$. Note that by treating $v$ and $w$ as vectors over GF($2^s$), we can compute an inner product with $n$-bit output efficiently. Also note that it is easy to see that for every $v, w_1,$ and $w_2$, $\langle v, w_1 \rangle \oplus \langle v, w_2 \rangle = \langle v, w_1 \oplus w_2 \rangle$. 

178
4.1 Security Evaluation of Scheme in [Tu05] under Classification of Adversaries

We evaluate the security of the scheme in [Tu05] with secure channel connection between the client and the server, because their scheme requires only a biometric characteristic for identification and is suitable for our model. A straightforward adoption of their scheme to our model is as follows.

ENROLL: The client $C$ with a biometric sample $b$ generates a feature set $z$ by $\text{FE}(b)$. Then, $C$ chooses a random $r \in \{0,1\}^s$ and computes $W_1 = \text{ ENCODE}(r) \oplus z$ and $W_2 = H(r)$. $C$ sends $(W_1,W_2)$ to the server $S$ as his $\text{PI}$. Note that we assume $W_1$ is held in the server $S$, while no explicit designation in [Tu05].

IDENTIFY: $C$ with a biometric sample $b'$ generates a feature set $z'$ by $\text{FE}(b')$. Then, $C$ sends $z'$ to $S$ as $\text{QD}$.

$s$. Our basic idea of construction is as follows: (1) We replace the hash function $H$ of the scheme in [Tu05] with an xor homomorphic function $h$ such that $h(s) \oplus h(s') = h(s \oplus s')$. (2) In ENROLL protocol, the same as the scheme in [Tu05], the proposed scheme protects a feature set $z$ by xor-ing with a codeword $c = \text{ ENCODE}(r)$, for some random number $r$. (3) In IDENTIFY protocol, the proposed scheme protects a transmitted feature set $z'$ by xor-ing with a codeword $c' = \text{ ENCODE}(r')$, for another random number $r'$. Then, we can verify whether $z$ and $z'$ are close enough by checking if $h(\text{DECODE}((z \oplus c) \oplus (z' \oplus c')))$ is equal to $h(r) \oplus h(r')$, without recovering $z, z', r, r'$.

4.2 Construction of Proposed Scheme

We improve the scheme in [Tu05] in several points for achieving irreversibility against attacks with $\text{PI}s$ and $\text{QD}s$. Our basic idea of construction is as follows: (1) We replace the hash function $H$ of the scheme in [Tu05] with an xor homomorphic function $h$ such that $h(s) \oplus h(s') = h(s \oplus s')$. (2) In ENROLL protocol, the same as the scheme in [Tu05], the proposed scheme protects a feature set $z$ by xor-ing with a codeword $c = \text{ ENCODE}(r)$, for some random number $r$. (3) In IDENTIFY protocol, the proposed scheme protects a transmitted feature set $z'$ by xor-ing with a codeword $c' = \text{ ENCODE}(r')$, for another random number $r'$. Then, we can verify whether $z$ and $z'$ are close enough by checking if $h(\text{DECODE}((z \oplus c) \oplus (z' \oplus c')))$ is equal to $h(r) \oplus h(r')$, without recovering $z, z', r, r'$.

We describe the proposed scheme. Let $2K$ be the length of a feature set, $p$ an order of a group $G$, and $\ell$ the security parameter. The proposed scheme uses an ECC with parameters $(K,s,d)$ and two cryptographic hash functions $H_1$ and $H_2$ such that $H_1 : \{0,1\}^* \times \mathbb{F}_p \rightarrow \{0,1\}^s$ and $H_2 : \{0,1\}^* \rightarrow \{0,1\}^\ell$. We assume that the client $C$ and the server $S$ have the common parameter $c \in \{0,1\}^K$ and $g$ that is a generator of $G$. We also assume that FE possesses the following properties $^2$: (1) If the biometric samples $b$ and $b'$ are generated by the same claimant, then $d_H(z_{1},z'_{1}) \leq d$ and $d_H(z_{2},z'_{2}) \leq d$, where $z_{1},z_{2} \leftarrow \text{FE}(b)$ and $z'_{1},z'_{2} \leftarrow \text{FE}(b')$. (2) If the biometric samples $b$ and $b'$ are generated by the different claimants, then $d_H(z_{1},z'_{1}) > d$ and $d_H(z_{2},z'_{2}) > d$.

\footnote{This assumption on the feature extractor is stronger than that in [Tu05]. However, the feature extractor we need can be easily constructed by using the feature extractor in [Tu05] twice.}
ENROLL: The client $C$ with a biometric sample $b$ wants to be registered in the biometric system of the server $S$. We set $param = (H_1, H_2, K, s, d, c)$. Then, the enrollment protocol $ENROLL[C(b), S(\varnothing)](param)$ is as follows:

1. $C$ generates a feature set $z_1 \| z_2 \in \{0, 1\}^{2K}$ by $\text{FE}(b)$.

2. $C$ runs the algorithm $\text{PIE}(z_1 \| z_2)$ as follows:
   (a) $C$ chooses $r_1 \in \{0, 1\}^s$ uniformly at random and computes $\text{ENCODE}(r_1)$. $C$ also computes $V_1 = \text{ENCODE}(r_1) \oplus z_1$.
   (b) $C$ chooses $r_2 \in \mathbb{Z}_p$ uniformly at random and computes $V_2 = H_2(<c, \text{ENCODE}(r_1)> \oplus H_1(V_1, g^r_2))$. $C$ also computes $V_3 = \text{ENCODE}(H_1(V_1, g^r_2)) \oplus z_2$.
   (c) $C$ sets and sends $PI = (V_1, V_2, V_3)$ to $S$ as a protected template.

3. $S$ stores them with original identity reference $ID$ in its storage and outputs $ID$.

IDENTIFY: We assume that $S$ has the set of protected template $PI = \{PI_1, PI_2, \ldots, PI_N\}$, where $N$ is the number of enrolled users and each $PI_i = (V_{1,i}, V_{2,i}, V_{3,i})$.

Before the identification phase starts, $C$ and $S$ collaboratively generate a shared secret $g_{sc}$ by using a key exchange protocol such as the Diffie-Hellman key exchange protocol [DH76]. We set $SSD = g_{sc}$.

Then, the identification protocol $\text{IDENTIFY}[C(b'), S(PI)](param, \varnothing, g_{sc})$ is as follows:

1. $C$ who has a biometric sample $b'$ generates feature sets $z'_1 \| z'_2 \in \{0, 1\}^{2K}$ by $\text{FE}(b)$.

2. $C$ executes the algorithm $\text{PIR}(z'_1 \| z'_2, \varnothing, g_{sc})$ as follows:
   (a) $C$ chooses $r_3, r_4 \in \{0, 1\}^s$ uniformly at random and computes $r_5 = r_3 \oplus r_4$.
   (b) $C$ computes $W_1 = \text{ENCODE}(r_5) \oplus z'_1$, $W_2 = <c, \text{ENCODE}(r_3)> \oplus H_1(W_1, g_{sc})$, and $W_3 = \text{ENCODE}(<c, \text{ENCODE}(r_4)> ) \oplus z'_2$.
   (c) $C$ sets and sends $QD = (W_1, W_2, W_3)$ to $S$.

3. $S$ runs the algorithm $\text{PIC}(QD, PI, g_{sc})$ as follows:
   (a) $S$ computes $H_1(W_1, g_{sc})$ and sets $i = 1$. For $PI_i = (V_{1,i}, V_{2,i}, V_{3,i})$, $S$ computes as follows:
      i. $S$ computes $W_1 \oplus V_{1,i}$ and $W_3 \oplus V_{3,i}$.
      ii. $S$ generates $WV_{1,i} = \text{DECODE}(W_1 \oplus V_{1,i})$ and $WV_{3,i} = \text{DECODE}(W_3 \oplus V_{3,i})$. If $WV_{1,i}$ or $WV_{3,i}$ is $\bot$, then sets $i = i + 1$ and goes to the step 3(a)i. Otherwise, goes to the next step.
      iii. $S$ computes $WV_{2,i} = H_2(H_1(W_1, g_{sc}) \oplus <c, \text{ENCODE}(WV_{1,i})> \oplus WV_{3,i} \oplus W_2$.
      iv. If $V_{2,i} = WV_{2,i}$, then outputs $res = \text{accept}$ and corresponding $ID_i$. Otherwise, if $i = N$, then outputs $res = \bot$. If not, sets $i = i + 1$ and goes to the step 3(a)i.
(b) $S$ outputs $res$. If $res = accept$ then $S$ also outputs $ID_i$.

**Correctness.** Let $PI = (V_1, V_2, V_3)$ be a protected template from the biometric samples $b$ and $QD = (W_1, W_2, W_3)$ be verification data from $b'$ in the proposed system. We can prove that if $b$ and $b'$ are generated by the same client, then the following equation holds.

$$V_2 = H_2 (H_1 (W_1, g_{sc}) \oplus \langle c, \text{ENCODE} (WV_1) \rangle \oplus WV_3 \oplus W_2), \quad (1)$$

where $WV_1 = \text{DECODE} (W_1 \oplus V_1)$ and $WV_3 = \text{DECODE} (W_3 \oplus V_3)$. From the assumption on the algorithm $\text{FE}$, $d_H (z_1, z'_1) \leq d$ and $d_H (z_2, z'_2) \leq d$ where $z_1 \parallel z_2 \leftarrow \text{FE} (b)$ and $z'_1 \parallel z'_2 \leftarrow \text{FE} (b')$. Therefore, $WV_1 = r_1 \oplus r_5$ and $WV_3 = H_1 (V_1, g^{r_2}) \oplus \langle c, \text{ENCODE} (r_4) \rangle$. Then, the right-hand side of the equation (1) is

$$H_2 (H_1 (W_1, g_{sc}) \oplus \langle c, \text{ENCODE} (r_1 \oplus r_5) \rangle \oplus H_1 (V_1, g^{r_2}) \oplus \langle c, \text{ENCODE} (r_4) \rangle \oplus \langle c, \text{ENCODE} (r_3) \rangle \oplus H_1 (W_1, g_{sc}))$$

$$= H_2 (\langle c, \text{ENCODE} (r_1 \oplus r_3 \oplus r_4 \oplus r_5) \rangle \oplus H_1 (V_1, g^{r_2}))$$

$$= V_2.$$

### 4.3 Intuition behind the Proposed Scheme

The security of the proposed scheme has not been proven to be connected to any infeasible problems. Here, we intuitively explain the security of the scheme instead.

Let the hash functions $H_1$ and $H_2$ be collision resistant. We assume that it is difficult to compute $g^{ab}$ from $g^a$ and $g^b$ over $G$. Then, the security of the proposed scheme can be discussed as follows.

To compute $z_1 \parallel z_2$ from $PI$, an adversary needs to compute $r_1$ and $H_1 (V_1, g^{r_2})$ from $PI$. However, since $H_2$ is a cryptographic hash function, the adversary cannot compute $r_1$ and $H_1 (V_1, g^{r_2})$. Therefore, we can say that the proposed scheme is irreversible against the attacks with $PI$s.

Similarly, we can say that the proposed scheme is irreversible against the attacks with $QD$s. To compute $z'_1 \parallel z'_2$ from $QD$, an adversary needs to compute $r_4$ and $r_5$. Then the adversary cannot compute $g_{sc}$ and $r_4$ and $r_5$ are uniformly random.

The compromised servers obtain not only $PI$s and $QD$s, but also $SSD(= g_{sc})$. From some $QD = (W_1, W_2, W_3)$ and $g_{sc}$, the adversary obtains the following equation:

$$\text{ENCODE} (\langle c, W_1 \rangle) \oplus \text{ENCODE} (W_2) \oplus W_3 \oplus \text{ENCODE} (H_1 (W_1, g_{sc})) = \text{ENCODE} (\langle c, z'_1 \rangle) \oplus z'_2.$$

The above equation can be seen as a system of at most $K$ equations with $2K$ unknowns (since both $z'_1$ and $z'_2$ are $K$ bits). Therefore, we can say that the proposed scheme is irreversible against the attacks with the compromised servers.

---

3This is a standard assumption in cryptography: *Computational Diffie-Hellman (CDH)* assumption.
5 Conclusion

We proposed a classification of adversaries on biometric authentication with template protection into three types in accordance with their knowledge: (1) PI's, (2) QD's, and (3) PI's and QD's. We also presented the security definitions against each type of adversary and showed that the scheme of Tuyls et al. [Tu05] is not irreversible against the third. Then we proposed a biometric authentication scheme that is irreversible against all types of adversaries.

Constructing a scheme possessing unforgeability is our future work.

References


BIOSIG 2013

Further Conference Contributions
A template privacy protection scheme for fingerprint minutiae descriptors

Leila Mirmohamadsadeghi and Andrzej Drygajlo

School of Engineering, Swiss Federal Institute of Technology Lausanne (EPFL)
Lausanne, Switzerland
leila.mirmohamadsadeghi@epfl.ch, andrzej.drygajlo@epfl.ch

Abstract: It is important in biometric person recognition systems to protect personal data and privacy of users. This paper introduces a new mechanism to revoke and protect fingerprint minutiae information, which can be used in today’s security-aware society. The recently developed minutiae cylinder code (MCC), which provides rotation and translation invariant descriptors for accurate fingerprint recognition by describing minutiae neighborhoods with respect to each other, is used as baseline fingerprint descriptor. A hybrid scheme combining a transformation and a user key is designed to provide the MCC-based fingerprint representation with revocability and irreversibility properties for template privacy protection across multiple applications. Furthermore, using the publicly available FVC datasets, it is demonstrated that the designed scheme improves the baseline accuracy of fingerprint recognition using the MCC method.

1 Introduction

Biometric template privacy protection is gaining importance with the widespread use of biometric person recognition and an increase in the awareness on related privacy issues. Biometric template privacy protection should ensure that an individual’s biometric characteristic is only available in the form of a template which is diverse and thus revocable if the template is compromised. Furthermore, the template must be irreversible to the original capture and must preserve the accuracy of the underlying recognition system [JNN08]. Recently, a specific scheme was designed to provide irreversibility for the minutiae cylinder code (MCC) representation, which involves the quantization of the Karhunen-Loeve transform [FMC12]. However, this scheme is not designed to provide diversity and revocability.

In the present paper, a novel method is proposed to provide template privacy protection for the minutiae cylinder code (MCC) representation [CFM10], which provides diversity, revocability and irreversibility properties, without degrading the baseline recognition accuracy. Fingerprint recognition using the MCC representation is chosen because it is alignment-free and computationally light.

The remainder of this paper is organized as follows. In Section 2, a brief description of the baseline fingerprint recognition system is provided. In Section 3, the proposed privacy protection scheme is presented. Experiments to assess performance of the protection
scheme in terms of accuracy are depicted in Section 4, while security aspects are addressed in Section 4.2. Conclusions are drawn in Section 5. Note: a complete version of this paper can be found at http://infoscience.epfl.ch/.

2 Fingerprint minutiae descriptors based on the minutiae cylinder code (MCC)

The minutiae cylinder code (MCC) [CFM10] is a recent fingerprint description method, which presents the advantages of both nearest-neighbor-based and fixed-radius-based minutiae description methods and is considered as the state-of-the-art in minutiae descriptor design [FZ11].

This method takes as input a set of standard ISO minutiae [ISO05], and creates for each minutia, a descriptor based on its distance to neighboring minutiae and their angular differences. This descriptor is of fixed length, robust to rotations and translations and skin distortions and is computed in a fast manner. The output consists of a template, which contains a descriptor for each minutia. This descriptor is a linearized cylinder whose discretized volume represents weighted spatial and angular distances of each minutiae to its neighbors. In order to compare two such templates within a recognition system, several comparison measures were originally introduced. In this paper, the local similarity sort (LSS) method is chosen among the others because it requires the least extra information about the original minutiae set when performing cylinder set comparison. The LSS comparison method computes all two by two distances of the cylinders and provides a similarity score based on the closest cylinder matches and angular distances of minutiae pairs.

A slightly different version of the MCC method is used in this work. With respect to the original method, cylinder and cell validities are not considered and the weighting functions used to compute the spatial and angular contributions of neighboring minutiae are discretized.

3 A privacy protection scheme for the MCC templates

According to template privacy protection requirements, it is desirable to create from a raw biometric sample, several diverse and revocable templates which are irreversible to the original biometric characteristic and support accurate recognition. The baseline accuracy corresponds to recognition without template protection. It is thus required that template protection does not degrade this accuracy. Revocability is achieved when including a revocable component into the template, as original biometric characteristics are not revocable. Therefore, the privacy protection scheme presented in this paper is a hybrid two-factor scheme which transforms the template with a user key in a revocable and non-reversible manner. The key is assumed secret in the baseline protected system, but it is shown that even if the key is compromised or lost, the original biometric characteristic remains protected. The proposed solution takes root from cryptographic primitives such as the hard problem of square root modulo a composite number [HPS08]. During the modulo operation the quotient is dismissed, it is impossible to reconstruct accurately the original value. Also, as in filtering techniques, e.g., mean filtering, where the underlying operation replaces one value by a value inferred from several original values, the output is more ho-
mogeneous with respect to the input and removes small variations such as noise. For this reason, the transformation presented here includes a step in which two original values are added to each other. The reduced intra-class variations are a positive step in working with biometric data, which suffer from inherent variations.

The output of the MCC descriptor creation is a template $T$ for every fingerprint, which contains a fixed length descriptor, referred to as cylinder $TC$, for each minutia. In order to dismantle the original structure of the cylinders, two-by-two elements are summed based on a user key, which is a permutation of the cylinder indexes. The remainder of the division of the square of this sum by a given parameter is then considered as a new revocable value as summarized in Equation 1. A binarization step is performed on the transformed values to introduce irreversibility and quantization, which is necessary in presence of the intrinsic intra-class variations of biometric data. Because of the small values of the baseline MCC template, a multiplication factor $A$ is used to adapt the size of the argument. A user key $k$, which is a random permutation of the cylinder indexes, is employed to specify the order of the summations of two-by-two elements. Changing this key ensures that it is possible to create several diverse instances of one biometric characteristic and allows to implement revocability through key management. The revocable template, which is the output of the method presented in this paper, is denoted by $RT$ and its cylinders as $RTC$s. The parameters of the protection operation include the multiplication factor $A$, the user key $k$ and the divisor $n$.

$$RTC[i] = B((A(TC[k(2i - 1)] + TC[k(2i)]))^2 \mod n)$$
for $i = 1, \ldots, nb_{\text{elements}}(TC)/2$, and $\forall TC \in T$, (1)

where $B(v)$ binarizes each element of the descriptor by means of a threshold:

$$B(v) = \begin{cases} 1, & \text{if } v > t \\ 0, & \text{if } v \leq t. \end{cases}$$

(2)

The values of $A$ and $n$ must be in accordance with each other in order for the modulo operation to be meaningful. It is observed that $\max(TC[k(i)] + TC[k(i + 1)]) = 2$, because the output of the MCC method yields descriptors whose values are normalized between 0 and 1. In order for the modulo operation to yield meaningful results, $A$ and $n$ must be chosen such that $n < (2A)^2$. Furthermore, if the argument (i.e., $(2A)^2$) is very large compared to the divisor $n$, the discriminatory power of the biometric information is lost. Empirically, it is observed that the argument must not be more than three orders of magnitude larger than the divisor, i.e., $n < (2A)^2 < 10^3 n$. The threshold $t$ is determined based on the values of the transformed descriptors (which in turn depend on $A$ and $n$) and is empirically fine-tuned.

Given that the same transformation is applied on every cylinder in the template, it is possible to use the original LSS comparison measure of the MCC method [CFM10]. The LSS matching is based on computing the Euclidean distance of two by two cylinders. Given that by using the same key, the elements of the cylinders are shuffled in the same manner, their element by element distance does not vary. Due to the properties of Euclidean spaces and the nature of the transformation in Equation 1, a correlation exists between the distance of two descriptors before and after the transformation. This correlation is later
In order to evaluate the proposed privacy protection scheme, the public and widely used FVC2002 [MMC+02] and 2004 [MMC+04] databases are used, which contain each, 8 impressions from 100 fingers. The minutiae of every fingerprint, formatted according to the ISO standard [ISO05], are extracted using the open source FingerJetFXOSE software by DigitalPersona [dig].

The original FVC protocol is used to generate genuine and impostor scores. The MCC algorithm is implemented according to Section 2 and is hereafter referred to as the "baseline MCC”. The MCC double-valued cylinder creation and matching parameters are $n_s = 16$, $n_d = 8$, $R = 75$, $\sigma_s = 6$, $\mu_\psi = 0.005$, $\sigma_D = 0.4363$, $\min_{VC} = 20$, $\min_M = 1$, $\min_{ME} = 20\%$, $\min_{np} = 3$, $\max_{np} = 10$, $\mu_P = 10$, $\tau_P = 0.4$ and $\Delta \Theta = 2.35$. The size of each descriptor in the template using these parameters is 2048 elements. The minutiae extractor is modified to allow template creation for images with any number of minutiae.

## 4 Experiments, performance results and discussion

In order to assess accuracy changes after applying the template privacy protection method introduced in this paper, the genuine and impostor distributions are displayed in Figure 1 (a) for the two cases of recognition using the baseline MCC templates and recognition using the protected MCC templates on the FVC2002 DB1 in the case where each identity is assigned a different key. On the corresponding detection error trade-off (DET) curves in Figure 1 (b), it can be observed that recognition using protected MCC templates yields better overall separation and lower false accept rates (FAR) and false rejection rates (FRR) than recognition using the baseline MCC templates. Consequently, the equal error rate (EER) (the operating point at which $\text{FAR} = \text{FRR}$) is lowered as well. Corresponding error rates are reported in Table 1 for three operating points, as well as genuine/impostor class separation computed using the characteristics of Gaussian curves fitted to the actual distributions [MR05]. The EERs for the FVC2002 and 2004 databases (unseen data) are reported in Table 3. The parameters $A$ and $n$ (Equation 1) are set to $5 \times 10^3$ and $10^6$ respectively in order for the modulo operation to be meaningful while preserving the discriminatory power of the biometric information. Furthermore, in order to verify the validity of the conditions given for the values of $A$ and $n$ with respect to each other, several combinations were empirically chosen and tested. The threshold $t$ is set to $10^5$. However, this value is not optimal and other values within the range of $n$ are empirically tested in Table 2.

### 4.1 Accuracy of the transformed templates

<table>
<thead>
<tr>
<th>Method</th>
<th>EER</th>
<th>FRR@FAR 1%</th>
<th>FRR@FAR 0.1%</th>
<th>Class separation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline MCC</td>
<td>1.21%</td>
<td>1.21%</td>
<td>2.25%</td>
<td>0.89</td>
</tr>
<tr>
<td>Protected MCC</td>
<td>0.72%</td>
<td>0.67%</td>
<td>1.39%</td>
<td>1.43</td>
</tr>
</tbody>
</table>

Table 1: Recognition results. FVC2002 DB1 images.
From the results presented in Figure 1, it can be concluded that the protection scheme presented in this paper improves the overall verification performance. This observation is made through the lowered error rates as well as an increased genuine/impostor class separation. This phenomenon is explained by the two-factor authentication. The user key provides extra discriminative information to the templates. It must be noted that experiments further in this paper (Section 4.2) show that by setting all keys to one unique and universal key, it is the biometric information being recognized and not the key. Another positive aspect of this privacy protection scheme is that the transformation reduces the size of the template by half. This property is beneficial when considering large databases as well as applications with reduced resources such as smart cards. Note: a comparison with the current state of the art can be found in the full version of this paper (http://infoscience.epfl.ch/).

4.2 Diversity and revocability of transformed templates

In targeted working conditions of the biometric recognition system, where the protection scheme is implemented and working in its normal mode, each user has his own key. This key is first used to enroll (enr) his template in the database and is later reused to reproduce the transformation during verification (ver). The operating point decision thresholds are determined in this scenario, which is referred to as same key enr/ver.

If a template is revoked, the corresponding key is black-listed and a new template is generated using a different key. The two templates must be different in order for the old template to be nullified. This scenario is referred to as different key enr/ver and corresponds to the case where a user needs several different instances of his biometric for use in different applications. Given that the original template consists of descriptors with $m = nb_{\text{elements}}(TC)$ elements ($m = 2048$ in this paper), $m!$ different keys can be gener-
ated. However, since the two elements of $TC$ specified by $k$ are added to each other before undergoing any other operations, the order of the elements does not matter. For example, if $n_{\text{elements}}(TC) = 8$, then keys $k_1 = 37814256$ and $k_2 = 73814256$ yield the same protected template. They are equivalent to all other keys which have the same indices in the consecutive odd and even positions, regardless of the order of the indices within the pair. A “pair” here refers to an odd and even position in the key, for example $k[1]$ and $k[2]$ are a pair, as well as $k[3]$ and $k[4]$. If $k$ has $m$ elements, then there are $\frac{m}{2}$ pairs. There are $2^{m/2}$ possible permutations within the pairs since every pair has two permutations. The $2^{m/2}$ permutations yield equivalent keys as only the position within a pair varies from one key to another, not affecting the outcome of the transformation. These $2^{m/2}$ keys are a category of equivalent keys. Therefore in order to compute the number of distinct keys (which yield different protected templates through the transformation), the total number of permutations must be divided by the number of categories of equivalent keys, which results in $\frac{m!}{2^{m/2}}$ different keys. In order to ensure two diverse templates of one person are not similar, the pseudo-impostor accept rate (PIAR) is introduced to evaluate the proportion of successful pseudo-impostor attempts. A pseudo-impostor comparison is the comparison of two templates of one individual, which are generated using two distinct keys. It can be seen in Table 4 and Figure 2, that it is difficult to use a template generated with the wrong key for positive verification. If a user loses his key, it is known by an adversary. This scenario is simulated by using the same key for enrollment and verification, for all users, and is referred to as the stolen-token scenario. Furthermore, this scenario shows that it is the biometric information being verified, and not the key. These diversity and security testing scenarios are implemented according to [WH12].

<table>
<thead>
<tr>
<th>Threshold</th>
<th>same key enz/ver.</th>
<th>different key enz/ver.</th>
<th>stolen token</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>FAR</td>
<td>FRR</td>
<td>FAR</td>
</tr>
<tr>
<td>0.4489</td>
<td>1%</td>
<td>0.67%</td>
<td>1.03%</td>
</tr>
<tr>
<td>0.4518</td>
<td>0.72%</td>
<td>0.72%</td>
<td>0.66%</td>
</tr>
<tr>
<td>0.4673</td>
<td>0.1%</td>
<td>1.39%</td>
<td>0.06%</td>
</tr>
<tr>
<td>0.4710</td>
<td>0.04%</td>
<td>1.6%</td>
<td>0.06%</td>
</tr>
<tr>
<td>0.4934</td>
<td>0</td>
<td>2.71%</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4: Error rates at different operating points under different diversity and security assumptions. FVC2002 DB1 images.

It can be concluded from the observations in Table 4, that it is possible to define an operating point for the privacy protection scheme presented in this paper, such that, a very limited number of impostors (and pseudo-impostors) gain access to the system with an unauthorized key. It must be noted that the security of this scheme is dependent on the decision threshold and offers flexibility in various operational scenarios.
4.3 Irreversibility of the transformed templates

In order to effectively protect a person’s biometric characteristic, it must not be possible to deduce said characteristic from the transformed template, even when all parameters, including the key, are known. In the transformation presented in this paper, non-reversibility is provided by the fact that during the transformation in Equation 1, part of the data is discarded through the modulo and quantization, which does not allow exact reconstruction even if all parameters are known. However, it is possible to create an approximation of a binary MCC template given a transformed template and the corresponding key as described in Equation 3:

\[ TC_R[k(2i)] = \begin{cases} 
1, & \text{if } RTC[i] = 1 \\
0, & \text{otherwise and} \\
\end{cases} \]

\[ TC_R[k(2i + 1)] = \begin{cases} 
1, & \text{if } RTC[i] = 1 \\
0, & \text{otherwise.} \\
\end{cases} \]

\[ i = 1, \ldots, nb_{\text{elements}}(RTC), \text{ and } \forall RTC \in RT. \]  (3)

\( TC_R \) denotes the approximated MCC templates and is compared to the baseline binary MCC template, for all users of a database in Figure 2 (b). It can be seen and concluded that these approximated templates do not resemble the baseline MCC templates and thus cannot be systematically used to extract useful minutiae information. In fact, when compared with genuine templates, the reconstructed templates are similar to impostor templates.

5 Conclusions and Future Work

In this paper, a novel template privacy protection technique for the MCC representation of fingerprint minutiae templates was presented. The proposed hybrid, two-factor technique combining a transformation and a user key, provides diversity, revocability, and irreversibility for the MCC descriptors with respect to the original minutiae information while improving the accuracy in recognition. Furthermore, the proposed technique reduces the template size by half.
Future work will include study and modification of the transformation in order to extend of the ideas presented in this paper to other modalities than fingerprints.

References


Abstract: Since the past decade biometric technologies are field-proven, facilitating reliable and secure access control. Numerous successful deployments on large-scale systems, e.g. airports, confirm the feasibility of biometric recognition. However, applications of biometric systems involve privacy concerns, i.e. debates on social and ethical acceptance of biometrics reached levels never previously witnessed.

In this work a comprehensive questionnaire regarding social acceptance of biometric technologies in Germany is presented. Results are obtained from a total number of 140 respondents, allowing a representative analysis of citizens’ attitudes towards biometric technologies. Relevant questions are put into view and perceptions of German citizens regarding the rise of biometric technologies are discussed in detail and interesting conclusions are drawn.

1 Introduction

Biometric recognition [JRP04] refers to automatic authentication of humans by their physiological or behavioral characteristics or traits. Several biometric characteristics, e.g. fingerprints, iris, or face, have been discovered for robust and reliable recognition and reveal impressive performance in terms of recognition accuracy holding tremendous promise for applying biometric technologies in diverse application scenarios [JFR08]. So far, the vast majority of existing research conducted in biometrics is mainly focused on technical aspects. Without a doubt, biometric technologies are on the rise enforcing an integration of biometrics into daily life, e.g. passports or door-locks, which often leaves citizens with no choice but to accept biometrics. While majorities are convinced of the need for improved authentication controls and biometrics are claimed to provide secure long-time solutions for fundamental problems, public issues arise which citizens need to be heard on, e.g. convenience or transparency. Contemporary attitudes of citizens towards biometric recognition must be considered important, since a wide-spread use of biometrics may cause even more issues. For instance, attacks on biometric systems [RCB01], e.g. spoofing, hacking, reconstructing biometric information from templates, insider attacks, or even theft of body parts may all seem far-fetched now, but they could become common if the use of biometrics and the value of the information protected by biometrics increases [Pat08]. While research already tackles some issues, e.g. privacy-protecting technologies [RU11, JNN08], others may not be considered a technical or scientific question. In con-
trast, issues such as social acceptance can only be taken into account by understanding the ‘people’-side of biometrics [CSR04].

The contribution of this work is the investigation of social acceptance of biometric technologies within Germany. Based on a comprehensive questionnaire, which focuses on current issues regarding deployments of biometric technologies, attitudes of 140 citizens of all over Germany are aggregated. Obtained results are interpreted and discussed in detail, whereat interesting conclusions emerge.

The remainder of this paper is organized as follows: in Section 2 related studies are briefly summarized. The proposed questionnaire, obtained results and a discussion of these are presented in Section 3. In Section 4 conclusions are given.

2 Related Studies

As previously mentioned, only a few scientific works regarding biometric recognition focus on social acceptance of users. Furnell et al. [FDMR00] proposed a questionnaire on the user-acceptance of biometric technologies which was completed by 175 respondents living in the UK. Interestingly, the survey has shown that, although demonstrably weak, typed passwords remained the most popular form of authentication in the minds of users. In [PW06] a survey is proposed in order to measure perception based on various uses of biometric technology as well as implantable RFID-chips in the human body as an enhanced biometric method. It was found that the 141 respondents in this survey were most willing to employ biometric identifiers into the United States Passport system (almost half of respondents). Conversely, respondents were least willing to employ biometric identifiers into a system to obtain a credit card with results showing nearly two-thirds of respondents unwilling. With respect to the implantable RFID-chips only less than half of respondents are unwilling to implant these. Heckle et al. [HPO07] presented a study, in which 24 participants were asked to role-play the use of a fingerprint biometric identification system when making purchases at an online bookstore. The results showed that 21 out of 24 participants found it beneficial to use the fingerprint system. Participants were also asked why they might prefer to use a biometric system when making an online purchase. The most frequent response was that the system would be easier to use than a traditional username and password (60% of responses). Only 35% of the participants said they would prefer to use a biometric system because of an increase in security. The authors found that the participants relayed a sense of confusion when it came to assessing the security strength of biometrics. Interestingly, subjects stated that they would accept biometrics if it was the social norm. Jones et al. [JAE07] investigated the users’ perceptions of biometrics with respect to various application scenarios, e.g. building access, computer access, or financial transactions. Overall, the use of biometrics did not receive more acceptability compared to conventional authentication mechanisms, e.g. password-based authentication. In contrast, regarding biometrics the majority of the 115 respondents raised concerns about misuse with respect to fraud and tracking. Similar results were obtained in [FE07], in particular, ∼38% of 209 respondents were far from confident that biometric information will only be used for authentication purposes. In [EAGHR10] 70 volunteers were enrolled.
in biometric systems based on keystroke dynamics and face recognition. Interestingly, the majority of respondents found that the system based on keystroke dynamics outperformed the face recognition system, i.e. they were more satisfied with the system based on keystroke dynamics, although it actually revealed higher error rates compared to the face recognition system. Furthermore, the authors identified significant correlations between education level and respondents’ opinions about secret-based solutions against fraud and their concerns about privacy issues. In [BS00] the procedure of selecting Passfaces from a grid of faces displayed on the screen is compared to conventional passwords. In this study, which was carried out with 34 student participants in a 3-month field trial, password caused substantial login failure rates compared to the Passface approach, i.e. the latter achieved improved usability. Recently, Mok and Kumar [MK12] investigated privacy related concerns in the deployment of biometrics and data protection technologies in China. In a survey in which 305 subjects participated it was found that the most acceptable biometric characteristics are fingerprint, iris, and face. In addition, it was found that Airports and Banks are the most preferred venues of deploying biometrics technologies.

3 Social Acceptance of Biometric Technologies in Germany

In order to give an insight to the social acceptance of biometric technologies in Germany, a comprehensive questionnaire was proposed. In the following subsections the questionnaire, according evaluations are described in detail, obtained results are presented and the most interesting findings are discussed.

3.1 Proposed Questionnaire

The survey which consists of 56 question was completed by a total number of 140 respondents, the distribution of respondents across different states of Germany is summarized in Table 1. As can be seen nearly all states are represented which forms an adequate basis for an representative investigation. This national distribution was achieved by publishing

<table>
<thead>
<tr>
<th>State</th>
<th>Respondents</th>
<th>State</th>
<th>Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baden-Württemberg</td>
<td>18</td>
<td>Niedersachsen</td>
<td>13</td>
</tr>
<tr>
<td>Bayern</td>
<td>19</td>
<td>Nordrhein-Westfalen</td>
<td>30</td>
</tr>
<tr>
<td>Berlin</td>
<td>4</td>
<td>Rheinland-Pfalz</td>
<td>7</td>
</tr>
<tr>
<td>Brandenburg</td>
<td>2</td>
<td>Saarland</td>
<td>1</td>
</tr>
<tr>
<td>Bremen</td>
<td>–</td>
<td>Sachsen</td>
<td>6</td>
</tr>
<tr>
<td>Hamburg</td>
<td>2</td>
<td>Sachsen-Anhalt</td>
<td>4</td>
</tr>
<tr>
<td>Hessen</td>
<td>12</td>
<td>Schleswig-Holstein</td>
<td>13</td>
</tr>
<tr>
<td>Mecklenburg-Vorpommern</td>
<td>3</td>
<td>Thüringen</td>
<td>6</td>
</tr>
</tbody>
</table>
the survey via diverse Internet platforms, e.g. Facebook\(^1\) and Twitter\(^2\). The distribution of participants across different age-groups is depicted in Fig. 1 (a), i.e. the vast majority of respondents is between 20 and 29 years old. The entire group consisted of 90 males and 50 females and the majority of respondents were either students or employees. In this work only the most important and most interesting results will be presented.

### 3.2 Obtained Results

Within initial questions emphasis was put on the convenience of knowledge-based authentication mechanism, in particular PINs and passwords. As shown in Fig. 1 (b) the majority of respondents finds it hard to remember PINs or passwords. As a consequence, see Fig. 1 (c), people tend to forget PINs and passwords, interestingly, passwords are forgotten more often than PINs. Fig. 2 (a) illustrates the amount of PINs or passwords, physical keys,
and chip cards German citizens have to maintain on average. While most respondents posses only 4-5 physical keys, the majority of participants has to remember more than 7 passwords. Based on the fact that most respondents find it hard to remember PINs and passwords these are re-used at various application scenarios while the majority does not frequently change their PINs or passwords, as shown in Fig. 2 (b)-(c). Based on these results, knowledge-based authentication systems requiring PINs or passwords appear inconvenient.

Focusing on biometrics, which may represent a suitable solution to these issues, the most well-known biometric characteristics are fingerprints, iris, face and speaker/voice recognition, see Fig. 3 (a). In addition, the acceptability of biometric characteristics is depicted in Fig. 3 (a). Obviously, the highest acceptability is gained for fingerprint recognition which is also the most well-known technology, which coincides with the findings in [HPO07], where participants state that a more common use of technology brings about more social acceptance. In contrast, while the majority of respondents is familiar with face and speaker/voice, recognition based on these characteristics is only accepted by a few. The majority of respondents answered in the negative when they were asked whether they consider biometric recognition as too personal, intimate or even frightening, as illustrated in Fig. 3 (b).

While biometrics are deployed at a great variety of applications still only \( \sim 45\% \) of respondents have already registered with at least one biometric system, see Fig. 4 (a) (in most cases fingerprint recognition systems). While only a slight majority believes that biometric systems facilitate everyday lives \( \sim 75\% \) are convinced that biometric access control systems are necessary at distinct venues, as illustrated in Fig. 4 (a). The effects of aging on biometric recognition [Lan10] has been investigated in past years and numerous biometric characteristics, e.g. fingerprints [HGL+11, DAB04] and iris [FB12], were found to be highly influenced by aging. When asking participants what they consider an adequate time lapse after which a re-registration with the biometric system is required, rather short time lapses of 3 or 6 months are acceptable for only a few respondents, as shown in Fig 4 (b), again, convenience plays an important role.
(a) Have you used biometrics? Do biometrics facilitate daily life/are they necessary?  (b) What do you consider as an adequate time lapse after which re-registration is required?

Figure 4: Amount of respondents which have already used biometric systems, the common attitude with respect to daily life, and time-lapses considered adequate for re-registration.

(a) How do you face the future development of biometric technologies? (b) Do you consider biometrics or knowledge-based authentication more tamper-proof?  (c) Which type of authentication do you think provides more advantages?

Figure 5: Future perspectives of respondents regarding biometric technologies and the comparison between knowledge-based authentication and biometric systems.

While ~25% of respondents are not concerned about the future of biometric technologies a majority of ~45% face a positive future for the development of biometric systems, see Fig. 5 (a). Compared to knowledge-based authentication schemes most respondents believe that biometrics are more tamper-proof while improved security may not be considered as a major advantage as shown in Fig. 5 (b)-(c).

3.3 Discussion

Users seem to be annoyed by maintaining several PINs and passwords, i.e. biometrics, which can not be lost or forgotten, bring about substantial benefits with respect to usability. However, while citizens are familiar with most biometric characteristics, in general the acceptance of biometrics appears quite disillusioning. For instance, while ~25% (!) of respondents did not accept biometric recognition at all, only one third or less of the
respondents which are familiar with face and speaker/voice recognition accept these technologies. These results may be influenced by negative recent press releases, e.g. regarding Facebook’s face recognition software.

Generally speaking, based on the obtained results deployments of biometric systems need not face refusal due to physical invasion of privacy or even fear. While less than half of the participants have experience with biometric systems these are not considered as more convenient than, for instance, knowledge-based approaches. This implies biometric systems require improvements with respect to usability (which is also related to biometric performance rates). Usability is influenced by aging effects as well, while most participants accept only long time lapses ($\geq 1$ year) between re-enrollments. Overall, attitudes towards future developments of biometric technologies are positive while at the time citizens do not discover major advantages of biometrics over knowledge-based authentication mechanisms, despite improved security.

Obtained results confirm that general awareness regarding biometric technologies requires major improvement in Germany, which may also improve the social acceptance of biometrics which appears still unsatisfying.

4 Conclusion

In this paper results obtained from a survey on the social acceptance of biometric technologies, which was completed by 140 German citizens, are presented. Results of the comprehensive questionnaire, which was completed by people of almost all states of Germany, reflect the current general attitudes towards different biometrics-related topics within Germany, providing interesting insights. Nevertheless, the evaluation of the presented survey appears rather disillusioning confirming the fact that general awareness of biometric recognition technologies needs to be improved.

Acknowledgment

This work has been partially supported by the the Center for Advanced Security Research Darmstadt (CASED), Germany.

References


Quality Filtering of EEG Signals for Enhanced Biometric Recognition

Su Yang, Farzin Deravi

School of Engineering and Digital Arts
University of Kent
Canterbury, Kent
CT2 7NZ
Sy91@kent.ac.uk
F.Deravi@kent.ac.uk

Abstract: In this paper we present a biometric person recognition system based on EEG signals incorporating a novel strategy to find and utilize the most informative data segments using the concept of Sample Entropy. The users are presented with a stimulus that prompts a motor-imagery response. This is then measured using an array of EEG sensors. A sliding-window segmentation scheme and Wavelet Packet Decomposition are adopted for primary feature extraction before the quality measurement stage. The quality-filtered feature windows are then used to extract secondary features that are in turn classified using a linear discriminant classifier. The proposed system is tested using a publicly available EEG database and it shows that entropy filtering results in a significant improvement on performance. An average identification accuracy rate of more than 90% is achieved for 109 subjects using only eight electrodes, utilizing only the highest quality for each subject.

1 Introduction & Related work

Biometric person recognition technologies have been an active area of research in recent years driven by advances in machine learning techniques and availability of low-cost sensors. While this has resulted in significant deployments in a range of applications, important challenges still remain to their widespread adoption and acceptance [JSSL.A04]. Because of this, the search for new biometric modalities continues. In this paper, we investigate a relatively new source of biometric information with the potential to overcome some of these challenges in some applications – Electroencephalogram (EEG) signals.

