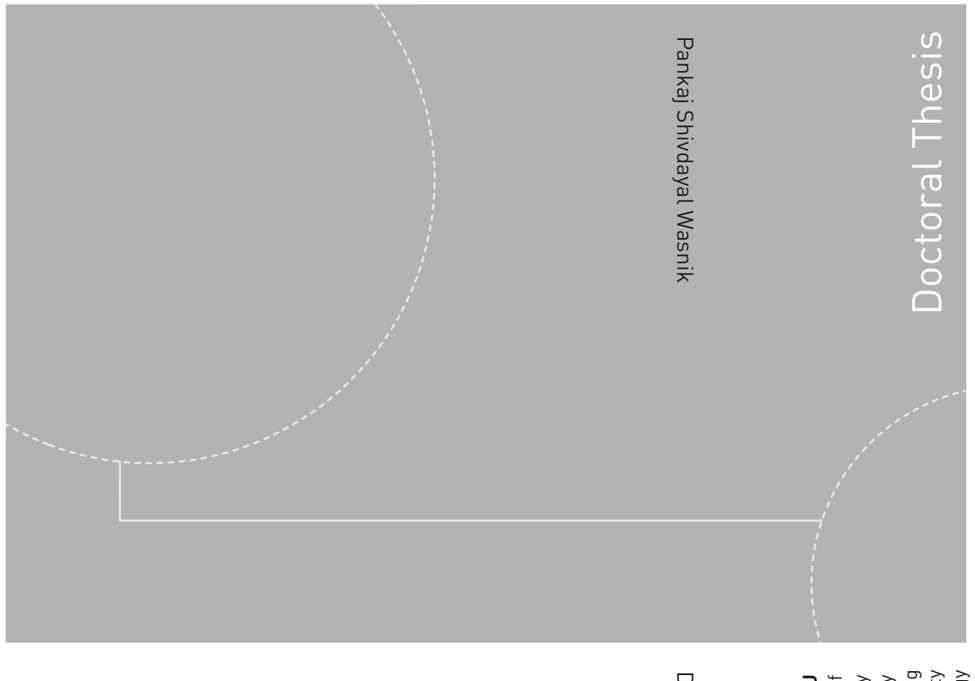
ISBN 978-82-326-3860-4 (printed version)
ISBN 978-82-326-3861-1 (electronic version)
ISSN 1503-8181




Doctoral theses at NTNU, 2019:131

Pankaj Shivdayal Wasnik

# Robust Biometrics on Smartphones

## Using Quality Assessment, Presentation Attack Detection, and Biometric Fusion

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Information Security and Communication Technology

Doctoral theses at NTNU, 2019:131

NTNU
Norwegian University of Science and Technology

NTNU
Norwegian University of Science and Technology

Pankaj Shivdayal Wasnik

# Robust Biometrics on Smartphones

## Using Quality Assessment, Presentation Attack Detection, and Biometric Fusion

Thesis for the degree of Philosophiae Doctor

Gjøvik, May 2019

Norwegian University of Science and Technology
Faculty of Information Technology
and Electrical Engineering
Department of Information Security and Communication Technology

NTNU
Norwegian University of
Science and Technology

**NTNU**
Norwegian University of Science and Technology

Thesis for the degree of Philosophiae Doctor

Faculty of Information Technology
and Electrical Engineering
Department of Information Security and Communication
Technology



ISBN 978-82-326-3860-4 (printed version)
ISBN 978-82-326-3861-1 (electronic version)
ISSN 1503-8181

Doctoral theses at NTNU, 2019:131

Printed by Skipnes Kommunikasjon as

*Dedicated to the beautiful women of my life my mother, Shyamla, my wife, Soniya and my daughter, Iyovi without them it would not be this much fun completing this thesis.*

## Statement of authorship

I declare that I completed and wrote this thesis on my own and the information which has been directly or indirectly taken from other sources has been noted as such. Neither this nor a similar work has been presented to an examination committee elsewhere.

Gjøvik, 8th May 2019 .....

# Abstract

With the technological advancements in mobile technology, there is a massive adoption of biometrics as a security measure in today1s smartphones. Smartphones are used in all day to day activities such as online banking, accessing official and personal emails, social networking and also to store personal data. Although smartphones provide high user convenience, there is an inherent security threat as losing such a device could lead to a loss of such sensitive data. This could cause disastrous effects on the smartphone user. In order to reduce the privacy and security threats, basic solutions are provided with every smartphone. However such solutions could cause user inconvenience sometimes, for example, it is hard to remember complex lock patterns, longer pin codes; also such patterns and pins could be easily hacked. Thus, an inherent need of added security measure is there and which could be conveniently fulfilled by biometrics on smartphones. As a result of which, recently, most of the smartphones are manufactured with inbuilt fingerprint sensor, or state-of-the-art face or iris recognition system.

Today, we can say that for any smartphone, a biometric system is one of an essential component just like the front and rear cameras. However, the inclusion of such a biometric system comes with a cost such as the performance of a biometric system depends on several factors such as the input sample quality, systematic and random errors. Moreover, biometric systems are highly vulnerable to direct and indirect attacks. The direct attacks aka presentation attacks are carried out at the biometric sensor level by presenting a fake biometric sample. If a biometric system does not have an attack detection module also known as presentation attack detection module, it is trivial to spoof any biometric system.