Using EEG signals as a biometric modality is a relatively new area of research. Palaniappan et al. reported that by measuring 40 Hz oscillations from 61 electrodes a classification rate of 95.25% can be achieved [PDA07]. Their experiment concentrated on investigating the Visual Evoke Potential (VEP) when 20 subjects were stimulated by viewing sets of picture, the stimulus duration of each picture was 300 ms with an inter-trial interval of 5.1s. Their later experiments showed that by processing VEP of the EEG
gamma band signal, even when the number of subjects increased to more than 100, the accuracy rate achieved was 98.12%; and after the number of electrodes was optimized (35 out of 61 were kept) the median accuracy rate still reached 97.62% for 40 subjects [PDb07]. However, these performance levels were achieved at a cost: several electrodes (at least 35) were needed even after algorithm optimizations. 

For practical use as a biometric modality the EEG sensor equipment must be easy to set up: ideally the number of electrodes should be as small as possible. Recently some researchers have tested the identification performance with only one electrode. Fei Su et al. [SLAYJ10, SLAJ10] reported a system that could achieve an average accuracy of 97.5% for 40 subjects acquiring the signal from a single electrode only (FP1 in the international 10-20 system [NF05]), subjects were not required to perform any mental or physical tasks. The features were based on combining the coefficients of an auto-regressive model and Power Spectrum Density (PSD), and these were classified using a kNN-FDA (k-Nearest Neighbour and Fisher’s linear Discriminant Analysis) combined classifier. The recording of their experiment contains 480 records and each record lasts for 5 minutes, ten-fold cross validation was used for testing. Palaniappan et al. [PGRS11] claimed a 100% accuracy rate with three subjects using visual stimuli consisting of different numbers depicted on a screen. The signal was recorded using a single sensor at Cz position [Ch85] and was classified using a single hidden layer neural network. Despite these encouraging accuracy rates, the limitation of these experiments are either their relatively long signal recording duration needed for training and testing, or the small number of subjects involved. 

In this paper we propose a new processing scheme that is designed to only utilize the most information-rich data, and hence achieve acceptable performance by using relatively small number of electrodes (up to eight electrodes) on 109 subjects. The paper has the following structure: the general scheme for EEG data acquisition will be presented in Section 2, along with the electrodes positioning and the block diagram of the system. The principal algorithms used in the proposed system will be introduced in Section 3. Section 4 outlines the novel entropy-based method for quality measurement followed by the experimental results as well as the tests for optimizing the system parameters. Section 5 provides a summary and suggestions for further work. 

2 Experiment scheme

The proposed biometric system is based on measuring the evoked response of users when they are confronted with a stimulus and asked to perform a task. The EEG signals from the user are obtained by a sensor system attached to their scalp. The user is presented with a stimulus (e.g. visual target on the screen) which requests them to perform or imagine motor tasks. The EEG signals thus generated are then gathered, and processed to provide identity information for the particular user. The “EEG Motor Movement/Imagery Dataset”, supplied by the developers of the BCI2000 instrumentation system [SJHBR04, GAG’00] was used for evaluating the proposed algorithm and the performance of the system in an identification (one-to-many recognition) scenario. The sampling rate of the sensor was 160 Hz, and EEG data from
109 subjects were recorded. The database includes a number of recordings for each individual recorded in a single session and separated by short intervals. These include two-minute baseline recordings, one with the eyes open and one with the eyes closed. Three two-minute recording runs (separated by a “a couple of minutes”) are also made (Runs 1 to 3) for each of four different motor/imagery tasks (Tasks 1 to 4) [EMD09].

The proposed system is trained and evaluated using data obtained from Task 4 which was a motor imagery task for both hands and feet. The reason for adopting Task 4 is that the motor imagery task might better avoid the contamination of the EEG signal by other bioelectrical signals such as electromyography (EMG). Due to the need for ease of deployment as a biometric modality, no more than eight electrodes are used and their positions are clustered around the centre of the motor cortex: FC1, FC2, C3, CZ, C4, CP3, CPz, and CP4 [Ch85]. The electrode positions are depicted in Figure 1.

![Fig. 1 Tested electrode positions, modified from the figure in [EMD09] (10-10 system [Ch85])](image)

The block diagram of the proposed system is shown in Figure 2. The Wavelet Packet Decomposition [Da92] is used to generate the primary features. An entropy-based measurement method is designed to select optimal primary feature windows for generating secondary features which are then passed to a classifier. $M$, $K$ and $L$ are three parameters for controlling the system performance and are described in the following section.

![Fig. 2 Block diagram of the system](image)
3 Pre-processing and Feature Extraction

One goal of pre-processing and feature extraction, especially when EEG signals are used for biometric recognition, is to remove the unwanted signal components (such as signals generated by eye blinks and heartbeats) and only include the essential identity-bearing information for classification. For the proposed system, Wavelet Packet Decomposition (WPD) is utilized for the first stage of feature generation. This is followed by a novel method for quality measurement, designed to select and use only a small amount of data segments which are most likely to provide a correct recognition before feature extraction.

3.1 Sample Entropy for Quality Measurement

This section is devoted to the novel entropy-based filtering method. The motivation of adopting Sample Entropy for quality measurement is presented first, followed by a description of the scheme for using it in EEG filtering in the proposed system.

3.1.1 Motivation for using entropy as a measure of quality

Like human speech, the EEG signal produced by the brain is non-stationary [CBEV95]. As a stochastic time series, the frequencies of the typical EEG signal vary over time depending on what brain functions are being performed. Also similar to speaker recognition, it is unlikely the whole time series is equally informative for the purpose of identity recognition. In order to reduce the amount of the data used for processing and improve its quality, it is necessary to find a strategy to extract the most useful segments of the data and discard the relatively less-informative portions of it. Entropy of the EEG signal has been used as a feature to identify the seizures in epileptic patients. It has been reported that during seizure patients’ brains generate lower entropy EEG signals than for healthy people. This implies that the healthy brain signal possess less regularity than a brain during seizure [AMK09]. In that experiment they tested three different kinds of entropies calculated from EEG segments (Shannon entropy, Sample entropy and Log entropy) and all of them showed such a trend. Liang et al. later investigated whether the entropy of the EEG signal could also identify sleeping stages. They measured the EEG signal on an epoch-by-epoch basis, using multi-scale entropy analysis (MSE) and noticed that the “entropy values monotonically decrease from awake to deep sleep” [SCY+12]. These results suggest that entropy may be used to measure the level of brain activity from EEG signals in healthy human brain functions. The more active the brain is with cognitive/motor functions the more unpredictable the EEG signal is likely to be, hence the higher the entropy value. Based on this hypothesis, we propose to use the Sample Entropy as a measure of EEG signal quality for biometric applications as described below.

3.1.2 Using sample entropy to filter EEG data

After calculating the WPD for each window of the time series, the sample entropy of each wavelet coefficients window is computed. In the experiments reported in the next section, each recording run of approximately 2 minutes is segmented into windows of
960 samples (6 seconds duration) using a sliding window approach with a shift of 24 samples between windows, thus producing 760 windows for every EEG band per electrode. More generally, the number of windows generated per band per electrode can be defined as a system parameter, \( M \), as shown in Figure 2. These coefficient windows are then fed to the SampEn calculation module which ranks the windows in order of their entropy values listing them from the highest entropy window to the lowest. For each band, \( K \) out of \( M \) windows, starting from the \( L \)-th value in the entropy-sorted list of entropy values, are preserved in order to reduce the quantity of data needed for further processing and remove the information-poor windows. In the experiments that follow, only about 1/10 of the data (80 windows) is used for secondary feature extraction and classification. The standard deviation of the wavelet coefficients from the selected windows are then calculated to serve as the secondary features for classification. It has been suggested that the choice of the tolerance threshold \( r \) is important in the calculation of SampEn: if it is set at too high a value, detailed system information may be lost and if it is set at too small a value, poor conditional probability estimates might result [RR00]. In the experiments reported below \( r \) is set to 1 and run length \( m \) is set to 2.

4 Experiment Results and Evaluations

During the training phase, for every electrode per subject \( M=760 \) windows are fed to the WPD stage. Each window is decomposed into nine bands of wavelet coefficients. Since eight electrodes are used in the experiments, before the entropy screening stage a total of \( 8 \times 760 \times 9 = 54720 \) coefficient windows are generated for each subject. Next, the Sample Entropy is calculated for each window and the windows are sorted in descending order of SampEn. Out of the \( M \) windows for each electrode and band, \( K \) (=80) windows (hereby referred to as “observations”) are retained, starting from the \( L \)-th ranked window \( (L = 1, \ldots, M-K) \), as a contiguous range from the SampEn-sorted list of windows to preserve the most information-bearing part of the data. This amounts to roughly 10% of the whole data. After this quality measurement and screening stage, for every subject only 5760 out of 54720 windows are kept and the standard deviation (\( \sigma \)) of each coefficient window is calculated and used as features for classification using a normalised Linear Discriminant Classifier (LDC, [DJD+04]). This choice of classifier was based on tests and comparisons with several other classifiers (Support Vector Machines with different kernels, k-Nearest Neighbour classifiers, kernel-LDC and kernel-k-NN) using the database described above.

4.1 Entropy filtering optimization (optimising \( L \))

Different contiguous ranges of windows from the entropy-sorted list of coefficient windows are extracted and used to filter the training data and only those windows within the selected range are used for classifier training. The first range tested is for the windows with the highest entropy values (rank 1 to 80). The system is trained and tested with the same range of entropy value ranks. Data from Run 2 is used for testing the identification accuracy. A range of high-entropy ranks (130th to 360th range) is identified which provides high biometric performance. It could be that the highest
ranking windows correspond to activities that do not carry identity information. As each window lasts 6 seconds, it could include 5 to 6 cycles of motor actions (e.g. opening and closing of hands). This could be considered as a relatively regular function with moderate SampEn values. Tests with different parameter L, suggests that the windows corresponding to approximately the highest 15% rankings (L = 1 to 180) Sample Entropy values should be discarded to improve performance.

4.2 Performance as a function of training data volume (optimising K)

The amount of data used for training was varied to assess how a reduced training data volume affects the accuracy of the system. The number K of observations used for classification is reduced in a number of steps from 80 to 1 and the identification test results are noted. The results are relatively stable up to K=10. Still more than 70% accuracy rate can be achieved while only 2 observations are kept. However, the performance significantly degrades when only 1 observation is preserved. A compromise setting may be K=10 roughly using 1.3% of the data is utilized, and still achieving more than 90% identification rate for 109 subjects.

4.3 Performance as a function of test duration

Figure 3 depicts the degradation of the accuracy rate for identification when the testing duration t is reduced. These tests are all based on number of observations K=10 and starting value L=200, utilizing Run 2 (or part of it) for testing and the other two Runs for training. Hence, 4 minutes of recording is used for training the classifier and each point on the horizontal axis refers to testing with test durations of different length. Dropping the test duration by a factor of four from 120 seconds to 30 seconds results in a loss of mean accuracy of only 3.48%; dropping the test duration all the way to just 6 seconds, results in a drop in accuracy of less than 9% compared with using the whole two minutes of data for testing.

Fig. 3 Impact of the testing duration on average identification accuracy with K=10 and L=131
5 Conclusions and Future Work

In this section to illustrate the impact of the proposed entropy filtering method a number of different schemes are compared as shown in Table 1. The schemes I to III employ entropy filtering while schemes IV and V are methods only utilizing wavelet decomposition. The results presented in Table 1 indicate that, given the chosen system parameters, the entropy filtering method improves the recognition performance by around 5% compared to using no entropy filtering at all. The results for Scheme III suggests that the low entropy-windows contain significantly less biometric information.

<table>
<thead>
<tr>
<th>Schemes</th>
<th>L: Rank starting value</th>
<th>K: Preserved observations</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Highest entropy</td>
<td>1</td>
<td>80</td>
<td>87.0%</td>
</tr>
<tr>
<td>II. Highest performance</td>
<td>131</td>
<td>80</td>
<td>90.4%</td>
</tr>
<tr>
<td>III. Lowest entropy</td>
<td>681</td>
<td>80</td>
<td>74.8%</td>
</tr>
<tr>
<td>IV. No entropy filtering</td>
<td>1</td>
<td>760</td>
<td>86.5%</td>
</tr>
<tr>
<td>V. No entropy filtering</td>
<td>1</td>
<td>80</td>
<td>85.6%</td>
</tr>
</tbody>
</table>

This paper explored the notion of quality for EEG signals used for biometric person identification. A novel system was presented where a measure of signal quality, the Sample Entropy, was used to filter the data available for biometric recognition. For a 109 subject database an identification rate of more than 90% was achieved. The results indicate comparative performance with other published methods while promising the possibility of being able to handle large number of subjects using data from fewer electrodes. Further work will focus on optimizing the system parameters separately for different frequency bands and increasing the amount of data used for system evaluation.

References


[GAG00] Goldberger, Ary L., Luis AN Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang


Spoo\fing 2D Face Recognition Systems with 3D Masks

Nesli Erdogmus, Sébastien Marcel
Idiap Research Institute
Centre du Parc - rue Marconi 19 CH-1920 Martigny, Suisse
{nesli.erdogmus,sebastien.marcel}@idiap.ch

Abstract:
Vulnerability to spoofing attacks is a serious drawback for many biometric systems. Among all biometric traits, face is the one that is exposed to the most serious threat, since it is exceptionally easy to access. The limited work on fraud detection capabilities for face mainly shapes around 2D attacks forged by displaying printed photos or replaying recorded videos on mobile devices. A significant portion of this work is based on the flatness of the facial surface in front of the sensor. In this study, we complicate the spoofing problem further by introducing the 3rd dimension and examine possible 3D attack instruments. A small database is constructed with six different types of 3D facial masks and experimented on to determine the right direction to study 3D attacks. Spoofing performance for each type of mask is assessed and analysed thoroughly using two Gabor-wavelet-based algorithms.

1 Introduction

Automatic recognition of human biometric traits has numerous important advantages over conventional methods like passwords or ID cards [JR08] and hence, it has become a vast research field today as the need and investment for access control systems grow continuously. Among these traits, face recognition stands out with its favourable reconciliation between convenience and reliability. Unfortunately it suffers security vulnerabilities because it is easy to access to face samples and subsequently, to devise spoofing attacks.

Spoofing attack is the act of presenting a fake biometric evidence to a system in order to achieve authentication [NAR08]. Forging such an attack is relatively simple for facial recognition systems since the photographs or videos of valid users can be acquired from internet or captured at a distance. Attackers can attempt to penetrate by displaying printed photos or replaying recorded videos on mobile devices in front of the sensors. Since these are the most common, easiest and cheapest methods to circumvent face recognition systems [BAF+12], the counter measure studies primarily shape around them. Recently, several papers that enable comparison of various counter attack algorithms are published, providing public databases and reproducible works [CAM+11, AM11, CAM12].

Many anti-spoofing techniques analyse the texture of the captured image, mainly based on the assumption of printing artifacts [BNGS10] and/or blurring [LWTJ04] which rely on the printed image or display quality.

Another group of methods that appears as a separate or complementary measure, tries
to detect liveness of the face based on live-face specific movements such as eye blinking [PSWL07]. However, this kind of methods will definitely fail in the case of video replay attacks or even more simply, photographic masks which are high resolution photographs worn on face with eyes and mouth regions cut out, as illustrated in [KFB08].

Finally, there exist motion analysis techniques to detect spoofing attacks based on the fact that planar objects like papers or screens move in a significantly different way than the real faces. For instance, in [KFB05], the trajectories of single parts of faces are analysed to distinguish between live and spoofed ones. In a similar manner, Marsico et al. [DMNRD12] exploit the facial shape and detect attacks by computing geometric invariants of a set of automatically located facial points. The 3D nature of the captured face can also be perceived by employing additional devices. Even though they were mostly considered to be expensive, this view is bound to grow obsolete with the introduction of affordable consumer depth cameras. In [TDM07], 3D data acquired with a low-cost sensor is utilized to localize face and at the same time to test the "face-ness", rendering the system invulnerable to spoofing attacks. This claim is pretty valid, since differentiating a real face from a planar surface with a depth sensor is quite straightforward.

On the other hand, the advancements in depth sensing technologies are counteracted by similar progress trends in 3D facial mask manufacture. Taking the face spoofing attacks one step, namely one dimension, further, 3D masks introduce new challenges for countermeasure studies. As discussed above, the majority of the counter-measures developed for 2D spoofing fails to function properly in the case of masks.

In [ZYL+12] Zhang et al. states that since usually it is too expensive to produce client-like masks, massive usage of masks rarely appears in the literature. To the best of our knowledge, there have been three papers published in this field, which aim to detect 3D mask attacks using multi-spectral lighting [KNYY09, ZDLL11]. In both approaches, the authors try to classify human skin and non-skin by using Lambertian model to analyse multi-spectral reflectance properties of face. In [KNYY09], reflectance of human skin and different mask materials (silicon, latex and skin-jell) is measured on the forehead region for visible and infra-red light at different wavelengths. A 2D feature vector is proposed under 850 and 685 nm illumination which reaches a classification accuracy of 96.77%. In a similar manner, Zhang et al. [ZDLL11] select two most discriminative wavelengths and train an SVM classifier to learn real face and mask distributions at multi-distances. An average detection accuracy of 89.18% is achieved. Lastly, in a recent publication [KD13], the authors propose an LBP-based counter-measure against 3D printed masks. The spoofing detection accuracies are reported to be 88% and 86% for color and depth images. There are two main shortcomings in those studies: As later stated by the authors of [ZDLL11] in [YZL+12], the first limitation of the first two techniques is that they are not very convenient due to their special expensive hardware requirements. Secondly and most importantly, no analysis on the spoofing performances of the 3D masks is included.

In this paper, our purpose is to fill this gap by shedding some light on how high the spoofing performances can get for different types of masks. We believe that developing counter-measures for an impractical attack would be a waste of resources. Our experiments show that silicone masks whose reflectance properties are analysed in depth [KNYY09, ZDLL11] and 3D printed masks which are examined in [KD13] to develop anti-spoofing
measures, in fact, can not deceive 2D face recognition systems that well. It is important to emphasize that our study merely focuses on examination of spoofing performances and counter measure design is out of its scope. Despite its small size, the collected dataset offers crucial preliminary findings which a larger-scale research can be based on.

2 3D facial masks

The production of 3D facial masks spans a wide variety of materials and methods. It can range from an easy/low-cost way of printing a 2D photo on a deformable surface like cloth to a more difficult/expensive method of obtaining and printing a 3D model of a person.

Printing photo on a cloth is not a very promising technique to spoof a 2D face recognition system since the image will be distorted when applied to the attacker's face. In our studies, we analyse a more sophisticated way of creating 3D masks out of printed 2D face images. But first, we like to talk about a recent service called ThatsMyFace.com which specializes in facial reconstruction and in transforming 2D portraiture into 3D sculptures. Using this service one only has to upload 2D mugshots (one frontal and one profile) of a person and he can receive this person's constructed 3D face in his mailbox, either in a nice picture frame as a present, or on an action figure as a toy, or as a wearable life-size mask that can be easily used in a spoofing attack.

In our research, we include four types of masks from ThatsMyFace.com. The first type (M1) comes as a PDF paper-craft layout of the face in 3D and can be printed, cut, folded and glued together into a 3D face mask. The other three types of masks are made out of a hard resin composite in full 24-bit colour with a matte varnish. One is $\frac{1}{2}$ life-size (M2) while the other two are of real size (M3 and M4). In addition to these masks, we analyze two more mask types. The first one is printed from a real 3D scan of a person (M5). For the time being, the acquisition of a detailed 3D face model of a person without his permission is not very feasible since the scanning process mostly requires subject cooperation. But with the pace of advancements in 3D sensing technologies, it is still beneficial to analyse this kind of masks. The last type of mask is made of silicone that is produced using face mould (M6) and painted realistically.

3 Experiments and results

3.1 Face recognition systems

In order to test the spoofing performances of 3D face masks, an open source framework for standardized comparisons of face recognition algorithms is utilized [GWM12]. Among a variety of implemented methods in the framework, two gabor-wavelet-based algorithms are selected which do not require any training of a prior model, naturally avoiding the generalizability problem: Local Gabor binary pattern histogram sequences
Table 1: Number of samples in the utilized database for each subject (S1,S2,S3,S4) and attack type (masks: M1,M2,M3,M4,M5,M6 and photo attack: PA) together with example face images from the database: The first row is composed of real faces. Samples of different life-size masks are given in the second row. The first two are ordered from ThatsMyFace.com with and without holes at the eyes and the nostrils, and the third is directly printed from a real 3D face model. Finally in the last row, paper-cut, half-size (from ThatsMyFace.com) and silicon masks produced using face mould are presented from left to right.

<table>
<thead>
<tr>
<th></th>
<th>M1</th>
<th>M2</th>
<th>M3</th>
<th>M4</th>
<th>M5</th>
<th>M6</th>
<th>PA</th>
<th>Real</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>7</td>
<td>-</td>
<td>19</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>S2</td>
<td>12</td>
<td>32</td>
<td>-</td>
<td>25</td>
<td>-</td>
<td>-</td>
<td>20</td>
<td>17</td>
</tr>
<tr>
<td>S3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>60</td>
<td>22</td>
<td>9</td>
</tr>
<tr>
<td>S4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>16</td>
<td>-</td>
<td>21</td>
<td>25</td>
</tr>
</tbody>
</table>

(LGBP[HS]) [ZSG+05] and Gabor graphs [WFKvdM97] with a Gabor phase based similarity measure [GHW12]. Additionally, due to the robustness of Gabor features against local distortions caused by variance of illumination, expression and pose, they have been successfully and extensively applied for face recognition.

For both systems, the faces are firstly geometrically normalized using manually labelled eye positions and histogram equalization is applied. For the LGBP[HS] algorithm, facial images are convoluted with a set of 40 Gabor wavelets and Local Binary Pattern histograms [AHP04] are calculated for non-overlapping 8 × 8 pixel blocks. Finally, 64 resulting histograms are concatenated into a final feature vector and compared using the $\chi^2$ metric. In the second method, the Gabor jets collected in grid graphs are compared using both the magnitude and phase of the Gabor wavelet response. For more details, interested readers can also refer to open-source implementations of these methods in BOB [AESW+12].

3.2 Database

For this study, we constructed a small database that consists of 4 subjects. Apart from 4 gallery samples, 59 face images are collected for real attempts under controlled illumination conditions. Low similarity scores can be observed for genuine and spoofing samples under adverse conditions, clouding the judgement on the exact impact of the masks. This behaviour is undesirable when assessing the spoofing performances.

In order to perform the attacks, we obtained different types of masks for each subject. There are no types of mask which were manufactured for more than one person, except M1. For comparison reasons, we also included photo attack samples (PA). The number of samples for each type of attack and real and spoofing samples for all subjects with each type of attack are given in Table 1.
3.3 Spoofing performances

Before evaluating the 3D mask attacks, two face recognition systems are analysed and the operating point for the score threshold is decided to be at the equal error rate (EER) where false accept and false reject rates are equal\(^1\). With these settings, at the score threshold of 0.59, Gabor graphs algorithm achieves 1.64% EER. Falling a little behind, LGBPHTS algorithm reaches 5.19% at the threshold of 1.69. The photo attacks at these thresholds achieve 78.31% and 97.59% success for Gabor graphs and LGBPHTS algorithms, respectively.

The similarity scores for masks are analysed separately. For M1, the score distributions for genuine, impostor and attacker scores are obtained and analyzed. At the EER threshold, this type of masks reach 5.26% attack success rate. This is expectedly far behind the photo attacks. The paper-craft masks have distorted shapes and noisy textures. The edges of the patches are clearly visible.

If we break down the results even further and analyse the two M1 masks separately, it is observed that the mask for subject 2 performs much better (14.29%) and in fact none of the samples from subject 2 manage to penetrate. This may be due to several reasons: The first factor is illumination conditions. The planar facets on facial surface lead to specular reflections and create susceptibility to lighting. M1 for subject 2 is sampled under more "ideal" conditions. The second factor is the process of production for these masks. The 3D shape for the masks are calculated from the 2D photos of the clients. Additionally, crafting them requires manual skills. Hence, the accuracy of the outcome may not be the same for all cases. Still, it does not change the fact that the chances are very low for these type of masks to spoof a good 2D face recognition system.

Scale invariance in 2D face recognition gives a remarkable advantage to M2 masks. Thanks to their compactness, attackers can easily carry and use them with a high chance of success and without drawing too much attention. At EER threshold these masks reach 78.12% success rate in both recognition systems, drawing very close to photo attacks.

Oddly, the realistic life-size mask M3 do not meet the expectations and perform very poorly. It reaches spoofing success rates of 21.05% and 15.79% for Gabor graph and LGBPHTS systems, respectively. As is the case with M1, these masks highly depend on the input 2D images that are utilized to reconstruct the 3D shape and the quality and correctness of the final product.

M4, on the other hand, is the most successful attack among all types, reaching 100% success rate for both systems. In fact, the holes at the eyes are expected to be detrimental, since the attackers are not always able to align their eyes correctly. But, this argument is much strongly disproved by the obtained results. Instead, they made it clear that the spoofing success for 3D life-size masks is highly contingent on degree of accuracy in mask manufacture. In addition, again there is a substantial difference between the lighting conditions of M3 and M4. M4 is photographed in better conditions with respect to M3 and this may also have impact on the performances.

M5 is different from the previous masks in the sense that it has the exact shape of the client

\(^1\)Longer version of this paper with additional figures and illustrations can be found at http://publications.idiap.ch/index.php/publications/show/2658
face, because it is directly printed from a real face scan using a 3D printer. In the case of a 3D face recognition system, it would be expected to achieve high attack success rates. However, in 2D face recognition, the texture is of higher importance. The printed mask is the first face model in the training set of the FRGC database [PFS+05]. The texture on the printed mask did not turn out to be very realistic. Still, the attack success rates are better than M3; 25.00% and 31.25% for Gabor graphs and LGBP,HS algorithms, respectively.

Among all masks, M6 is the only one that has been studied to develop counter measures in the literature [KNYY09, ZDLL11]. Unfortunately, it is also the least successful. None of the spoofing attempts with the silicone mask manages to break into the two authentication systems. In score distributions, the silicon mask is observed to behave similarly to impostor (zero-effort) trials. The silicone masks may be used for evasion where the purpose is to hide ones identity. But this does not hold for spoofing, because in order to have a silicone mask of a valid user, the attacker needs to know the 3D shape of the user’s face. Even if he did so, it requires painting skills to create a realistic texture on the mask surface. Briefly, for the time being, it is highly impractical to forge a 3D silicone mask attack and gain illegitimate access through 2D face recognition systems.

### 3.4 Conclusion

Even though the appearance of affordable customer depth cameras in the market facilitates face-ness detection and anti-spoofing for photo and video replay attacks, this advantage is counteracted with strong advancement trends in facial mask manufacture. As the thriving 3D mask production industry brings in new challenges for counter measure studies, the emphasis on this issue is still very limited. Two existing works [KNYY09, ZDLL11] try to differentiate human skin and silicon mask using spectral lighting, however neither of the papers contains an analysis on spoofing performances of the masks under discussion. The scarcity of works in this area is mainly due to the difficulty and costliness of constructing an extensive 3D mask database. In this paper, we present a preliminary study on different types of masks available in the market, paving the way for such a database and eventually more detailed exploration of the subject.

According the obtained results, spoofing attacks with silicone masks for which counter measures were proposed previously in [KNYY09, ZDLL11, KD13] are found to be the most ineffective and impractical. This outcome is solely enough to justify our purpose in this analytic study. The most successful attacks are performed by real life-size masks from ThatsMyFace.com with holes at the eyes and nostril (M4). Reaching 100% attack success rate, they are proven to impose the highest threat on 2D face recognition systems.

Based on these findings, as future work, we plan to collect a database of 3D spoofing attacks using masks of type M4. Further investigation is needed using a larger database with higher number of subjects and more real/fake access attempts. Additionally, in longer term, we aim to develop counter measures for this type of attacks. It would also be interesting to assess how these masks perform in front of depth sensors and investigate if it is possible to make use of the shape data in anti-spoofing. For this purpose, we aim to
employ both 2D and 3D devices in creating our 3D spoofing attack database.

Acknowledgements
The authors would like to express their thanks to the FP7 European TABULA RASA Project (257289) for its financial support and to Ralph Breithaupt for his collaboration.

References


The influence of dataset quality on the results of behavioral biometric experiments

Pawel Kasprowski
Institute of Informatics
Silesian University of Technology
44-100 Gliwice, Poland
kasprowski@polsl.pl

Ioannis Rigas
Physics Department
University of Patras
26504 Rio, Patras, Greece
rigas@upatras.gr

Abstract: This paper explores some aspects that are involved during the construction of reliable benchmark sample databases for novel behavioral biometric identification methods, such as the data quality, the recording patterns and the post processing procedures that may be applied on the data. A large collection of eye movement samples was employed as a test case. It was recorded under a variety of settings and processed with a number of different approaches. Our analysis reveals that there are specific features during the construction of a database that may significantly influence the final identification performance. It also leads on the establishment of some guidelines, which can be generalized on other behavioral biometric methods, regarding the factors that should be taken into consideration during the creation, the description and the processing of a database of biometric samples.

1 Introduction

Whenever a novel biometric identification method is proposed, two of the most important factors that need to be evaluated are its distinctiveness and repeatability. Distinctiveness refers to the extent of difference in measurements among different people. Repeatability measures the degree of similarity in biometric samples taken from the same person. The latter is especially problematic in the case of biometric identification based on behavioral cues. Human behavior is susceptible to endogenous and exogenous factors that affect its characteristics over time. For example, the way the persons behave is influenced by their mood - people behave differently when they are sad, angry or just tired. Such type of dependencies should be taken into consideration during the experiments for the construction of a biometric database, since they may affect considerably both the performance and the consistency of a behavioral biometric identification system.

One of the main problems encountered in novel biometrics is the absence of reference databases on which the methods can be evaluated, contrary to the case of well-established methods [Ca11][Or03]. Hence, the prospective researchers need to set their own experiments and create their own database of measurements, which is often a
difficult and time consuming task. Only after creating the database of measurements, the researchers are able to perform their identification or authorization algorithms and publish their performance rates. As the published results are dependent on the dataset collected, it is very important for the samples database to have been appropriately constructed. Two main distinctive features that need to be considered during the construction of such a database are: the overall number of samples and the number of different individuals that participated in the experiments. A large number of individuals gives the opportunity to properly evaluate the distinctiveness of a method. Similarly, a large number of samples per individual helps in the evaluation of repeatability. In most papers these two factors are used as the most important and - unfortunately in many cases - the only metrics for measuring the dataset quality. As we will show, there are several other aspects worth to be inspected, since their significance to the final results is not negligible.

The quality of measurements is reportedly one of the major problems of biometric research [Wa02][Ga05]. There are general recommendations on how to control factors that influence performance i.e. volunteers selection or test size. It is also well known that noisy, low quality samples influence the overall result [GT07][Ja00]. It is worth mentioning that quality of samples (named also ‘corpora quality’ [Wa02]) is something different than quality of the database (meta data quality). There are several papers that analyse quality of dataset i.e. registration procedures, storage procedures, removing odd samples procedures [MW02][HK06]. On the contrast, the impact of the quality of measurements for the case of behavioral biometrics is a less researched domain [De95]. As behavioral biometric methods measure person’s activity in a given time, the time delay between successive subject measurements also appears as an important property during the dataset construction. The problem is somehow similar to a ‘template aging’ effect, well known in biometrics but it concerns much shorter periods of time.

In our paper, we emphasize that commonly published properties such as the number of samples and the number of participants are not enough for a complete assessment of the quality of a biometric database. We propose other possible factors that should also be taken into consideration (especially time related metrics) and we demonstrate the impact of these factors on the overall identification results. To the best of our knowledge, there are no generally accepted guidelines of dataset preparation for behavioral biometric identification experiments so this paper aspires to present the influence of certain parameters during the construction of samples datasets and serve as a useful tool during the setup of novel experiments (focused, but not limited to eye movement biometrics) in the future.

2 Eye movements biometrics

The idea of using eye movements for human identification is almost ten years old, with several publications showing the promising perspectives of the field [Be05][KO04]. However, it is still on a very early research stage and may be considered as ‘novel’.
Collection of eye movement samples is a relatively challenging task, since it requires specialized devices (eye trackers) and carefully planned experiments to ensure the correct recording of the signals. Until recently, due to lack of publicly available datasets, every research team that developed a new scheme for the extraction of biometric features from the eye movements needed to conduct its own experiment and construct its own dataset. In this way though, a comparison among different methods that analyze eye movements for biometric purposes was very difficult to occur. The First Eye Movement Verification and Identification Competition (EMVIC) organized in 2012 as an official BTAS conference competition [KKK12] was the first to establish a common environment for the comparison of different approaches for the identification of individuals on the basis of their eye movements. The organizers prepared four different datasets of eye movements collected with different stimuli patterns and different eye trackers. There were about 50 competitors with over 500 separate submissions. An oddity that arouse during the EMVIC was that the identification results were inconsistent for the different prepared datasets. For datasets A and B the identification rates were better than 90% whereas for datasets C and D the best results approximated 60%. A question arose as to the reasons of such differences. In [KKK12] authors suggested several possible reasons. The first one was a number of recordings per person, because there were only 4 recordings per individual in datasets C and D and up to 100 recordings per individual in datasets A and B. The second reason was that data from both eyes were available for datasets A and B and data for one eye only were available for datasets C and D. The last two reasons were calibration and recording patterns impacts. It was pointed out that data in datasets A and B were not calibrated and most data were gathered using very close time proximity. Regarding the first reason, it is quite obvious that a bigger pool of samples per individual allows for a more complete evaluation of the similarities in the characteristics extracted from the experimental subjects. Detailed analysis regarding the second reason was given in [REF12][Ng12]. In our study we focus on the two latter reasons investigating how calibration, data quality and time proximity between samples may influence the final results.

3 Data preparation

The main objective during our experiments was to examine the influence of two different aspects on the final classification results: a) preprocessing of raw eye movements signals in order to improve samples quality, and b) temporal proximity of the recorded samples during the experiments. For this reason we have constructed several different subsets from a dataset of eye movements and used them with different algorithms in order to perform biometrical identification.

3.1 Dataset

The dataset on which we conducted our research was the publicly available dataset B from the EMVIC 2012 originally published in [BO05]. It consisted of 4168 samples taken from 75 subjects. The data were collected within a period of 9 consecutive months. The reason that led on the adoption of dataset B was the relatively large amount of data
it provided. This made possible the extraction of different subsets of samples from the whole dataset and the application of different methodologies on these subsets in order to evaluate the impact of the inspected parameters on the identification performances.

3.2 Preprocessing of raw eye movements signals

During the recording phase of dataset B no calibration procedure occurred, so we decided to apply a post-calibration scheme and in the sequence a cleaning algorithm on the samples and inspect the impact of these data processing procedures on the final performance. Three groups of samples were created: raw samples dataset, which consisted of the samples in the exact form that they were recorded, i.e. without any calibration or cleaning. An analysis of the samples revealed that the quality of the data was relatively low and the amplitudes of signals may differ significantly among samples. Due to the nature of the stimulus it was easy to identify required fixation locations (RFLs) in the samples in every moment of registration [HH02]. This allowed us to perform a post-calibration procedure for the recorded raw signals. The post-calibration of a sample was implemented using the algorithm available in [Ka04]. The samples calibrated using the algorithm described above formed a calibrated samples dataset. Although for many samples the calibration was successful, there were cases where the resulting signal after the calibration was very noisy. To remove such bad conditioned samples a rejection threshold was added to the calibration algorithm. Signal levels calculated for different RFLs were compared and if the difference for two levels was lower than the rejection threshold, the sample was removed. The samples that passed that rejection test were used for the formation of the third group of samples, the cleaned samples dataset.

3.3 Temporal proximity of the recorded samples

The second parameter that was investigated regards the degree to which the time proximity among the recorded samples may influence the identification rates. Four different groups of datasets were constructed, each one corresponding to a different time interval between sample recordings. In order to perform a comprehensive evaluation procedure every dataset was created containing the same number of samples and the same distribution of subject identifiers, according to algorithm described in [Ka13]. The datasets denominated as ‘no interval’ consisted of samples with no minimal time separation condition. Since the samples in dataset B were mostly taken sequentially (from 1 to 15 samples per experiment) these datasets consisted mainly of samples taken at less than one minute’s intervals. Datasets denominated as ‘10 min’ were constructed with a minimal interval of 10 minutes between two samples taken from the same person. Similarly, for ‘1 h’ datasets the minimal interval between two samples was one hour and for ‘1 d’ datasets it was one day. A total of twelve separate datasets were created (four datasets for each of three groups of samples). Every dataset consisted of 275 samples taken from 38 individuals (38 classes), offering a sufficient amount of data for the evaluation of the inspected parameters significance on the identification performances of the employed classification methods.
4 Results

The datasets that were created with the preparation process described in Section 3 were employed in an ensemble of classification experiments. In our research we have used four different methods in order to classify the collected data and perform biometric identification via the eye movements. The first two methods are a spectral processing scheme (MEL) introduced in [Ng12] and a graph based approach (GRAPH) [REF12]. Both methodologies were developed specifically for the extraction of biometric traits from eye movement data and were successfully used during the EMVIC 2012. On the other hand we employed two universal methods, the J48 (Java implementation of C4.5) and the Random Forest (RF) algorithms, that are generally known for their efficiency in classification tasks. In this section the identification results are presented and conclusions are drawn with regard to the influence of the investigated parameters on the biometric performance of the benchmark methods.

4.1 Influence of data quality

The first factor that will be analyzed concerns the influence of the calibration and the cleaning procedure on the identification performances. The averaged results for the different methods that were used may be observed in Figure 1. For the three processing scenarios (raw, calibrated and cleaned) the rank-1 accuracy - i.e. number of correctly identified samples to the overall number of samples - is demonstrated for every method.

![Figure 1: Average rank-1 identification accuracy for every method for the three types of samples processing (raw, calibrated and cleaned).](image)

A close inspection of the identification rates shows that the *calibration* procedure significantly worsens results for the GRAPH method. The results are also worse (but not significantly) for MEL method, which may be surprising since this method works in the frequency domain. *Calibration* does not affect classic methods J48 and RF. In fact, in most cases results for calibrated data are even better although the difference is not significant. A possible explanation of the observed results may be the following: Samples in *raw* format have different amplitudes. Those amplitudes may somehow
enclose individually distinctive information. For instance it may be connected with the shape of the face or the way a person wears eye tracking equipment.

Regarding now the data cleaning procedure, we can see that it has no significant effect on MEL and J48 methods. It improves results for RF but the improvement is not significant (p>0.05). Yet, it significantly improves results for GRAPH method. This leads to the conclusion that the framework in which the data are processed by the latter method makes it more sensitive to low quality data.

4.2 Influence of time intervals

The other analyzed parameter during our experiments regarded the influence of time intervals between samples recording on the identification performance. In Table 1 we can observe the effect of time-interval between the eye movement recordings on the performance of each method, for the raw samples (i.e. without calibration and cleaning).

<table>
<thead>
<tr>
<th>interval</th>
<th>MEL</th>
<th>GRAPH</th>
<th>J48</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>86.2</td>
<td>76.4</td>
<td>26.5</td>
<td>37.8</td>
</tr>
<tr>
<td>10 min</td>
<td>66.9</td>
<td>41.2</td>
<td>17.1</td>
<td>24.4</td>
</tr>
<tr>
<td>1 h</td>
<td>49.8</td>
<td>37.1</td>
<td>12.4</td>
<td>16.7</td>
</tr>
<tr>
<td>1 d</td>
<td>48.4</td>
<td>33.0</td>
<td>10.5</td>
<td>12.7</td>
</tr>
</tbody>
</table>

As it can be observed, time interval between samples has a significant impact on the final results. For all methods, a longer interval gradually worsens the identification rates. The difference is the most significant during the transition from no-interval to 10 minutes interval. The results for datasets from the second group (calibrated samples) are presented in Table 2. A similar behavior between time-interval and accuracy exists in this case too.

<table>
<thead>
<tr>
<th>interval</th>
<th>MEL</th>
<th>GRAPH</th>
<th>J48</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>73.6</td>
<td>46.3</td>
<td>26.7</td>
<td>37.7</td>
</tr>
<tr>
<td>10 min</td>
<td>60.1</td>
<td>28.0</td>
<td>20.5</td>
<td>26.7</td>
</tr>
<tr>
<td>1 h</td>
<td>45.1</td>
<td>20.6</td>
<td>14.7</td>
<td>17.6</td>
</tr>
<tr>
<td>1 d</td>
<td>40.3</td>
<td>17.0</td>
<td>13.2</td>
<td>16.1</td>
</tr>
</tbody>
</table>

Samples in the third group were calibrated and cleaned to remove outliers - i.e. samples with low quality. The results for datasets from the third group are presented in Table 3. We may observe the same negative correlation between time-interval and accuracy.

<table>
<thead>
<tr>
<th>interval</th>
<th>MEL</th>
<th>GRAPH</th>
<th>J48</th>
<th>RF</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>74.6</td>
<td>53.8</td>
<td>24.6</td>
<td>41.0</td>
</tr>
<tr>
<td>10 min</td>
<td>62.9</td>
<td>39.6</td>
<td>19.1</td>
<td>30.1</td>
</tr>
<tr>
<td>1 h</td>
<td>49.2</td>
<td>31.4</td>
<td>16.8</td>
<td>27.0</td>
</tr>
<tr>
<td>1 d</td>
<td>40.6</td>
<td>26.0</td>
<td>17.2</td>
<td>20.7</td>
</tr>
</tbody>
</table>

The results presented above lead to the conclusion that the time recording pattern (that is time between subsequent samples) seems to have a great impact on the identification
results. The impact is more significant for short intervals (i.e. between no interval and 10 minutes interval) and is not so important (yet visible) for longer intervals (i.e. between 1 hour and 1 day). Consequently, the time interval during the recording of the samples is a factor that should be always taken into consideration during the inspection of the identification rates of any behavioral biometric identification method, in order to have a more complete view regarding the operational framework of the method. It should also be always taken into account that several factors as the participants’ attitude, mood or physical condition may have an influence on the measurements.

5 Conclusions

In our experiments we have tested two very important parameters that affect the classification accuracy in behavioral biometrics methods, the quality of the samples and the time interval between consequent enrollments for every subject. Regarding the impact of data quality, it was observed that when the samples are employed in raw format - without any calibration and cleaning - it is probable that some biases may arise which can influence the results. Such phenomena may artificially change the resulting rates, so they should be considered during the setting of an experiment, or during the samples processing procedure. Another important factor that should always be taken into consideration is the time interval between the samples recordings. Our recommendation is that data taken in short term series may be used for classification experiments only when samples from the same series are not mixed in both training and testing sets. Unlike in most physiological biometrics, behavioral biometrics experiments are not mutually independent and we observed that the dependency between samples is inversely proportional to the time interval between the samples. In our opinion this dependency deserves more investigation in future.

Finally, under the light of the presented experimental findings we offer a list of suggestions regarding possible meta-information that may be considered during reporting research experiments results that concern behavioral biometrics: General information about the data: (1) number of samples, (2) number of subjects and additional information: (3) distribution of number of samples per subject, (4) minimal time between subsequent samples, (5) minimal time between subsequent samples of one subject. If the dataset is designed to publication, it is preferred to publish it without any preprocessing. If only classification results are published, it is important to add information about: (6) algorithm used for data calibration (if applied), (7) algorithm used to remove samples with low quality.

Our research shows that all parameters mentioned above may have a strong impact on classification results and therefore should not be omitted. As all experiments during the research were performed only for one modality - eye movements – we believe that our findings could be generalized to all behavioral methods. Naturally, the latter statement deserves further study.

Acknowledgements. Authors would like to thank Nguyen Viet Cuong for his contribution in performing Mel-Frequency classification and Oleg Komogortsev for his initial inspiration.
References


Behavioral biometrics for DARPA's active authentication program

Ingo Deutschmann, Johan Lindholm

Research & Development
Behaviometrics AB
Aurorum Science Park 8
SE-977 75 Lulea
ideutschmann@behaviosec.com, jlindholm@behaviosec.com

Abstract: The aim of the US Defense Advance Research Project Agency’s (DARPA) Active Authentication program is the continuous authentication of users by using behavioral biometrics authentication systems, which does not depend on specific hardware or sensors. This paper presents how such a continuous authentication system would perform in an office like environment. The analysis is performed on a data set captured from 99 users over a 10 week period. Our continuous authentication system builds a behavior biometric profile of the user by observing mouse movement, keystrokes and application usage. The user is then actively matched against his profile. The goal was, as DARPA is mentioning in their Active Authentication program, "This means the system would, potentially have to falsely reject the user more than five times in a row during continuous usage over a 40 hour period to fail to meet this target. The technologies developed under this solicitation should be able to work invisibly to the user unless five false positives are reached". The results of our study indicate that the correct user can work through a regular workday without being falsely rejected, while the incorrect user would be detected within 18 seconds using keyboard or 2.4 minutes using mouse. Application process usage results show that the incorrect user would be detected in just over 1.5 minutes.

1 Introduction

Behaviometrics, or behavioral biometrics, is the identification of humans by their measurable behavior. It focuses on behavioral patterns rather than physical attributes. Behavioral biometrics utilizes the characteristics of the user’s keystoke, mouse input and application usage to create virtual fingerprints of their behavior. It can efficiently prevent intrusions on laptops or workstations by continuously verifying that it is the authorized user that is accessing the computer. The user is monitored during the whole working session to create an ongoing authentication process. When using behavioral biometrics rather than static biometrics such as fingerprint, the normal biometric approach is not ideal. One of the core strengths of behavioral biometrics is that it can continuously improve and adapt to the user. The reason that behavioral authentication is so beneficial is, that it is very non-intrusive and provides continuous protection independent of what application is being used. In this study we are using passive attacks to calculate the
accuracy. This mean that we are using different users normal behavior and cross compare them to determine the performance of the system.

3 General Continuous Authentication System

An authentication system generally consists of four major parts:

- A monitor, which collects relevant behavioral data from the underlying system. In our research we gathered mouse, keyboard data and application usage.
- A classifier that sorting and filtering the data into different categories.
- A user profile which is storing condensed timings and is continuously updated when the input is assumed to be from the correct user. The system has to perform a initial training when the behavioral profile is empty. During this phase our system will assume that it is the correct user that is interacting with the system.
- An evaluator that continuously verifying the filtered data. The evaluator uses the stored profile to run its tests on the data from the classifiers to calculate a similarity score, which is used to verify the integrity of the user.

4 Calculating FAR/FRR/EER

To calculate the FAR/FRR/EER of the behavioral system, a test group of 99 users were selected. They worked in DoD-like environment and their behavior was captured for three months by a monitor installed on every machine. We monitored applications as well as keyboard and mouse interactions.

4.1 Interactions

In order to assess the likelihood that the originally authenticated user is still in control of the keyboard, the collection of the various independent interactions from typing (press, flight, sequence), mouse movement (movement, click, drag and drop) and application usage are combined into a single similarity score.

4.1.1 Keyboard interactions

Keystroke patterns are collected by the way users type at the keyboard. We count one keystroke as one interaction.
4.1.2 Mouse interactions

Mouse dynamics can be divided into different categories of actions, such as: main action types: Mouse move, click, drag and drop and action characteristics: Angle, velocity, direct distance.

4.1.3 Application interactions

Keyboard or mouse events are tied to the application that is active when they are performed. The system uses application specific profiles to enhance performance and accuracy. We analyzed information such as time open, apps open in parallel, how they are opened and closed, number of open views, memory and CPU consumption. We tested our data without taking into account holidays and daily fluctuations of use.

4.2 Training

A profile starts empty and the system has to train it to learn the behavior of the user. At the beginning an early stage it is difficult to differentiate between different users, so our system initially assumes that it is the correct user handling the computer. As users evolve their behavior over time the profile is continuously updated when the system is sure, that the right user is using it.

5 Trust Model

The trust model is intended to enhance or replace such simple detection mechanisms as score/threshold models, where a score is matched against a threshold for verification. One issue with the score/threshold model is that small deviations in behavior may cause false rejects; i.e. the user’s similarity score dips below the current threshold and is considered to be incorrect. The idea behind a trust model is, that the correct user will generate better scores over time than the incorrect user, thus smoothing out small mistakes by the correct user, while still detecting the incorrect user. The trust of a user is a value calculated from a set of evaluated scores, which could be the similarity of the current user’s behavior against the stored profile, as well as the system current state. The system triggers detection if the trust reaches a certain threshold.

![Figure 5.1: Simplified Score-Trust mapping example](image)

Figure 5.1: Simplified Score-Trust mapping example
The similarity score is mapped against the trust model by using a trust-threshold. If the user is above the threshold the trust will increase, and if the user is below the threshold, the trust will decrease. Staying above the threshold improves the trust level to its maximum; the higher the user is above the threshold, the quicker the trust reaches its maximum.

The detection speed between users may vary because of:

1. How much the behavior differs between the active user and the profile. Larger difference gives faster detection time.

2. How much trust the previous user has achieved. If the previous user has worked him up to be fully trusted, it will take longer time for the incorrect user to reach the not trusted level. Note that the trust will decay over time; so that a user that leaves his computer will have his trust set back to the start value (see 5.1).

Abnormal actions can trigger a decrease in trust, such as using key combinations that have never been used before or repeatedly providing immeasurable samples (compare to pushing your forehead against a fingerprint scanner).

### 5.1 BehavioSec model

The trust in the BehavioSec model is represented as value between 0 and 100, where higher value means higher trust. The initial trust value is configurable and sets how aggressive the system should be from start. Trust is updated using equation 5.1, with input from the different continuous tests (keyboard, mouse, application usage etc.). A detection is made when the trust decreases below the trust threshold.

\[
C := \begin{cases} 
\text{startup value} \\
\max\left(\left(C - \frac{T - P}{Z'}\right), 0\right), P < T \\
\min\left(\left(C + \frac{P - T}{1 - T}\right), 100\right), P \geq T
\end{cases} \tag{5.1}
\]

- \( C = \) Trust of the user
- \( T = \) Similarity score threshold between trusting and not trusting the user based on one single test
- \( P = \) Similarity Score from the last test of the users input against the template
- \( Z' = Z/100 \)
Equation (5.2) calculates the increase in C which is proportional to the difference between |P-T| by a variable factor Z. The decrease function works in the same way as the increase function, when P is strictly less than T. The Z variable can be set independently for the increase (5.2) and decrease functions. The value of Z sets the slope of the curve for how much a similarity score alters the trust C. The model increases the trust C if the score P is equal or greater than the threshold T. The trust level can never increase beyond 100. Another type of trust model, by Patrick Bours, can be found in [Ha01].

5.2 A practical example of trust

The trust model can be compared against a score-threshold model to display the differences between the two. We have used the mentioned dataset of behavior on two different models of detection, score-threshold and trust, to create a comparison between the two. Note that the trust models parameters in these tests are not tweaked to be the best possible, one could for example allow more false rejects to detect the incorrect user faster. We have selected users that are quite similar in score on purpose, to demonstrate how the trust model can take small variances in score into account. The data from figure 5.2 and figure 5.3 are taken from real user input, where the upper diagram is the similarity score and the lower is the trust, calculated from the score. The correct user, overtime, generates a much better trust value than the incorrect user. Note that the trust value is reset after each detection (value goes to zero), that is why the incorrect user trust value seems to oscillate. One can see why a score-threshold has the drawback of more false rejects. Connecting the score to our trust model, would produce only one false rejection for the right user. While it takes a bit longer to detect an intruder, we believe it is within reasonable time. Also note that these values and graphs do not reflect the accuracy of the solutions, they are merely here to show the correlation between score and trust, and how a score-threshold system is different from a trust model system.

Figure 5.2: Correct user score and trust for three months
6 Methodology

6.1 The test bed

The test bed was consisting of 99 users, which have been using the computer for regular work for at least 20 hours per week, for a total of 10 weeks on standard Windows 7 desktops with office suite, web browser, antivirus etc. The data was collected by software, which tracked various events that could be interesting regarding their behavior. The aim was that the data collector should be as invisible as possible for the test subjects, even though the users know it is collecting the data, they should be carefree so that they can keep working as usual without any attention. The data had been collected bi-weekly. The users where asked not to share their working space with other people to avoid a “contamination” of the data. The data collection could be stopped by the users at any time, if they shared their working space or wanted to obtain privacy.

6.2 Analysis methodology

The collected behaviors (behavioral data files) had been fed into the system, which built behavioral profiles that were used to compare against themselves and others (posing as passive attackers). The scores for the correct user where generated as the behavioral profile built up, meaning that the results reflects how the system would work in a continuous setting. As the attacking simulations took a long time, we had to restrict the amount of attackers. To calculate how long an attacker could use a user’s workstation, five different users where randomly selected to act as passive attackers (comparing the attacking users’ normal behavior). The statistical data, which can be viewed as a list of raw scores for each comparison, was then analyzed in our FAR/FRR tool that calculated thresholds, active time, training times etc. If a parameter adjustment allowed the correct
user to use the computer for longer time, without significantly increasing the detection time for the incorrect user, that setting was be picked instead.

7 Performance Metrics

We have proposed a model for determining the performance of a continuous authentication system by using the metric of the total number of detections after a given amount of interactions. Detections are places in the interaction stream, where the system decides that the latest subset of interactions is not a behavioral match with the profile. It is both hard and unpractical, to use time as metric for determining the performance of a continuous authentication system. We instead propose the use of number of interactions as the metrical time factor.

7.1 Proposed usage of model

The model we propose can be used on an offline version of the system, where one build profiles of users from stored logs, simulating the actual continuous authentication system, to compare users against themselves and others. We have U users with a log-file L associated with each user, where the log file contains all the users’ interactions over a period of time. The simulation builds a profile P for each user; using a subset of L (we took the first 40% of the log). A research of the needed training times will be discussed in a later paper. The rest of the log file was used to test the user against himself, as well as updating the profile. The system notes how many detections (if any) are made. The next step was to test each profile against the other logs, also noting how many detections that are made. To test how good a system performs, we measured the amount of interactions for each input that is required to catch an intruder on average, while not rejecting the right user. Based on the results, a global optimal threshold was calculated.

8 Results

The detection times that are presented in table 8.1 shows the median for each modality. We differentiated between 'active' time and wall clock time. As a user is not pushing buttons on the keyboard every second of the day, nor moving the mouse, we counted only the active time. We chose a timeout period of 10 seconds. So after this timeout period the active time counter stopped counting. The table 8.1 shows the number of interactions the correct user can perform before rejected, and an incorrect user can perform before getting detected and how many interactions the users where performing during a typical work day for each modality. Please note, that the interactions are counted only in the 'active' time window.
<table>
<thead>
<tr>
<th>Modality</th>
<th>Typical day *</th>
<th>Incorrect User</th>
<th>Detection times ***</th>
<th>Correct User</th>
<th>False rejection times ***</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interactions</td>
<td>Minutes</td>
<td>Interactions</td>
<td>Minutes</td>
<td></td>
</tr>
<tr>
<td>Keyboard**</td>
<td>400-1300</td>
<td>15-50</td>
<td>0.3</td>
<td>10000-32000</td>
<td>138</td>
</tr>
<tr>
<td>Mouse</td>
<td>743</td>
<td>59</td>
<td>2.4</td>
<td>1596</td>
<td>66</td>
</tr>
<tr>
<td>Application</td>
<td>7690</td>
<td>67</td>
<td>1.5</td>
<td>17181</td>
<td>384</td>
</tr>
</tbody>
</table>

Table 8.1: Performance Results for continuous usage

* The amount of interactions for the modality that the users on our test bed performed during a regular work day.

** for keyboard one interaction is one keystroke (typing “test” counts for 4 interactions)

*** The time is calculated from continuous active usage (no pauses included).

9 Discussion & Conclusions

In this paper we have explored the performance of a continuous authentication system, as a second factor for a DoD personnel. The correct user can work through at least a regular workday without being falsely rejected, the incorrect user would be detected within 18 seconds using keyboard (15-50 keystrokes) or 2.4 minutes using mouse (66 interactions). Application process usage results show that the incorrect user would be detected in 1.5 minutes. What is missing in our study is the test with real attackers. We think that an attacker would switch much quicker between applications as well as starting a lot of applications after another, so would be more easy to recognize. A further study will target a continuous authentication system on mobile devices under DARPA-BAA 13-16.

References


[CS01] Crowd Sourced. Biometrics

Abstract: Temperature distribution on the eyes of sober and drunk persons is studied and preliminary results are given regarding temperature changes. It is observed using simple image processing algorithms and histogram modification techniques that the temperature difference between the sclera and the iris increases when somebody consumes alcohol. Iris becomes darker, which means that its temperature decreases compared to sclera temperature.

1 Introduction

Biometrics is a research area with numerous publications in recognition or identification of persons. Applications are met in medicine, financial transactions, face identification and mainly in security issues. Research is carried out [DeM11][Ch10][Be10][JBP99] in several biometric problems such as face and fingerprint recognition, facial expression classification and iris identification (Fig. 1.1) with high rate of success.

Due to the availability of low cost cameras in the visible spectrum and the fact that face recognition is one of the primary activities of the human visual system, research in face recognition [Zh03] has been biased towards the visible spectrum. However, machine recognition using visible light is complicated due to the fact that the acquired images change with the conditions of illumination.

Recently, emphasis has been given to acquiring information from faces in thermal infrared spectrum [SS02][Bu07][ZG05]. The main reason for this is that the temperature of the face depends mainly on the physiological condition of the person [KIW06]. The human face being in a mean temperature around 300° Kelvin, radiates, according to the Wien Law, as a perfect black body [DB08], with maximum at 10 μm wavelength. Thus, this region of electromagnetic spectrum (7 μm – 13 μm) is the most appropriate to acquire information from the face.

Drunkenness is a challenging physiological condition to be investigated using infrared imagery. This is because arteries and vessels on the face of a drunk person, increase
activity with the consumption of alcohol. However, most of the publications refer only to automotive anti-drunk driving systems, which utilize electrical signals from the heart or brain [Wu09].

Figure 1.1 The parts of the eye. Sclera is surrounding iris which is actually a muscle controlled part of the eye to adjust the size of the pupil. Sclera lies on a net of blood vessels.

Iris identification has been extensively used in the past, and research results are on the market in robust and high performance commercial products. In [Pi11] a unified framework based on random projections and sparse representations, is presented in relation to iris biometrics. The work in [Mi06] presents an implementation of iris recognition algorithm using phase-based image matching — an image matching technique using phase components in 2D Discrete Fourier Transforms (DFTs) of given images. The authors in [Da07] present four advances in iris recognition: 1) more disciplined methods for detecting and faithfully modeling the iris inner and outer boundaries with active contours, leading to more flexible embedded coordinate systems; 2) Fourier-based methods for solving problems in iris trigonometry and projective geometry; 3) statistical inference methods for detecting and excluding eyelashes; and 4) exploration of score normalizations, depending on the amount of iris data that is available in images and the required scale of database search. In [Da04] algorithms suitable for recognizing persons by their iris patterns have been tested in many field and laboratory trials, producing no false matches in several million comparison tests. Finally, in [Ar12] matching iris images captured before and after alcohol consumption is investigated. Due to alcohol consumption, the pupil dilates/constricts which causes deformation in iris pattern, possibly affecting iris recognition performance. Near infrared technology is used for image capturing.

In this paper a preliminary study is provided regarding the temperature distribution on the whole eye before and after alcohol consumption. Experimentation is based on thermal infrared acquisitions in the range of 7-13 microns, by means of a passive thermal infrared camera. Half a litre of red wine was given to 40 volunteers in order to create our data base. It is observed using simple image processing algorithms and histogram modification
techniques that the temperature difference between the sclera and the iris increases when somebody consumes alcohol. Iris becomes darker, which means that its temperature compared to sclera temperature decreases.

The work layout is as follows. In section 2 is explained the way, the infrared images used in the experiment were acquired. In section 3 are described the histogram modification algorithms used and the experimental results obtained. The conclusions are drawn in section 4.

### 2 Infrared Data Used

The infrared images used in this work were acquired by means of the Thermo Vision Micron/A10 Model infrared camera (18 mm, f/1.6) of FLIR Company. The operating wavelengths are from 7.5 to 13.0 microns, which means that the obtained information is in the thermal infrared where we have the maximum of the Wien curve at 9.5 microns for 300o Kelvin. Actually, the human skin emits electromagnetic radiation as an almost black body in this exact region of wavelengths [DB08].

In this experimental procedure forty-one people were involved, 31 males and 10 females. Each person consumed four glasses of red wine, 120 ml each (13% vol.), in one hour - time period (total of 480 ml wine i.e. 62.4 ml alcohol). In our experiment, wine was used for alcohol consumption compared to the experimental procedures in [KA12] and [KA11]. In those cases a number of twenty people had been employed and all persons participated in that experiment consumed beer. The quantity of beer required in order to consume the same amount of alcohol is almost three times larger (330x4=1320 ml) compared to that of wine. This is the reason why wine was used for drunk identification.

The first acquisition of 50 frames for each specific person was obtained just before starting alcohol consumption. The second acquisition of 50 frames was obtained 30 minutes after drinking the fourth glass of wine. Similar acquisitions were obtained one hour as well as one hour and a half after finishing alcohol consumption. Thus a total of four acquisitions were obtained for each person. In each acquisition, a sequence of 50 frames was acquired from each person with a sampling period of 100 msec between the frames. The mean value of the 50 frames of each acquisition was evaluated using MATLAB. This mean value of the eyes of a specific drunk person is demonstrated in Figure 2.1. The resolution of the infrared images is 128x160 pixels. The images were acquired in such a way so that all the image is occupied by the eyes of the specific person.
Figure 2.1 The mean value of 50 frames for a sober person (left) and a drunk person (right). No preprocessing has been applied on the images. It is evident for this person that the sclera becomes hotter compared to the iris when the person consumes alcohol.

The term “drunk” is attributed to the person that has consumed four glasses of red wine or a total of 62.4 ml of alcohol. We consider that the number of forty people employed for this experiment is the maximum we could gather, so that infrared images in both cases, sober and drunk, are acquired. Furthermore, we consider this quantity of alcohol the maximum that our researchers could consume and participate in our experiment. No blood tests were conducted in order the blood alcohol content be known. According to various tests available to the web, only three glasses of wine are enough for any person to go beyond the limit of 0.5 gr/(litre of blood), a quantity set for secure driving [Wi13].

We noticed that, with the same quality of alcohol the participants are affected differently. This was realized by the measurements carried out by the police using an alcoholmeter. With the quality of 62.4 ml alcohol given to all persons the breath alcohol content was between 0.25 and 0.9 mg/lit. It was found that this was the maximum concentration and was reached half an hour after the consumption of the last glass of wine. This is why in our experimental procedure we used only the corresponding acquisition as far as the data regarding the drunk persons are concerned. After that, the breath alcohol content started to decrease. The females were affected more than the males. The heaviest participants were also affected less that the thinner ones.

All persons participated in the experiment were aware about the risk they were undertaking. However, all of them were healthy without any problem that could bring them in difficult position. All persons accepted their personal data to be available in the Internet for use by the scientific community. In the specific data base all data such as age, weight and sex are recorded.

Finally, it is worth mentioning that the people employed in our experiment were calm and in normal physical and psychological condition during the experiment. No illness, no psychological stress or any kind of body exercises were recorded for any one of the participants. They were asked to be present in the room of the experiment half an hour earlier and to keep calm till the first acquisition of frames.

During the acquisition procedure, the temperature and the lights in the room were kept unchanged. Actually, a very dim light was available from a neighboring room for the researchers to be able to work. This light did not affect the operation of the infrared camera or the acquired infrared images. The distance of each face from the camera was around 30 cm and was kept constant from acquisition to acquisition. This results in a face which occupies the whole area of the frame and simultaneously it gives the images of the same person the capability of been easily compared.

3 Experimental Results

Among the forty persons participated in our experiment, it was evident for 28 of them that the sclera becomes hotter than the iris when the person consumes alcohol. In all these cases, in the original thermal image of the eye corresponding to the sober person the sclera
and the iris are almost of the same temperature appearing with almost the same gray level as shown on the left image of Figure 3.1. When these persons consumed alcohol, the sclera becomes hotter and brighter compared to the iris, as shown in the right image of Figure 3.1 (images are shown in their original form without any kind of preprocessing).

Figure 3.1 Thermal Infrared images of the same eye of a person sober (left) and drunk (right) (person 5). The images are shown in their original form without any kind of preprocessing. The sclera is evidently hotter (brighter) than the iris for the drunk person.

In all these cases the iris was darker than the sclera for the drunk person or this difference simply became prominent by means of a histogram equalization algorithm. Such cases are shown in Figure 3.2 (right images).

Moreover, another 8 persons presented significant temperature difference between the sclera and the iris for the drunk person using other histogram modification algorithms. These cases are depicted in Figures 3.3, 3.4 and 3.5 respectively. Particularly, in Figure 3.3 is shown the original images of the eye of person 21, for the sober and drunk cases respectively. This images have been modified using a histogram modification algorithm which clips all values lower than 0.5 and higher than 0.75, and stretches the rest to occupy the whole histogram range (MATLAB imadjust ([0.5 0.75],[0 1])).

Figure 3.2 Using a histogram equalization algorithm the eyes of two different persons (persons 11 and 13) present an iris with lower temperature after alcohol consumption. The algorithm was applied in the thermal images of the eye for the sober persons (left) as well as for the corresponding images in the case of the drunk persons (right).

In Figure 3.4, the results for the same person 21 are obtained from a contrast stretching algorithm which was applied in the original images. In Figure 7, results are shown when a logarithmic transformation is applied to the original data of person 37. In all these cases the sclera becomes hotter after consuming alcohol and this is evident from all images on
the right of Figures 3.3, 3.4 and 3.5.

Figure 3.3 (a) The original images of the eye of person 21, for the sober (left) and drunk (right) cases respectively. A histogram modification algorithm which clips all values lower than 0.5 and higher than 0.75 and stretches the rest to occupy the whole histogram range (MATLAB imadjust ([0.5 0.75],[0 1])) gives the images in (b).

Figure 3.4 The results obtained from a contrast stretching algorithm which was applied in the original images of person 21.

Figure 3.5 The results obtained from the logarithmic transformation applied to the original data of person 37.

Finally, we have to mention that for 5 persons no algorithm applied to both sober and drunk data succeeded to show off increased temperature for the sclera in the case of drunk persons.
4 Conclusions

In this work preliminary experimental results are presented which describe the temperature changes on the human eyes when somebody consumes alcohol. Thermal images are used for this purpose. The basic evidence is that the iris remains in the same temperature while the sclera increases its temperature with alcohol consumption. Consequently the iris appears darker in the thermal imagery. A physical explanation is that the sclera is full of blood vessels which increase the temperature of sclera with alcohol consumption. In our experiments 36 among forty-one persons which consumed alcohol presented darker iris in their thermal imagery.

5 Acknowledgments

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

6 References


Using the conformal embedding analysis to compensate the channel effect in the i-vector based speaker verification system

Z. Boulkenafet¹, M. Bengherabi¹, O. Nouali², M. Cheriet³

Centre de Développement des Technologies Avancées (CDTA) Algeria¹
Centre de Recherche sur l’Information Scientifique et Technique (CERIST) Algeria²
Ecole de Technologie Supérieure ETS, Canada

Abstract: The I-vector approach to speaker recognition has become the prevalent paradigm over the past 2 years, showing top performance in NIST evaluations. This success is due mainly to the capability of the I-vector to capture and compress the speaker characteristics at low dimension and the subsequent channel compensation techniques that minimize channel variability. The Linear Discriminative Analysis (LDA) followed by Within-Class Covariance Normalization (WCCN) and Cosine Similarity Scoring (CSS) represents the best compromise between performance and computational complexity. In this paper, we propose to use Conformal Embedding Analysis (CEA); a recently proposed manifold leaning technique; to tackle the main limitations of LDA which are: the Gaussian assumption on the classes distribution, the inability to preserve the local geometric relationships of the data-space and its reliance on the Euclidean distance for characterizing the relationships between feature vectors. Experimental results on the challenging MOBIO-voice database show that CEA+WCCN outperforms LDA+WCCN for both male and female speakers at all operating points.

1 Introduction

The Gaussian Mixture Model-Universal Background Model (GMM-UBM) framework [RQD00] forms the core of the state-of-the-art speaker recognition systems. Starting from the Joint Factor Analysis (JFA), proposed by Kenny et al [KBOD04] to model jointly speaker and session components to the recently proposed total variability paradigm dubbed I-vector [DKD+11]. The experiences performed by Dehak et al [DKD+11] demonstrated that the space of channel component in the JFA method contains information which can be used to discriminate between speakers. Consequently, they proposed a new representation called Identity vector or Intermediate vector (I-vector), in which the speaker and the channel sub-spaces are represented by a single total variability space. Unlike the JFA method which models the channel variability during the training stage, I-vector takes into account
the channel compensation during the scoring stage.

In the original I-vector system [DKD+11], the channel compensation is done by using Linear Discriminant Analysis (LDA) projection followed by Within-Class Covariance Normalization (WCCN). The motivation for using LDA is to maximize the inter-speakers variability and minimize the intra-speaker variability, which is an important point in the speaker recognition. The LDA projection measures the euclidean distance between the input vectors and assumes that each class vectors have a Gaussian distribution. In general, this approach suppose that the testing data drawn from the same underlying distribution as the training data. Unfortunately, it is usually hard to guarantee this assumption. Therefore, recent studies reveal that the local features and intrinsic geometric structures [SR03] [WH00] in the input data can further improve the discriminative power. Such techniques suppose that the targeted space is a sub-manifold of low dimension embedded in a high dimensional ambient space.

The representative non-linear manifold learning such as Local Linear Embedding (LLE) [RS00], Isometric feature mapping (ISOMAP) [TSL00], Laplacian Eigenmaps (LE) [BN03], etc. aims to map data into a low dimensional manifold which preserves the local topological structure of neighbors connections. These embedding methods are designed to describe a fixed set of data and not to generalize to novel data (test data). To cover the new data some techniques suggest to use the linear approximation of the non-linear method such as Locality Preserving Projections (LPP) [HYH+05], Locally Embedded Analysis (LEA) [FH05] and Neighborhood Preserving Embedding (NPE) [HCYZ05]. In these techniques the projection from a high dimensional space to a low dimensional space is described by a transformation matrix instead of using a nonlinear mapping method defined on the training set. Hence, it is easy to apply the transformation to unseen data. However, these transformations focus on preserving data localities and similarities so the discrimination between classes can not be sufficiently guaranteed. To deal with this problem, some methods such as Local Discriminant Embedding(LDE) [CCL05] and Locality Sensitive Discriminant Analysis (LSDA) [CHZ+07] propose to use Fisher criterion and Kernel transformation to boost the discrimination power. Furthermore, and starting from the fact that the Euclidean metric is incapable of capturing the intrinsic similarities. Some recent researches have suggested to use the cosine distance for better discrimination and robustness [FLH07].

In this work, we propose the use of Conformal Embedding Analysis (CEA ) [FLH07]; a recently proposed manifold leaning technique; as an alternative of LDA. The main motivations for using this dimensionality reduction technique are: 1) The CEA has no assumption about the distribution of the input data. 2) The CEA preserves the local geometric relationships of the data-space and increase the inter-class discrimination using the cosine distance for characterizing the relationships between feature vectors projected on a unit sphere. Knowing that the cosine similarity scoring is found to be the most appropriate for I-vector, we could expect that CEA will boost the system performance.

The rest of this paper is organized as follows. In Section 2 we present the I-vector system. The CEA approach is described in Section 3. Section 4 is dedicated to experimental results while the main conclusions and possible research perspectives will be presented in section 5.

242
2 Overview of the I-vector system

Inspired by the earlier use of JFA speaker factors directly as features for SVM classification, Dehak et al [DKD+11] have recently proposed a new approach of speaker modeling called I-vector. Unlike the JFA, the I-vector method represents the speaker and the channel subspaces with a single total variability space (equation 1). This representation is motivated by the fact that the channel space of JFA system contains information which can be used to discriminate between speakers.

\[ m(s) = M + Tw(s) \]  

(1)

where \( M \) is the UBM supervector (a supervector is constructed by concatenating all the mean vectors of the GMM model), \( w \) is a latent variable with a standard normal distribution and \( T \) is a low rank variability matrix.

As the total variability space represented by \( T \) contains both speaker and channel information, the I-vector method requires additional techniques to attenuate the effect of session variability. Using the Linear Discriminant Analysis (LDA) followed by the Within Class Covariance Normalization (WCNN) in [DKD+11] has given the best performances.

2.1 Scoring

In i-vector system, we use a simple method of similarity measure, which calculates the Cosine Similarity Score (CSS) between the enrollment speaker I-vector \( (w_E) \) and the test I-vector \( (w_T) \). With the use of the LDA and WCNN projection matrices \( (A \) and \( S \) respectively) the cosine similarity will be given by:

\[
\text{score}(w_E, w_T) = \frac{(A^t w_E)^t}{\sqrt{(A^t w_E)^t S^{-1} (A^t w_E)}} S^{-1} \frac{(A^t w_T)}{\sqrt{(A^t w_T)^t S^{-1} (A^t w_T)}}
\]  

(2)

The use of the angle between the two vectors make this scoring method more robust to the channel and the session effects.

3 Conformal Embedding Analysis CEA

Given a data set of I-vectors \( X = [x_1, x_2, ..., x_n] \) with a high dimensionality and the corresponding class labels \( L = [l_1, l_2, ..., l_n] \), where each \( x_i \) belongs to a class \( l_i \). The CEA objectives are:

1. Preserve the same-class conformal affinity while keeping away the diff-class conformal affinity after the embedding.

2. If two original high dimensional i-vectors are close (large conformal affinity), then the embedded low-dimensional points are close as well.
3. The embedded sub-manifold can better reflect the class relations with respect to the labeling information.

To achieve these objectives the computation of the CEA projection matrix $P$ passes by the following steps:

- Scale each i-vector $x_i$ to be a norm-one:
  \[ x_i = \frac{x_i}{||x_i||} \]  
  \( (3) \)

- Construct intrinsic graph $G_s$ and penalty graph $G_d$ both with $n$ nodes (each node corresponds to an i-vector). For $G_s$, we only consider each pair of data $x_i$ and $x_j$ from the same class ($l_i = l_j$). An edge is constructed between nodes $i$ and $j$ if $x_j$ is among the $k_s$ largest conformal neighbors of $x_i$ and vice versa. For $G_p$, we only consider each pair of data $x_i$ and $x_j$ from different classes ($l_i \neq l_j$). An edge is constructed between nodes $i$ and $j$ if $x_j$ is among the $k_d$ largest conformal neighbors of $x_i$ and vice versa.

- Define the conformal affinity weight matrices $W_s$ and $W_d$ for $G_s$ and $G_d$, respectively. If the two nodes $i$ and $j$ are connected then the weight of the edge between $i$ and $j$ is set by: $w_{i,j} = w_1$, Otherwise, if $i$ and $j$ are not connected the weight $w_{i,j} = w_2$.

In [FLH07], the authors present three types of weight: Balanced Rigid Weights, Unbalanced Soft Weights and Balanced Soft Weights. The last one is used in our study and it is described by:

\[ w_{i,j} = \begin{cases} 
  \exp\left(\frac{(\cos(\theta_{(i,j)}) - 1)}{t}\right) & \text{if nodes } i \text{ and } j \text{ are connected} \\
  0 & \text{Otherwise}
\end{cases} \]  

(4)

where $t$ is a constant and $\theta_{(i,j)}$ is the angle between the vectors $i$ and $j$

- Compute the CEA projection matrix $P = [p_1, p_2, ..., p_N]$ by finding the $N$ eigenvectors corresponding to the $N$ largest eigenvalues of the matrix:

\[ B = ((X(D_i - W_i)X^t))^{-1}(X(D_p - W_p)X^t). \]  

(5)

where $D$ is a diagonal matrix: $D(i, i) = \sum_j w_{i,j}$.

4 Experiments

In this section, we give a brief description of the database and the associated benchmarking protocol, followed by the feature extraction module. Finally, we present and discuss the obtained results.
4.1 Database description

The MOBIO database [MMH+12] contains audiovisual recordings of 152 people (100 males and 52 females) from five European countries. These recordings were registered through two phases, each one consists of 6 sessions. In the first phase data, the speakers were asked to answer a set of 21 questions with the question types ranging from: Short Response Questions, Short Response Free Speech, Set Speech, and Free Speech wherein the second phase data the speakers were asked to answer a set 11 questions with the question types ranging from: Short Response Questions, Set Speech, and Free Speech.

The MOBIO database was recorded using two mobile devices: NOKIA N93i mobile phone and MacBook (2008) laptop computer. Since this database was acquired with mobile devices, it had a significant amount of noise [SCP+10]. About 10% of utterances had an SNR less than 5 dB, while 60% had SNR between 5 to 10 dB. Also the utterances had limited amount of speech. About 25% of utterances had less than 2 seconds of speech while 35% had between 2 to 3 seconds of speech.

The MOBIO database was partitioned in three sets: the training set, the development set and the evaluation set. The training set utterances were used to learn the background parameters such as the UBM model and the subspace matrices. The development set data are used to tune the meta-parameters such as number of Gaussians and the subspace dimensions, this data set is divided into two parts, the enrollment set which is used to generate the client models (5 utterances for each speaker) and the probe set which is used in scoring. The evaluation set has the same structure as the the development and it is used for compute the final evaluation performance.

4.2 Experimental Setup

In our experiments, we use 19 MFCC coefficients extracted using 20 ms Hamming window taken every 10 ms. These features were augmented with the log energy, delta and double delta coefficients to produce 60 dimensional feature vector. To reduce the channel effect, we apply the Cepstral Mean Subtraction (CMS) normalization [Ata74] to the features. We eliminated the no speech segments using the voice activity detection (VAD) algorithm described in [RSB+04].

In our case, the UBM is gender independent. It was trained using 100 utterances from each speaker, and it is composed with 256 Gaussian components with diagonal covariance matrices. For the i-vector experiments, the dimension of the total variability subspace is 400 while the dimensions of the CEA projection matrix is 200. In the Balanced Soft Weights described in section 4, the constant $t = 0.1$ yields the best performance.

System performance is assessed using both equal error rate (EER) of the development set and the half total error rate (HTER) of the evaluation set. The EER corresponds to the point defined by some threshold $\theta$, where false acceptance rate (FAR) is equal to false rejection rate (FRR). HTER is the mean of FAR and FRR of the evaluation set, at the threshold $\theta$ previously tuned in the development set. Further more we plot DET curves which allow
Table 1: EER and HTER of CEA+WCCN and LDA+WCCN compensation techniques on MOBIO database.

<table>
<thead>
<tr>
<th>Method</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EER</td>
<td>HTER</td>
</tr>
<tr>
<td>CEA+WCCN</td>
<td>10.992%</td>
<td>20.651%</td>
</tr>
<tr>
<td>LDA+WCCN</td>
<td>12.337%</td>
<td>20.982%</td>
</tr>
</tbody>
</table>

the comparison of many systems at different operating points.

4.3 Results and discussion

In this section we present the results of the verification system using the two techniques of channel compensation: LDA+ WCCN and CEA+ WCCN. The performances are given by the EER and HTER values (table 1) and represented by the DET curves (figures 1 and 2).

We notice from table 1 that in the development set the EER value of the male gender decreases from 12.337% with the LDA+WCCN technique to 10.992% with the CEA+WCCN technique. Thus, the CEA+WCCN yields a relative error reduction rate of 10.90% compared to the baseline LDA+WCCN. For female speakers, the reduction is more pronounced (13.38%), where the EER decreases from 14.402% with LDA+WCCN to 12.474% with CEA+WCCN. The same in the evaluation set the HTER of the male gender decreases from 20.982% with LDA+WCCN technique to 20.651% with CEA+WCCN technique (reduction of 1.57%) and for the female gender the HTER decreases from 28.614% with LDA+WCCN to 25.298% with CEA+WCCN (reduction of 11.58%).

The comparison of the two systems at different operating points are illustrated through the DET curves in figures 1 and 2 for the male and female cases respectively. We notice that the CEA+WCCN outperforms the LDA+WCCN at all operating points.

5 Conclusion

In this work we investigated the use of the CEA sub-manifold learning method to compensate the channel variability in the i-vectors based speaker verification systems. Unlike LDA, CEA has no assumption about the distribution of the input data and it uses both Conformal embedding nature and discriminating criterion to compute the projection matrix. Experiments on the MOBIO-voice database shows that the CEA+WCCN technique performs better than the LDA+WCCN technique. Future work includes further experimental analysis on other databases to confirm obtained results.
Figure 1: DET curves of CEA+WCCN and LDA+WCCN techniques of male and female for the development set.

Figure 2: DET curves of CEA+WCCN and LDA+WCCN techniques of male and female for the evaluation set.

References


Protection of Fingerprint Data with the Glass Maze Algorithm

Markus Springer
Department of Computer Science
Hochschule Darmstadt
Haardtring 100
64295 Darmstadt
markus.springer@stud.h-da.de

Abstract: This work proposes an implementation of the glass maze algorithm proposed by Carlo A. Trugenberger and presents a short summary of its theoretical basis. The implementation is roughly tested with synthetic fingerprints and the experimental results and findings during the testing phase regarding security issues and performance are discussed. It could be seen, that even though from a theoretical standpoint the system seems to be stable against brute force attacks, under certain circumstances the system may still be vulnerable to brute force attacks.

1 Introduction

With the help of cryptography it is possible to provide confidentiality, non-repudiation, integrity and authenticity. One of the big problems of traditional cryptography is the size of its keys. The size of the keys limits the possible applications of cryptographic functions. While on the one hand, longer keys often guarantee for a higher level of security and a wider search space for brute force attacks, they are also in most cases not usable for memorization by a human. This problem gets intensified by the fact that most service providers use different cryptographic techniques and most of the time issue their own keys which makes it even harder for the user to remember the keys. One way around this problem is the idea of biometric cryptosystems. Biometric cryptosystems combine biometric and cryptographic systems and methods in order to make them more human usable. Biometric features, depending on the type that is used, can under normal circumstances neither be forgotten nor lost. The biometric trait is used as the key of the system. These so called templates of the user need to be saved in order to use them. This fact leads to a new problem: How can be made sure, that the biometric template data can not be stolen or abused in any way? The protection of these templates is not only a privacy issue but also an issue of security. In order to solve this problem the fuzzy fingerprint vault was developed [CKL03]. The fuzzy fingerprint vault is used to protect the minutiae data of a fingerprint. The security of the fuzzy vault is based on the difficulty of polynomial reconstruction [Tru11]. While the fuzzy fingerprint vault helps in decreasing the risk of stolen biometric template data and reduces privacy issues, it was shown to be vulnerable to brute force [Mih07] as
well as cross-matching attacks [SB07]. Due to these problems another scheme is needed in order to replace the fingerprint vault. Because of this, Carlo A. Trugenberger proposed the idea of the glass maze algorithm as a new way for template protection.

2 The Hopfield Model

The Hopfield model is a neural network which is based on the Spin Glass Model. The Hopfield model was designed to model the associative memory of the human brain. It is able to recognize patterns which it previously learned [AGS85]. A mechanic that can be used for key retrieval as proposed by Trugenberger in [Tru11]. The Hopfield model consists of N binary neurons which may take one of two different states: firing (1) and resting (−1). The states of all neurons combined is considered the current state of the system. The neurons are connected by symmetric synapses with coupling strength \( w_{ij} = w_{ji} \) and \( w_{ii} = 0 \) which makes them interact with each other. Depending on the sign of the coupling strength it can either be interpreted as inhibitory (\(< 0\)) or excitatory (\(> 0\)). The process of recognizing a pattern is done via evolution of the network state. The dynamic evolution of the current state by random sequential updating of neurons is defined as follows:

\[
s_i(t + 1) = \text{sign}[h_i(t)]
\]

\[
h_i(t) = \sum_{i \neq j} w_{ij} s_j(t) + \theta_i
\]

\( h_i \) is called the local magnetization of neuron \( s_i \). The value \( \theta_i \) is a threshold value that can be used to control when a neuron should change its state. Trugenberger proposes this factor should be set to 0 [Tru11]. The synaptic coupling strength \( w_{ij} \) is chosen according to the Hebbian learning rule [Mov90] which is described by the following formula:

\[
w_{ij} = \frac{1}{N} \sum_{\mu=1...p} \sigma_\mu^i \sigma_\mu^j
\]

The \( \sigma_\mu^i, \mu = 1...p \) are binary patterns that should be memorized by the neural network. The associative memory is defined as a dynamical memory that, upon preparing the network in an initial state \( s_0^i \), retrieves the stored pattern \( \sigma_\lambda^i \) that most closely resembles the presented pattern \( s_0^i \). Resemblance is defined by minimizing the Hamming distance between both of these configurations. The described mechanic encodes all the information that is stored in the neural network into the synaptic coupling strengths [Tru11]. As stated before, the neural network will try to retrieve patterns by dynamically updating the network state with the previously defined functions. With these neuron updates the neural network uses a hill climbing dynamic to retrieve patterns. The patterns which were stored in the neural network correspond to local minima of the system. This means that the stored patterns
are attractors for the dynamic network updates which in turn means that the system will evolve till it overlaps with the closest stored pattern. At that point the state will not change anymore. The behavior of the Hopfield model depends on the so called loading factor \( a = p/N \) which is the ratio between the stored patterns and the number of neurons of the system. By analyzing this factor in the thermodynamic limit \( p \to \infty, N \to \infty \) it is possible to find three distinct sectors which are further described in [Tru11]:

\[ a < 0.051: \text{In this sector the system is in a ferromagnetic phase. This means that there are global minima that correspond to all stored pattern. This means that an exhaustive search for all stored patterns may be organized.} \]

\[ a > 0.138: \text{The system is in a state of chaos where all retrieval capabilities are lost and no information can be retrieved successfully.} \]

\[ 0.051 < a < 0.138: \text{In this sector the system is in a mixed spin and ferromagnetic phase. This is the ideal sector for hiding and retrieving the key because the number of minima is increased exponentially and network states that are close to a stored pattern will converge to the closest (in Hamming distance) stored pattern. If the state is sufficiently different from the stored patterns it will converge to a minimum that was not stored.} \]

3 The Glass Maze Algorithm

The enrollment of a fingerprint consists of a number of non-trivial steps. The first of these steps is the so called quantization. In this step the fingerprint is provided in the form of \( M \) pixel coordinates \((x_i, y_i)\), where \( M \) corresponds to the amount of minutiae of the fingerprint. These coordinates are now used to generate a system configuration for the Hopfield model to remember. This is done by dividing the image of the fingerprint into \( N \) squares of almost equal size. Every one of these \( N \) squares corresponds to a neuron. A neuron can take two states: -1 if there is no minutiae in the corresponding square and 1 if at least one minutiae is in the square. The configuration that is generated out of this is called \( \sigma^{fp} \). The next step is to generate the key configuration \( \sigma^{key} \). The key is created by randomly flipping \( k \) bits of \( \sigma^{fp} \). This means that \( \sigma^{key} \) and \( \sigma^{fp} \) have a Hamming distance of \( k \). This way the fingerprint of the user is never directly stored in the system. After generating the key the next step is to generate \( p - 1 \) random bit patterns. These patterns are then used to bind the key. The key binding mechanism is described by the following function:

\[
w_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \sigma_{i}^{\mu} \sigma_{j}^{\mu} \quad (4)
\]

\[
\sigma_{i}^{1} = \sigma_{i}^{key} \quad (5)
\]

\[
\sigma_{i}^{\mu} \in \{-1, 1\}^N, \mu = 2...p \quad (6)
\]

The number of patterns must be chosen according to loading factor \( 0.051 < a < 0.138 \) as
it was described in section 2. Trugenberger proposes $a = 0.1$ as a good choice. In the last step the it must be assured that the fingerprint can actually retrieve the generated key. This is done by simply trying to retrieve it via key retrieval. If key retrieval fails, the key must be generated by lowering the value $k$ and trying it again. The key retrieval corresponds to the dynamic network updates of the Hopfield model. The key retrieval is achieved by first quantizing the fingerprint and the resulting configuration $\sigma^{fp}$ is then chosen as the systems initial state $s^0$. The state is then dynamically evolved as described in 2 until the system reaches a minimum. If the provided fingerprint was part of the keys basin of attraction, the key will be retrieved successfully [Tru11].

4 Implementation

The implementation of the glass maze algorithm was mostly done the way it was proposed by Trugenberger in [Tru11]. Some parts were altered and will be discussed in further detail. The enrollment process was split up into five different non-trivial steps:

1. Quantization of the fingerprint
2. Key generation
3. Random pattern generation
4. Coefficient matrix calculation
5. Enrollment verification

The quantization (1) of a fingerprint is the process of creating $\sigma^{fp}$ from raw fingerprint data. The fingerprint data is provided in the form of a list of minutiae coordinate pairs M. Coordinate pairs $(x, y)$ represent minutiae. Other necessary parameters are the dimension $D$ of the fingerprint scanner image, as well as the number of neurons $N$. During quantization the minutiae are mapped directly to the neurons which are represented by a vector $S$ of size $N$. The overall number of pixels on one image of the scanner can be calculated with the given dimension parameter which are interpreted as width and height. These two are used to calculate the number of pixels the image consists of. With the help of the given number of neurons $N$ we can compute the amount of pixels that represent the same neuron field. All pixels together are then interpreted as a long string of points. The minutiae are placed on this string and afterwards are mapped onto their corresponding field in the neuron vector via the following formulas:

\[ p_i = y_i \cdot \text{width} + x_i, \quad i = 1...M \]  
\[ S_i = \left\lfloor \frac{N \cdot (p_i + 1)}{\text{width} \cdot \text{height}} \right\rfloor, \quad i = 1...M \]
The described method is not the best way to quantize a fingerprint, but it is one of the easier and faster ways to implement it. This part of the implementation is probably the part that can be improved the most. The big problem with this method of quantization is that it does not make use of the fingerprints properties. There are positions that are a lot more likely to contain minutiae as others. The edges of an image for example will be a lot less likely to contain minutiae than the rest of the image. This makes it easier for an attacker since he may exclude those areas in brute force attacks which decreases the overall search space. The main decision for this approach was the ease of implementation and flexibility for different image sizes.

The key generation (2) depends on the quantized fingerprint $\sigma_{fp}$ and the number of flipped bits $k$. The flipping of bits is done in rounds. In each round a value between 1 and $N$ is chosen at random and used as the index of the neuron that will be flipped. The algorithm repeats the rounds until the Hamming distance between the currently derived key $\sigma_{key}$ and $\sigma_{fp}$ is $k$.

The random pattern generation (3) takes the number of neurons $N$ and the loading factor $a$. It uses $a$ to calculate the number of random patterns $p - 1$ and generates a $N \times (p - 1)$ matrix $R$ of random patterns. Each field of these patterns represents a neuron as it was described in section 2. A random pattern is generated by creating a vector of $N$ random values between 0 and 1. The values in this vector are then rounded and mapped in a way that values $\geq 0.5$ are mapped to 1 and values $< 0.5$ are mapped to -1. Each of these vectors represents one column of the matrix.

The coefficient matrix calculation (4) was not directly mentioned in [Tru11] but in [Tru12]. This step is done in order to hide the stored patterns. This is achieved by applying the Hebbian learning rule, which was mentioned in section 2, to the previously generated random patterns and the key. The function takes the parameters $N$, $\sigma_{key}$ as well as the previously created matrix $R$ and outputs a $N \times p$ matrix $W$ containing the coupling strengths.

In the last step the enrollment verification (5) is done to ensure that the provided fingerprint is able to retrieve the previously generated key. This is done by calling the key retrieval mechanism with the fingerprint $M$ and the matrix $W$ and check if it returns the key $\sigma_{key}$. If it works, the enrollment is finished. If it did not work, $k$ is decreased by one and from step (2) onwards all steps are repeated. The overall output of the enrollment is the matrix $W$ which contains the key as well as the random patterns.

The key retrieval mechanism was mostly implemented the way it was proposed by Trugenberger. The key retrieval takes the parameters $W$, the fingerprint coordinates $M$, the dimension $D$ of the scanner image as well as the number of neurons $N$. It starts by quantizing the given fingerprint coordinates $M$. The quantized fingerprint $\sigma_{fp}$ is then used as the initial state of the network. The neurons are then evolved as described in 3 until the Hamming distance between the old neuron state $s_{i-1}$ equals the newly evolved neuron state $s_i$. An important thing to note is, that Trugenberger defines the updating of neurons as random and sequential. In order to ease the implementation, it was decided to update each neuron in strict rotation. Such an update may be considered as a round. After each round it is checked, whether the system has reached a minimum or if further updating is required. If updating was done randomly it would be harder or more error prone to deter-
mine if the system has reached a minimum. During testing no significant differences could be found between different rotations, though this would need further testing.

5 Experimental Results

For all experiments described only synthetic fingerprints were used to avoid problems due to alignment and in order to test the algorithms on a more general level. During the implementation and testing phases a first glimpse on possible problems could be examined pretty early on. One problem was the resemblance between quantized fingerprint, key and randomly generated patterns. In relation to the number of neurons the fingerprints and the key resemble each other a lot more then they would with the randomly generated patterns. This would mean that the FAR can be quite high. This stems from the fact that the used random generator most of the time creates a rather uniform pattern while fingerprints often do not tend to have a uniform pattern. The associative memory will relate fingerprints to each other more often then it would with random patterns or any other minimum at all.

Another factor that plays into this can be seen in figure 1. The greater we chose the factor $N$ of neurons, the more likely it is that key retrieval will actually return a previously trained pattern rather then a random minimum. The test was done with 1000 synthetic fingerprints per value of $N$. The factor $a$ was kept at 0.1 which was proposed by Trugenberger to be close to a good choice. This also makes the hiding of patterns within the coefficient matrix a rather useless trial, because an attacker may brute force and guess most of the saved patterns in minimal time given that $N$ was chosen big enough. A way to circumvent this problem could be in choosing parameter $k$ according to the randomly generated patterns to make the pattern more uniform or by improving the quantization step to generate a more uniform distribution which will make the quantized fingerprint more similar to the saved patterns.

Another experiment was done testing the false acceptance rate for fixed values of $N$. In these experiments the loading factor $a$ was varied. For each chosen value of $a$ 5000 key retrievals were executed. What can be seen is that for low values of $a$ the system often
falsely accepts the presented fingerprint while for rather high values of $a$ the system rarely falsely accepts a fingerprint. What is to note though is, that the higher the value $N$ is chosen, the better the false acceptance rate gets for smaller values of $a$. Figure 2 displays this asymmetric observation. For $N = 128$ the curve is a lot steeper then for $N = 64$. As Trugenberger described in [Tru11], higher values of $a$ set the system in a state of chaos so that saved patterns are not retrievable. Further testing this effect may be able to loosen the lower bound for higher values of $N$.

As for performance the system was tested according to the number of iterations the key retrieval mechanism needs to find a minimum, as well as the average execution time of finding a minimum. The results of the experiment can be seen in figure 3. For every chosen value of $N$ 5000 key retrievals were executed. For the testing a simple Intel i5 processor with 2.26Ghz was used. Something to note is, that per 16 additional neurons the average number of iterations is increased by $38\%-40\%$. But while the system is exponential and not linear, it is still growing rather slow. The average execution time on the other hand is growing fast and is also exponential. This seems logical, as the number of computations for one iteration grows with the number of neurons, as well as the number of iterations.

6 Conclusion and Future Work

Overall the proposed idea of Trugenberger seems promising. In order to protect the templates the system is required to make it hard for the attacker to get the stored pattern. The privacy issue may be solved, but the security issues of the system can not be dismissed. A big problems seems to be that, while the number of minima increases with higher numbers of neurons, the actual search space seems to decrease because these minima are scarcer reached. This limits the system to smaller numbers of neurons, which in return reduces the time of its usefulness a lot. Further testing is needed to determine if this problem can be circumvented and check the influence of the loading factor, $k$ and neurons more. Also a
better method for quantization should be implemented to increase the overall search space by distributing the minutiae more. Further tests have to be conducted with real minutiae data in order to check for FAR and FRR with real user data and check for the influence of alignment.

References


Embedded Face Detection Implementation

Laurentiu Acasandrei, Angel Barriga

Instituto de Microelectronic de Sevilla
IMSE-CNMC-SIC
C/ Américo Vespucio, s/n (esquina Leonardo da Vinci)
Sevilla, Spain
laurentiu@imse-cnmc.csic.es
barriga@imse-cnmc.csic.es

Abstract: In this communication an embedded implementation of the Viola-Jones
face detection algorithm targeting low frequency, low memory, and low power
consumption, is presented. The design methodology, performance analysis and
algorithm optimization in order to accelerate the face detection process, will be
described. The resulted implementation is platform independent and achieves on
average a 3 times detection speed up.

1 Introduction

Face detection is an important aspect for biometrics, video surveillance and human
computer interaction. Detection systems require huge computational and memory
resources due to the complexity of detection algorithms. Implementation difficulties
appear when it is required to apply face detection techniques in embedded systems,
where there are restrictions in size and power consumption.

Automatic face detection systems have experienced a remarkable progress in last decade.
However, for large image dimensions, most of applications of these systems are based
on software realizations running on computers. Usually on embedded systems like
photo/video cameras, smartphones and tablets the detection is done on images scaled to
lower resolutions, but this introduces the disadvantage of not being able to detect smaller
to medium dimension faces that appear in the original image. To detect faces from full
size images or video frames is necessary the adaptation of face detection algorithms to
those embedded systems.

The purpose of this communication is to describe a face detection application with high
performance that can be run smoothly on any embedded system. Taking into
consideration that the OpenCV library comes with a baseline application for video face
detection we have used the sources of that application as starting point in developing the
video detection application for low resources embedded systems (with reduced memory
resources, without floating point arithmetic unit, etc).
The structure of the paper is the following. Section 2 introduces the face detection technique. Section 3 presents the design methodology of the face detection embedded system. Section 4 exposes the embedded software face detection implementation. Section 5 focuses in the analysis of the proposed solution, while in section 6 some optimizations in order to accelerate the detection algorithm have been proposed. Finally, Section 7 summarizes the main conclusions of this research work.

2 Face Detection Technique

The face detection technique is based on the face detection framework proposed by Viola-Jones [VJ04]. The proposed framework is capable of processing images extremely rapidly while achieving high detection rates. The speed of the face detection framework relies on three important key components. Firstly, the image is transformed into “Integral Image” which allows the features, used by the detector, to be computed very quickly. Secondly, the used classifier is simple and efficient and is build using the AdaBoost learning algorithm [SFB98] to select a small number of critical visual features from a very large set of potential features. And thirdly, the classifier is formed by combining weak classifiers in a “cascade” which allows background regions of the image to be quickly discarded while spending more computation on promising face-like regions. Viola-Jones technique is based on exploring the image by means of a window looking for features. This window is scaled to find faces of different sizes. The system architecture is based on a cascade of detectors. The first stages consist of simple detectors, very fast and low cost, that allow eliminating those windows that do not contain faces. In the successive stages the complexity of detectors are increased in order to make a more detailed analysis of features. A face is detected only if it makes it through the entire cascade. The Haar-like features used by the classifier consist of rectangular areas whose processing requires simple arithmetical operations. The calculation is based on the sum of the pixels of each rectangular region weighed by a weight. At all scales, these features form the “raw material” that will be used by the detector. The set of rectangle features in the image is quite large and overcomplete, so to reduce that number the AdaBoost learning algorithm [SFB98] is applied. The Viola-Jones classifier employs AdaBoost at each node in the cascade to learn a high detection rate at the cost of low rejection rate multistage (mostly multistump) classifier at each node of the cascade. To facilitate the processing of the features, the operations are not made on the original image but on an integral image. Therefore the detection algorithm requires a preprocessing step that calculates this integral image. The advantage of the integral image is that it allows calculating the sum of any rectangle in constant time. The integration of the image consists of adding for each pixel the values of the previous pixels.

3 Design methodology

The starting point of the design methodology of the embedded system is the OpenCV face detection framework application. OpenCV (Open Source Computer Vision), started
by Intel in 1999, is a library of programming functions for real time computer vision [Op13]. OpenCV is released under a BSD license and hence it’s free for both academic and commercial use. It is written in C/C++ and was designed for computational efficiency and with a strong focus on real-time applications. OpenCV comes with a face detection application which is the starting point of the proposed methodology. The target of the proposed face detection system is an embedded implementation using an AMBA bus processor. The embedded architecture of the face detection system is based on a software implementation within an operating environment running on the LEON3 processor. The LEON3 is a synthesizable VHDL model of a 32-bit processor compliant with the SPARC V8 architecture [Leo13]. The model is highly configurable, and particularly suitable for System-on-Chip (SOC) designs.

The design flow is based on three stages, as shown in Figure 1. In the first stage an adaptation of the software application to execute on the embedded system has been made. In the next stage an analysis of the new embedded software application is performed, in order to detect "bottlenecks" and those tasks that are suitable to accelerate through hardware implementation. In the third phase, as result of the previous analysis, solutions to optimize and accelerate some of the tasks of the face detection process will be proposed. At each stage of the design process the performance of the face detection system is analyzed, in terms of speed and quality.

![Design methodology](image)

**Figure 1: Design methodology**

### 4 Embedded Software Face Detection Implementation

Viola and Jones organized each boosted classifier group into nodes of a rejection cascade. Each of the nodes contains an entire boosted cascade of groups of decision stumps (or trees) trained on the Haar-like features from faces and nonfaces (or other objects the user has chosen to train on). Typically, the nodes are ordered from least to most complex so that computations are minimized (simple nodes are tried first) when rejecting easy regions of the image. Typically, the boosting in each node is tuned to have
a very high detection rate (at the usual cost of many false positives). When training on faces, for example, almost all (99.9%) of the faces are found [BK08] but many (about 50%) of the nonfaces are erroneously “detected” at each node. But this is satisfactory because using (say) 20 nodes will still yield a face detection rate (through the whole cascade) of $0.999^{20} \approx 98\%$ with a false positive rate of only $0.5^{20} \approx 0.000001\%$. During the run mode, a search window of different sizes is swept over the original image. In practice, 70–80% of nonfaces are rejected in the first two nodes of the rejection cascade, where each node uses about ten decision stumps. This quick and early “attentional reject” vastly speeds up face detection. The Haar-like features are trained to be applied for evaluating rectangular window of 20x20 pixels. For other dimensions of the evaluating window the Haar-like features must be scaled correspondingly. The face detection system consists of 22 cascade detectors, containing 2135 Haar like features.

The first task of software/hardware codesign was adapting and optimizing the OpenCV baseline application for an embedded environment. We have considered that the majority of embedded environments are capable of running C/C++ applications with or without Operating System (OS) support. This means that the resulting application code has to be compatible for both C, C++ compilers and in the same time platform independent. Another consideration made is the fact that most of the SoC have no floating point support. For it, the resulting application uses integer operations instead of floating point operations in order to preserve the generality of the application for the embedded system world. An important moment in this step was finding an acceptable scaling coefficient of the floating point variables and data to integer variables and data. After trying different values and comparing the resulted integer application with the floating point application we found that by scaling with 20 bits (precision of 20 bits for the floating point decimals) the integer and floating point applications obtain identical results. Also the floating point squared root function necessary to calculate variance of the evaluating window was replaced with fast integer squared root function.

In the end it was obtained a face detection standalone application compatible with C/C++, using only integer type operations and data, where the cascade of classifiers and the image can be loaded from a desired memory location which is previously initialized with a raw RGB image. The accuracy of the resulted standalone application is the same as the OpenCV face detection application.

5 Embedded Software Face Detection Analysis

The next step in application development was analyzing different modes of detection in order to find the run time bottlenecks and optimize the detection. For the embedded target we have decided to use the detection mode where the detector (Haar-like feature) is scaled and the biggest regions containing faces are searched within an image. In this mode the detections starts with the biggest evaluation window and biggest evaluation step and progressively, the window together with the evaluation step are decreased until a region containing a face is detected. In the case that a region with face or multiple faces is detected, the attention of the detector concentrates in that region.
The trained classifier cascade (Haar-like feature) is provided by OpenCV in an XML file format. Using this XML format in an embedded system will produce memory and run time overhead. In order to avoid the overhead we have developed an application that receives a XML file, interprets the data and saves it in a simpler format to a C header file. The resulted embedded application can be compiled with the cascade of classifier or the data can be transmitted during the execution of the application via an appropriate interface. In order to obtain relevant insight about which parts (or functions) of the face detection program are taking most of the execution time GNU gprof tool was used. This tool permits one to learn where the program spends its time and the function calling tree during the execution. It can also tell which functions are being called more or less than are expected. Table I shows the obtained results.

<table>
<thead>
<tr>
<th>Time %</th>
<th># calls</th>
<th>Function name</th>
</tr>
</thead>
<tbody>
<tr>
<td>24.64</td>
<td>17</td>
<td>SetMatZero()</td>
</tr>
<tr>
<td>20.16</td>
<td>3201</td>
<td>RunHaarClassifierCascadeEmbed()</td>
</tr>
<tr>
<td>14.81</td>
<td>16</td>
<td>SetImagesForHaarClassifierCascadeEmbed()</td>
</tr>
<tr>
<td>13.39</td>
<td>1</td>
<td>Integral()</td>
</tr>
<tr>
<td>11.35</td>
<td>1</td>
<td>LinkDataToEmbedClassifierCascade()</td>
</tr>
<tr>
<td>4.22</td>
<td>511</td>
<td>HResizeLinear()</td>
</tr>
<tr>
<td>3.11</td>
<td>262144</td>
<td>saturate_uchar()</td>
</tr>
<tr>
<td>2.62</td>
<td>512</td>
<td>VResizeLinear()</td>
</tr>
<tr>
<td>1.80</td>
<td>296384</td>
<td>sum_elem_ptr()</td>
</tr>
<tr>
<td>1.10</td>
<td>3201</td>
<td>sqrt64()</td>
</tr>
</tbody>
</table>

As we can see the function SetMatZero() uses 24.64% of the executing time even surpassing the time spent applying the cascade Haar like-features (20.16%) for the entire image. The function SetMatZero() is used to set to zero all the elements of a temporary matrix having the same dimension of the image. In this matrix the top left coordinates of a detected face are flagged with value 1. In this mode the detections starts with the biggest evaluation window and biggest step and progressively, the window and the step are decreased until a region containing a face is detected. The detection is done in two steps. In the first step the image is scanned with the evaluation window by applying only the first two Haar-like feature stages in order to rapidly detect regions containing potential faces. If a region is found to have a potential face then the coordinates are set to value one in the temporary matrix. In the second step each potential face (starting with their coordinates) from the temporary matrix is evaluated with the remaining Haar-like feature stages. If a true face is detected then the coordinates, width and height are stored in a list. At the end of the second step, the temporary matrix is set to zero in preparation for the next image where the evaluating window has smaller dimensions.

We can improve the speed by not using the function SetMatZero() at the end of the second step and instead each time after we have detected a face during the second step and that face is stored into a list, we set to zero the coordinates inside the temporary matrix. After applying this change, the detection time is improved with 24.64%.
6 Embedded Face Detection Optimization

The OpenCV face detection baseline application implements detection in two distinct modes (see Figure 2). In mode 1 the image is scaled using linear interpolation until it reaches a predefined minimal dimension. Each time the image is scaled the two integral images (normal=$\sum x$ and squared=$\sum x^2$), needed for variance, are recalculated for the scaled image. The search window has fixed dimension during the detection process. In mode 2 the integral images(normal=$\sum x$ and squared=$\sum x^2$), needed for variance, are calculated only once for the original image but the Haar-Like features form the classifier are scaled progressively until their dimensions are close to the dimension of the original window. This mode lacks the interpolator used in mode 1. The search window has a variable dimension during detection process.

The mode 1 and mode 2 have different scaling and search window control. In both detection mode (mode 1 and mode 2) the Haar Like-features components (weights and dimensions) are scaled proportionally with the dimensions of search window. If we do not scale the Haar-like feature weights and adjust the variance computation

$$\sigma = \sqrt{\frac{1}{W \cdot H} \left( \sum_{x} x^2 - \left( \frac{\sum_{x} x}{W \cdot H} \right)^2 \right)}$$

by using the formula

$$\sigma_{\text{adjusted}} = \sqrt{W \cdot H \sum_{x} x^2 - \left(\sum_{x} x\right)^2}$$

it results that the number of arithmetic operations (like division and multiplication) and memory accesses are decreased substantially. This will make the algorithm perform faster due to the reduced number of computation for the adjusted variance of the search window [AB11]. Figure 2 shows the proposed optimization. In order to compare the performance of the OpenCV 2.2 baseline face detection with the accelerated Viola-Jones algorithm, and to analyze the influence of the configuration parameters, both implementations have been compiled and speed optimized for 64 bit Win7 OS using Visual Studio 2010 Professional edition. Both implementation (OpenCV’s implementation and accelerated Viola-Jones version) have received the same test VGA(640x480) images and the same configuration parameters. In Table II we have different scale factor (sf) and minimum search window dimensions (swd): sf=1.1 and swd=20x20 (Conf1); sf=1.2 and swd=20x20 (Conf2); sf=1.1 and swd=30x30 (Conf3).

Table II: Performances of OpenCV and accelerated Viola-Jones implementation for different parameters

<table>
<thead>
<tr>
<th></th>
<th>Mode 1: Img scaling</th>
<th></th>
<th>Mode 2: Haar scaling</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OpenCV Baseline</td>
<td>Optimized version</td>
<td>OpenCV Baseline</td>
</tr>
<tr>
<td>Conf1</td>
<td>speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>708.8 ms</td>
<td>185.7 ms</td>
<td>843.9 ms</td>
</tr>
<tr>
<td></td>
<td>1.41 FPS</td>
<td>5.38 FPS</td>
<td>1.18 FPS</td>
</tr>
<tr>
<td>Search windows</td>
<td>602348</td>
<td>599816</td>
<td>697582</td>
</tr>
<tr>
<td>Conf2</td>
<td>speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>409.5 ms</td>
<td>164.8 ms</td>
<td>479.7 ms</td>
</tr>
<tr>
<td></td>
<td>2.44 FPS</td>
<td>6.06 FPS</td>
<td>2.08 FPS</td>
</tr>
<tr>
<td>Search windows</td>
<td>354321</td>
<td>353935</td>
<td>381474</td>
</tr>
<tr>
<td>Conf3</td>
<td>speed</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>348.1 ms</td>
<td>99.4 ms</td>
<td>456.7 ms</td>
</tr>
<tr>
<td></td>
<td>2.87 FPS</td>
<td>10 FPS</td>
<td>2.18 FPS</td>
</tr>
<tr>
<td>Search windows</td>
<td>352718</td>
<td>351184</td>
<td>402519</td>
</tr>
</tbody>
</table>
Figure 2: Proposed face detection acceleration algorithm

As the proposed implementation has kept the control mechanism for search windows identical with the one form the OpenCV baseline it is difficult to make a comparison with previous work done in accelerating detection due to the lack of setup information and how many search windows are evaluated. There is one exception Cho et all [CMO09] where they use model1 with a scaling factor of 1.2, minimal search window dimension of 20x20 and the search window is applied with a vertical horizontal step of 1. For the Cho et all setup, we measure for the OpenCV implementation a execution time of 972.3 ms for a total of 881484 search windows and the accelerated versions of Viola-Jones has an execution time of 451.2 ms for 880585 search windows. As shown in Table II the number of search windows depends heavily on the configuration setup and also of the control mechanism. Measuring the number of searched windows performed by the detection system gives more realistic information about the detection performance. The speeds obtained by the accelerated Viola-Jones implementation in some configuration are faster that any single GPU acceleration [H09] of 4.2 FPS or FPGA implementation [CMO09] of 6.5 FPS, and for the lowest number of search windows (see Table II) it has closer performances to [HOT10] of 15.2 FPS. The accuracy of the accelerated Viola-Jones version is the same as the OpenCV face detection application. In order to analyze the detection accuracy a test software was developed. A PC based test bench software configures LEON3 based detection system and sends the test images. Then it receives the detection results for further analysis. The test setup is based on 2409 frontal face images from the color FERET database [PWH98].
7 Conclusions

This communication presents an embedded software implementation of the Viola-Jones face detection algorithm targeted for low frequency, low memory and low power embedded systems. The starting point was the trained classifier cascade (Haar-like feature) provided by OpenCV library resources for video face detection. For it some modifications has been make adapting and optimizing the OpenCV baseline application for an embedded environment. This modifications can be summarized in changing the floating point operations by integer ones, analyzing the performance in order to detect the system bottlenecks, and algorithmic speed-up by not scaling the Haar-Like feature weights and adjusting the computation of variance.

Acknowledgement

This work was supported in part by Spanish Ministerio de Ciencia y Tecnología under the Project TEC2011-24319, and by Junta de Andalucía under the Project P08-TIC-03674, and by Spanish Ministerio de Economía y Competitividad under the Project IPT-2012-0695-390000. Co-financed by FEDER.

References


Landmark-based Model-free 3D Face Shape Reconstruction from Video Sequences

Chris van Dam, Raymond Veldhuis, Luuk Spreeuwers
Faculty of EEMCS
University of Twente, The Netherlands
P.O. Box 217
7500 AE Enschede
c.vandam@utwente.nl
r.n.j.veldhuis@utwente.nl
l.j.spreeuwers@utwente.nl

Abstract: In forensic comparison of facial video data, often only the best quality frontal face frames are selected, and hence potentially useful video data is ignored. To improve 2D facial comparison for law enforcement and forensic investigation, we introduce a model-free 3D shape reconstruction algorithm based on 2D landmarks. The algorithm uses around 20 landmarks on the face and combines the structure information of multiple frames. Model based 3D reconstruction methods, such as Morphable Models, reconstruct a 3D face shape model that is strongly biased towards the average face. Therefore, we don’t use statistical face shape models in our model-free approach. The 3D landmark reconstruction algorithm simultaneously estimates the shape, pose and position of the face, based only on the fact that all images in the sequence are recorded using a single calibrated camera. The algorithm iteratively updates the reconstruction by including new frames, while maintaining the consistency of the reconstruction. We demonstrate the convergence properties of the method reflected in the 2D reprojection error and the 3D error with respect to a ground truth model. We show that the quality of the reconstruction depends on the level of noise in the landmarks. In follow-up experiments we show that our method is able to reconstruct the 3D structure of a face, using a styrofoam head and real video data. The results of the real face data show the same behavior as the results of the simulated data, which indicates that our method is capable of reconstructing real facial structures, depending on the noise of the landmarks.

1 Introduction

One of the unsolved issues in forensic comparison of facial data is the comparison with ‘wild’ photo or video data. Law enforcement services are constrained to work with the case material provided, and unlike researchers, they are not able to use recordings from a controlled environment. Among the most difficult problems of ‘wild’ photo materials are the non-frontal poses of faces and low resolution facial images, because often material of overview cameras is used for facial comparison. Automatic face recognition software can only handle 2D facial data under a small pose angle. At the moment the accuracy of automatic face comparison algorithms degrades quickly for faces under large pose.
As a consequence often only the best quality frontal face frames are selected, and hence much video data is ignored. Law enforcement services are still in search of the ‘tools’ to compare non frontal faces. However, these ‘tools’ should treat the video data in such a way that no supplementary information is added to the video data. Reconstruction methods, such as Morphable Models [BV99], reconstruct a 3D face shape model that is strongly biased towards the average face. Such reconstructions could lead to unacceptable forensic conclusions. In the proposed method we try to avoid this situation caused by facial models.

In this paper we introduce a model-free 3D shape reconstruction algorithm based on 2D landmarks, so no additional statistical face models or average face models will be used. We assume that the calibration parameters of the camera, such as focal length, principal point and skew, are available. Any recording is assumed to contain a subset of frames with different views of a face without variation in facial expression. Our final goal is to reconstruct the face in 3D. We use around 20 landmarks on the face to estimate the shape of the face together with the pose and the position of the face for each view. We present three different experiments. In our first experiment we use simulated data to demonstrate the convergence properties of the method reflected in the 2D reprojection error and the 3D error with respect to a ground truth model. In the second experiment we continue our work in [DSV13] and we explore the strength of our method more extensively on realistic face shape data with a styrofoam head model. In our last experiment we use real video sequences for our reconstruction. Note that our reconstructed 3D models only contain shape information and no texture information. This paper continues with section 2 where we give a background on the methods and notations used in this paper. In section 3 we introduce and explain our proposed algorithm. In section 4 we show the performance of our algorithm in several experiments. Then we end up with the conclusion in section 5.

2 Background

Our problem, in which the face of the suspect is moving in front of a static camera, is equivalent to a problem where the camera is moving and the suspect is static. So for each view \( i = 1..N \) we have to find the external camera parameters of that specific view. The static shape of the face can be described by \( j = 1..M \) 3D landmarks. We will use \( M \) 2D landmarks with known correspondences to the 3D landmarks in all \( N \) views to obtain a 3D reconstruction of the landmarks on the face. Our camera is described by the pinhole camera model [HZ04], where a 3D point \( Q \) is projected on the image plane in 2D point \( q \). The point projection equation is usually written as \( q = P \cdot Q \), where \( P \) contains both the calibration parameters of the camera and the rotation and translation of a view.

We prefer a method in which we can add additional views to the current solution to improve the reconstruction. To be able to find such a method, we should search for a method that starts with one pair of views and then provides an iterative solution or a solution that merges groups of views. The method described in [Har93] is able to estimate the rotation and translation parameters for one pair of views. This method expresses the relation between calibrated views in the essential matrix. The essential matrix can be estimated from corresponding landmarks in two views using a robust MSAC method (M-estimator SAm-
ple Consensus) method [TZ00]. Once we determined the relation between two views, the relative rotations and translation parameters can be estimated for both views. This method provides four solutions for the rotation and translation parameters, see Equation 2.1, but only one of these solutions is posing the points in front of the camera:

\[
\hat{P}_1 = [UWV^T | +u_3] \quad \hat{P}_2 = [UW^T V^T | +u_3] \\
\hat{P}_3 = [UWV^T | -u_3] \quad \hat{P}_4 = [UW^T V^T | -u_3]
\] (2.1)

where the rotation matrix defined by \(U, W\) and \(V\) is based on the result of a singular value decomposition of the essential matrix. The matrix \(W\) is a matrix that mirrors one of the axes. The translation \(u_3\) is the last column of \(U\), see [HZ04] [Har93]. This solution has 5 degrees of freedom, 3 for the rotation and only 2 for the translation, because the equation is determined up to an unknown scale. The rotation and translation parameters are extracted directly from the essential matrix of one pair of views. Then, we can estimate the structure by linear triangulation of one pair of views [HZ04]:

\[
A = \begin{bmatrix}
  xP^3 - p^1 \\
  yP^3 - p^2 \\
  x'P^3 - p'^1 \\
  y'P^3 - p'^2
\end{bmatrix}
\] (2.2)

where \(P^i\) are the rows of \(P\) in the first view and \(x, y\) are the x- and y-values of the projection of point \(Q\) in the first view. The other parameters are the corresponding values of the second view. The point \(Q\) can be found by solving \(AQ = 0\). This method reconstructs only the visible points in one pair of views. The method can be extended to more than 2 views by including more equations from additional views in \(A\). In our case we have a low number of landmarks, so the reconstruction based on two views gives a poor estimation of the shape. Therefore, we extend the algorithm using multiple views to overcome the problems of noise and the low number of landmarks. We introduce an algorithm that iteratively updates the reconstruction by including new views, while maintaining the consistency of the reconstruction for a low number of landmarks. The quality of the reconstruction can be determined by the 2D RMS reprojection error \(E_{2D}\):

\[
E_{2D} = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \|q_{ij} - \hat{P}_i \cdot \hat{Q}_j\|^2}
\] (2.3)

where index \(i\) represents a view \(1..N\) and \(j\) represents a point \(1..M\), \(\hat{P}\) contains the external camera parameters of each view and \(\hat{Q}\) contains a collection of homogeneous 3D points. The homogeneous 2D vector \(q_{ij}\) represents the known projections including the noise on the landmarks.

3 Reconstruction Algorithm

In this section we describe the proposed algorithm for the reconstruction of the structure of the face based on 2D projections. In short the algorithm finds an initial pair of views
with a low reprojection error. Based on this pair of views we obtain a linear estimation of the structure. Then we start an iterative procedure in which we add one new view in every step of the procedure. After adding the new view, the current selection of views and the current structure estimation are optimized. The result of the reconstruction algorithm is an estimation of the 3D positions of the landmarks and an estimation of the rotation and translation parameters of each view.

The best initial estimate for the structure is found by calculating the reprojection error for every possible pair of views in the dataset and to select the pair with the minimum reprojection error. To calculate the reprojection error we need to know the rotation and translation parameters of each view. These values (except for the scale) can be extracted from the Essential Matrix, see Equation 2.1. The essential matrix can be estimated, in turn, from the projections using a robust MSAC method (M-estimator SAmple Consensus) [TZ00]. Knowing the rotation and translation parameters of a pair of views, allows us to estimate the structure for this pair of views. Based on this structure we can calculate the reprojection error for this pair of views. However, also the reprojection errors of the other views are important for consistency during the optimization. So, to find the best pair of views we choose a reference view and calculate all rotations and translation relative to the reference view. Then we calculate the reprojection error of the total set of views for every view as reference. A second criterion for the selection of the best pair of views is the number of landmarks that could be reconstructed, because not only the reprojection error is important, but also the number of visible corresponding landmarks in the initial pair of views. Our selection criterion is now to find the pair of views with the maximum number of corresponding landmarks in two views and a minimal reprojection error for the total set of views. We choose to obtain the subset of 25% of the solutions containing the most reconstructed points over all views. From this subset we select the pair with the lowest reprojection error over all views. This solution provides us a solution that is sufficient for initialization of our iterative optimization. We calculate an initial linear estimation of the structure based on the selected pair of views.

In the optimization step one new view is added in each iteration to keep all views in our current estimation consistent with the estimated structure. The selection of the new candidate view is based on the convergence behavior of the candidate view. The view with the lowest reprojection error after 10 optimization iterations, is chosen as the next view. This candidate selection is necessary to prevent the algorithm from failing in the first few iterations. Based on the new selection of views, a linear estimation of the structure is obtained, see Equation 2.2. Then the reprojection error of both the rotation and translation parameters and the structure are minimized using the Levenberg-Marquardt algorithm. To prevent overfitting, we used only 30 Levenberg-Marquardt iterations for each optimization step, which performs properly for the minimization. Finally, the rotation and translation parameters of the views that were not in the selection set are optimized to maintain the consistency of the total set of views. The iterative optimization procedure continues until all views are added and optimized.
4 Experiments

The goal of the first experiment on simulated data is to determine the influence of the number of views on the reconstruction, and to investigate the convergence properties of our algorithm. We create a random point cloud of 25 3D points and obtain a set of 100 projections of this point cloud with variation in rotation and translation. The calibration information and a random selection of the projections are used in the reconstruction algorithm. We performed two experiments in which we added a different level of Gaussian zero-mean noise to the projections, with a standard deviation of 1.0 and 2.0 pixels respectively. The size of the face in each frame is around 250-350 pixels. The noise is added independently to the x- and y-coordinates of the projections. Finally we used a random mask to hide 30% of the data to imitate the hidden landmarks on a face. We use our reconstruction algorithm to estimate the 3D structure. The quality of the reconstruction will be determined by the 2D RMS reprojection error $E_{2D}$, see Equation 2.3. All landmarks that were not visible, were left out of the equation, so $M$ is defined as the total number of visible landmarks summed over all views. After reconstruction the 3D RMS error $E_{3D}$ between the reconstruction and the ground truth point cloud can be calculated with:

$$E_{3D} = \arg\min_{\mathcal{H}} \sqrt{\frac{1}{M} \sum_{j=1}^{M} \|Q_j - \mathcal{H}\hat{Q}_j\|^2}$$

(4.1)

where $\mathcal{H}$ is a rigid 3D transformation which aligns the ground truth point cloud $Q$ with the reconstruction $\hat{Q}$ and $j$ is the index of a point. The experiment is repeated 100 times with different instances of noise to investigate the robustness of the algorithm.

Figure 1: 2D and 3D error of the reconstructions using noise with a standard deviation of 1.0 pixels.

The graphs in Figure 1 show the expected behavior for Gaussian noise with a standard deviation of 1.0 pixels. The more views are added, the more robust the reconstruction is. If the shape is estimated perfectly, then we would expect the 2D reprojection error to converge to the level of noise added. The 2D reprojection error converges to an asymptote of $\sqrt{2} \approx 1.41$, which is the expected level of noise, see the left graph of Figure 1. Another observation we make is that the number of views above 30 has little influence on both the 2D and the 3D average error. The robustness of the algorithm is only slowly increasing for more than 30 views, see the right graph of Figure 1. So adding more than 30 views seems to have only a small impact on both the quality and robustness of the algorithm.
If the level of Gaussian noise is doubled to a standard deviation of 2.0 pixels, the behavior is similar to the previous experiment. The asymptote here is $\sqrt{8} \approx 2.83$, see the left graph in Figure 2. Adding more views has less effect on the robustness of the reconstruction algorithm, but it still has a decreasing effect on the average reprojection error. When more views are added, the average 3D error also decreases slowly, though the robustness of the algorithm seems not to increase. For more than 35 views, the system shows even more variation in the 3D errors than for 35 views, see Figure 2. This can be explained by the fact that the more views are added, the higher the chance for heavy outliers in the projections. Since none of the selected views are skipped, outliers might severely decrease the result of the reconstruction. The reconstruction is assumed to be failed, if the reprojection error is above 5.0 pixels. For the experiment with Gaussian noise with a standard deviation of 1.0 pixels and more than 30 views, the algorithm converges to a solution in about 99% of the cases. In the case of a standard deviation of 2.0 pixels, the algorithm only converges in 75% of the cases. So the algorithm seems stable for Gaussian noise with a standard deviation of 1.0 pixels, but becomes less stable for Gaussian noise with a standard deviation of 2.0 pixels or above.

The goal of the second experiment with the styrofoam head is to determine whether the algorithm is capable of working with manually labeled face data. We acquired a 3D model of a styrofoam head with 22 colored pins located on the face. An orthogonal view of the styrofoam model can be seen in the left image in Figure 3. We choose a virtual camera and we extract the calibration data from this camera. We created 51 renderings of the model with different rotation and translation parameters, see Figure 3. All visible landmarks are labeled manually in all renderings. In contrast to the previous experiment, no noise was added to the projections, leaving us with only the noise of the manual landmarking. The reconstruction is based on the calibration data and subsets of the renderings. The 3D points of the ground truth model are also manually labeled on the 3D model of the styrofoam head, which, in contrast to the previous experiment, could influence the 3D error. The experiment is repeated 100 times for each number of views to determine the robustness of the algorithm.

The second experiment shows the same behavior as the experiment with Gaussian noise with a standard deviation of 1.0 pixels. Adding more views increases the quality of the 3D reconstruction, but for more than 40 views, in this case, the gain is very low for both the 2D and 3D error. The asymptote for the 2D error is around 2.0 pixels, which is somewhere
between the results of the first two experiments. This noise level is similar to a $\sqrt{2} \approx 1.41$ pixels error in both x- and y-coordinates, which is probably the accuracy of the manual landmarking of the 2D dataset. A rough estimation gives us a head size of 300 mm and the size of the head in the frames is around 500 pixels. So each pixel represents 0.6 mm. Our method is able to estimate the landmarks with $(2.16 \cdot 0.6 =) 1.3$ mm precision on average. The average 3D error is 1.22, which is around 0.7% of the size of the head. The results are in line with the results of the first experiment on simulated data. This second experiment shows that our algorithm has similar convergence properties and errors to the experiment on the simulated data, and can therefore be applied on manually labeled realistic face data.

In the last experiment we show that our algorithm can handle real video data using a calibrated camera. We acquired 100 frames of several volunteers in which they slowly moved and rotated their heads in front of a camera. We annotated 20 landmarks in each frame in a semi-automatic manner. Finally we calibrated our camera with 20 frames of a planar calibration board, which provided us the camera calibration data. In the next experiment we use a selection of 50 frames to reconstruct the structure of the face. Since we don’t have 3D ground truth data of our landmarks, we will only use the 2D reprojection error and visual inspection to express the quality of the reconstruction. We ran the experiment two times, with different subsets of views: one using the 50 even frames and another using the 50 odd frames of the first volunteer.
The 2D reprojection error for the even set was 2.13 and the reprojection error of the odd set was 2.54, where the size of the frames is similar to the styrofoam experiment. These results are completely in line with the results of the styrofoam experiment, see the left graph in Figure 4. There is a small variation in the 2D error, but nevertheless the variation seems acceptable compared to the previous results. Visual inspection of the 3D structure shows that both 3D structures are close to each other, see Figure 5. So even with real video data, including calibration, and landmarking, our algorithm is able to reconstruct the position of the 3D landmarks of the face.

5 Conclusion and Future Work

The experiment on the simulated point cloud shows that the quality of our reconstruction depends on the level of noise in the projections. For a small level of noise, around 1.0 pixels, the convergence and robustness of the algorithm seem sufficient. For a larger level of noise the system might become unstable, and even not converge to a useful solution. For Gaussian noise with a standard deviation of 2.0 pixels the algorithm only converges in 75% of the cases. The minimum number of views needed to get sufficient quality for the reconstruction is around 30 views. More views can improve the reconstruction, but this will only give a small improvement. In the second experiment, we showed that manual landmarking leads to an error comparable to a Gaussian noise with a standard deviation of 1.4 pixels. The results of the styrofoam experiment were in line with the simulated reconstructions with Gaussian noise. The third experiment with real video data shows results similar to the styrofoam experiment. The visual inspection of the 3D structure and the 2D reprojection errors indicate that the algorithm is capable of reconstructing real facial structures. In future work we will include the texture information in the reconstruction to get a full 3D model of the face. The full reconstruction allows us to perform facial recognition experiments on 2D faces under pose.

References


Solving Terminal Revocation in EAC by Augmenting Terminal Authentication*

Rafik Chaabouni\textsuperscript{1,2}

\textsuperscript{1}EPFL  
CH-1015 Lausanne, Switzerland  
rafik.chaabouni@epfl.ch

\textsuperscript{2}University of Tartu  
Ülikooli 18, 50090 Tartu, ESTONIA  
rafik@ut.ee

Abstract: In this paper we propose a solution to enable an accurate terminal revocation in the Extended Access Control (EAC). Chaabouni and Vaudenay in [CV09] pointed out the need for an accurate revocation procedure, but failed to provide a complete solution description. We aim at filling this gap. Our solution relies on augmenting terminal authentication with a $t$-out-of-$\ell$ threshold signature provided by neighboring terminals. These terminals will be in charge of checking the revocation status of the requested terminal. As Terminals have a real clock embedded and more computational power than Machine Readable Travel Documents (MRTDs), they are better suited for checking revocation status.

1 Introduction

In response to the initial weak standard for Machine Readable Travel Documents (MRTDs), produced by the International Civil Aviation Organization (ICAO), the European Union has mandated the Federal Office for Information Security (BSI) to provide and maintain a stronger standard for MRTDs. In that regard, the BSI has issued the Extended Access Control (EAC) which provides a stronger privacy protection for MRTDs. Its first initial release [BfSidI12a] was made in 2006, while the last version [BfSidI12b, BfSidI12c, BfSidI12d] was published in 2012. It was believed that with the introduction of EACv2 in 2009, the majority of threats were solved. Unfortunately, Chaabouni and Vaudenay [CV09] pointed out several remaining flaws and threats. The major flaw pointed out was the absence of a good terminal revocation. The other issues are now considered marginal as they are or will be gradually solved with the evolution of previous standards (notably the one from the ICAO [ICAO08, ICAO13]). However, no progress has been made regarding terminal revocation nor terminal authentication.

*This work has been supported from research theme IUT2-1 and European Regional Development Fund through the Estonian Center of Excellence in Computer Science, EXCS.
suggested a solution for terminal revocation but they omitted to give a detailed description. We aim at filling this gap by providing an efficient and secure solution.

Our concern in this paper targets two types of threats. We are first concerned by the threat of a stolen integrated terminal device. These are considered to be Portable Computing Devices (PCD) in the Technical Guideline TR-03110 [BfSidI12b, BfSidI12c, BfSidI12d], when terminal key pairs are explained. An integrated terminal, as explained in [BfSidI09], consists of a single reader with an integrated hardware security module and a proximity coupling device. Moreover a stolen integrated terminal could still be used to read MRTD, as long as its certificate is not expired. This threat even applies with an expired certificate if the date approximated in the MRTD is outdated. Hence there is no real revocation system present for terminals. This is a known problem and is even mentioned by the BSI in [BfSidI09], section 1.2.1:

The disadvantage of this architecture is, that a stolen reader can be used to perform Terminal Authentication at least as long as the current CV certificate is valid.

Secondly, we have to keep in mind that threats come often from an inside attack. This pushes us to study the threat of a compromised terminal that is remained in place, acting maliciously. With the actual standard, a stolen or compromised terminal could be used to target a group of person (e.g. by nationality), or a specific person (e.g. important politicians).

Furthermore, we need to take into account efficiency. In [Fri], it is mentioned that more than 56 millions passengers traveled through Frankfurt airport in 2011. As around half of them are only transfer passengers, and thus do not necessarily need a passport control, we can see that big hubs need to process more than 2 millions passport checks per month.

1.1 Prior and Related Work

Terminal revocation has received little amount of interest as the BSI community is convinced that the Password Authenticated Connection Establishment (PACE) protocol mitigate this threat, as explained in [BDFK12]. Indeed, when executing EACv2, PACE is the initial phase before Terminal Authentication. After its successful completion, the MRTD is ensured that the terminal has knowledge of a shared password, and can proceed with Terminal Authentication. Moreover, the ISO/IEC JTC1 SC17 WG3 mentioned in [ICAO13] that:

In its meeting on 19-21 February 2013 the NTWG concluded that as of the date 01 January 2018 eMRTDs supporting only PACE will be considered to be ICAO compliant.

However no guarantees are provided in the obtention of this password. If the shared password has been obtained by social engineering, or read directly by eavesdropping on the
MRTD, then a successful terminal authentication will allow the stolen terminal to access all sensitive data contained in the MRTD. This issue has been raised by Belguechi et al. in [BLR12]. Unfortunately they concentrate on the protection of biometric data and do not provide a solution for terminal revocation. Li et al. in [LZJX10] also mention the threat of terminal revocation. However they concentrate on presenting the Singapore solution that implicates Authorized Smartcard with Identity Based Cryptography. Hence to solve the terminal revocation issue they require heavy hardware modifications.

1.2 Contribution

Our main idea is to introduce terminal collaboration in order to achieve terminal authentication. Terminal revocation will thus be verified with the help of neighboring terminals. We make use of threshold cryptography to enforce terminal collaboration. We assume that Document Verifiers (DV) in the EAC standard are trusted participants. In our general case, several terminals are present. If the number of terminals is considered too low, our scheme can easily be modified to provide equivalent properties. Moreover the required modifications to enable this method are solely software upgrades and the existence assumption of a communication channel between terminals. Hence no hardware modification is needed in MRTDs. Due to space limitation, we assume that readers are familiar with several topics: the EAC standard [BfsidI12b, BfsidI12c, BfsidI12d], Terminal Authentication, Terminal Revocation, Shamir’s secret sharing scheme [Sha79], Non-Interactive Zero-Knowledge (NIZK) Proofs, Threshold Signature schemes and more specifically the Threshold RSA signature scheme presented by Shoup in [Sho00]. For the sake of completeness, a full version of this paper explaining these topics as well can be found in [Cha13].

1.3 Organization

Section 2 will precise our security assumptions. In section 3, we explain how terminal authentication should be augmented to achieve a realistic terminal revocation. Section 4 will provide the security outcomes and we will finish by some closing remarks in section 5.

2 Security Assumptions

We assume the same structure of participants than the EAC model. However we make some precisions. Each DV is responsible for \( \ell \) terminals (\( \ell \) differs from one DV to the other). DVs play the role of trusted authority amongst their terminals. We assume the existence of secure and authenticated channels between all \( \ell \) terminals. This is easily achieved with public key encryption as it is the same DV, i.e. a trusted party, that issued every terminal key pairs. When a terminal is stolen, its certificate will be revoked. This revocation will disable its use. Moreover, the lack of online connectivity should affect
only CVCAs and DVs as they are Public Key Generators. As such they should be turned offline once their keys setup generation has been achieved ([Sha84]). This is not the case for terminals.

Furthermore, we assume attackers to be computationally bounded. We will focus on threats targeting terminals, as they are somehow neglected in the current EAC. Nevertheless, we assume CVCAs and DVs to be honest. We consider a threshold security assumption, i.e. cases where the adversary can corrupt up to \( t \) terminals among \( \ell \geq 2t + 1 \). We will expect adversaries to be either passive adversaries, where attackers corrupt targets by reading their contents and secrets, or active adversaries, where attackers will additionally change the behavior of corrupted terminals. Lastly we restrict ourselves to static adversaries, meaning that the adversary will select which terminals to corrupt before the start of the protocol. Moreover, the adversary is free to corrupt them when he wants to. When a terminal is corrupted, all his communications will be revealed to the adversary. We set aside cases of dynamic adversaries as the corresponding solutions will induce a high loss in efficiency.

### 3 Augmented Terminal Authentication

Figure 1 gives a sketch of the general structure of our additional part to the current terminal authentication protocol. Our Setup phase is very similar to the original EAC one. DVs have to contact CVCAs from every other country, in order to obtain their DV certificate. The main difference is that now, certificates will contain an additional public key  \( PK_{DV} \) corresponding to a secret key  \( SK_{DV} \) only known by the DV and that will be shared among terminals. Moreover, certificates will contain additional information regarding how many terminals are required to collaborate in order to authenticate themselves (parameters \( t \) and \( \ell \)). When a DV will set up his terminals, he will additionally give them a share \( d_i \) of his secret such that every terminal authentication will require the collaboration of at least \( t+1 \) of them. Hence our scheme tolerates up to \( t \) corrupted terminals. As long as \( t+1 \) honest terminals are available, terminal authentication will be able to proceed. Once the Setup phase has been completed, only terminals and MRTDs are present in the interactions. Hence the DV can be used offline as described in the EAC standard.

DVs are in charge of the setup phase. They will run the key generation algorithm and distribute to each terminal its corresponding secret key, the public key  \( pk \) of the system and the verification keys of all participants. After this step, DVs can be turned offline.

During the terminal authentication and just after the Certificate Chain Validation process, a MRTD will first select a random challenge  \( M \) in the message space \( \mathcal{M} \). He will then challenge the terminal with \( (M||\widetilde{date}) \) where \( || \) denotes concatenation and \( \widetilde{date} \) is the approximation of the current date stored in the MRTD. Moreover  \( M \) must be independent from the MRTD identity, otherwise a tracking privacy threat would rise. Indeed, in this case the signature will prove that a given identity was at a given specific location and time. In order to sign the challenge, the terminal will have to collaborate with at least \( t \) other terminals. The revocation process takes place during the terminal collaboration. It will be
the role of other terminals to determine whether the requesting terminal $T_r$ is revoked or not. As terminals have real clocks and better computation capabilities than MRTDs, they will be able to check this revocation status much more efficiently. Any standard strong revocation mechanism can then be used here. The basic solution is to apply Certificate Validation as described in section 2.5 of [BfSid12d], but with a real clock. More complex solution can also be used such as Certificate Revocation Lists (CRL) or with an Online Certificate Status Protocol (OCSP) if an OCSP responder is set up for terminals. If the requesting terminal is revoked, then his request can be simply ignored. If $T_r$ status is not revoked, then a partial signature $\sigma_i$ can be computed from the partial signature algorithm $\Sigma_i$ and sent to him, possibly with a verification proof $\pi_i$. At this stage, $T_r$ will check, with the partial signature verification algorithm $\Sigma_v$, the validity of each $\sigma_i$. Then, $T_r$ will combine with the combining share algorithm $\Sigma_c$, $t$ valid partial signatures together with its own to create a global signature $\sigma$ on the MRTD challenge. The latter will be sent to the MRTD as a proof of authenticity and non-revocation.

Once the MRTD receives the global threshold signature, he will have to verify it with the global public key of the DV. If the check is successful, he can be ensured that either the terminal knows the DV secret or that he has gone through a threshold signature involving some revocation checks. As we assume the DV to have correctly achieved the initial setup, the MRTD is ensured on the non-revocation status of the terminal.

\[
\begin{align*}
    M \in_R \mathcal{M} & \rightarrow (M||date) \rightarrow \sigma_r = \Sigma_{i=r}^{d_r} (M||\tilde{date}) & \text{Check revocation status} \\
    \text{Check} \Sigma_v(\sigma_i, \pi_i) & \leftrightarrow (\sigma_i, \pi_i) = \Sigma_{i}^{d_i} (M||\tilde{date}) \\
    \text{Check} V_\sigma(M||\tilde{date}) & \leftrightarrow \sigma = \Sigma_c(\{\sigma_i\}_\Psi)
\end{align*}
\]

Figure 1: Terminal Authentication with Revocation

At this point, any efficient and secure threshold signature scheme can be used. In that regard, we suggest to use Shoup’s threshold RSA signature [Sho00]. In this case, the MRTD computation will be dominated by one single exponentiation. The terminal communicating directly with the MRTD and in charge of combining the partial signatures, will have a computational complexity dominated by $(5t + 4)$ exponentiations. However this computational cost can be reduced to $(t + 5)$ exponentiations as explained in section 5.1. For the collaborating terminals, the computational cost is dominated by 3 exponentiations. More details can be found in [Cha13].
4 Security Outcome

Due to space limitation, we refer to [Cha13] for a complete analysis of the security outcomes. Following are the main conclusions.

As our augmented terminal authentication is enforced with threshold signatures, the security achieved is highly dependent on the security of the threshold signature scheme used. We assume a threshold signature scheme that is robust, unforgeable and threshold secure, as the one from [Sho00]. Hence any computationally bounded adversary corrupting at most $t$ terminals will not be able to learn the master secret of the threshold signature scheme ($sk = d_0 = f(0)$). Moreover adversaries will not be able to forge valid signatures on chosen messages.

A stolen terminal will not be able to authenticate itself. A corrupted collaborating terminal will learn no information except that a MRTD with some approximation date has requested an authentication process. However, a corrupted requesting terminal interacting with a MRTD will be granted access to the MRTD sensitive data if the terminal behaves honestly with the other collaborating terminals. As long as at most $t$ terminals are corrupted, the secret key used to authenticate terminals remains protected. Furthermore, the leakage of the secret key can be achieved only if at least $t + 1$ key shares are compromised. These security properties are desirable as they improve the current state of the EAC. By lowering the trust in terminals, we increase the DV level of trust. This is an acceptable change as DVs are less exposed than terminals.

Proactive security can be achieved by frequently renewing the global secret of the threshold signature scheme. This can be done efficiently by resharin the same secret with the means of sharing the “secret” value ’0’ and adding the obtained partial secrets to the previous ones. This method reduces the threat of terminal keys being exposed. In order to compromise the general secret key, an adversary will have to obtain $t + 1$ key shares during the same time frame of a resharin phase. This allows DV certificates to protect their general secret used for threshold signature throughout their entire time validity. Notice that this step is highly efficient if performed by the DV, i.e. the DV generates the additional secret key shares and distribute them to their corresponding terminal. Verification keys will also have to be redistributed to every participants. However, this can be achieved without the need of the DV with secure multiparty computation.

5 Closing Remarks

5.1 Efficiency Tuning

Regarding computational costs, several modifications can be brought to reduce them. We refer readers to [Cha13] for complete details. First, the terminal in charge of combining partial signatures could perform the robustness checks solely if the resulting combined signature is not valid. Hence instead of computing $4t$ exponentiations he would first check the validity of the signature with one exponentiation. Furthermore, minor enhancements
are possible by letting the DV perform some precomputations and storing results in each terminals during the setup phase. The drawback of this method is that it will require a storage space in terminals. In the case of a large $\ell$ (e.g. $\ell > 100$), the threshold signature scheme of Gennaro et al. [GHKR08] will be preferable than the one from Shoup [Sho00] as it will be more efficient.

Furthermore, a small efficiency gain could be obtained by using the threshold signatures of King [Kin00] which is itself derived from the Desmedt-Frankel [DF94] scheme. However, the gain in efficiency is achieved by an increased difficulty to implement them and a higher storage requirement.

5.2 Remarks

Let us mention the existence of multisignatures. These are a type of threshold signature where the identity of signers is provided in the general signature. However, even the latest result in multisignatures that we could use, namely the scheme from Boldyreva [Bol03], would imply an important efficiency decrease.

The overhead in time of our suggested solution should be less than 0.1 seconds, assuming 30 MHz CPU for MRTDs, 520 MHz CPU for terminals, 802.11g wireless communication between terminals (net average of 22 Mbit/s) and 200 Kbit/s communication speed between MRTDs and terminals.

References


Soft Biometrics Database: a Benchmark for Keystroke Dynamics Biometric Systems

Syed Zulkarnain Syed Idrus\textsuperscript{1,2}, Estelle Cherrier\textsuperscript{2}, Christophe Rosenberger\textsuperscript{2}, Patrick Bours\textsuperscript{3}

\textsuperscript{1}Universiti Malaysia Perlis, 01000 Kangar, Perlis, Malaysia
\textsuperscript{2}Université de Caen Basse-Normandie, UMR 6072 GREYC, F-14032 Caen, France
ENSICAEN, UMR 6072 GREYC, F-14032 Caen, France
CNRS, UMR 6072 GREYC, F-14032 Caen, France
\{syed-zulkarnain.syed-idrus,estelle.cherrier,christophe.rosenberger\}@ensicaen.fr
\textsuperscript{3}NISlab, Gjøvik University College, Gjøvik, Norway
patrick.bours@hig.no

Abstract: Among all the existing biometric modalities, authentication systems based on keystroke dynamics are particularly interesting for usability reasons. Many researchers proposed in the last decades some algorithms to increase the efficiency of this biometric modality. Propose in this paper: a benchmark testing suite composed of a database containing multiple data (keystroke dynamics templates, soft biometric traits ...), which will be made available for the research community and a software that is already available for the scientific community for the evaluation of keystroke dynamics based systems. We also built the proposed biometric database on soft biometric traits for keystroke dynamics to suit the experiment. 110 people had voluntarily participated and gave their soft biometrics data i.e. the way of typing, gender, age and handedness.

1 Introduction

Soft biometric traits are physical, behavioural or biological human characteristics, classifiable in pre-defined human compliant categories, which have been derived from the way human beings normally distinguish their peers (e.g. height, gender, hair colour etc ...). Those attributes have a low discriminating power, thus not capable of identification performance. Additionally, they are fully available to everyone which makes them privacy-safe. Keystroke dynamics is a viable and practical way as an addition to security for identity verification. It can be combined with passphrases authentication resulting in a more secure verification system. Soft biometrics allow a refinement of the search of the genuine user in the database, resulting in a computing time reduction. For example, if the capture corresponds to a male according to a soft biometric module, then the standard biometric authentication system can restrict its research area to male users, without considering female ones. Since this work of Jain et al. [JDN04], there have been several other articles dedicated to soft biometrics that can be found in the literature, some of which, can be
mentioned here. The paper [AVL+06], focusing on body weight and fat measurements to enhance a fingerprint based biometric system. An overview can be found in the paper [DVDD10] about soft biometrics, under the form of a “Bag of Soft Biometrics”: the authors make a comparison with the pioneering work of Alphonse Bertillon, whose anthropometric criteria gave rise to soft biometrics, see [Rho56]. This paper proposes some facial soft biometrics and also body soft biometrics, namely weight and clothes colour detection.

Keystroke dynamics is an interesting and a low cost biometric modality as it enables the biometric system to authenticate or identify an individual based on a person’s way of typing a password or a passphrase on a keyboard [GEAR09]. An original approach is presented in the work of Epp et al. [ELM11], strongly linked with the behavioural feature of keystroke dynamics. The authors show that it is possible to detect the emotional state of an individual through a person’s way of typing. In this case, detecting anger and excitation is possible in 84% of the cases. Gender recognition is dealt in the work of Giot et al. in [GR12]: they show that it is possible to detect the gender of an individual through the typing of a fixed text. The gender recognition rate is more than 90% and the use of this information in association to the keystroke dynamics authentication reduces the Equal Error Rate (EER) by 20%. We also did an experiment in [SICRB13] and obtained some interesting and promising results. Our results show that given at least 10 keystroke dynamics templates of users, it is possible to detect their way of typing (using one/two hands), gender, age category and handedness between 65% to 96% correct recognition accuracies performed on a dataset with five passphrases.

The objective of this paper is to present a new data collection of 110 users, both from France and Norway. This new benchmark will be released to the scientific community. We are interested in the criteria that can influence the way of typing of the users. We test if it is possible to predict if the user:

1. types with one or two hands
2. is a male or a female
3. belongs to a particular age category
4. is right- or left-handed

The aim of this paper is threefold. Section 2 deals with the state-of-the-art and the existing keystroke databases. In Section 3, it is devoted to the description of the creation of the database and its main features. The results of the analysis from the database created are discuss in Section 4.

2 State-of-the-art: Public Keystroke Dynamics Dataset

In most studies, researchers use their own dataset, which most of the time, suffers from lack of number of users and sessions. Some keystroke dynamics databases are publicly available in the literature [GEAR09, KM09, GEAR12]. In [GEAR09], several users typed the passphrase “greyc laboratory” on two different keyboards on the same computer during
several sessions. 100 users have provided at least 60 samples each on 5 different sessions spaced of one week (most of the time). In [KM09], several users have typed the password “.tie5Roanl” on a single computer during several sessions. 51 users have provided 400 samples each on 8 different sessions spaced of, at least, one day. This database contains a huge number of samples, but the time interval may be too small to track the variability over a long period. These two databases are the only ones containing enough samples and users to give statistically significant results. In [GEAR12], each user was asked to key-in different logins and passwords. This is the most realistic scenario for keystroke dynamics as real users use different logins and passwords. 83 users have provided 5185 genuine samples (pair of login, password typed by its owner); 5754 impostor samples (pair of login, password typed by a user different of its owner); and 5439 imposed samples (pair of imposed login and password). This database is not the largest in terms of number of users involved, however, it is the only public keystroke dynamics providing different logins and passwords per users. Table 1 summarises this information.

<table>
<thead>
<tr>
<th>Study</th>
<th># users</th>
<th># samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>[GEAR09]</td>
<td>100</td>
<td>60,000</td>
</tr>
<tr>
<td>[KM09]</td>
<td>51</td>
<td>20,400</td>
</tr>
<tr>
<td>[GEAR12]</td>
<td>83</td>
<td>5185 + 5754 / 5439</td>
</tr>
<tr>
<td>Proposed</td>
<td>110</td>
<td>11,000</td>
</tr>
</tbody>
</table>

3 Benchmark Features: Biometric Database

3.1 Requirements

Hardware devices were pre-prepared such as a laptop with two external keyboards (French keyboard for users in France and Norwegian keyboard for users in Norway) i.e. AZERTY and QWERTY, respectively - the layouts are shown in Figures 1(a) and 1(b). An application to collect the keystroke dynamics data was also available. The location and position of the hardware are in a fixed position and immovable throughout the session for the authenticity of the outcomes.

3.2 Acquisition Protocol

An experiment has been performed in two locations: France and Norway, but in fact the subjects originate from 24 different countries who are either studying or residing in one of
the concerned countries. A total of 110 individuals had volunteered to participate in this experiment where 70 of them were located in France and 40 in Norway. They are among students, researchers, faculty members, administration staff, and others. Table 2 shows the statistics repartition of gender, age and handedness among males/females and the number of each category with respect to the studied categories.

According to experts, the best password is a sentence [Dur11]. Having said that, since our study takes place both in France and Norway, we have chosen passphrases instead of sentences, well-known in both countries. Hence, for the purpose of our study for keystroke dynamics in [SICRB13], we present 5 passphrases as shown in Table 3, which are between 17 and 24 characters (including spaces) long, chosen from some of the well-known or popular names, denoted $P_1$ to $P_5$. All the participants are asked to type these 5 different passphrases 20 times. Thanks to GREYC Keystroke software developed at GREYC Laboratory (downloadable at the following address: http://www.ecole.ensicaen.fr/~rosenber/keystroke.html), we are able to capture biometric data. In [GEAR09], the authors describe the GREYC-Keystroke, which is a software developed for allowing the creation of a keystroke dynamics database and highlighted its functionalities. The keystroke application allows to add users to the application; capture the keystroke dynamics of one user several times; change the attended password; and verify the user authentication (when he/she has at least 5 captures in order to define his/her reference). Here, we define two classes of the way of typing; gender category; age category; and handedness category denoted $C_1$ and $C_2$, respectively as follows:

- **Way of typing**: $C_1 = One$ Hand: only one hand is used (right/left depends if the user is right/left-handed person); $C_2 = Two$ Hands: both hands are used.

- **Gender**: $C_1 = Male$; $C_2 = Female$.

- **Age**: $C_1 = < 30$ years old; $C_2 = \geq 30$ years old.

- **Handedness**: $C_1 = Right$-handed; $C_2 = Left$-handed.

### 3.3 Keystroke Data Capture

For any keystroke capture, the captured data are the (i) code of the key, (ii) the type of event (press or release), and (iii) the time of the event. All this information is stored in
Table 2: Repartition of samples

<table>
<thead>
<tr>
<th>User</th>
<th>70 (France); 40 (Norway)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>78 males (47 from France, 31 from Norway); 32 females (23 from France, 9 from Norway)</td>
</tr>
<tr>
<td>Age Category</td>
<td>&lt; 30 years old (37 men, 14 women); ≥ 30 years old (41 men, 18 women)</td>
</tr>
<tr>
<td>Handedness</td>
<td>98 right-handed (70 men, 28 women); 12 left-handed (8 men, 4 women)</td>
</tr>
</tbody>
</table>

Table 3: Passphrases

<table>
<thead>
<tr>
<th>Password</th>
<th>Description</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_1$</td>
<td>leonardo dicaprio</td>
<td>17-char</td>
</tr>
<tr>
<td>$P_2$</td>
<td>the rolling stones</td>
<td>18-char</td>
</tr>
<tr>
<td>$P_3$</td>
<td>michael schumacher</td>
<td>18-char</td>
</tr>
<tr>
<td>$P_4$</td>
<td>red hot chilli peppers</td>
<td>22-char</td>
</tr>
<tr>
<td>$P_5$</td>
<td>united states of america</td>
<td>24-char</td>
</tr>
</tbody>
</table>

The keystroke_data table in the fields rawPress and rawRelease, for respectively press and release events, for each keystroke typing of an entire and correctly typed password. The data are saved following this scheme: code of the key, followed by a space, followed by the time-stamp of the event, followed by a new line and so on, for each events. The interest of storing these raw data, is to permit other researchers to create their own feature extracted data if our data does not fit their requirements. The extracted data features stored in the database are the timing differences between two events of these kinds: press/press, release/release, press/release and release/press, an additional vector resulting of the concatenation of the previous ones and the total typing timing of the password. They are stored in the fields ppTime, rrTime, prTime, rpTime and vector of the table keystroke_data and time_to_type of the table keystroke_typing. The following are keystroke dynamics data consist of information containing the timing values of keystrokes [GEAR09], (see Figure 2):

- $ppTime$ ($PP$): the latencies of when the two buttons (keys) are pressed;
- $rrTime$ ($RR$): the latencies of when the two buttons (keys) are released;
- $prTime$ ($PR$): the durations of when one button (key) is pressed and the other is released;
- $rpTime$ ($RP$): the latencies of when one button (key) is released and the other is pressed;
- $vector$ ($V$): the concatenation of the four previous timing values.
The keystroke template $V$ was used here for the analysis, which is the concatenation of the four mentioned timing values to perform the data analysis by classifying two classes for each category. Hence, five different features/patterns or timing vectors are extracted from each typing sample i.e. $PP, RR, PR, RP, V$. Since these extracted features are already available in the database, we can re-use them directly without having to compute it again.

![Figure 2: Keystroke typing features](image)

### 3.4 Data Collection Process

To begin the process of data acquisition, firstly, metadata such as gender, age, handedness, and country of origin were collected. Note that also information on the keyboard is available (AZERTY or QWERTY), although that was not explicitly asked. Then, after all those data have been obtained, each user has to type each passphrase $P_j$, $j = 1..5$ for each hand class $C_i$ ($i$ = the way of typing: one/two hands), $i = 1, 2$ ($1$ = one hand, $2$ = two hands), 10 times without errors. If there are typing errors, the current entry has to be cancelled and the user have to resume until 10 successful entries for both classes of hand have been recorded into the system. If the user is a right-handed person, he/she only need to use the right hand to key-in the passphrases in a normal typing pace, and similarly for the left-handed people. At the end of the data collection, a total of 11000 data samples (= 5 passphrases x 2 classes of hand x 110 users x 10 entries) are in the proposed biometric database. For each user, 7 out of 10 samples are used for both training and test data. The first three entries for each user were not taken into account because leeway was given to the users to allow them to train themselves for each of the given passphrases.

### 4 Results

Several simulations have been performed with SVM (Support Vector Machine) for computations on several different aspects of the data namely hand recognition, gender recog-
nition, age category recognition, and handedness recognition, where the results have been published in [SICRB13]. However, we further analyse the two countries separately, both users in France and Norway, to see if there are any differences in term of their performances. Here, with substantial amount of data, we only analysed two soft biometrics information namely Hand Recognition and Gender Recognition.

Figure 3(a) and Figure 3(c) illustrate the results of the recognition rates for France and Norway, respectively on different learning ratios with one hand ($C_1$) and two hands ($C_2$) for five different passphrases $P_1$ to $P_5$. In this experiments, the results are promising, since from the ratio of 50% of total data used for training the SVM, the recognition rate for France is between 89% and 96%, and over 90% for Norway. Figure 3(b) and Figure 3(d) illustrate the results of the recognition rates for France and Norway, respectively on different learning ratios with males ($C_1$) and females ($C_2$) for passphrases $P_1$ to $P_5$. The recognition rate, depending on the considered passphrase, is between 66.4% and 68% for France, and between 76.5% and 78.2% for Norway for a ratio over or equal to 50%.

From the previous results, notice that the performances differ between the two soft categories because of the different criteria involved in the analysis mentioned earlier in the article. Generally, the recognition performances for all soft categories have the same trend: at the initial learning ratio, the recognition rates are quite low but then gradually increase after the learning ratios become greater i.e having more data for the learning.
5 Conclusions and Perspectives

Presented here is a new dataset for keystroke dynamics, which will soon be publicly available. This dataset is composed of several soft biometrics data of users. It consists of data on the user’s way of typing by defining the number of hands used to type (one or two), gender, age and handedness. This work is however, the creation of a substantial database, with 110 users, from France and Norway, with 100 samples per user. The results illustrated here could be useful and used as a reference model to assist the biometric systems to better recognise a user by a way he/she types on a keyboard.

References


Abstract: This paper proposes a combined approach for robust face recognition from low resolution images captured by a low-budget commercial depth camera. The low resolution of the facial region of interest is compensated via oversampling techniques and efficient trimming algorithms for the generation of an accurate 3D facial model. Two state of the art algorithms for geometric feature extraction are then utilized, i.e. the estimation of the Directional Indices between all the isogeodasic stripes of the same facial surface via the $3D$ Weighted Walkthroughs ($3D$WW) transformation and the estimation of the Spherical Face Representation ($SFR$). The biometric signature is then enhanced via user-specific cohort biometric templates for each feature, respectively. The experiments have been carried out on the demanding “BIOTAFTOTITA” dataset and the results are very promising even under difficult scenarios (e.g. looking away instances, grimace, etc.). Despite the obvious superiority of the $3D$WW transformation over the $SFR$, it has been noted that the score level fusion of both algorithms improves the authentication performance of the system. On the contrary, only the $3D$WW transformation should be preferred in identification scenarios. Indicatively, the experimental validation on the aforementioned dataset containing 54 subjects illustrates significant succeeds an identification performance of $\sim 100\%$ in Rank-1 and Equal Error Rate of $0.25\%$ regarding the authentication performance in the neutral face experiment.

1 Introduction

It is a common place that security in computer systems is an increasingly critical issue that affects a series of diverse applications, ranging from granting access control in restricted infrastructures to e-commerce transactions. Such applications require reliable personal recognition schemes to either confirm or determine the ID of an individual requesting their services. To this extent, biometrics have been proven to provide unique and powerful advantages over other traditional technologies for ID verification (e.g. PINs, tokens, etc.) that can be easily forgotten, lost or stolen.

Human recognition systems have long been developed based on biometric characteristics of the person, such as in [Ross2003a] and in [Tsalakanidou07]. Although, researchers have long been working on a wide range of biometric traits of several major categories (e.g. hard and soft biometrics, the static biometrics, the activity-related ones etc.), only specific modalities have been proven sufficient to support robust and accurate recogni-
tion performance up to now, i.e. fingerprint-, palmprint-, iris, face- and to an extent gait recognition.

1.1 Current Approaches

From the above, only face and gait recognition methods can be claimed to be less obtrusive since the subjects do not either directly interact with the recording sensor nor do they come in contact with it. Moreover, face recognition, that has long been and still one of the most active research areas in the recent years exhibits much higher recognition results than gait.

In this extend, face recognition has long been and still is one of the most active research areas in the recent years. Performance of 2D face matching systems depends on their capability of being insensitive to critical factors such as facial expressions, makeup, and aging, but mainly hinges upon extrinsic factors such as illumination differences, camera viewpoint, and scene geometry [Zhao2003]. However, provided the inherent limitations of 2D face matching, many researchers have stood for more environmental invariant recognition approaches. As such, the exploitation of the geometry of the anatomical structure of the face rather than its projective appearance has been a growing field of research recently [Smeets2010], via the implementation of efficient 3D transformation techniques [Mian2007] and corresponding face matching algorithms and systems [Berretti2010].

In order to improve the performance of typical biometric systems several methods have been proposed in the literature suggesting either multi-modal fusion [Ross2003b], generation of the user-specific fusion factors [Aggarwal2008], or seamless combination of heterogenous characteristic feature (e.g. anthropometric characteristics) in bayesian inference frameworks [Drosou2012].

1.2 Motivation

The motivation behind the current work is the need for robust and efficient face recognition systems that can address everyday security applications (e.g. PC login, gaming, etc.) with low-cost cameras. Most of the proposed approached have been evaluated with high resolution images or samples dense 3D point clouds, thus making it difficult to verify their validity and applicability in regular real world conditions, where the recorded images are not only captured in low resolution but also in noisy environments.

In this respect, the current paper proposes a thorough preprocessing of the recorded images, so as to improve their quality via efficient denoising and trimming techniques towards increasing their recognition capacity. Moreover, existing geometric feature extraction techniques effectively fused, so as to deliver multi-feature facial recognition with augmented performance under demanding scenarios of real world applications.

2 System Overview

An overview of the structure of the paper is illustrated in Figure 1. Initially, the raw 3D facial information is recorded and iteratively processed, so as to restore the face in frontal view and to deliver a smooth facial surface with no holes in it. Then, two state of the art algorithms (i.e. the 3DWW transform and the Spherical Face Representation) are

---

1Due to the restricted length of the current paper, the Figures referenced herein can be found under the following repository http://www.iti.gr/~drosou/RobustFaceRecLR
utilized for extracting discriminative geometrical features, while the produced biometric signature is enhanced via the generation of the corresponding cohort coefficients. Finally, the recognition decision is based on the comparison of this signature with the gallery template that refers to the claimed ID.

The current paper is organized as follows. A detailed description of the pre-processing algorithms for the generation and smoothing of the reconstructed facial surface is included in Section 3, while the core processing algorithm dealing with the extraction of the facial geometric characteristics is discussed in Sections 4.1 and 4.1, correspondingly, so as to make the paper self-consistent. Next, the estimation of the supplementary cohort coefficients follows in Section 4.3. A short description of the experiment and the utilized database follows in Section 5.1, and the experimental results and the contribution to the recognition performance of the combined approach proposed is exhibited in Section 5.2. Finally, the conclusions are drawn in Section 6.

3 Preprocessing

Initially, a point cloud of the whole captured setting is generated from depth and colour images provided by the Microsoft Kinect Sensor. The facial point cloud that contains useful information for user recognition is segmented from the rest of the image (i.e. noisy information from the facial images, such as areas with hairs or background areas), by preserving these 3D points that are included within the sphere with a radius of $\sim 10\text{cm}$ that is centered on the location of the nose tip (see Section 3.1).

This way, the work of Berretti et al. [Berretti2010] has been enriched with a pre-processing step for drawing face-specific ellipses, as it can be shown in the 1st column of Figure 2(a). Once the facial region is segmented from the background, the triangulation of the remaining point cloud follows.

Yet, before extracting the geometric facial features that will be used as biometric descriptors, some further preprocessing is required. In particular, due to possible occlusions during the image capturing or due to noise induced by the materials of the sensor and the environmental context, some missing facial information may occur. In order to compensate this, a moving average window of $5 \times 5$ pixels is iteratively applied on the surface until all gaps are filled (Figure 2(b)).

Then, the rotation of the facial surface follows via the PCA algorithm on the 3D points, followed by the application of a uniform resampling algorithm. This procedure is iteratively applied until convergence (i.e. no further rotation of the point cloud occurs), as shown in Figure 2(c). The original face 3D points are placed in a perspective field due to the camera lens distortion. This results into facial point clouds with different resolutions. Via uniform sampling it is ensured that the faces have the same resolution before they compared. Differences in resolution of the faces can bias the similarity scores in favour of faces that are more densely sampled. The uniform sampling uses a $2d$ grid on the $x$ and $y$ planes, in which each cell is placed $1\text{mm}$ apart from its neighbours. In our experiments the average gap distance before the point cloud was about $3\text{mm}$, thus the uniform sampling meant also an oversampling operation.

Next, the resulted surface undergoes a final smoothing step via a moving median window with a size $3 \times 3$ pixels. The finally trimmed facial surface $f$ is shown in Figure 2(e).
3.1 Nose tip detection

The step for the detection of the nose tip precedes the background segmentation and is initially based on an initial detection of the location of the nose tip ($N'_{\text{kinect}}$) from the coloured image, as delivered by the Kinect SDK toolkit. However, since the detection accuracy of this algorithm is not sufficient (red spots in Figure 3) a post processing algorithm has been initiated. In particular, all points within a sphere of 4cm around $N'_{\text{kinect}}$ are undergone a PCA transformation and the new nose tip location is calculated as the median value of the $M$ closest points to the origin of the depth axis. Then, by mapping this depth value on the initial surface, one can easily estimate a good approximation of the actual nose tip location ($N(x_0, y_0, z_0)$), as indicated by the blue spots in Figure 3.

4 Geometric Features Extraction

The 3D geometrical characteristics of the face of the user are extracted according to the algorithms presented in [Berretti2010] and in [Mian2007], respectively. In order to deliver a self consistent paper, a short description of these algorithms is included hereafter.

4.1 Intrafacial Directional Indices

The intrafacial Directional Indices of a 3D surface are extracted by estimating the 3D Walking Walkthroughs ($3DW W$) on it, as described below. Initially, the shortest geodesic distances of each point on the facial surface $f$ with respect to the detected nose tip location $N(x_0, y_0, z_0)$ is estimated via the Dijkstra algorithm. This way, isogeodesic stripes of equal width (i.e. 1cm) are formed, concentric and centered on the nose tip (1st row in Figure 4). Thereafter, the so-called $3DW W$ are computed between all pairs of isogeodesic stripes (interstripe $3DW W$) and between each stripe and itself (intrastripe $3DW W$), as described in [Berretti2010]. In particular, the $3DW W s$ are computed as aggregate measures (i.e. Directional Indices) that provide a representation for the mutual displacement between the set of points of two spatial entities (i.e. isogeodesic stripes). Finally, these Directional Indices are cast to a graph representation $x_{3DW W}$, where stripes are used to label the graph nodes and $3DW W s$ to label the graph edges. This way, the face recognition problem is reduced to a graph matching issue, suitable for very large data sets. Thereby, the similarity score $S(x_{3DW W}, \omega)$ between two face-related graphs is the combination of both the inter- and intra-stripe $3DW W s$ similarity measures.

4.2 Spherical Face Representation

The Spherical Face Representation (SFR) [Mian2007] is an integral non-invertible transform that can be seen as an extension of Circular Integration Transformation (CIT) in the 3D space. The SFR is used due to its aptitude to represent meaningful shape characteristics. The location of the nose tip ($x_0, y_0, z_0$) is detected following the approach in Paragraph 3.1 and is used as the center of integration for the utilized transformation method as shown in the 2nd row of Figure 4.

In particular the 3D vector representing the facial surface $f$ is transformed as shown by the following equation to an 1D vector, each element of which represents the number of pixels within the boundaries defined by two successive spheres (i.e. “rings”) with radius
\[ k \Delta \rho \text{ and } (k + 1) \Delta \rho, \text{ respectively.} \]

\[ x_{SFR} = SFR(\Delta \rho, t_1, t_2) = \frac{1}{T_1 T_2} \sum_{k=1}^{K} \sum_{t_1=1}^{T_1} \sum_{t_2=1}^{T_2} f(x_0 + k \Delta \rho \cos(t_1 \Delta \theta) \sin(t_2 \Delta \varphi), y_0 + k \Delta \rho \sin(t_1 \Delta \theta) \sin(t_2 \Delta \varphi), z_0 + k \Delta \rho \cos(t_2 \Delta \varphi)) \]

(1)

for \( k = 1, \ldots, K \) with \( T_1 = 360^\circ / \Delta \theta \) with \( T_2 = 360^\circ / \Delta \phi \), where \( \Delta \rho, \Delta \theta \) and \( \Delta \phi \) are the constant step sizes of the radius and angles variables and finally \( K \Delta \rho \) is the radius of the smallest sphere that encloses the facial surface \( f \).

The similarity measure between two facial surfaces is computed as the \( L_1 \)-distance score \( S(x_{SFR}, \omega) \) between the current signature \( x_{SFR} \) and the template of the claimed ID \( \omega \).

### 4.3 Cohort

An important issue that may lower the performance of fusion based approaches deals with the biometric classes that are not compact with respect to the inter-class distances and not similarly distributed. In particular, when their distributions vary across identities, the recognition threshold may become too stringent for a few classes or too lenient for others. Moreover, their anisotropic distribution around the available samples renders it difficult to set a robust threshold separately for each class.

In this respect, the cohort biometric templates (i.e. neighboring signatures in terms of high similarity factor) have been suggested in [Aggarwal2008], so as to initiate an efficient fusion approach via the definition of user-specific scaling factors that are based on the inter-similarity scores of a genuine signature of the user with the most similar impostor signatures of a training dataset.

Thus, having knowledge of the cohort of each enrolled identity, the similarity of a query with the claimed identity is computed as the ratio of its raw similarity with the claimed identity divided by the raw similarity with the cohort of the claimed identity \( \omega \)

\[ S(x, \omega) = \frac{s(x, \omega)}{s(x, \bar{\omega})} \]

(2)

where \( s(x, \bar{\omega}) \) is the similarity score of the query with the cohort. The raw similarity with the claimed identity can directly be determined using the available matcher. Assuming the cohort set to be of size \( k \), \( s(x, \bar{\omega}) \) is determined using the following max-rule

\[ s(x, \bar{\omega}) = \max\{s(x, \omega^1), s(x, \omega^2), \ldots, s(x, \omega^k)\} \]

(3)

where \( s(x, \omega^1), s(x, \omega^2), \ldots, s(x, \omega^k) \) is the set of similarity scores of the query with the cohort for the enrolled identity \( \omega \).

Herein, the two types of utilized face-related biometric features (i.e. 3DWW and SFR) are treated independently for cohort normalization and 5 cohort signatures are utilized for each separate biometric feature. Finally, the combined score is obtained by fusing the final cohort normalized scores of individual biometrics according to the late fusion technique proposed in [Aggarwal2008], where separate cohort sets for each biometric can be independently selected.
\[ S_f(x, \omega) = f(S_{3DW}(x, \omega), S_{SFR}(x, \omega)) \] (4)

where \( S_f(x, \omega) \) denotes the final combined score of the two biometrics and \( f \) is a fusion function like simple sum rule or product rule. \( S_{3DW}(x, \omega) \) and \( S_{SFR}(x, \omega) \) denotes the cohort normalized scores of the two individual biometrics as determined using Eq.(2).

5 Experimental Results

The selected recognition protocols along with the utilized dataset are described below (Section 5.1), while the performance of the system as a whole, as well as its performance when enabling each geometric feature separately are discussed in Section 5.2.

5.1 Dataset

The evaluation presented in this paragraph refers to the first session of the proprietary dataset “BIOTAFTOTITA”. This database was captured in an indoor environment and includes various poses, angles (e.g. \(-90^\circ, 45^\circ\)), and grimaces (e.g. neutral, smile, scream, etc.) of the 3D recorded faces, under different lightning conditions, for both enrollment (“gallery”) and authentication (“probe”) procedures. Moreover, the first recorded session of the database consists includes 54 subjects. All 3D related recordings have been exclusively performed via the Microsoft Kinect Sensor®. Although the utilized 3D face matching algorithm exhibits high robustness in difficult environmental conditions and strange poses, herein, frames with neutral poses in \(0^\circ\) have been selected for both gallery and probe for the initial evaluation of the proposed algorithm. In particular, for each scenario (e.g. neutral, smile, scream, etc.) multiple sessions have been recorded. Only one of the “neutral” sessions is selected to be used as the gallery and specifically only the five most discriminative frames of it. The distinctiveness is evaluated by creating a confusion matrix with the similarity measures between all frames. This way, 5 frames of this session are selected to be included in the biometric signature, while all other sessions and scenarios are used only for testing.

5.2 Results

In the current session the behaviour of the proposed system is presented as it is evaluated under a series of different scenarios. For reasons of brevity, from now on the term “gallery” will refer to the set of reference recorded images, whereas the term “probe” will stand for the test frames to be verified or identified.

It should be noted that different settings (e.g. stripes number for \(3DW\) estimation or \(k\)-step for the \(SFR\) algorithm, etc.) have been used when the system was functioning for identification purposes (i.e. Cumulative Match Characteristic (CMC) Curves) than when it was functioning in authentication mode (i.e. Receiver Operating Characteristic (ROC) Curves\(^2\)).

In order to exhibit the contribution of the proposed preprocessing steps in the recognition performance of the current biometric system, the following experiments have been conducted. Specifically, the feature extraction algorithm and the matching process have been

\(^2\)In the current paper, a ROC Curve is plotted, without loss of generality, as the function of the False Rejection Rate (y-axis) to the False Acceptance Rate (x-axis).
applied on (i) the raw reconstructed face as depicted in Figure 2(a), (ii) the reconstructed 3D face after the gaps have been filled as depicted in Figure 2(b), (iii) the reconstructed 3D face after face alignment, uniform sampling, gap filling and after the first iteration of the rotating and uniform resampling algorithm and (iv) the fully rotated reconstructed face (Figure 2(d)) without the final smoothing step.

The improvements in the performance of the two state-of-the-art algorithms presented in [Mian2007] and [Berretti2010], as well as the proposed combined approach are presented in terms of authentication and identification capacity in Figure 5(a), Figure 5(b), Figure 5(c) and Figure 6(a), Figure 6(b), Figure 6(c), respectively.

Some slight advances of the no-preprocessing performance over the after-gap-filling preprocessing performance can be explained by the fact, that in the neutral face in frontal-view recording protocol, that is examined herein, the gaps on the reconstructed face of the gallery recordings coincide with the ones in the probe recordings due to similar view angle and do not reflect the general case.

The most common scenario for face recognition refers to the capturing of images depicting the neutral (i.e. no grimace) face of a user in frontal view. Provided that the gallery images have been captured under the same conditions and protocol, it is expected that the performance of the system will be at its maximum. Indeed, as the reader can notice in Figure 7(a) and in Figure 7(b) the identification and the authentication rates are 100% and 0.25%, respectively. As expected, the combination of both types of the aforementioned geometric features (see Section 4) improves both the authentication and identification performance of the system.

The corresponding improvement in the performance of the combined system can be noted in the distributions of the matching scores of the clients (i.e. blue coloured bars) and the corresponding ones (i.e. red coloured bars) of the impostors, as shown in Figure 8(c). The corresponding distribution when each feature is utilized independently are shown in Figure 8(a) and Figure 8(b), for the Directional Indices and the SFR transformation, respectively.

The evaluation of the recognition performance of the proposed system in some more difficult scenarios for user is presented hereafter. At this point, it should be noted that the gallery images are always the same and have been recorded according to the neutral face in frontal view protocol, described above.

In this respect, Figure 9(a) and Figure 9(b) present the identification and authentication performance of the system in two more difficult scenarios. Specifically, one protocol indicates that the users should wear glasses when they undergo a recognition process, while the other one indicates a bad illumination in the environment when the recording is performed.

Two protocols that address cases of facial deformation with respect to the enrollment recordings indicate the yawning and the smiling of the user when his face is recorded in frontal view. The identification and authentication performance of the system in this cases are illustrated in Figure 10(a) and Figure 10(b), respectively.

Finally, a very demanding protocol is the one that deals with the face rotated with respect to the frontal view. The recognition potential falls even more when a grimace is performed by the users in parallel with the rotation of their heads. The identification and authentication performance of the system under these scenarios is analytically illustrated in Figure 11(a) and Figure 11(b), respectively.
6 Conclusion

An efficient and fast (real-time) methodology for robust user recognition based on two different types of geometric features related to the facial surface of the users has been proposed. Hereby, two fast transformations (i.e. $3DWW$ and $SFR$) of the $3D$ facial surface were utilized are seamlessly combined along with the corresponding cohort templates, so that the recognition performance of the system exhibits high potential even in the most difficult scenarios. The most important contribution of the current work refers to the preprocessing of the extremely low resolution facial image (i.e. $\sim 120 \times 90$ pixels per frame), so as to produce a smooth and trimmed continuous $3D$ facial surface. Future work of the current paper include the application of the proposed algorithm on the full version of the “BIOTAFTOTITA” database (Session 1 & 2), which includes 80 different subjects in total. Moreover, the system will be benchmarked in larger databases, so as to verify its robustness for real-world applications.

Acknowledgment

This work was supported by the Greek, nationally funded, research project “BIOTAFTOTITA”.

References


Selecting Discriminative Features with Discriminative Multiple Canonical Correlation Analysis for Multi-Feature Information Fusion

Lei Gao1, Lin Qi1, Ling Guan1,2
1 School of Information Engineering             2 Electrical and Computer Engineering
Zhengzhou University                                         Ryerson University
450001, Zhengzhou, China                               M5B 2K3, Toronto, Canada
email- iegaolei@gmail.com;  ielqi@zzu.edu.cn
lguan@ee.ryerson.ca

Abstract: In this paper, it presents a novel approach for selecting discriminative features in multimodal information fusion based discriminative multiple canonical correlation analysis (DMCCA), which is the generalized form of canonical correlation analysis (CCA), multiple canonical correlation analysis (MCCA) and discriminative canonical correlation analysis (DCCA). The proposed approach identifies the discriminative features from the multi-feature in Fractional Fourier Transform (FRFT) domain, which are capable of simultaneously maximizing the within-class correlation and minimizing the between-class correlation, leading to better utilization of the multi-feature information and producing more effective pattern recognition results. The effectiveness of the introduced solution is demonstrated through extensive experimentation on a visual based emotion recognition problem.

1 Introduction

The effective utilization and integration of multiple information contents presented in different media sources are becoming an increasingly important research topic in many applications with the proliferation of multimedia and the advances in sensing technology. Since single information based pattern analysis and recognition systems only afford low level of performance due to the drastic variation and noisy nature of the acquired signals, it leads to insufficient and inaccurate pattern representation of the perception of interest. However, multiple data contains more information about the semantics presented in the media. The combination of multiple data may potentially provide a more complete and discriminatory description of the intrinsic characteristics of the pattern, and produce improved system performance compared with single information only [LYRYAM-2010].

Therefore, information fusion becomes an emerging and challenging research area in multimedia processing. The major difficulties lie in the identification of the inherent relationship between different information, and the design of a fusion strategy that can effectively utilize the complementary information presented in different channels. A wide variety of methods have been proposed in the literature to address the difficulties.

In general, there are three levels of information fusion: feature/data level, score level and decision level [AA-2003]. Compared with other methods, the advantage of the feature level fusion is as follows. As different feature vectors extracted from the same pattern tend to reflect different characteristics of the pattern, optimally combining these features not only keeps the effective discriminant information of multi-feature, but also eliminates the redundant information to certain degree, which is especially important to classification and recognition of large scale database in high dimensional feature space. It is also the focus of this paper.

Generally, there exist two traditional classes of feature fusion methods which are serial feature fusion and parallel feature fusion [JJDJ-2003]. Recently, there has been extensive interest in the analysis of correlation based approaches for multi-feature information
fusion such as canonical correlation analysis (CCA) [XK-1999], kernel canonical correlation analysis (KCCA) [CP-2001], discriminative CCA (DCCA) [TSZJ-2007] and multi-set canonical correlation analysis (MCCA) [A-2002], which have been applied to audiovisual based talking-face biometric verification [HG-2007], medical imaging analysis [NTYV-2010], handwriting recognition [QSYPD-2005], audio-visual synchronization [MYEA-2007], joint blind source separation [TWV-2009], blind single-input and multiple-output (SIMO) channels equalization [JID-2005]. However, as CCA, KCCA or DCCA method only could deal with the mutual relationships between two random vectors and it lacks discriminative character for MCCA. In order to address the mentioned problems, the approach of DMCCA is introduced in [LLEL-2012], which extracts more discriminative characteristics of multi-feature information. Nevertheless, one important yet not well studied problem in DMCCA is that there is not any reliable approach to select the discriminative features from multi-feature information to achieve better recognition results.

In this paper, we conduct a novel approach for selecting discriminative features in multimodal information fusion based discriminative multiple canonical correlation analysis (DMCCA). The proposed approach identifies the more discriminative features in FRFT domain on a visual based emotion recognition problem. The effectiveness of the introduced solution is demonstrated through extensive experimentation. The remainder of this paper is organized as follows: the conception of FRFT and 2D-FRFT is briefly introduced in Section 2. In Section 3, the analysis and derivation of the proposed approach is presented. The emotion recognition system and experimental results are given in Section 4. Conclusions are drawn in Section 5.

2 FRFT and 2D-FRFT

FRFT is a generalized form of the FT, which can be interpreted as a rotation of the signal in the time-frequency plane [L-1994]. It contains simultaneity the time-frequency information of the signal, and is considered as a new tool for time-frequency analysis, especially in the area of image representation [HD-1993, A-1993]. In the area of facial expression recognition, some researches have shown its superiority with respect to other feature extraction tools [LLEXL-2010]. As different order features of 2D-FRFT contain different time-frequency information [CMH-2000], and the previous studies only focus on the single order FRFT features. Thus, it is reasonable to fuse the different features to improve the recognition results.

In this section, we introduce the definition and properties of the FRFT and the two-dimensional form 2D-FRFT.

Given a signal , its FRFT is defined as:

\[ X_u(t) = \mathcal{F}(x(t)) = \int_{-\infty}^{\infty} x(t) K_{\alpha}(t, u) dt \]  

where \( \alpha = (p*\pi)/2 \) is the rotation angle in FRFT domain and \( p \) is the transform order.

The FRFT can be extended to its two-dimensional form. For a two-dimensional signal \( x(s, t) \), its 2D-FRFT is defined as:
\[X_{\alpha,\beta}(u,v) = F^{\alpha-\omega}_{\beta} \{F^{\alpha}_{\omega}[x(t,s)]\}\]  

where \(\alpha = (p^*\pi)/2\), \(\beta = (q^*\pi)/2\) are the rotation angles, and p, q are the transform orders in the 2D-FRFT. It satisfies the property of additivity in 2D-FRFT [SH-1998]:

\[X_{\alpha,\beta}(u,v) = F^{\alpha-\omega}_{\beta} \{F^{\alpha}_{\omega}[X_{\alpha,\beta}(u',v')]\}\]

In the fields of digital image processing, the two-dimensional discrete FRFT can be implemented by row-column computation as shown in [SH-1998].

3 The Approach of Selecting Discriminative Feature for DMCCA

Let \(X_1, X_2, \cdots, X_Q\) be Q sets zero-mean random samples as:

\[X_i = \{x_{i1}^{(1)}, x_{i2}^{(1)}, \cdots, x_{im}^{(1)}, \cdots, x_{i1}^{(c)}, x_{i2}^{(c)}, \cdots, x_{im}^{(c)}\} \in \mathbb{R}^{m \times N}\]

where \(i\) is the number sequence of the random samples, \(x_{ij}^{(m)}\) denotes the \(j\)th sample in the \(m\)th class, respectively, and \(n_{ij}\) is the number of samples in the \(l\)th class of \(X_i\) set.

\[\sum_{l=1}^{c} n_{ij} = N\]

where \(c\) is the total number of classes.

The aim of DMCCA is to seek the projection vectors \(\omega = [\omega_1, \omega_2, \cdots, \omega_N]\) for feature extraction so that the within-class correlation is maximized and the between-class correlation is minimized. With the definition of DMCCA in [LLEL-2012], it can be written as:

\[\frac{1}{N-1} (C - D)\omega = \rho D\omega\]  

where

\[C = \begin{bmatrix} X_{x1}^{\top} & \cdots & X_{xN}^{\top} \\ \vdots & \ddots & \vdots \\ X_{xN}^{\top} & \cdots & X_{xN}^{\top} \end{bmatrix}, \quad D = \begin{bmatrix} X_{x1}^{\top} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & X_{xN}^{\top} \end{bmatrix}, \quad A = \begin{bmatrix} I_{n_{x1} \times n_{x1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & I_{n_{xN} \times n_{xN}} \end{bmatrix}\]

with \(I = [1,1,\cdots,1]^T \in \mathbb{R}^n\) and \(\omega = [\omega_1^T, \omega_2^T, \cdots, \omega_N^T]^T\).

Therefore, the solution obtained is the eigenvector associated to the eigenvalue of equation (7). That is:

\[\frac{1}{N-1} \text{inv}(D)^*(C - D)\omega = \rho \omega\]  

where \(\text{inv}\) means inverse transform of matrix. However, unless the covariance matrices \(D\) have full rank, the block matrix in Eq. (8) will become singular. An approach [TMH-2003] to dealing with singular covariance matrices and to controlling complexity is to add a multiple of the identity matrix \(\lambda I\), \(\lambda > 0\) to \(D\).
Thus, the general form of equation (8) is written:

\[ \frac{1}{N-1} \text{inv}(D^*)(C - D) \omega = \rho \omega \]  

\[(9)\]

where

\[ D^* = \begin{cases} 
D & \text{when } D \text{ is insingular matrix} \\
D + \lambda I & \text{when } D \text{ is singular matrix} 
\end{cases} \]

From equation (7), the \( \rho \) is the criterion to seek the projection vectors for feature extraction. That is to say, the value of \( \rho \) is the key parameter to the effect of selecting discriminative features, and the larger \( \rho \) corresponds to the more discriminative features, which effectively maximize the within-class correlation and minimize the between-class correlation. While the smaller \( \rho \) corresponds to the less discriminative features. Thus, it is reasonable to evaluate the final information fusion results by judgment \( J(\rho_i) \):

\[ J(\rho_i) = \sum_{i=1}^{N} \rho_i \]  

\[(10)\]

where \( \rho_i \) is the \( i \)th eigenvalue of equation (9).

Furthermore, since

\[ \text{trance}\left( \frac{1}{N-1} \text{inv}(D^*)(C - D) \right) = \sum_{i=1}^{N} \rho_i \]  

\[(11)\]

where \( \text{trance}(\cdot) \) represents the trace of the matrix. Thus, we can calculate \( J(\rho_i) \) through \( \text{trance} \) of the matrix instead of the eigenvalue decomposition when the number of features is large, which could significantly speed up the calculation process.

4 Emotion Recognition System and Experimental Results Analysis

4.1 Emotion recognition system

In this paper, we demonstrate the effectiveness of the proposed method on a visual based emotion recognition problem with FRFT features. Figure 1 depicts a general block diagram of the proposed emotion recognition system.

For facial expression representation, a key image which corresponds to the frame with largest speech amplitude is first identified from the audiovisual. Then we detect the face region from the image frame using a color based method [YL-2008]. Then the face
region is normalized to an image of size of $112 \times 96$. As the large dimensionality of the coefficients, we downsample each subband to a size of $32 \times 32$ to enhance the recognition process. The extracted FRFT features in each of the time segments are then analyzed and selected using the proposed analysis approach. Subsequently, the newly generated features, which represent the multi-set information among different patterns, are concatenated into projected vectors for classification through the method of DMCCA. Then, the nearest-neighbor classifier is used for emotion recognition with visual features.

4.2 Experimental results analysis

To evaluate the performance of the proposed method, we conduct experiments on the RML audiovisual emotion database [YL-2008], which consists of video samples from eight human subjects, speaking six different languages (English, Mandarin, Urdu, Punjabi, Persian, and Italian). In the experiment, a total of 288 samples are selected from the RML audiovisual database, each belonging to one of the six universal emotional states-anger(AN), disgust(DI), fear(FE), sadness(SA), surprise(SU) and happiness(HA). Among the samples, 192 samples are chosen for training set and 96 are chosen for evaluation. As a benchmark, the performances of using single FRFT features in different transform orders (from 0.1 to 1.0) are first evaluated as Figure 2. From Figure 2, it is observed that recognition results are dissimilar in different transform orders due to different time-frequency information, which could be applied in the field of information fusion with expecting to reach higher recognition accuracy. In the experiment, we select the five single features ($\rho=0.3, 0.7, 0.8, 0.9, 1.0$) with better recognition accuracy from Figure 2 as the fused multi-feature. Based on equation (14) (15), the values of $J(\rho)$ fused with different 2D-FRFT features in DMCCA method are shown as Table 1, which means the discrimination of the fused features.

![Figure 2: Results of emotion recognition with single FRFT feature](image)

<table>
<thead>
<tr>
<th>FRFT Features</th>
<th>$0.8, 0.9$</th>
<th>$0.8, 0.9, 1.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J(\rho)$</td>
<td>-0.24</td>
<td>0.12</td>
</tr>
</tbody>
</table>

From Table 1, it can be observed that the value of $J(\rho)$ with four 2D-FRFT features($0.7, 0.8, 0.9, 1.0$) obtains the maximum $8.93$, while the result with two 2D-FRFT features($0.8, 0.9$) arrives the minimum $-0.24$. Based on our analysis, as $J(\rho)$ means the discrimination of the fused features, the recognition result fused with four 2D-FRFT features($0.7, 0.8, 0.9, 1.0$) should outperform others. In order to demonstrate the effectiveness of the proposed method on 2D-FRFT features domain and our
mathematical analysis, the experiments on visual emotion recognition using the method of DMCCA performed with the different number of 2D-FRFT features are shown as follows. In the experiments, it is worthwhile to point out that, as there are 32*32 dimensions in the visual recognition problem, we adopt the first 50 projected dimensions by the approach of DMCCA. The overall recognition rates are shown in Figure 3. Furthermore, we also perform the highest recognition accuracy under different number of features, which can be shown as Table 2. Besides, in order to demonstrate the efficiency of the proposed method effectively, the relation between \( J(\rho) \) and highest recognition results is shown as Figure 4. It can be seen that the experimental results comply with our mathematical analysis presented in the previous section.

![Figure 3](image3.png)

**Figure 3** Emotion recognition experimental results of different number FRFT features

**Table 2** Highest emotion recognition results of different FRFT features

<table>
<thead>
<tr>
<th>FRFT Features</th>
<th>0.8, 0.9</th>
<th>0.8, 0.9, 1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest results(%)</td>
<td>83.33</td>
<td>85.42</td>
</tr>
<tr>
<td>FRFT Features</td>
<td>0.7, 0.8, 0.9, 1.0</td>
<td>0.3, 0.7, 0.8, 0.9, 1.0</td>
</tr>
<tr>
<td>Highest results(%)</td>
<td>90.67</td>
<td>84.38</td>
</tr>
</tbody>
</table>

![Figure 4](image4.png)

**Figure 4** Relation between highest recognition results and values of \( J(\rho) \)

Furthermore, in order to demonstrate the effectiveness of the proposed method in 2D-FRFT features domain, we also implemented feature fusion with the state-of-the-art methods such as CCA, MCCA and DCAA with different 2D-FRFT features for the purpose of comparison. The recognition rates with different projected dimensions are shown in Figure 5 and the highest recognition results with mentioned methods are shown as Figure 6. Based on Figure 5 and Figure 6, it can be seen that the highest recognition rates were achieved with the proposed method on selecting discriminative features in 2D-FRFT features with DMCCA.
mathematical analysis, the experiments on visual emotion recognition using the method of DMCCA performed with the different number of 2D-FRFT features are shown as follows. In the experiments, it is worthwhile to point out that, as there are 32*32 dimensions in the visual recognition problem, we adopt the first 50 projected dimensions by the approach of DMCCA. The overall recognition rates are shown in Figure 3. Furthermore, we also perform the highest recognition accuracy under different number of features, which can be shown as Table 2. Besides, in order to demonstrate the efficiency of the proposed method effectively, the relation between $J(\rho_i)$ and highest recognition results is shown as Figure 4. It can be seen that the experimental results comply with our mathematical analysis presented in the previous section.

### Table 2  Highest emotion recognition results of different FRFT features

<table>
<thead>
<tr>
<th>FRFT Features</th>
<th>Highest results(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.8, 0.9$</td>
<td>83.33</td>
</tr>
<tr>
<td>$0.8, 0.9, 1.0$</td>
<td>85.42</td>
</tr>
<tr>
<td>$0.7, 0.8, 0.9, 1.0$</td>
<td>90.67</td>
</tr>
<tr>
<td>$0.3, 0.7, 0.8, 0.9, 1.0$</td>
<td>84.38</td>
</tr>
</tbody>
</table>

5 Conclusion

This paper has introduced a new approach for effectively selecting discriminative features in multimodal information fusion based the method of DMCCA. The proposed solution identifies the discriminative features from multiple features for information fusion. The effective combination of multi-feature data potentially provides a more complete and discriminatory description of the intrinsic characteristics of the pattern, and produce improved system performance. Experimental results on a visual based emotion recognition problem in FRFT domain demonstrate the proposed method selects more discriminative features effectively, improves the recognition performance, and is computationally efficient. Although we focus on an emotion recognition problem in FRFT domain in this paper, the proposed solution can also be applied to other multimodal or multi-feature related multimedia analysis problems. Another interesting and challenging topic will be investigating kernelized version of the DMCCA based on the Kernel canonical correlation analysis (KCCA) in order to obtain a more effective method to solve nonlinear fusion problems which are more prevalent in information fusion.

6 Acknowledgement

This work is supported by the National Natural Science Foundation of China, No.61071211 and the Canada Research Chair Program.
References


BioHashing with Fingerprint Spectral Minutiae

Berkay Topcu¹,², Hakan Erdogan², Cagatay Karabat¹, Berrin Yanikoglu²

¹ eID Program
TUBITAK - BILGEM - UEKAE
Gebze, Kocaeli, TURKEY 41470
{berk.topcu, cagatay.karabat}@tubitak.gov.tr

² Sabancı University
Tuzla, Istanbul, TURKEY 34956
{haerdogan, berrin@sabanciuniv.edu}

Abstract: In recent years, the interest in human authentication has been increasing. Biometrics are one of the easy authentication schemes, however, security and privacy problems limit their widespread usage. Following the interest in privacy protecting biometric authentication, template protection schemes for biometric modalities has increased significantly in order to cope with security and privacy issues. BioHashing, which is based on transforming the biometric template using pseudo-random projections that are generated using a user-specified key or token, has received much attention as it improves verification accuracies over using only the biometric data, allows template revocation and preserves privacy. In our work, we develop a new BioHashing scheme for fingerprints. A fixed-length feature vector is required in order to design a BioHashing scheme. In the literature, most of the studies on fingerprint BioHashing uses features extracted from fingerprint texture. On the other hand, our new BioHashing scheme is based on minutiae based feature vectors. We use the spectral minutiae representation for obtaining a fixed-length feature vector for a fingerprint sample. Then, we use a random projection matrix, which is generated from user’s key/token, in order to generate a BioHash vector. We propose to randomly project each column of the spectral minutiae feature matrix via a single matrix which allows fast bit string extraction and adaptive quantization. Experiments on FVC2002 databases show the promise of the proposed system for fast and secure verification.

1 Introduction

Personal authentication using biometric systems that use physiological or behavioral characteristic are becoming widespread. Among various biometric modalities, fingerprint is preferred in many settings, due to its distinctiveness and performance, as well as the practicality and low cost of the fingerprint readers. Most fingerprint recognition systems are based on matching fingerprint minutiae which are the endpoints and bifurcations of fingerprint ridges. They are known to remain unchanged over an individual’s lifetime and allow a very discriminative classification of fingerprints [MMJP09].

Increasing widespread use of fingerprint identification, as well as other biometric modal-
ities, raises privacy concerns significantly [JNN08] and protecting biometric fingerprint templates (or mostly minutiae templates) becomes a requirement. Combining fingerprint recognition with template protection puts two important constraints to a fingerprint recognition system [XV10]. No relative pose-alignment of two fingerprints is possible due to the encrypted storage and a fixed-length feature vector is required as input of template protection schemes such as fuzzy commitment and helper data schemes [TAK+05, Jue07].

Extracting fixed-length feature vector from fingerprints has been an interesting research topic in the last decade and several texture-based methods are proposed. In [JPHP00], Jain et al. presented FingerCode that is based on the fingerprint texture. Using a bank of Gabor filters to capture both local and global information, a compact fixed-length feature vector (FingerCode) is formed. The author concluded that FingerCodes are not as distinctive as minutiae. Similar to FingerCode, [TAK+05] suggested a quantization algorithm based on local orientation of ridges. Some other studies that follow the same idea (texture patterns of fingerprints) are [BCK08] and [CV11], and they require several samples per user at the enrollment to extract stable feature vectors.

Apart from texture-based approaches, several methods for fixed-length feature extraction from minutiae have also been investigated. Park et al. used histogram of the quantized distances between all possible minutiae in the ROI determined by the core point as a reference [PSBL05]. The method is very sensitive to minutiae insertions and deletions and minutiae orientations are not taken into account; therefore, the performance is not satisfying. [DKM+07] presented a statistical model of the relationship between the enrollment biometric and noisy biometric measurement taken during authentication and designed specific encoding and decoding algorithms to deal with displacement, erasure and insertion of minutiae using some stored public information about the biometric template (helper data). In [NRV10], numbers of minutiae in local cuboids are used for binary representation. However, in addition to minutiae information, ridge orientation map and ridge frequency is also used in this study and a preliminary registration step before comparison is required. In [BD10], a fingerprint is characterized by its similarity with a fixed number set of representative local minutiae vicinities. This approach by representative leads to a fixed length binary representation, and, as the approach is local, it enables to deal with local distortions that may occur between two acquisitions. Capelli et al. recently presented the Minutiae Cylinder-Code [CFM10] where a minutiae cylinder record the neighborhood information of a minutiae as a 3D function. A cylinder contains several layers and each layer represents the density of neighboring minutiae along the corresponding direction. Another study represents minutiae in spectral domain (Spectral Minutiae Representation [XV10]) and creates a fixed-length feature vector using log-polar transform.

In this work, we use the spectral minutiae representation to create a fixed-length feature vector, using minutiae location and direction information $(x, y$ and $\theta$) in ISO 19794-2 standard [ISO05]. The obtained fixed-length template for a fingerprint sample is then combined with pseudo-random data, generated from a user specified key or token used as a seed, to generate a unique code per person, using the BioHash scheme [JLG04]. Mixing of pseudo-random number and biometric data - BioHashing - leads to protection of the biometric template against biometric fabrication without possession of the corresponding token or knowledge of the randomization. Token-based randomization also enables
revocation of one’s biometric template via token replacement.

This paper is organized as follows. First we review the spectral minutiae representation and give details of spectral minutiae representation in Section 2. Section 3 presents the basic BioHashing idea and our approach for spectral minutiae representation. In Section 4, we discuss the experimental results. Finally we draw our conclusion in Section 5.

2 Spectral Minutiae Representation

The spectral minutiae representation of a minutiae set is a fixed-length feature vector that is invariant to translation, rotation and scaling [XV10]. These characteristics enable the combination of fingerprint recognition systems with template protection schemes and allow for fast minutiae-based matching. The spectral minutiae representation can be applied on minutiae sets without any other requirement, therefore it is compatible with most of the existing fingerprint databases and minutiae-based fingerprint verification systems.

Complex spectral minutiae (SMC) is one of the three possible spectral minutiae representations, proposed by Xu et al. in order to obtain a fixed-length feature vector using minutiae location and orientation [XV10]. The other two alternatives are location based spectral minutiae (SML) and orientation based spectral minutiae (SMO). In SMC, each minutiae is represented by a Dirac pulse and in order to reduce the sensitivity to small variations in minutiae locations in the spatial domain, a Gaussian low-pass filter is used to attenuate the higher frequencies. This corresponds to a convolution in the spatial domain where every minutia is now represented by an isotropic two-dimensional Gaussian function with standard deviation $\sigma_C$. Minutiae locations on a fingerprint image together with the Gaussian functions are illustrated in Figure 1. The minutiae orientation is incorporated into this representation by assigning each Gaussian a complex amplitude $e^{j\theta}$, where $\theta_i$ is the orientation of the corresponding minutiae. For a set of $Z$ minutiae with locations $(x_i, y_i)_{i=1}^{Z}$, by evaluating the magnitude of the Fourier spectrum (1) on a polar-logarithmic grid, we
obtain a complex spectral representation $M_C(w_x, w_y; \sigma_C^2)$:

$$M_C(w_x, w_y; \sigma_C^2) = \exp \left( -\frac{w_x^2 + w_y^2}{2\sigma_C^2} \right) \sum_{i=1}^{Z} \exp(-j(w_x x_i + w_y y_i) + j\theta_i) \right|$$

(1)

where $w_x$ and $w_y$ are the spatial frequencies in the $x$ and $y$ directions.

The Fourier spectral magnitude is mapped onto a polar-logarithmic coordinate system as

$$\lambda = \sqrt{w_x^2 + w_y^2}$$

and

$$\beta = \arctan\left( \frac{w_y}{w_x} \right)$$

where $\lambda$ corresponds to the radial direction and $\beta$ corresponds to the angular direction. In the radial direction $M = 128$ samples are used between $\lambda_l = 0.05$ and $\lambda_h = 0.63$. In the angular direction $N = 256$ samples are used between $\beta = 0$ and $\beta = 2\pi$. The resulting complex spectral representation of a minutiae set is a $128 \times 256$ matrix.

3 BioHash for Protecting the SMC Template

BioHashing, applied to fingerprint biometric by Jin et al. [JLG04], is a two factor authentication approach that combines fingerprint feature with a user specified key/token and generates a unique compact code per person. A bit string from biometric data is created by inner product between the pseudo-random number sequence generated using the key as the seed and fixed-length fingerprint feature vector and deciding each bit on the sign of the result after subtracting a threshold.

A fixed-length biometric feature vector, $f \in \mathbb{R}^d$ with length $d$, is reduced down to a bit string $b \in \{0, 1\}^p$, with $p$ the length of the bit string ($p < d$), via a pseudo random pattern, $r \in \mathbb{R}^p$ whose entries are uniformly distributed between $-1$ and $1$. Details of this operation can be found in [JLG04].

In our study, to apply the BioHash scheme to complex spectral minutiae features, we reduce an $M \times N$ spectral fingerprint feature ($M_C$) down to a bit string, $b \in \{0, 1\}^p$. Each column of $M_C$ is a $M$-dimensional column vector. Randomly projecting each column of $M_C$ to $k$ dimensions and then thresholding the resulting vector, we obtain a $k$-length bit string. Mean value of the $k$-dimensional feature vector is used as the threshold for the quantization. We apply the same procedure to each column of $M_C$ and concatenate the bit strings to create $p$-length bit string, where $p = k \times N$.

Each column $f_n$ of this matrix is a 128-dimensional column vector and it is reduced to $k$ dimensions by calculating its multiplication ($R \cdot f_n$) with the random projection matrix, $R$ (which is a $k \times 128$ matrix). Thresholding the resulting $k$-dimensional feature vector by using its mean value as the threshold, we obtain a $k$-length bit string. The outputs of each column of $M_C$ are then concatenated in order to create a bit string of length $k \times 256$. In this work, we evaluated different values of $k$ to obtain a high verification accuracy with the smallest feature vector and used $k = 4$ resulting in a 1024-bit final feature vector (Figure 2).
4 Experimental Results

4.1 Experimental Settings

The proposed algorithm has been evaluated on publicly available FVC2002 fingerprint databases, namely DB1A, DB2A and DB3A [MMC+02]. DB1 and DB2 consist of fingerprint images captured with optical sensors and images in DB3 are captured with a capacitive fingerprint sensor. We have selected these databases in order to evaluate the performance of the proposed method for different image capturing technologies and left DB4 out in our experiments because it is a synthetic fingerprint database.

For the performance evaluation we adopt the equal error rate (EER). The minutiae sets are obtained by a commercial automatic minutiae extractor (Verifinger 4.4 SDK). We propose to use our algorithm in a high security scenario. In FVC2002 databases, some of the samples were obtained by requesting the users to provide fingerprint with exaggerated displacement and rotation. In a high security scenario where the user is aware that cooperation is crucial for security reasons, (s)he will be cooperative. Therefore, only four out of eight samples are chosen for each subject (1-2-6-7 for DB1 and DB3, 1-2-7-8 for DB2). Following the verification setting described in FVC competitions, we used all possible combinations for matching genuine pairs and the first sample of each subject is chosen for imposter matches (a total of 600 genuine and 4950 imposter matches for 100 subjects).
4.2 Results

We tested the BioHash of spectral minutiae representation on three databases. For comparison, we also included results from two other matching methods: i) matching two fingerprints based on the correlation of their complex spectral minutiae (SMC-Correlation) and ii) the minutiae-based commercial matcher which is also used for minutiae extraction. Equal error rates (EER) for both methods together with minutiae-based matching results on three databases are given in Table 1. As can be seen in this table, we obtain 0% EER for all the databases, when we apply BioHashing over the spectral minutiae features.

<table>
<thead>
<tr>
<th></th>
<th>SMC-Correlation</th>
<th>Minutiae Matching</th>
<th>SMC-BioHash</th>
<th>Stolen Key</th>
</tr>
</thead>
<tbody>
<tr>
<td>DB1</td>
<td>6.50%</td>
<td>0.50%</td>
<td>0.00%</td>
<td>14.77%</td>
</tr>
<tr>
<td>DB2</td>
<td>6.47%</td>
<td>0.83%</td>
<td>0.00%</td>
<td>13.10%</td>
</tr>
<tr>
<td>DB3</td>
<td>11.68%</td>
<td>2.50%</td>
<td>0.00%</td>
<td>26.46%</td>
</tr>
</tbody>
</table>

In addition, we evaluated the performance of the proposed scheme on a stolen key scenario, where an unauthorized imposter acquires the secret key/token of a genuine user but does not have the claimed person’s fingerprint information. In this case, the imposter sends his/her fingerprint template and the secret key/token of the genuine user in order to be authenticated as the genuine user. This is a serious threat to the system as the pseudo-random vectors generated using the secret key has a considerable influence on the generated bit string, therefore on the matching score.

Assuming the key to be unknown at all times (never stolen) makes using the biometric unnecessary for real authentication scenarios. In order to analyze the effect of the key/token on the resulting bit strings, we have conducted experiments with stolen key scenario where an imposter attempt has the same secret key with the user that (s)he is intended to authenticate as (Table 1). While the error rates are considerably high in this case, they are in the same range as other results obtained with fingerprint BioHash implementations. For instance the straightforward BioHash implementation using FingerCode ([JPHP00]) reported in [LN06] achieves 15%, 15% and 27% EER on FVC2002 DB1-DB3 databases respectively for the stolen key scenario (see BASE row in Table 5 of the reference). Our error rates for this case are slightly better in the same scenario. The same authors report improved results (7%, 6.8% and 22% on FVC2002 DB1-DB3 respectively) with a classifier combination approach that aims to reduce the stability issues of BioHash, presumably with a system significantly slower and larger than ours [LN06].

5 Conclusion

In this study, we proposed a BioHashing approach for fingerprint identification based on minutiae information. Using the spectral minutiae representation of a fingerprint minutiae
set, we create a fixed-length bit string by randomly projecting spectral minutiae feature vectors. With this approach, one can obtain perfect separation between genuine and imposter population and the system provides 0% equal error rate, which is desired for all identity verification systems. In addition, in case the secret key of a valid user is stolen, the system allows acceptable error rates for imposter authentication attempts with a valid secret key. Also, biometric revocation becomes feasible through secret key (token) replacement, which addresses the cancellability issue.

Our main contribution is providing the first implementation of the BioHash scheme with the spectral minutiae representation. The proposed scheme is computationally fast as it only uses column-wise random projection of the spectral minutiae matrix, while achieving the 0% EER in the verification scenario. The original spectral minutiae features are 8096-dimensional (128 × 256) and in order to create a 1024-bits string, one needs to generate a random projection matrix of size 1024 × 8096. Instead, we propose to use a single 4 × 128 random projection matrix for multiplying with each column of SMC (which are 128-dimensional column vectors). This results in a computationally low random projection operation as well as adaptive thresholding for each column of SMC, instead of generating a larger projection matrix (which takes much time to generate as orthonormalization of vectors is required for higher number of vectors - 1024 instead of 4) and using a single threshold for quantization.

6 Acknowledgments

This work has been performed by the BEAT project 7th Framework Research Programme of the European Union (EU), grant agreement number: 284989. The authors would like to thank the EU for the financial support and the partners within the consortium for a fruitful collaboration. For more information about the BEAT consortium please visit http://www.beat-eu.org.

References


Usability Analysis of Dynamic Signature Verification in Mobile Environments

Ramon Blanco-Gonzalo, Raul Sanchez-Reillo, Oscar Miguel-Hurtado, Judith Liu-Jimenez
Carlos III University of Madrid
University Group for Identification Technologies (GUTI)
Avda. Universidad, 30; 28911 Leganes (Madrid), SPAIN
{rbgonzal, rsreillo, omiguel, jliu}@ing.uc3m.es

Abstract: Biometrics has experienced significant advances during the last years becoming involved in several systems. Nevertheless, many of these systems are rejected by users due to a lack of usability. This indicates that the user-system interaction is a factor which needs to be improved in biometric environments. Even less studied, are the dynamic modalities (gait, signature, voice, etc.) which include behavioural conditions which exhibit greater variability from person to person. This paper showcases some of the latest efforts made by authors in the analysis of usability in dynamic signature verification (DSV). Several outcomes (such as the importance of the styluses used or stress in users) were extracted in order to establish guidelines for future developments.

1. Introduction

The use of biometric recognition is extended worldwide as a trustable way to identify individuals and guarantee security. Biometrics systems are used in many places such as airports, points of sales, institutions or companies and its use is being increased. However, in the intent to develop systems with high performance the users’ satisfaction is most of the times put aside. A non-usable system has not only repercussions in performance but in users’ acceptance of the technology also. Therefore, it is necessary to involve users from the first stages of the development for designing user-centric systems and improve the whole throughput in consequence [AS99]. At this point, various works were done in the line of decrease the misuse of biometrics and increase users’ satisfaction. The ISO 13407:1999 [ISO99] was taken as a basis for measuring usability in most of those works. They defined usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use”. The National Institute of Standards and Technology (NIST) made some experiments emphasizing in ergonomics to better capture user traits. For instance, in [TSS08] they measured the usability of the face image capturing system at the US ports of entry. Kukula et. al designed a model, the HBSI (Human Biometric System Interaction) [Ku08], where the interaction between the user and the system is studied through ergonomics, usability and signal processing. Also, guidelines for applying HBSI to DSV were published in [BGE11]. DSV is gaining
popularity especially for processes where the user is familiar with signing such as administration processes or points of sales.

Our framework pursuits to categorize the factors which could influence the usability of systems in biometrics (especially in dynamic modalities). In DSV, due to the variability of solutions and the increasing interest this modality arouses in users and developers, authors have carried out a series of works including the analysis of different devices, platforms, technologies, scenarios, etc. In addition, other influential factors such as visual feedback or user stress were also studied. As long as there are only a few usability works made in dynamic modalities, the research showed in this paper supposes several novelties (such as the effects of the stress in users). Most of these works were gathered, analysed thoroughly and summarized for this paper. In Table 1 the features studied in each work are shown. This paper is divided in 3 sections. In section 2 the usability concepts of user interaction in DSV are studied. Finally, all the conclusions extracted are given in section 3.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Visual Feedback</th>
<th>Posture</th>
<th>Devices</th>
<th>Styluses</th>
<th>Stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experiment 1 [Sa12]</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment 2 [Bl13]</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Experiment 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

2. Usability concepts of interaction in DSV

Measuring usability in behavioural modalities tends to be harder than in static biometrics. This is because presentations require a bigger effort by users and the quantity of possible misuses increases substantially. Furthermore, in many dynamic modalities there is not a defined “correct” way to present the trait (e.g. there are not specific, complete and commonly approved guidelines to sign in DSV, apart from accomplishing some criteria such as performing the signature within the signing area boundaries). Thus, the range of possible user-system interaction errors is bigger than in the static modalities (i.e. fingerprint), where the process is more straightforward. Specifically, for DSV, the user-system interaction errors accounted for efficiency and effectiveness in the works presented here are: signing in the air (over the device without touching the screen), signing out of bounds, repeating strokes and leaning the wrist on the signing area. All of these errors influence negatively in both performance and usability. They were measured by video recordings and by operators’ notes. Some of these errors are automatically measured during Experiment 3.

In this section, works made by authors in the field of usability in DSV are detailed. Furthermore, the main accessibility concerns affecting this modality are also described.

1 Not yet published
There are multiple possible influential factors in usability concerns in the DSV, including not only devices but users’ mood or the ceremony of the occasion. Works explained below include the study of the influence of the device, stylus, visual feedback, users’ position and stress situations as some of the most decisive factors. In Table 2 are the devices used, their features and the obtained performance (the superscript represents the experiment showed in Table 1). The performance results were obtained with a DTW (Dynamic Time Warping)-based algorithm [PCV09]. No skilled forgeries were used in these experiments, using random forgeries for obtaining the EER. The captured signatures were stored in a database during the evaluations for later comparisons.

Table 2. Devices used, their features and performance

<table>
<thead>
<tr>
<th>Device</th>
<th>Experiment</th>
<th>Note$^{1,3}$</th>
<th>STU$^1$</th>
<th>Intuos$^1$</th>
<th>Asus$^1$</th>
<th>iPad$^{1,2}$</th>
<th>Playbook$^1$</th>
<th>Tab$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual Feedback</td>
<td></td>
<td>Excellent</td>
<td>Excellent</td>
<td>-</td>
<td>Bad</td>
<td>Good</td>
<td>Bad</td>
<td>Good</td>
</tr>
<tr>
<td>Stylus or Finger based</td>
<td></td>
<td>Stylus$^{1,3}$</td>
<td>Finger$^1$</td>
<td>Stylus</td>
<td>Stylus</td>
<td>Stylus</td>
<td>Stylus$^2$</td>
<td>Finger</td>
</tr>
<tr>
<td>EER (%)</td>
<td></td>
<td>0.58$^1$(finger)</td>
<td>1.62$^1$(stylus)</td>
<td>0.84$^3$</td>
<td>0.63</td>
<td>1.45</td>
<td>1.10</td>
<td>0.21$^2$</td>
</tr>
</tbody>
</table>

2.1 Effectiveness, Efficiency and Satisfaction

These factors applied for measuring usability are provided by the ISO 13407:1999 and defined in this section. The learnability is included also as it is considered as quite relevant by authors. Effectiveness and efficiency are quantitative factors calculated through objective measures but satisfaction is a qualitative metric obtained from the satisfaction forms and users’ impressions and reactions. These three factors are closely interrelated: a high number of effectiveness or efficiency errors involve a decrease in the user satisfaction.

-Effectiveness: Is the task completion by users. In order to calculate the effectiveness errors in these works all the wrong signatures are divided by the total amount of them. This factor influences directly on the FTA (each wrong signature involves a FTA but not all the FTAs are produced by an effectiveness error) and therefore in the system throughput also.
Table 3. Accessibility concerns in DSV and possible solutions tested in the different experiments

<table>
<thead>
<tr>
<th>Problem</th>
<th>Description</th>
<th>Solution experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Visual</strong></td>
<td>Visual feedback</td>
<td>Devices have to show the signature ¹</td>
</tr>
<tr>
<td></td>
<td>Sign out of bounds</td>
<td>-Vibration</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-Beeps</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-Text messages as warnings ¹, 2, 3</td>
</tr>
<tr>
<td></td>
<td>Small screen</td>
<td>-Use a bigger device</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-Increase the text size</td>
</tr>
<tr>
<td><strong>Auditory</strong></td>
<td>Instructions</td>
<td>-Text messages ¹, 2, 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-Vibration</td>
</tr>
<tr>
<td><strong>Physic</strong></td>
<td>Hands shivering</td>
<td>-Place the device on a table / on the knees ¹, 2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-Choose stylus or fingertip ¹, 2</td>
</tr>
<tr>
<td><strong>Cognitive</strong></td>
<td>Difficulties to understand the software</td>
<td>-Previous detailed instructions ¹, 2, 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-Basic and intuitive software ¹, 2</td>
</tr>
</tbody>
</table>

Efficiency: Is the completion of tasks on time. The more complicated the task to perform, the more time it requires. Furthermore employ too much time in the biometric system is frustrating for users. This parameter is measured through the time employed signing.

Satisfaction: Is the user experience in the evaluation. It was measured through satisfaction forms and users impressions during the evaluations. When users are not comfortable interacting with the system they deliver bad quality signatures decreasing the system throughput. Learnability: Is the users’ ability to recall tasks. Due to all the experiments were divided by sessions it is feasible to measure whether the user remember how to proceed or not. A high learnability degree involves higher satisfaction, less errors and consequently better performance results.

2.2 Accessibility

Regarding accessibility, the DSV in mobile devices could be a solution for those users with different types of disabilities or diseases affecting the motor skills (such as Parkinson’s disease) or the elderly. For instance, as it was demonstrated in this work, scenarios where users handle the device with their hands offer reliable results. The architectural barriers can be reduced with the use of the mobile device to authenticate the user. Accessibility concerns and its possible solutions (according to the obtained results) are shown in Table 3.

2.3 Devices and styluses influence

The devices used to acquire signatures have to capture time series coordinates in order to apply a DSV algorithm. The Experiment 1 includes signatures gathered with 5 common mobile devices such as smartphones or tablets (Samsung Galaxy Note and Tab, Asus Eee PC TM101, iPad and Playbook) and 2 specific digitizers (Wacom STU 500 and Intuos 4). This experiment was a usability evaluation made to study the influence of
various devices features such as technology (capacitive / resistive), screen size or operative system. In addition a comparison between signing with a stylus or directly with the fingertip was made in order to check if the DSV should be split in two sub-modalities (stylus and finger). As a result authors did not find, according to the performance outcomes (Table 2), any decisive influential parameter apart from the technology: capacitive devices offer always better performance. Stylus-based devices returns slightly better results but it is not consistent in all the devices tested, so that a unique modality can be considered. The best results were obtained with the stylus-based device used as reference (STU - EER = 0.63%) and with the iPad (EER = 0.21%) coinciding with the most preferred by users (extracted from the satisfaction form).

As the results obtained with the iPad (both performance and usability) were more than acceptable, authors used this device for the next experiment, but including different styluses for signing (all of them conductive, as it is required to print in the iPad). Generally, users feel more comfortable using a stylus (As pointed out in Experiment 1 where users expressed it through a satisfaction form). Therefore, in the Experiment 2 three of the most common styluses showed in Figure 2 were used to sign on an iPad within a usability evaluation of 3 sessions. The efficiency and effectiveness results, calculated as in ISO 13407:1999, show a clear influence of the styluses order (overall at the first session, where users have not acquired enough skills yet). The styluses order was rotated to avoid worst results in some of them. The results show that the preferred styluses (1 and 2) offer the better performance. Regarding the learnability, users became more habituated to the process between session 2 and 3 (once both timing and FTA stabilised) indicating that a training process is necessary. Users scored the experiment (including the 3 styluses) with 4.28/5 in the satisfaction form (Figure 3) at the end of the evaluation, showing predisposition to use this kind of systems daily for signing procedures.
2.4 Users’ position influence

The signing process can be completed in several ways and there is not an ideal position or way to sign as demonstrated in Experiment 1 and Experiment 2. That is one of the reasons why measuring usability is challenging in DSV (and in the rest of biometric dynamic modalities): the variability of movements and reactions is high. Though users consider staying seated and having the paper/device over a table as the most common and comfortable way to sign, the range of possibilities have increased significantly with the launch of mobile devices. According to this, authors made a usability evaluation in Experiment 1 to test which scenario (user-device-position) is the most appropriated in DSV. The proposed scenarios are shown in Figure 3. Regarding performance, the best results were obtained in different scenarios for the different devices, being the ones where the user has to handle the device by herself/himself the best in performance for the lighter devices (e.g. iPad EER = 0.92% Figure4-04 / iPad EER = 0.6% Figure4-05 in the Experiment 1). By contrast, users prefer the scenarios where the device rests on a table in all the cases. The scenarios order was influential also during the evaluation, obtaining the first completed the worst performance results, due to the lack of habituation at the beginning. The effectiveness and efficiency errors increase in the scenarios where the device is handled by users: without a fix support for devices the number of failed signatures increases. Experiment 2 presents another study of the influence of the users’ position, though only considering 3 scenarios (in Figure 4: 01, 02 and 03). The best results, including performance (EER = 0.13%), efficiency and effectiveness were achieved in scenario 03, where the device was resting over a slope surface. This scenario design, typical from many points of sales was also well accepted by users. However in the Experiment 1, signing with the iPad but with the finger the best performance results were achieved in the scenarios where users have to handle the device. It indicates stylus conditions the way of sign and the final signature.

2.5 Visual feedback

The visual feedback is defined as the image of the signature that users’ receive while performing that signature. The ideal visual feedback is the one received signing with a pen on a paper as it is the traditional way and the one expected at the time to sign. With the use of digitizers and mobile devices for signing new concerns such as latency appeared increasing errors and misuses. The visual feedback scores shown in Table 2 (bad, good, or excellent) were extracted from the satisfaction forms of the Experiment 1 (Figure 3). Within the Experiment 1 and Experiment 2 this effect was tested over different mobile devices where the screen refreshing frequency varies meaning different visual feedbacks as stated in Table 2. Analysing the performance achieved and the
satisfaction forms, it is shown that visual feedback is highly influential. Those devices that provide excellent or good visual feedback are better scored by users and achieve better performance than the rest. The only device that does not provide visual feedback (Intuos 4) is the worst in performance. Furthermore, most of the users expressed their low confidence in this “blind signature”.

2.6 Study of the stress on users

Users’ mood seems to be one of the most influential parameters at the time to present the biometric trait to the system (nerves or time constraints, among other factors, can involve a different user-system interaction). In DSV one of the most considered influential parameter is user stress. In many situations such as in malls or in post offices, people are encouraged to sign fast and carelessly, which involves stress and anxiety. Therefore, in order to test how these users’ reactions affect the results and the whole system throughput a usability evaluation (not yet published) was carried out. In the Experiment 3 users signed in a Samsung Galaxy Note with a stylus (Samsung proprietary) in a typical interface first (a blank space) and secondly the environment became more stressful: the screen blinks from yellow to red, a countdown from 5 starts and an annoying beep sounds loudly. If the user take more than 5 seconds to provide a signature a warning text message appears “you are too slow, go faster”. Regarding usability, the results were as expected. The time spent during the stress test was much less, the number of errors during the process was much higher and all the users remarked in the satisfaction form (Figure 3) that they had felt indeed stressed. Performance results were worst also (1.26% of EER in the stress scenario and 0.94% of EER under normal conditions), demonstrating the bad influence of stress in these systems. Nevertheless, this performance decrease is affordable in DSV, being less than 0.5% in EER in absolute value. Being these stressful scenarios habitual and unavoidable in most cases, these results are significantly good and position the DSV as a trustable modality to be launched in malls, post offices, administrations, etc.

3. Lessons learned and future work

Some of the most influential parameters in usability were analysed in these works giving a users’ perspective of the DSV in mobile devices. Performance results were correlated with users’ reactions and opinions obtaining usability outcomes, which are considered really useful for future implementations. The most relevant results are the following:

People feel confident using DSV in mobile devices. Moreover, the migration of DSV seems to be welcomed by people as most users participating in these usability evaluations would use DSV in their daily life. Regarding accessibility, the results obtained show the feasibility of using DSV under different scenarios, styluses and devices offering reliability. There is not an ideal scenario for signing. The stylus based devices obtain better results in performance when the user is sat on a chair and the device is resting on a table (as the most common situation for signing). It is also the most preferred scenario by users. On the other hand, the finger based devices (designed for being supported by users) return the best results in performance within the scenarios
where the user has to handle the device without support. The better visual feedback the better performance and usability. Latency at the time to perform the signature involves annoyance in users and it also affects the performance. Moreover, users do not feel comfortable when not any visual feedback is provided. Then, even being the obtained performance acceptable it involves usability concerns. The stress influences negatively both performance and usability. Nevertheless, probably due to the increasing stressing situations at the time to sign, users become used to sign fast and careless. So that, performance results are close to the ones obtained without stress.

Future works include extending these experiments to other dynamic modalities such as gait or voice recognition. Furthermore, a complete categorization of the influential factors in the usability of biometrics needs to be made in order to better understand the human-biometrics interaction.

References


FPGA based palmprint and palm vein biometric system

Mihails Pudzs, Rihards Fuksis, Rinalds Ruskuls, Teodors Eglitis, Arturs Kadikis, Modris Greitans
Institute of Electronics and Computer Science
14 Dzerbenes Street, Riga, LV1006, Latvia
Mihails.Pudzs@edi.lv; Rihards.Fuksis@edi.lv; Rinalds.Ruskuls@edi.lv
Teodors.Eglitis@edi.lv; Arturs.Kadikis@edi.lv; Modris.Greitans@edi.lv

Abstract: This paper presents an FPGA based multimodal palm biometric system. System is prototyped on an Altera DE2-115 board from Terasic, using an additional hardware for palm image acquisition in two light spectrums, communication with smart card, and debugging. System captures person’s palmprint and palm vein images, extracts biometric data, encrypts it, and prepares for comparison. The comparison of the biometric data is performed on smart card for additional security. The proposed multimodal palm biometric system achieves a matching accuracy of EER of 16.65 % and a verification processing time of 0.8 seconds.

1 Introduction

Increasing identity fraud in recent years stimulates a development of new biometric systems. To obtain higher precision, developers more often employ more than one biometric parameter by developing multimodal biometric systems. Some recent papers discuss how to develop an FPGA based biometric system, for example, FPGA-based finger vein biometric system [KHE10]. However, this system is implemented in soft processor using Nios2-Linux Real Time Operating System (RTOS). In this paper, authors propose the implementation of an FPGA-based multimodal palm biometric system that does not use soft processor, instead all of the functionality is exclusively developed for the current task using FPGA logic.

Authors propose to use palm vein and palmprint images for human authentication. It is easy to use palm as a biometric feature because it is convenient to present the palm to a reading device. By acquiring images in infrared and visible light spectrum, the real palm can be distinguished from photography. To extract valuable information from the captured images Non-Halo Complex Matched Filter (NH-CMF) [PGF11a] is used. We also reduce the amount of the extracted data by eliminating the non-biometric information. This step further simplifies the encryption and comparison, because less data is used for processing. For biometric data encryption, we use BioHash to obtain unique biometric code. Biometric data comparison is performed on the smart card by employing the Match on Card concept.

Paper is organized as follows. First, the architecture of the biometric system is presented. Second, the processing techniques performed on the biometric data are explained, and
implementation of the techniques in the FPGA are discussed. Third, experimental setup is described and experimental results in terms of processing time and recognition accuracy are presented. Paper finishes with conclusions and future work.

2 System architecture

The block diagram of the proposed FPGA based biometric system is shown in Fig. 1. The biometric system can be divided into three main parts: image acquisition module, FPGA, and smart card module.

![Figure 1: The architecture of the proposed biometric system](image)

Because of the hemoglobin absorption properties described in [FGP10], palm print and palm vein images must be acquired in a different light spectrum - visible light and near infrared light, respectively. To be able to acquire both images using one image sensor an imaging acquisition module has been developed. This module consists of Aptina MT9V024 image sensor, which has quantum efficiency of more than 20% in the region of 400 - 900 nm. Two types of LEDs are used to illuminate palm during image acquisition - white LEDs and infrared LEDs of 850nm wavelength. To minimize the influence of ambient light on acquired image, two optical filters are used - one that transmits only the near infrared light (700-1400nm), and one that transmits only the visible light (400-700nm). An electromechanical switching device is used to toggle between the mentioned filters.

User authentication starts by connecting a smart card and presenting a palm to the image acquisition device. After the image is acquired, it is filtered using Non-Halo Complex Matched Filter (NH-CMF), which extracts information about visible line-like objects that include palmprint and palm veins. To acquire palmprint/palm vein biometric information, feature selection follows. Then biometric data is encrypted using the advanced BioHashing algorithm. The encrypted biometric data is sent to smart card to be compared with data stored in it. Smart card responds with similarity between compared data vectors and FPGA decides whether the person that tries to authenticate is the owner of presented smart card.
More details about data processing are presented in the next section.

Additional module is developed for debugging. This module uses USB interface to control the system from a PC. ASRAM is used only to display the acquired image.

3 Data processing

Data processing is divided into three parts: feature extraction and selection, data encryption using BioHash, and data comparison on smart card.

3.1 Feature extraction and selection

Feature extraction is accomplished using Non-Halo Complex Matched Filtering (NH-CMF) approach, which was firstly introduced in [PGF11b]. It is angle invariant line extraction filter that obtains angle of the extracted lines.

We use four rotated line extraction kernels for NH-CMF. Each kernel is convolved with the input image \( f(x, y) \) using equation (1):

\[
s_{n}(x_0, y_0; \varphi_n = n\pi/4) = \int \int_{D} f(x, y) \cdot M((x-x_0, y-y_0; n\pi/4) dx\,dy, \quad (1)
\]

where \( D \) is the area of image \( f(x, y) \) overlapped with the mask \( M \), and \( n \) is the index of mask rotation angle \( \varphi_n \). To prevent filter from generating Halo artifacts, negative values of mask correlation with overlaid areas of the image are set to zero by using ramp function \( R[x] \).

The corrected values of mask correlation are transformed to complex form, using:

\[
\bar{c}_{n}(x, y; \varphi_n) = R[s_{n}(x_n, y_n; \varphi_n)] \cdot e^{j2\varphi_n} \quad (2)
\]

Four mask rotation angles \((n = 0, 1, 2, 3)\) are used for the following reasons:

- it is the minimum amount of rotated kernels, therefore, also convolution operations, without filter losing it extraction abilities,
- the multiplier \( e^{j2n\pi/4} \) in equation (2) can be conveniently simplified to \( e^{jn\pi/2} = \{+1, -1, +j, -j\} \), which doesn’t require any embedded multipliers to perform.

After values are transformed into complex form, they are summed together to acquire a cumulative vector \( \bar{c}(x, y) \):

\[
\bar{c}(x, y) = \sum_{n=0}^{3} \bar{c}_{n}(x, y, \varphi_n). \quad (3)
\]
The region of interest is divided into 64 subregions, and the most intense cumulative vector is selected from each subregion. Each selected cumulative vector $\vec{c}(x_i, y_i)$ is described by 4 parameters: its origin coordinates $(x_i, y_i)$ and its projections $Re(\vec{c}(x_i, y_i))$, $Im(\vec{c}(x_i, y_i))$. The concatenation of $m = 64 \cdot 4$ parameters is used as a biometric feature vector in our system.

### 3.2 Data encryption

Data encryption is performed by using a one-way hash function – biohash [BRAA10] – a member of cancelable biometrics introduced in [RCB01]. Biohash performs a random orthonormal transform $H$ to biometric data $X$, obtaining vector

$$Y = H \cdot X.$$  \hspace{1cm} (4)

Elements of $Y$ are then thresholded to acquire binary data sequence called BioCode. A token is a number that is used as a seed to generate random orthonormal vectors for transform matrix $H$. In the presented biometric system each user has his own private token that is stored on smart card and sent to biometric system at the beginning of the authentication process.

State of the art implementation of Biohash involves following steps: an $m \times m$ matrix consisting of pseudo-random values is generated using provided token; Gram-Schmidt process is applied to orthonormalize generated matrix vectors; matrix multiplication according to (4) is performed.

However, described approach can not be efficiently implemented on FPGA for the following reasons: it is necessary to store a minimum of $m \cdot m$ values in the memory (1Mb for 16bit coefficients if $m = 256$); a large amount of arithmetic operations must be performed, which can take more than a second and extend the authentication time.

The transformation (4) can be performed more efficiently if transformation matrix $H$ is replaced by a product of simpler orthonormal random transformation matrices $(B_m)_i$. Stairs-like Orthonormal Generalized Rotation Matrices (SOGRM) introduced in [MT05] were chosen for this purpose:

$$H = \prod_{i=0}^{(\log_2 m) - 1} (B_m)_i,$$  \hspace{1cm} (5)

where $B_m$ is SOGRM with size $m$, which is defined using (6).
\[ (B_m)_i = \begin{bmatrix}
    a\left(\Theta_i \cdot \frac{m}{2}\right) & b\left(\Theta_i \cdot \frac{m}{2}\right) & 0 & 0 & \cdots \\
    0 & 0 & a\left(\Theta_i \cdot \frac{m}{2} + 1\right) & b\left(\Theta_i \cdot \frac{m}{2} + 1\right) & \cdots \\
    \vdots & \vdots & \vdots & \vdots & \ddots \\
    b\left(\Theta_i \cdot \frac{m}{2}\right) & -a\left(\Theta_i \cdot \frac{m}{2}\right) & 0 & 0 & \cdots \\
    0 & 0 & b\left(\Theta_i \cdot \frac{m}{2} + 1\right) & -a\left(\Theta_i \cdot \frac{m}{2} + 1\right) & \cdots \\
    \vdots & \vdots & \vdots & \vdots & \ddots 
\end{bmatrix} \quad (6) \]

\[ \Theta_i, i \in \left[0 \ldots \frac{m \cdot \log_2 \text{m}}{2} - 1\right] \] are pseudo-random numbers that are generated using provided token. To preserve orthonormality, \(a()\) and \(b()\) must satisfy the following requirement:

\[ (a(\Theta))^2 + (b(\Theta))^2 - 1 \rightarrow 0. \]

It is convenient to use trigonometric functions, such as \(\sin/\cos\), as \(a()\) and \(b()\), because they satisfy all necessary requirements and can be calculated using CORDIC algorithm. In this case, each subsequent pair of \((B_m)_i\) columns form an elementary rotation matrix. Because of the associative property of matrix product, elementary rotations might be applied to input vector \(X\) in an iterative way:

\[ Y = (B_{256})_7 \cdot ((B_{256})_6 \cdot \ldots \cdot ((B_{256})_1 \cdot ((B_{256})_0 \cdot X))). \quad (7) \]

Each multiplication of data vector with transform matrix \(B_m\) is equivalent to \(m/2\) elementary rotation operations, which can be performed in FPGA using the transposed FIR filter structure shown in Fig. 2. This unit is connected to 4 memory blocks: a pair for input data, and a pair for output data. Each pair contains samples 0 to \(m/2 - 1\) and \(m\) to \(m - 1\), respectively. This unit performs \(\log_2 \text{m}\) data processing cycles during which \(m/2\) elementary rotation operations are performed. Starting data encryption process, biometric data vector \(X\) is stored in memory blocks A and B. During first processing cycle data is read from A and B and written to memory blocks C and D. After each processing cycle, reading/writing directions are toggled.

Every two input samples \(x_n\) and \(x_{n+1}\) \((n = 0, 2, 4\ldots 254)\) and angle \(\Theta\) are used to calculate two output samples \(y_{n/2}\) and \(y_{n/2+m/2}\), therefore, a new \(\sin \Theta\) and \(\cos \Theta\) values must be loaded after every two input samples.

Blum Blum Shub (BBS) pseudo-random number generator is used to calculate rotation angles \(\Theta\).

### 3.3 Data comparison

To compare acquired biocode with the one stored in the smart card, Hamming distance is used. To improve recognition accuracy, both resulting similarities (for palmprint and palm vein patterns) need to be fused. Weighted sum method [RJ03] is used for this purpose. For
two similarity values - \( s_1 \) and \( s_2 \), fused similarity can be estimated as \( s_\Sigma = k \cdot s_1 + (1 - k) \cdot s_2 \), where parameter \( k \in [0, 1] \). After both similarities are fused together, the result is thresholded to decide whether to authenticate person, or not.

4 Experimental part

4.1 Test setup and procedure

To test proposed FPGA based biometric authentication system we have changed functionality of several system modules: image acquisition module was modified to receive images from PC; encryption module was modified to send biocodes to PC for further analysis. Images from CASIA Multi-Spectral Palmprint Database were used to test the performance of system’s data processing.

Images from database (100 persons, 6 images per person; two light spectrums - white and infrared 850nm) are sent to FPGA and corresponding biocodes are received. Further analysis, such as similarity calculation and fusion, is performed on PC. False accept and false reject ratios are calculated by mutually comparing all similarity values.

4.2 Experimental results

Figure 3 shows False Accept Rate (FAR) and False Reject Rate (FRR) curves for three tests that were performed. Equal Error Rate (EER) for each test is provided in the table below. Recognition time is 0.8 seconds.
Table 3: Experimental results

<table>
<thead>
<tr>
<th>Test</th>
<th>EER [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>palmprints</td>
<td>25.5</td>
</tr>
<tr>
<td>palm veins</td>
<td>19.9</td>
</tr>
<tr>
<td>fused data</td>
<td>16.65</td>
</tr>
</tbody>
</table>

5 Conclusions

In this paper we have discussed three main challenges, faced in the field of biometrics, such as acquisition, encryption, and comparison of biometric data. Each of the mentioned tasks are implemented in an FPGA-based system to achieve fast person authentication - it takes approximately 0.8 seconds from the moment when the smart card is inserted into the system until access is granted.

As expected, the least reliable parameter is palmprint, which is harder to distinguish from person to person, than palm vein pattern. Experimental results showed that by using more than one biometric parameter for person authentication, lower EER can be achieved, compared to cases when each of the parameters is used in isolation. The results showed that proposed system is not yet ready to be used in real life applications. However our motivation is to build an easy to use, safe and reliable biometric system for everyday usage, and we believe that this work is a step towards to achieving our goals. Future work involves improvement of a feature extraction algorithm.
Acknowledgments

This research is partially supported by European Regional Structural Funds project under the agreement No.2010/0285/2DP/2.1.1.1.0/10/APIA/VIAA/098.

References


Verification of Individuals from Accelerometer Measures of Cardiac Chest Movements

Esra Vural¹, Steven Simske², Stephanie Schuckers¹

¹Clarkson University
Potsdam, NY
evural@clarkson.edu

²Hewlett Packard
Palo Alto, CA
sschucke@clarkson.edu
steven.simske@hp.com

Abstract: Biometric verification is gaining popularity particularly for personal security during internet and mobile device usage. A novel approach for verification of individuals is proposed to measure mechanical cardiovascular activity through an accelerometer sensor placed on the surface of the chest above the sternum. Time frequency analysis methods are employed to evaluate biometric performance. Accelerometer measurements were acquired on two different sessions from ten subjects after delays ranging from 1 to 2 weeks. For individual subject verification, Gaussian mixture models were built per each individual and a background model was created for the remaining impostors. A likelihood ratio test with background model was employed for testing. In this study we found preliminary evidence for the use of the cardiovascular signal measured with an accelerometer placed on the sternum as a biometric sensor to verify individuals. Verification testing using this approach obtained a mean EER rate of 0.06 for inter-session testing.

1 Introduction

Human authentication technologies are commanding more attention recently mainly due to the increased need for personal security in hand-held devices. Biometric systems are used to identify or verify the identity of a person based on biological, physiological or behavioral characteristics. Fingerprints, iris scans, face images are examples of primary physiological characteristics that have been proposed as biometrics [LWJ98][n99]. Gait and keystroke are two biometric signatures that have primarily behavioral characteristics [MR99][DHS06]. Any biometric measure should have universality, uniqueness (discriminability), permanence (stability), measurability, resistance to circumvention and acceptability properties. Circumvention is a type of biometric forgery where spoof signals are being used to gain access to a system [s02]. The heart signal can be more robust to circumvention attacks as it is hard to mimic a person’s heart signature. In this paper we propose a new physiological biometric measurement that senses cardiac chest movements using an accelerometer. Our focus here is to assess the permanence and uniqueness traits of this biometric signal.

Conventional computer systems authenticate users only at the initial login stage. Continuous authentication has gained importance during recent years. In continuous authentication users are not only identified during initial login but are continuously monitored and verified for their identity. Keystroke biometrics and video of face are
some examples of a continuous authentication system [FR99][NUJ10]. Cardiac-based authentication can also be of use as a biometric signal for continuous authentication as the signal is present and can be readily measured continuously. Moreover, cardiac-based authentication has some advantages over other aforementioned biometric authentication systems as the heart is always beating whereas face may be obscured or users may stop typing.

Heart-based verification measurements can be summarized under three main techniques: The first group of techniques uses the mechanical activity of the heart, the second group of techniques measures the electrical activity of the heart and finally the third group of techniques measure the sound of the heart (phonocardiogram) for verification.

Mechanical activity of the heart in the literature is measured using displacement cardiograph techniques such as Ballistocardiogram (BCG) [Gu12], seismocardiogram (SCG) [Za90][Ca07], and finally Laser Doppler Vibrometry (LDV) [Ch10]. Sullivan et al developed a novel method to remotely sense mechanical activity related to the carotid pulse with Laser Doppler Vibrometry (LDV) [c09]. In SCG & BCG the minute movements caused by the beating heart are translated by a transducer into electric potential. In Guo et al’s work BCG is recorded using a BCG chair [Gu12].

Most of the cardiac identification research to date focus on electrical measurements of the heart using an electrocardiogram (ECG). For example Kyosoe et al has used ECG waveform features extracted from fiducial points of the ECG signal to identify subjects [KA01]. ECG verification and identification systems may require the use of electrodes to be attached to the surface of the body.

Beritelli et al examined the biometric characteristics of phonocardiogram (PCG) [BS07]. Phonocardiograph requires a high sampling rate. Using an accelerometer for measuring biometric cardiac signal has advantages in that accelerometers can be found cheaply and a single point of contact is needed for the measurement. LDV needs a large expensive laser which would not be appropriate for most applications.

Here we present a detailed examination of the accelerometer as a candidate biometric sensor measuring cardiorespiratory signal by measuring the chest movements over the sternum. The outline of the paper is as follows. First we describe the accelerometer data collection and signal properties. Next the application of time frequency analysis and feature selection is described. We then present how models are built for person verification. Finally results are reported and discussed.

2 Methods

2.1 Data Acquisition

Signals are recorded using a 3 axis MEMS accelerometer with a 256 Hz sampling rate. The accelerometer is coupled to the chest and placed on the center of the sternum. (See
Figure 1). In addition, a recording of the conventional ECG is obtained. Signal to noise ratio for ECG signal is approximately 17 dB. All data is digitized with a rate of 256 Hz and synchronized. Data is obtained during a continuous 4-min recording period, during which individuals were instructed to sit quietly and avoid voluntary movements. 10 individuals participated and subjects have an average of 304 heart beats per session. The ages of the individuals varies between 18 to 25 years. In order to assess the stability trait, these individuals are tested on two sessions (referred to as Sessions 1 and 2) with a one to two week interval in between the sessions.

Here the Richter vibrations are due to mechanical energy transmitted to the chest from various cardiac, respiratory, and other physiological sources. We focus here on the cardiac pulse sequence, which is referred to as accelerometer pulse signal.

2.2 Signal Basics

ECG arises from the electrical activity in the heart whereas the accelerometer signal carries mechanical activity. In the heart cycle electrical activity precedes mechanical activity. This is reflected in Figure 2 where QRS wave in the ECG precedes the dominant accelerometer wave, likely the contraction of the ventricle. Figure displays a 12 second segment of the continuously recorded data drawn from a 4 minute recording.

Figure 2: (a) ECG Signal (Blue) and Accelerometer Signal (Red) are displayed in time. (Right figure is zoomed version of the left figure). The Accelerometer Signal also carries the slow varying movements of the respiration signal.
2.3 Time Frequency Decomposition Based Analysis

In this experiment 900-ms windows were extracted for the analysis of each heart beat signal. The 900-ms epoch begins 300 milliseconds prior to the location of the detected ECG R peak point. Detection of R peak points can deteriorate under involuntary movements and on average 1 out of 300 beats were missed due to involuntary movements. Time frequency analysis based approach was used similar to techniques used in speech recognition and Laser Doppler Vibrometry [Ch10]. Time frequency analysis of accelerometer pulse signal was performed separately for the two peak regions (See Figure 3). The first peak region is from 270 to 470 milliseconds and the second peak region is 580 to 780 milliseconds. The accelerometer pulse signal around the peaks is normalized by subtracting the mean and dividing by the standard deviation of the signal patch. A Short Time Fourier Transform with a Hamming window of 96 milliseconds and a moving step size of 16 milliseconds is computed. The resultant spectrogram is a 7 (time) x 13 (frequency) matrix. Figure 3 displays the spectrograms of the accelerometer pulse signal obtained around the two peaks using these parameters. Later these two spectrogram matrices are combined into one matrix for feature selection.

![Figure 3: Spectrograms around the two peaks of the signal is displayed above.](image)

2.4 Selection of Important Spectral Bins / Statistical Model for Spectrogram with Informative Components Extraction

Chen et al developed an approach for identification of individuals from the vibration on the carotid artery measured using a Laser Doppler Vibrometry (LDV) [Ch10]. In their approach relative entropy of the spectrogram bins is estimated to select bins that carry significant information in identifying individuals. Freedman et al proved that the Fourier coefficients for stationary or asymptotically stationary random processes are independent complex Gaussian random variables [f80]. Chen et al used this theorem and stated that the magnitude of the Fourier coefficients is exponentially distributed [Ch10]. Chen et al calculates the relative entropy as in equation 1 where $X_{i,n}(l)$ is the training data in the $l$’th time frequency bin of the $n$’th spectrogram for the $i$’th individual. $M_l(l)$ is the $l$’th component of the background model, and $N$ is the number of training spectrograms.
Here we employ Chen’s approach for selecting significant spectrogram bins for verification using equation 1. The Relative Entropy for each bin is estimated with the assumption that the spectrogram bins are exponentially distributed. In equation 1, \( M_o(l) \) is estimated by calculating the overall mean spectrogram of the first session subjects excluding the individual test subject. The top 50 highest entropy bins of an individual were selected as features for use in verification.

2.4 Person Verification

Background models have been used in biometric verification and in particular speaker detection [MR99]. Reynolds et al. used a single speaker–independent background model trained to represent speaker independent distribution of features for speaker detection. Their system used a likelihood ratio test for verification [MR99]. In this paper we employ likelihood ratio test for verification of individuals where the numerator denotes the probability of the signal patch coming from the hypothesized individual (individual model) and the denominator denotes the probability of it coming from the background model. Both models are built using Gaussian Mixture Model (GMM) distributions with diagonal covariance matrices.

Given a set of training vectors, GMM maximum likelihood model parameters are estimated using the iterative expectation-maximization (EM) algorithm. Individual subject model is trained from the 1\(^{st}\) session of the individual subject data. The background model is trained using 1\(^{st}\) session data of all the subjects excluding the hypothesized subject (from a set of 9 subjects). Background GMM model for each individual is trained with 10, 20, 30, 40 mixtures and the model with the minimum Akaike Information criterion (AIC) is selected and stored. Similarly individual GMM models are trained with 1, 2, 3 and 4 mixtures and the model with the minimum AIC is selected. Testing is performed on the 2\(^{nd}\) session data using individual equal error rate thresholds (See Table 1). The EER threshold selection method and the results are reported in the next section.

3 Results

False acceptance and false rejection rates (FAR & FRR) are employed as a performance measure. The FAR and FRR is computed using a preselected threshold where the threshold is determined from the first session of the data as follows. For each individual subject the first session data is randomly partitioned into two where first half of the data is used in building models and the rest of the data is used for selecting the threshold as displayed in Table 1 (left table). Each subject’s individual and background models are built with the first half of the data. The second half of the data is used in testing the models and determining the value of the equal error rate thresholds which are stored for
the inter-session performance estimation. Table 1 right displays the inter-session training and testing for an individual using the stored thresholds.

<table>
<thead>
<tr>
<th>EER threshold determination for an individual</th>
<th>Inter-Session training and testing for an individual</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Impostor Model Training</strong></td>
<td><strong>Impostor Model Training</strong></td>
</tr>
<tr>
<td>Half of randomly selected first session data</td>
<td>9 subject’s combined first session data</td>
</tr>
<tr>
<td>9 subjects x ~300 instances/2 ~1350 training instances</td>
<td>9 x ~300 ~2700 instances</td>
</tr>
<tr>
<td><strong>Individual Model Training</strong></td>
<td><strong>Individual Model Training</strong></td>
</tr>
<tr>
<td>Half of randomly selected first session data</td>
<td>First session of the subject data</td>
</tr>
<tr>
<td>1 subject x ~300 instances /2 ~150 training instances</td>
<td>~300 instances</td>
</tr>
<tr>
<td><strong>EER threshold determination</strong></td>
<td><strong>Testing</strong></td>
</tr>
<tr>
<td>Rest of first session data</td>
<td>Second session of the data</td>
</tr>
<tr>
<td>~150 authentic instances, ~1350 impostor instances</td>
<td>~2700 impostor instances, ~300 authentic instances</td>
</tr>
</tbody>
</table>

Table 1: Data partitioning for EER threshold determination for an individual is displayed on the left. Right table displays partitioning of the data for intersession training and testing for an individual.

<table>
<thead>
<tr>
<th>Subject #</th>
<th>FAR</th>
<th>FRR</th>
<th>FAR sum 5</th>
<th>FRR sum 5</th>
<th>FAR sum 10</th>
<th>FRR sum 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject 1</td>
<td>0.04</td>
<td>0.03</td>
<td>0.01</td>
<td>0.04</td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>Subject 2</td>
<td>0.00</td>
<td>0.08</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Subject 3</td>
<td>0.03</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>Subject 4</td>
<td>0.15</td>
<td>0.15</td>
<td>0.07</td>
<td>0.01</td>
<td>0.05</td>
<td>0.00</td>
</tr>
<tr>
<td>Subject 5</td>
<td>0.02</td>
<td>0.53</td>
<td>0.02</td>
<td>0.62</td>
<td>0.00</td>
<td>0.67</td>
</tr>
<tr>
<td>Subject 6</td>
<td>0.27</td>
<td>0.22</td>
<td>0.16</td>
<td>0.11</td>
<td>0.13</td>
<td>0.10</td>
</tr>
<tr>
<td>Subject 7</td>
<td>0.24</td>
<td>0.04</td>
<td>0.16</td>
<td>0.07</td>
<td>0.13</td>
<td>0.04</td>
</tr>
<tr>
<td>Subject 8</td>
<td>0.02</td>
<td>0.29</td>
<td>0.00</td>
<td>0.24</td>
<td>0.00</td>
<td>0.24</td>
</tr>
<tr>
<td>Subject 9</td>
<td>0.29</td>
<td>0.03</td>
<td>0.29</td>
<td>0.03</td>
<td>0.28</td>
<td>0.01</td>
</tr>
<tr>
<td>Subject 10</td>
<td>0.32</td>
<td>0.16</td>
<td>0.26</td>
<td>0.09</td>
<td>0.25</td>
<td>0.08</td>
</tr>
<tr>
<td>Mean</td>
<td>0.14</td>
<td>0.15</td>
<td>0.09</td>
<td>0.12</td>
<td>0.08</td>
<td>0.13</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.12</td>
<td>0.15</td>
<td>0.11</td>
<td>0.18</td>
<td>0.10</td>
<td>0.19</td>
</tr>
</tbody>
</table>

Table 2: Inter-Session FAR and FRR single beat and sum FAR and FRR rates using 5 and 10 heart beats are displayed.

The results for the intra-session comparisons are not shown here to conserve space. Average inter-session FAR and FRR performance numbers for single beats and using the sum over 5 and 10 heart beats are reported in Table 2. A mean of 0.14 FAR and a mean of 0.15 FRR was achieved for inter-session person verification with a single heart beat. Moreover when the sum over 10 heart beats is employed our error rates decreased to a mean of 0.08 FAR and 0.13 FRR. Summation over more than 10 heart beats did not lead to a significant increase in performance.
Figure 4 displays the equal error rates for each individual by using single beat EER and 10 heart beat EER measures. The mean EER for 10 heart beats obtained a rate 0.06. This suggests a preliminary evidence for uniqueness in the heart signal measured with the accelerometer placed on the sternum using frequency spectrogram measures.

![Graph showing equal error rate performance results for single beat and 10 heart beat results](image)

Figure 4: Equal error rate performance results are displayed for single beat and 10 heart beat results

4 Conclusion

Using an accelerometer for measuring biometric cardiac signal has advantages in that accelerometers can be found cheaply and a single point of contact is needed. Moreover, accelerometer data is computationally inexpensive in comparison to methods that require a high sampling rate such as phonocardiogram. Cardiac measurements using an accelerometer does not require an expensive setup as opposed to Laser Doppler Vibrometry. Our studies confirm that the accelerometer is promising as a biometric sensor to measure cardiac pulse signal and has verification strength for individuals. More subjects are needed to evaluate this approach in larger populations. The signal may be affected by factors such as physical exercise and mental stress, hence, we plan future experiments to investigate this.

5 Acknowledgement

This research was supported by a grant from Hewlett Packard. The authors would like to thank to Hewlett Packard for their support.
References


Automatic Landmark Detection and Face Recognition for Side-View Face Images

Pinar Santemiz,* Luuk J. Spreeuwers, Raymond N.J. Veldhuis

Signals and Systems Group, Department of Electrical Engineering
University of Twente
Drienerlolaan 5 P.O.Box 217
7500AE Enschede, The Netherlands
P.Santemiz@utwente.nl
L.J.Spreeuwers@utwente.nl
R.N.J.Veldhuis@utwente.nl

Abstract: In real-life scenarios where pose variation is up to side-view positions, face recognition becomes a challenging task. In this paper we propose an automatic side-view face recognition system designed for home-safety applications. Our goal is to recognize people as they pass through doors in order to determine their location in the house. Here, we introduce a recognition method, where we detect facial landmarks automatically for registration and identify faces. We test our system on side-view face images from CMU-Multi PIE database. We achieve 95.95% accuracy on detecting landmarks, and 89.04% accuracy on identification.

1 Introduction

In applications dealing with identifying people from videos such as surveillance systems or smart homes, face recognition is the primary biometrics. One possible application area for face recognition are home-safety applications. Here, face recognition can be used to increase the situational awareness, and to prevent the factors that may cause further accidents. However, in real-life scenarios with uncontrolled environment, face recognition becomes a challenging task due to occlusion, expression, or pose variations.

In this paper we introduce a novel method for side-view face recognition to be used in house safety applications. Our aim is to identify people as they walk through doors, and estimate their location in the house. We design a system that uses video recordings from cameras attached to door posts under ambient illumination. The cameras have a limited view angle thus preserving the privacy of the people. Here, we test our system in a setting similar to this scenario. We use multiple still images that contain side-view face images, and we perform automatic landmark detection and recognition tests on these images.

Due to the complex structure of human face, face recognition under pose variation up to

*Contact Author: Pinar Santemiz, E-mail Address: P.Santemiz@utwente.nl
side-view is a difficult problem. In [ZG09], a literature survey on face recognition under pose variations can be found. In initial attempts to compare side-view face images, mainly profile curves or fiducial points on the profile curves were used. One such method is proposed by Bhanu and Zhou [BZ04], where they find nasion and throat point, and compare the curvature values using Dynamic Time Warping (DTW). They achieve a recognition accuracy of 90% on Bern database, which contains side-view face silhouettes of 30 people.

In video-based applications, people make use of the texture information in addition to profile curves. Tsalakanidou et al. [Tsa03] present a face recognition technique where they use the depth map for exploiting the 3D information, and apply Eigenfaces. They experiment on the XM2VTS database using 40 subjects, and recognize 87.5% of them correctly. In a recent study [SSV11], Santemiz et al. proposes a side-view face recognition method using manual landmarks. Here, they use local binary patterns to compare faces and achieve a recognition accuracy of 91.10% on a small subset from CMU-Multi PIE database [GMC+10], where they excluded the subjects wearing glasses.

In this study, we first find three landmark points automatically using Histogram of Oriented Gradients (HOG) [DT05] and train Support Vector Machines (SVM) [CC95]. We use these landmarks for registering images as presented in Section 2. Then we apply Principal Component Analysis (PCA) [TP91], Linear Discriminant Analysis (LDA) [BHK97], Local Binary Pattern (LBP) [AHP06], and Histogram of Oriented Gradients (HOG) [DT05] to describe the face images. The details of our feature extraction techniques are given in Section 3. We identify faces using nearest neighbor classifier and test our system on side-view face images of CMU-Multi PIE database [GMC+10]. We analyze our results in Section 4. Finally, we will give our conclusion in Section 5, and discuss our future work.

2 Automatic Landmark Detection and Registration

In our landmark detection approach, we aim to find three landmark points on the face, namely, the eye center, the tip of the nose, and the corner of the mouth. A visualization of these landmark points is given in Figure 1(c). In order to obtain an uncluttered profile, we first manually select large skin color regions including hair from training samples containing 50 subjects and 708 images, and learn the multivariate Gaussian distribution of the HSV color space. Using this distribution, we estimate the skin color masks of the remaining images, and extract the outer profile. Then, we compute the curvatures on facial profile, and use the curvatures having a local maxima as candidate points for the tip of the nose which is shown in Figure 1(a). Around each candidate point we extract a Region of Interest (ROI) of size $55 \times 55$ pixels. We assume that the ROI of the eye and the mouth is centered at a distance of $[-40, +40]$ pixels and $[-40, -40]$ pixels away from the tip of the nose, respectively. An example is given in Figure 1(b). For each nose tip candidate, we extract ROIs for the mouth corner and eye center, and scan all three ROIs to find the landmarks. An example is shown in Figure 1.

To find the landmark points, we train three separate SVMs. In training, for each landmark location we select nine positive and 16 negative samples of image patches of size $10 \times 10$
pixels. To select the positive samples, we use the manually labeled coordinates and eight neighboring coordinates for each landmark. The negative samples are chosen randomly from the ROIs of the landmarks. From all these image patches we extract the HOG features and train SVMs. Here we use the same training set as we use for training the skin colors. Using the SVMs, we compute scores for each candidate point and choose the three coordinates having the total maximum score as our landmarks.

For registration, we use Procrustes analysis [Goo91] to find the transformation parameters between each image. First, we align the landmarks of the images in the training set to the landmarks of the first image, and compute their mean to find the average landmarks. Then, we compute the transformation between each image and the average landmarks, and transform images, accordingly. Finally, in order to have fixed sized images we place a bounding rectangle around the face, and crop the image. Here, we use a fixed window for the bounding rectangle of size $200 \times 100$ pixels such that the right side of the rectangle is centered at the tip of the nose. Some examples can be seen in Figure 2.

![Figure 2: Automatic landmark detection results. The green points are ground truth, and the red points are the found landmark locations. (a) -90 degrees, (b) -75 degrees, (c) 75 degrees, (d) 90 degrees](image)

### 3 Feature Extraction

We describe the registered face images using two baseline algorithms, Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), and also using Local Binary Pattern (LBP), and Histogram of Oriented Gradients (HOG).

PCA [TP91] (Eigenface approach) is an algorithm for reducing dimensionality of a feature space by projecting it onto a space that spans the significant variations, and LDA [BHK97] (Fisherface approach) is a supervised method for classification problems. In our imple-
mentation, we learn the PCA parameters from the training set, project each image into PCA space, and from the projected values of the training samples we learn LDA parameters. For classification, we use nearest neighbor method using cosine similarity measure.

Local Binary Pattern is a method that describes the local spatial structure of an image [AHP06]. The most prominent advantages of LBP are its invariance against illumination changes, and its computational simplicity. In our system, we divide the images into 75 subregions, and compute the LBP histograms for each region. Then, we concatenate these histograms to form the feature vector of the image. For classification, we use nearest neighbor method using Chi square distance measure.

Histogram of Oriented Gradients (HOG) are mainly used in computer vision as feature descriptors in object detection and recognition [DT05]. HOG represents the shape via the distributions of local intensity gradients or edge directions. The main advantage of using HOG descriptors is that they offer some robustness to scene illumination changes, while capturing characteristic edge or gradient structure. We divide the image into cells with $10 \times 10$ pixels and for each cell, we form an orientation histogram having 32 bins. For classification, we use nearest neighbor method using Chi square distance measure.

4 Experimental Results

In CMU Multi-PIE database, each subject is recorded under 15 poses in up to four sessions, where 13 cameras are located at head height spaced at 15 degrees intervals. The images are acquired in a controlled environment with constant background and illumination, and have a resolution of $640 \times 480$ pixels. We select the images acquired from the four cameras that are located at $-90$, $-75, 75, \text{and} 90$ degrees as side-view images and use a total of 3684 side-view face images from all 337 subjects in our experiments.

4.1 Landmark Detection

In our landmark detection experiments, we divide the set into two subsets: a training set containing 50 subjects and 708 images, and a test set with 287 subject and 2976 images. The average distance between the eye center and the mouth corner in this set is 79.91 pixels. Therefore, an automatically detected point displaced 10-pixels distance from the ground truth is accepted as a correct detection. Using this threshold we detect 95.95% of the landmarks correctly, where the correct detection for the eye center, the tip of the nose, and the mouth corner separately are 94.79%, 96.34%, and 96.72%, respectively.

In our experiment, our skin color segmentation algorithm failed to detect the face in only one image where the subjects face is mostly covered by hair as seen in Figure 3(a). Other than this example, we were able to segment the skin color masks, but had cluttered profile curve on some images due to hair, facial hair, eyeglasses, or poor illumination. Yet, our approach to eliminate the false candidates using HOG and SVM proved to be successful in most of the examples.
When we observe the 363 images where our algorithm falsely detected landmarks, we see that the errors are mostly caused by occluded images due to hair or eyeglasses. Yet we also observe that in some images our algorithm falsely detect the upper lip location as the tip of the nose. Some false landmark detection examples are given in Figure 3.

![Figure 3: False landmark detection examples. The green points are ground truth, and the red points are the found landmark locations. (a) Failed skin color segmentation. (b) Falsely detected nose tip. (c) Occlusion of hair. (d) Occlusion of eyeglasses.](Image)

### 4.2 Recognition

In our identification experiments, we divide the database into three subsets: a training set containing 200 subjects and 2484 images, an enrollment set with 137 subjects and a total of 744 images consisting of six images for each subject, and a test set with a total of 456 images. Since we aim to use side-view face recognition to identify people from video recordings, here we keep the setting much similar to this scenario, and use multiple still images for enrollment. The enrollment images and the test images can have a 15 degrees pose variation, which we expect to be the case in a real life scenario.

We perform identification experiments using PCA, LDA, LBP, and HOG. We further applied sum-rule fusion to LBP and HOG. To compare our registration approach, we first applied registration using only the tip of nose without scaling and rotation, using three manually labeled landmarks, and using automatically detected landmarks. Our rank-one accuracies can be seen in Table 1, and the Cumulative Match Characteristic (CMC) curves for identification in Figure 4.

<table>
<thead>
<tr>
<th></th>
<th>Registered using One Manually Labeled Landmark</th>
<th>Registered using Manually Labeled Landmarks</th>
<th>Registered using Automatically Detected Landmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA</td>
<td>61.18%</td>
<td>60.96%</td>
<td>56.80%</td>
</tr>
<tr>
<td>LDA</td>
<td>62.06%</td>
<td>66.67%</td>
<td>56.58%</td>
</tr>
<tr>
<td>LBP</td>
<td>82.89%</td>
<td>88.82%</td>
<td>80.92%</td>
</tr>
<tr>
<td>HOG</td>
<td>85.75%</td>
<td>87.94%</td>
<td>82.89%</td>
</tr>
<tr>
<td>LBP+HOG</td>
<td>85.53%</td>
<td>89.04%</td>
<td>82.02%</td>
</tr>
</tbody>
</table>
When using one landmark we achieve our best performance using HOG features and obtain 85.75% recognition accuracy. Our highest accuracy for images registered with three manual landmarks is 89.04% which we obtain using sum-rule fusion of LBP and HOG. For images registered with automatically detected landmarks our best performance is 82.89% which is obtained using HOG features.

When we analyze these results, we see that LBP and HOG consistently perform better than PCA and LDA. It has been shown that compared to holistic methods, LBP is less sensitive against variations that occur due to illumination, expression, or pose [AHP06]. Both HOG, and LBP describe the image by dividing it into local regions, extracting texture descriptors for each region independently, and then combining these descriptors to form a global description of the image. Consequently, they are not affected by small local changes as much as PCA or LDA. When we compare HOG and LBP, we see that they achieve similar results for each registration method. Also when we look at the CMC curves, we see that on higher ranks both LBP and HOG have similar results. However, LBP is more affected by errors of automatic landmark detection which shows that HOG copes with local changes slightly better than LBP.

When we compare identification results of registration using one landmark and three landmarks, we see that we achieve better results with three landmarks except PCA. The results improve much significantly for LBP compared to HOG, which supports the robustness of HOG against local changes compared to LBP. We also see that our fusion approach does not improve our results.

We observe that our recognition accuracies drop significantly when we use automatically detected landmarks. To better understand the cause of this decline we perform another experiment using the samples for which the landmark is correctly found. For these samples images registered using manual landmarks give a rank-one recognition accuracy of 86.43%, where as using images registered with automatic landmarks the rank-one recognition accuracy increases to 87.62%. Based on this observation, we conclude that finding the landmarks within 10 pixels is accurate enough, and the decline we see in performances
is caused by the samples whose landmarks are falsely detected.

In order to better analyze these results, we also investigate the erroneous cases. Some misclassification examples caused by false landmark detection, and occlusion of hair or glasses can be seen in Figure 5. We observe that the misclassification errors for LBP and HOG are very similar based on the type of errors. We show two misclassification errors of LBP in Figures 5(a) and 5(b), and two misclassification errors of HOG in Figures 5(c) and 5(d).

![Figure 5: Misclassification examples: the test images (left), the nearest images found by the classifier (right). (a) and (b) Misclassification examples of LBP. (c) and (d) Misclassification examples of HOG. (a) and (c) Misclassification due to falsely detected nose tip. (b) Misclassification due to hair. (d) Misclassification due to glasses.](image)

In examples shown in Figures 5(a) and 5(c), the landmark detection algorithm falsely detects the upper lip location as the tip of the nose, and the faces are tilted upwards. We see that the pose and the shape of the faces are similar, but the difference in texture is significantly different. Especially, in the example shown in Figure 5(c), the test sample wears glasses and does not have beard, which is the opposite for the sample that is found as the most similar. When we compare the samples shown in Figure 5(b), the test sample wears a hat and the found sample has his forehead covered with hair in a similar way. In Figure 5(d), in both images the left eye of the sample is partly shown which shows that they both have the same head pose. Also, both the test sample and the found sample wear glasses.

5 Conclusion and Future Work

In this work we investigate automatic landmark detection and side-view face recognition to be used in house safety applications, where we aim to identify people as they walk through open doors, and estimate their location in a house. Here, we present our initial results that we achieved using side-view face images from the CMU-Multi PIE database. We automatically detect the landmarks with a detection accuracy of 95.95% and use these landmark points for registration. We test our system both with manually labeled landmarks and automatically detected landmarks using PCA, LDA, LBP, and HOG. We achieve 89.04% recognition accuracy using sum-rule fusion of LBP and HOG for manually labeled landmarks, and 82.89% recognition accuracy using HOG for automatically detected landmarks.
We see that, our automatic landmark detection method is effective, and shows high accuracy. Also, when we compare identification results using the samples for which our algorithm detects landmarks correctly, we see that the performance using automatic landmarks is higher than the performance using manual landmarks. Moreover, we achieve promising results with our recognition algorithm.

In the future, we aim to improve our landmark detection algorithm and increase the number of landmarks to better cope with images that are partially occluded due to hair or glasses. We also aim to include a higher pose variation in our experiments.

6 Acknowledgement

This work is supported by GUARANTEE (ITEA 2) 08018 project.

References


An Efficient 3D Facial Landmark Detection Algorithm with Haar-like Features and Anthropometric Constraints

Martin Böckeler, Xuebing Zhou

CASED - Center for Advanced Security Research Darmstadt
martin.boeckeler@googlemail.com
xuebing.zhou@cased.de

Abstract: In the last few years 3D face recognition has become more and more popular due to reducing cost of scanners and increasing computational power. The crucial and time-consuming step is landmark localization and normalization of facial surface. Due to acquisition, noise and other artifacts like spikes and holes occur. Most systems require computational intensive preprocessing steps to eliminate these artifacts. As a consequence, a trade-off between runtime or detection accuracy must be made. In contrast, we propose a landmark detection algorithm which uses the Viola & Jones classifier on gradient images. The algorithm is able to reliably detect landmarks in raw 3D data without complicated preprocessing. Additionally, selection of sub regions is exploited to limit search regions. It further reduces false detection rate and improves significantly detection accuracy.

1 Introduction

The field of biometric includes a wide range of applications driven approaches to automatically identify or verify a person. Different biometric characteristics like fingerprint, vein, hand geometry, iris, etc. can be used. The disadvantages of these modalities are that the subject usually has to cooperate with the system to gather a useful probe. Among those, the face is a very important biometric modality. The main advantages are that the subject doesn’t have to make direct physical contact with system and that face recognition is well accepted by users. Common 2D face recognition is sensitive to illumination changes and variation of pose and expression. 3D face recognition is more robust to illumination, because the surface of the face is more illumination invariant than the texture. Moreover, it can better overcome pose changes than 2D face recognition by correcting poses into a normalized position.

The accuracy of normalization has a strong influence on the recognition performance. The localization of landmarks is the key point in this process. Because of sensor noise and acquisition artifacts, the first step in 3D face recognition is a complex preprocessing to filtering noise and smooth facial surfaces. The preprocessing can increase detection accuracy, however it also enlarges the runtime. Therefore, many systems have to make a compromise between runtime and accuracy. This paper introduces a new landmark detection algorithm that achieve a high accuracy without complicated preprocessing.
This paper is organized as follows: Section 2 gives an overview of the existing techniques in the field of 3D landmark detection. Section 3 shows the details of the algorithm and also provides some background knowledge. Section 4 analyses the evaluation results. Section 5 gives conclusions of the paper and an outlook to the future.

2 The Existing Techniques

Falling sensor prices and increasing computational power have made the domain of 3D face recognition more and more popular in the past few years. Since long, 3D face recognition is an active research area. An early approach to recognize faces in 3D data was already done by using profile planes of the face in 1989 [CLR89]. In this early attempt, an iterative process extracted the profile planes from the range data using Gaussian curvature analysis. The reported performance reaches 100% with 18 datasets. Dibeklioglu et al. [DSA08] presented an algorithm based on statistical and heuristic localization. The statistical method using local features to determine the most likely location for each landmark. The localization of the nose tip was done in a heuristic way via curvature analysis. The cross database accuracy was between 48% and 100%. Perakis et al. [PPTK10] presented various methods for 3D landmark detection that were suitable to detect landmarks from frontal and side facial scans. They used local shape descriptors based on the shape index, the extrusion map and spin images. Detected landmarks were classified and labeled with help of a facial landmark model that was derived from the statistical mean shape of manually annotated landmarks. The reported cross database performance was between 83% and 97%.

The algorithm proposed in this paper is based on an existing approach of Ajmal Mian [Mia11]. Mian used gradient images and range images and detect landmarks with the Viola & Jones detection algorithm. The gradient images were directly generated from the raw 3D data. Classifier detected landmarks in the x-gradient, the y-gradient and the range images separately. Multiple landmark candidates were detected and clustered. The remaining candidates were further filtered due to a priori information of facial topology, namely the triangle of the two outer eye corners and the nose tip. Anthropometric constraints of the face were used to eliminate incorrect triangles. The performance was evaluated on 4007 face scans and reaches 99.9%.

3 An Efficient Algorithm based on Gradient Images

In this section we will show the details of the proposed algorithm. The algorithm aims to detect 10 important facial landmarks. These are the outer eye corners, the center of the pupils, the inner eye corners, the sides of the nose and the mouth corners, see figure 1.

The basic idea of this algorithm is to train the classifier for different landmark with the Viola & Jones object detection approach [VJ01]. Because the Viola & Jones is based on Haar liked features, the 3D data first has to be transformed into special gradient images to
Figure 1: The detectable landmarks from the left to the right. Eye: exocanthion right, centre of pupil right, endocanthion right. Nose: alare right. Mouth: cheilion right. The figure only shows the right landmarks of all 10 detectable.

enhance Haar like structures.

3.1 Gradient image

A gradient image is a special representation of the raw 3D data computed from depth information [Cro84]. The z-value of adjacent measure points will be compared either from left to right (x-gradient) or from the top to the bottom (y-gradient). If the second of the compared measure points is closer to the scanner than the first one, the pixel will be colored brighter. If the second measure point is farther away from the scanner, than the pixel will be colored darker. Figure 2 shows a plotted 3D face and the corresponding x- and y-gradient images. The advantages using gradient images are that edges on the 3D surface are enhanced and changing of 3D information are better described. It also allows to use image processing methods to detect landmarks.

3.2 Viola & Jones object detection

The Viola & Jones classifier is one of the most successful object detection algorithms. Its success relays on three main contributions. Firstly an efficient machine learning algorithm combining Adaboost with Haar like features is used. Haar like features describes relative illumination changes of adjust blocks. A vast number of Haar like features can be generalized. The Adaboost algorithm selects the most significant Haar-like features, which best describe the searched object.

Secondly a special image representation, called integral image, also known as summed-area table is used. An integral image contains the sum of gray scale values belonging to pixels. The average intensity of a certain rectangle can be easily calculated with only four values in an integral image. The last contribution of the Viola & Jones object detection is to use complex classifier as a multiple stages cascade. If the similarity between Haar like features derived from an inspected area and these from a classifier reaches a defined threshold, the area passes to the next stage of the classifier. Once the similarity is below the threshold, the area is rejected. Only when the inspected area passes all stages of the classifier, it is marked as a candidate of the searched object. The thresholds in the various
stages are determined during the training process.

3.3 The Proposed Algorithm

The presented algorithm is implemented in C++ and uses the Open Source Computer Vision library (OpenCV). As mentioned before, the raw 3D data has to be transformed into gradient images. In this approach the detection of the alare, the endocanthion and the cheilion is done on the y-gradient. The detection of the right exocanthion is done on the x-gradient, the detection of the left exocanthion is done on the mirrored-x-gradient and the detection of the pupils is done on a special gradient visualization, defined as x-abs-gradient. To generate the x-abs-gradient, the absolute values of the x-gradient plus an inversion of all black pixels are taken. Figure 2 visualizes different gradient images used for the landmark detection.

The eyes are the regions where artifacts often occur. Especially, regions of pupils cannot be captured and results holes. This property can be exploited to detect pupils inside the x-abs-gradient as the good visible, white filled disks in a dark region.

After the creation of the gradient images, the nose detector is applied. The reason for choosing the nose as the first detected modality is based on the facts, that the nose changes its shape only slightly due to expression. Additionally, the nose is normally not unintended covered by any other body part such as the pupils due to the closing eyes. Therefore, the nose detection reaches the highest detection rate with the trained classifier in comparison with other landmarks.

If the nose is detected, sub regions for the detection of the landmarks can be determined. The sub region for the pupils laying in a small rectangle above the detected nose. Its height is 130% of the detected nose. The preliminary sub region for the cheilions laying right under the detected nose with the same height as the detected nose. In the case that no nose was found, both sub regions have the size of the whole image. The landmark detection starts with the right and the left alare in the region of the detected nose. The second pair of landmarks, which are searched in its sub region, are both pupils. Every other sub region for the remaining landmarks is created dynamically. That means that their width is calculated as a result of already detected landmarks. In the following we describe

![Figure 2: from the left to the right: the plotted raw-3D data (just for better imagination of the real face), the y-gradient, the x-gradient, the mirrored-x-gradient and a new developed gradient visualization named x-abs-gradient.](image)
the detection process for the landmarks on the right side. The same process can be applied for the landmarks on the left side. The sub region for the endocanthions has always the same height as the sub region for detecting the pupils. The smallest region can be created, when the right pupil and both alares are detected. In this case, the width of the sub region shrinks between these two points. Figure 3 illustrates these process for all alternatives of already detected landmarks. The creation of sub regions for the exocanthions and cheilions follows the same scheme.

![Figure 3: sub regions for the right endocanthion. 1: from the right pupil to the midpoint of the alares, 2: from the right pupil to the left alare, 3: from the right edge of the image to the left alare, 4: from the right edge of the image to the left pupil, 5: whole search region as before for the pupils.](image)

4 Experimental Results

The proposed algorithm is tested with the Face Recognition Grand Challenge (FRGC) v2 database [Nat11]. The database contains 3D face scans with a resolution of $640 \times 480$ pixels. 500 scans are chosen to train the classifier and another 500 datasets are chosen for testing.

A big advantage by using sub regions is the lower runtime of the algorithm. It takes only 409ms to detect all 10 landmarks with sub regions and about 4900ms without sub regions. This implicit a speed up by nearly factor of 12. The measurements are the average of 100 program throughputs and taken with an Intel Core 2 Duo @ 2.16 GHz. But not only the runtime is reduced, additionally detection accuracy is improved with the use of sub regions. A single classifier detects many false positives on the whole image as shown in figure 4. After clustering, the resulting two landmarks are still far away from their actual position. By using sub regions, the false positive rate reduces and detected candidates are close to each other as well as to the actual position.

Individual classifier are derived from the training process. An important step in the training process is to prepare positive and negative samples. Positive samples are images that only contain the searched object while the content of negative samples can be arbitrary and only has the restriction, that it does not contain the searched object. The classifier are trained with the first 500 datasets out of the FRGC database. To double the number of positive samples, each classifier is trained for the right landmark representation and its mirrored
counter piece. So each classifier has nearly 1000 positive samples. Since the training data is also noisy or includes closed eyes, not every classifier has the amount of 1000 positive samples. Table 1 gives an overview of the sample sizes used in the training process.

<table>
<thead>
<tr>
<th>classifier</th>
<th>Nose</th>
<th>Alare</th>
<th>Pupil</th>
<th>Endocanthion</th>
<th>Exocanthion</th>
<th>Cheilion</th>
</tr>
</thead>
<tbody>
<tr>
<td># pos</td>
<td>1000</td>
<td>944</td>
<td>964</td>
<td>952</td>
<td>762</td>
<td>892</td>
</tr>
<tr>
<td># neg</td>
<td>2000</td>
<td>3776</td>
<td>3856</td>
<td>3808</td>
<td>4572</td>
<td>3568</td>
</tr>
<tr>
<td># pos : # neg</td>
<td>1:2</td>
<td>1:4</td>
<td>1:4</td>
<td>1:4</td>
<td>1:6</td>
<td>1:4</td>
</tr>
</tbody>
</table>

Table 1: the number of positive and negative samples that are used due the training process for each classifier. To have a minimum amount of training samples, the number of negative images were increased in cases of a lower count of positive samples.

The evaluation of classifier performance is shown in figure 5. Performance criteria is the distance, which was calculated out of the raw 3D data, between a detected landmark and its annotated ground truth counter-piece. The performance on testing data is nearly as good as the performance on training data. Only the detection accuracy of the exocanthion is significantly lower. This is caused by the training process of the classifier and the high deformation of the exocanthion during expression. The training of the exocanthion classifier was only done with neutral expression, so the classifier is unable to detect exocanthions that are highly distorted. This aggravation marks a big disadvantage in the general use of classifier that are based on the object detection approach. A classifier can only detect the objects which it was trained for. As soon as the object highly changes its appearance, the classifier is unable to detect the object in the picture.

The preexisting work of Ajmal Mian [Mia11] detected both exocanthions and the nose tip, so performance comparison can only be done for the exocanthion classifier. Figure 5 shows the performance comparison of Mians exocanthion classifier and the one generated in this approach. It must be mentioned, that the curve of Mians classifier was manually reconstructed out of the original paper. Up to a detection error of 5mm, the sub region approach proposed in this paper outperforms the one from Mian that uses anthropometric constrains. For detection faults bigger than 5mm, Mians approach has an obvious higher performance by nearly 20%.
In the paper we proposed a reliable landmark detection algorithm with low detection time. The usage of the $x$-abs-gradient image overcomes the well known pupil detection problem in 3D face recognition. The creation of sub regions not only dramatically reduce the detection time by nearly the factor of 10, they also increase the whole performance of the system. The designed algorithm has a big potential with the opportunity of further improvements. At first an enhancement of the classifier training should be done. An increase of training data would directly result a higher detection performance and would also boost the robustness against expression. Also a cross database development can be considered, because right now the algorithm runs only on the FRGC v2 database. In order to reduce the algorithm runtime time farther, a parallel detection can be implemented. The last improvement that will be listed here is an increase of the number landmarks to detect. That makes an implementation of an identification or verification system possible.

Figure 5: precision comparison between training data, testing data and the reconstructed performance curve of Ajmal Mian’s exocanthion classifier.

5 Conclusions and Future Work
References


GI-Edition Lecture Notes in Informatics


P-3 Ana M. Moreno, Reind P. van de Riet (Hrsg.): Applications of Natural Language to Information Systems, NLDB’2001.


P-5 Andy Schürr (Hg.): OMER – Object-Oriented Modeling of Embedded Real-Time Systems.


P-7 Andy Evans, Robert France, Ana Moreira, Bernhard Rumpe (Hrsg.): Practical UML-Based Rigorous Development Methods – Countering or Integrating the extremists, pUML’2001.


P-10 Mirjam Minor, Steffen Staab (Hrsg.): 1st German Workshop on Experience Management: Sharing Experiences about the Sharing Experience.


P-12 Martin Glinz, Günther Müller-Luschnat (Hrsg.): Modellierung 2002.

P-13 Jan von Knop, Peter Schirmbacher und Viljan Mahni (Hrsg.): The Changing Universities – The Role of Technology.


P-15 Hans-Bernd Bludau, Andreas Koop (Hrsg.): Mobile Computing in Medicine.


P-18 Elmar J. Sinz, Markus Plaha (Hrsg.): Modellierung betrieblicher Informationssysteme – MobilIS 2002.


P-21 Jörg Desel, Mathias Weske (Hrsg.): Promise 2002; Prozessorientierte Methoden und Werkzeuge für die Entwicklung von Informationssystemen.


P-23 Thorsten Spitta, Jens Borchers, Harry M. Sneed (Hrsg.): Software Management 2002 – Fortschritt durch Beständigkeit

P-24 Rainer Eckstein, Robert Tolksdorf (Hrsg.): XMI2003 – XML-Technologien für Middleware – Middleware für XML-Anwendungen

P-25 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Commerce – Anwendungen und Perspektiven – 3. Workshop Mobile Commerce, Universität Augsburg, 04.02.2003

P-26 Gerhard Weikum, Harald Schönig, Erhard Rahm (Hrsg.): BTW 2003: Datenbanksysteme für Business, Technologie und Web

P-27 Michael Kroll, Hans-Gerd Lipinski, Kay Melzer (Hrsg.): Mobiles Computing in der Medizin

P-28 Ulrich Reimer, Andreas Abecker, Steffen Staab, Gerd Stumme (Hrsg.): WM 2003: Professionelles Wissensmanagement – Er- fahrungen und Visionen

P-29 Antje Düsterhöft, Bernhard Thalheim (Eds.): NLDB’2003: Natural Language Processing and Information Systems

P-30 Mikhail Godlevsky, Stephen Liddle, Heinrich C. Mayr (Eds.): Information Systems Technology and its Applications

P-31 Arslan Brömme, Christoph Busch (Eds.): BIOSIG 2003: Biometrics and Electronic Signatures
P-32 Peter Hubwieser (Hrsg.): Informatische Fachkonzepte im Unterricht – INFOS 2003

P-33 Andreas Geyer-Schulz, Alfred Taudes (Hrsg.): Informationswirtschaft: Ein Sektor mit Zukunft

P-34 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 1)

P-35 Klaus Dittrich, Wolfgang König, Andreas Oberweis, Kai Rannenberg, Wolfgang Wahlster (Hrsg.): Informatik 2003 – Innovative Informatikanwendungen (Band 2)

P-36 Rüdiger Grimm, Hubert B. Keller, Kai Rannenberg (Hrsg.): Informatik 2003 – Mit Sicherheit Informatik

P-37 Arndt Bode, Jörg Desel, Sabine Rathmayer, Martin Wessner (Hrsg.): DeLi 2003: e-Learning Fachtagung Informatik

P-38 E.J. Sinz, M. Plaha, P. Neckel (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2003

P-39 Jens Nedon, Sandra Frings, Oliver Göbel (Hrsg.): IT-Incident Management & IT-Forensics – IMF 2003

P-40 Michael Rebstock (Hrsg.): Modellierung betrieblicher Informationssysteme – MobIS 2004

P-41 Uwe Brinkschulte, Jürgen Becker, Dietmar Fey, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle, Thomas Runkler (Edts.): ARCS 2004 – Organic and Pervasive Computing

P-42 Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Economy – Transaktionen und Prozesse, Anwendungen und Dienste

P-43 Birgitta König-Ries, Michael Klein, Philipp Obreiter (Hrsg.): Persistence, Scalability, Transactions – Database Mechanisms for Mobile Applications

P-44 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): Security, E-Learning, E-Services

P-45 Bernhard Rumpe, Wolfgang Hesse (Hrsg.): Modellierung 2004

P-46 Ulrich Flegel, Michael Meier (Hrsg.): Detection of Intrusions of Malware & Vulnerability Assessment

P-47 Alexander Prosser, Robert Krimmer (Hrsg.): Electronic Voting in Europe – Technology, Law, Politics and Society

P-48 Anatoly Doroshenko, Terry Halpin, Stephen W. Liddle, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications

P-49 G. Schiefer, P. Wagner, M. Morgenstern, U. Rickert (Hrsg.): Integration und Daten sicherheit – Anforderungen, Konflikte und Perspektiven

P-50 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 1) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm

P-51 Peter Dadam, Manfred Reichert (Hrsg.): INFORMATIK 2004 – Informatik verbindet (Band 2) Beiträge der 34. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 20.-24. September 2004 in Ulm

P-52 Gregor Engels, Silke Seehusen (Hrsg.): DELFI 2004 – Tagungsband der 2. e-Learning Fachtagung Informatik

P-53 Robert Giegerich, Jens Stoye (Hrsg.): German Conference on Bioinformatics – GCB 2004

P-54 Jens Borchers, Ralf Kneuper (Hrsg.): Softwaremanagement 2004 – Outsourcing und Integration

P-55 Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): E-Science and Grid Adhoc-Netz Medienintegration

P-56 Fernand Feltz, Andreas Oberweis, Benoit OjJacques (Hrsg.): EMISA 2004 – Informationssysteme im E-Business und E-Government

P-57 Klaus Turowski (Hrsg.): Architekturen, Komponenten, Anwendungen

P-58 Sami Beydada, Volker Gruhn, Johannes Mayer, Ralf Reussner, Franz Schweiggert (Hrsg.): Testing of Component-Based Systems and Software Quality

P-59 J. Felix Hampe, Franz Lehner, Key Pousttchi, Kai Rannenberg, Klaus Turowski (Hrsg.): Mobile Business – Processes, Platforms, Payments

P-60 Steffen Friedrich (Hrsg.): Unterrichtskonzepte für informatische Bildung

P-61 Paul Müller, Reinhard Gotzhein, Jens B. Schmitt (Hrsg.): Kommunikation in verteilten Systemen

P-62 Federrath, Hannes (Hrsg.): „Sicherheit 2005“ – Sicherheit – Schutz und Zuverlässigkeit

P-63 Roland Kaschek, Heinrich C. Mayr, Stephen Liddle (Hrsg.): Information Systems – Technology and ist Applications
P-64  Peter Liggesmeyer, Klaus Pohl, Michael Goedicke (Hrsg.): Software Engineering 2005

P-65  Gottfried Vossen, Frank Leymann, Peter Lockemann, Wolffried Stucky (Hrsg.): Datenbanksysteme in Business, Technologie und Web

P-66  Jörg M. Haake, Ulrike Lucke, Djamshid Tavangarian (Hrsg.): DeLFI 2005: 3. deutsche e-Learning Fachtagung Informatik

P-67  Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinlage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 1)

P-68  Armin B. Cremers, Rainer Manthey, Peter Martini, Volker Steinlage (Hrsg.): INFORMATIK 2005 – Informatik LIVE (Band 2)

P-69  Robert Hirschfeld, Ryszard Kowalczyk, Andreas Polze, Matthias Weske (Hrsg.): NODe 2005, GSEM 2005

P-70  Klaus Turowski, Johannes-Maria Zaha (Hrsg.): Component-oriented Enterprise Application (COAE 2005)

P-71  Andrew Torda, Stefan Kurz, Matthias Rarey (Hrsg.): German Conference on Bioinformatics 2005

P-72  Klaus P. Jantke, Klaus-Peter Führich, Wolfgang S. Wittig (Hrsg.): Marktplatz Internet: Von e-Learning bis e-Payment

P-73  Jan von Knop, Wilhelm Haverkamp, Eike Jessen (Hrsg.): “Heute schon das Morgen sehen”

P-74  Christopher Wolf, Stefan Lucks, Po-Wah Yau (Hrsg.): WEWoRC 2005 – Western European Workshop on Research in Cryptology

P-75  Jörg Desel, Ulrich Frank (Hrsg.): Enterprise Modelling and Information Systems Architecture

P-76  Thomas Kirste, Birgitta König-Ries, Key Pousttchi, Klaus Turowski (Hrsg.): Mobile Informationssysteme – Potentiale, Hindernisse, Einsatz

P-77  Jana Dittmann (Hrsg.): SICHERHEIT 2006

P-78  K.-O. Wenkel, P. Wagner, M. Morgenstern, K. Luzi, P. Eisermann (Hrsg.): Land- und Ernährungswirtschaft im Wandel

P-79  Bettina Biel, Matthias Book, Volker Gruhn (Hrsg.): Softwareengineering 2006

P-80  Mareike Schoop, Christian Huemer, Michael Rebstock, Martin Bichler (Hrsg.): Service-Oriented Electronic Commerce

P-81  Wolfgang Karl, Jürgen Becker, Karl-Erwin Großpietsch, Christian Hochberger, Erik Maehle (Hrsg.): ARCS ’06

P-82  Heinrich C. Mayr, Ruth Breu (Hrsg.): Modellierung 2006

P-83  Daniel Husen, Oliver Kohlbacher, Andrei Lupas, Kay Nieselt and Andreas Zell (eds.): German Conference on Bioinformatics

P-84  Dimitris Karagiannis, Heinrich C. Mayr (Hrsg.): Information Systems Technology and its Applications

P-85  Witold Abramowicz, Heinrich C. Mayr (Hrsg.): Business Information Systems

P-86  Robert Krimmer (Ed.): Electronic Voting 2006

P-87  Max Mühlhäuser, Guido Rößling, Ralf Steinmetz (Hrsg.): DELFI 2006: 4. e-Learning Fachtagung Informatik

P-88  Robert Hirschfeld, Andreas Polze, Ryszard Kowalczyk (Hrsg.): NODe 2006, GSEM 2006

P-90  Joachim Schelp, Robert Winter, Ulrich Frank, Bodo Rieger, Klaus Turowski (Hrsg.): Integration, Informationslogistik und Architektur

P-91  Henrik Stormer, Andreas Meier, Michael Schumacher (Eds.): European Conference on eHealth 2006

P-92  Fernand Feltz, Benoit Otjacques, Andreas Oberweis, Nicolas Poussing (Eds.): AIM 2006

P-93  Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 1

P-94  Christian Hochberger, Rüdiger Liskowsky (Eds.): INFORMATIK 2006 – Informatik für Menschen, Band 2

P-95  Matthias Weske, Markus Nüttgens (Eds.): EMISA 2005: Methoden, Konzepte und Technologien für die Entwicklung von dienstbasierten Informationssystemen

P-96  Saartje Brockmans, Jürgen Jung, York Sure (Eds.): Meta-Modelling and Ontologies

P-97  Oliver Göbel, Dirk Schadt, Sandra Frings, Haro Hase, Detlef Günther, Jens Neden (Eds.): IT-Incident Mangament & IT-Forensics – IMF 2006
P-98 Hans Brandt-Pook, Werner Simonsmeier und Thorsten Spitta (Hrsg.): Beratung in der Softwareentwicklung – Modelle, Methoden, Best Practices

P-99 Andreas Schwill, Carsten Schulte, Marco Thomas (Hrsg.): Didaktik der Informatik

P-100 Peter Forbrig, Günter Siegel, Markus Schneider (Hrsg.): HDI 2006: Hochschuldidaktik der Informatik

P-101 Stefan Böttinger, Ludwig Theuvsen, Susanne Rank, Marlies Morgenstern (Hrsg.): Agrarinformatik im Spannungsfeld zwischen Regionalisierung und globalen Wertschöpfungsketten

P-102 Otto Spaniol (Eds.): Mobile Services and Personalized Environments

P-103 Alfons Kemper, Harald Schöning, Thomas Rose, Matthias Jarke, Thomas Seidl, Christoph Quix, Christoph Brochhaus (Hrsg.): Datenbanksysteme in Business, Technologie und Web (BTW 2007)

P-104 Birgitta König-Ries, Franz Lehner, Rainer Malaka, Can Türker (Hrsg.): MMS 2007: Mobilität und mobile Informationssysteme

P-105 Wolf-Gideon Bleek, Jörg Raasch, Heinz Züllighoven (Hrsg.): Software Engineering 2007

P-106 Wolf-Gideon Bleek, Hemming Schwentner, Heinz Züllighoven (Hrsg.): Software Engineering 2007 – Beiträge zu den Workshops

P-107 Heinrich C. Mayr, Dimitris Karagiannis (eds.): Information Systems Technology and its Applications

P-108 Arslan Brömme, Christoph Busch, Detlef Hühnlein (eds.): BIOSIG 2007: Biometrics and Electronic Signatures

P-109 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.): INFORMATIK 2007 Informatik trifft Logistik Band 1

P-110 Rainer Koschke, Otthein Herzog, Karl-Heinz Rödiger, Marc Ronthaler (Hrsg.): INFORMATIK 2007 Informatik trifft Logistik Band 2

P-111 Christian Eibl, Johannes Magenheim, Sigrid Schubert, Martin Wessner (Hrsg.): DeLiF 2007: 5. e-Learning Fachtagung Informatik

P-112 Sigrid Schubert (Hrsg.): Didaktik der Informatik in Theorie und Praxis

P-113 Sören Auer, Christian Bizer, Claudia Müller, Anna V. Zhidanova (Eds.): The Social Semantic Web 2007 Proceedings of the 1st Conference on Social Semantic Web (CSSW)


P-115 Claudia Falter, Alexander Schliep, Joachim Selbig, Martin Vingron and Dirk Walther (Eds.): German conference on bioinformatics GCB 2007


P-117 Ryszard Kowalczyk (Ed.): Grid service engineering and management The 4th International Conference on Grid Service Engineering and Management GSEM 2007

P-118 Andreas Hein, Wilfried Thoben, Hans-Jürgen Appelrath, Peter Jensch (Eds.): European Conference on eHealth 2007

P-119 Manfred Reichert, Stefan Strecker, Klaus Turowski (Eds.): Enterprise Modelling and Information Systems Architectures Concepts and Applications

P-120 Adam Pawlak, Kurt Sandkuhl, Wojciech Cholewa, Leandro Soares Indrusiak (Eds.): Coordination of Collaborative Engineering - State of the Art and Future Challenges

P-121 Korbinian Hermann, Bernd Bruegge (Hrsg.): Software Engineering 2008 Fachtagung des GI-Fachbereichs Softwaretechnik

P-122 Walid Maalej, Bernd Bruegge (Hrsg.): Software Engineering 2008 - Workshopband Fachtagung des GI-Fachbereichs Softwaretechnik
P-208 Ursula Goltz, Marcus Magnor, Hans-Jürgen Appelrath, Herbert Matthies, Wolf-Tilo Balke, Lars Wolf (Hrsg.) INFORMATIK 2012
P-209 Hans Brandt-Pook, André Fleer, Thorsten Spitta, Malte Wattenberg (Hrsg.) Nachhaltiges Software Management
P-210 Erhard Plödereder, Peter Dencker, Herbert Klenk, Hubert B. Keller, Silke Spitzer (Hrsg.) Automotive – Safety & Security 2012 Sicherheit und Zuverlässigkeit für automobile Informationstechnik
P-212 Arslan Brömme, Christoph Busch (Eds.) BIOSIG 2013 Proceedings of the 12th International Conference of the Biometrics Special Interest Group 04.–06. September 2013 Darmstadt, Germany
P-213 Stefan Kowalewski, Bernhard Rumpe (Hrsg.) Software Engineering 2013 Fachtagung des GI-Fachbereichs Softwaretechnik
P-217 Paul Müller, Bernhard Neumair, Helmut Reiser, Gabi Dreo Rodosek (Hrsg.) 6. DFN-Forum Kommunikations-technologien Beiträge der Fachtagung 03.–04. Juni 2013, Erlangen

P-218 Andreas Breiter, Christoph Rensing (Hrsg.) DeLFI 2013: Die 11 e-Learning Fachtagung Informatik der Gesellschaft für Informatik e.V. (GI) 8.-11. September 2013, Bremen
P-222 Reinhard Jung, Manfred Reichert (Eds.) Enterprise Modelling and Information Systems Architectures (EMISA 2013) St. Gallen, Switzerland September 5 - 6, 2013

The titles can be purchased at:
Köllen Druck + Verlag GmbH
Ernst-Robert-Curtius-Str. 14 · D-53117 Bonn
Fax: +49 (0)228/9898222
E-Mail: druckverlag@koellen.de