Thus, the primary objectives of this thesis are to address the challenges of smartphone biometrics. The unconstrained nature of biometric samples captured in a smartphone environment could cause challenging input samples for the recognition system and results in a lower comparison score. Therefore, it is essential to assess the precise quality if the input samples. In this work, we present and compare several quality assessment algorithms to formulate a unified face recognition system. This thesis proposes two presentation attack detection techniques for smartphone-based face recognition systems and one for fingerphoto recognition systems. The thesis also extends the applications of some concepts from Subjective Logic to fuse the comparison scores from face and fingerprint recognition systems. Additionally, this thesis proposes a multi-biometric and multi-algorithmic fusion scheme to mitigate the effects of body weight variations for face recognition systems. Although the proposed framework does not use smartphone biometric data, the method could be easily adapted for the smartphone-based face recognition.

The validity of proposed frameworks for consistent performance is demonstrated through extensive experimentation on publicly available and newly created databases. We have also presented a new smartphone based multimodal biometric database as well as a presentation attack database in this work. Conclusively, the thesis proposes robust Biometric Quality Assessment (BQA), Presentation Attack Detection (PAD) and Biometric Fusion techniques to address the issue of sample quality assessment, presentation attacks, and multi-modal biometric fusion. A detailed experimental analysis and comprehensive studies have been executed to evaluate the proposed methods under the scope of this thesis work. The presented methods will help the researchers and users of smartphone biometrics to improve the robustness of the systems.

# Acknowledgment

First of all, I want to express my most profound gratitude to both of my supervisors Professor Raghavendra Ramachandra, and Professor Christoph Busch. Thank you for your constant support, opportunities and valuable feedback which I received from you. Being a fresh researcher, I have learned a lot from you academically as well as personally. I am honored to complete my Ph.D. research under your supervision. I have spent the most amazing last three years of my life working with you. I am indebted to all of my co-authors Dr. Kiran Raja, Prof. Raghavendra Ramachandra, Prof. Christoph Busch, Martin Stokkenes and others for sharing their valuable time, expertise, and guidance to achieve my research objectives. As a project funded Ph.D. student, I would like to thank the Research Council of Norway for their valuable funding and support for the SWAN project. I want to thank all the partners of the SWAN project, and it was a great experience working with such an international team of researchers.

I would like to express my gratitude to the faculties of the University of Oslo and NTNU with whom I completed my course works. I would also like to thank the Research School of Computer and Information Security (COINS) for funding me to attend Winter School and Ph.D. seminars, which helped me to communicate with fellow Ph.D. candidates from Norway. I would also like to acknowledge the financial support provided by NTNU for attending the conferences and seminars. During Ph.D., many participants helped me by providing their data. Thanks to students and faculty of S.G.G.S.I.E & T, Nanded as well as all the participants from NTNU who were the part of our various data collection. I am thankful for administration support received from Nils Karlstad Svendsen, Kathrine Huke Markengbakken, Urszula Nowostawska, Rachael McCallum, Hilde Bakke, Jingjing Yang, Stein Runar

Olsen, and Ingrid Schantz Bakka for making all the necessary official arrangements. Thanks to the IT department, who promptly provided their help during many technical issues.

I have made many excellent friends during my stay at NTNU. I am especially thankful to Dr. Vivek Agrawal for being a true friend and a great colleague, I will always remember the excellent time we had together. I would also like to thank Dr. Kiran Raja and Dr. Guoqiang Li for helping me to get settled in the initial days and helping me with many Ph.D. related things. Dr. Kiran is still an excellent mentor to me. I am very thankful to my best friend Martin Stokkkense for always being there and continually supporting me officially and personally, without him there would have been tough times in Gjøvik. I would also like to thank Adam Szekeres for helping and talking to me during stressed hours. I am thankful to all my friends especially to Aland (Bro), Hokan, Harish who always made my days beautiful.

Last but not least, I would like to thank my wife, Soniya Raipure for her unconditional love, encouragement, and support throughout my studies. Most importantly, I will always be grateful to her for blessing me with our daughter Iyovi, and I am very thankful to my mother-in-law Ms. Birla Ambade for helping us during pregnancy which hugely supported me to concentrate on writing this thesis. I am incredibly blessed to have my family members as my solid backbone. I am grateful to my father, Mr. Shivdayal Wasnik, mother, Mrs. Shyamla Wasnik, brother, Amit, and my sweet sisters, Manisha and Rhutuja.

# Contents

# List of Tables

# List of Figures

# List of Abbreviations

**Abbreviations**

ACER Average Classification Error Rate

APCER Attack Presentation Classification Error Rate

AUC Area Under Curve

BPCER Bona Fide Presentation Classification Error Rate

BQA Biometric Quality Assessment

CNN Convolutional Neural Network

COTS Commercial off-the-shelf system

DNN Deep Neural Network

ERC Error Versus Reject Curves

FC Fully Connected

FMR False Match Rate

FNMR False Non-Match Rate

FQA Face Quality Assessment

FTE Failure-to-Enrol

GMR Geniuen Match Rate

IAPMR Impostor Attack Presentation Match Rate

LBP Local Binary Patterns

MFR Maximum Filter Response

PA Presentation Attack

PAD Presentation Attack Detection

PAI Presentation Attack Instrument

PAUC Partial Area Under Curve

SL Subjective Logic

SLF Subjective Logic Fusion

SVM Support Vector Machine

SWAN Secure Access Control over Wide Area Network

# Part I

# Overview

# Chapter 1

# Introduction

Today, smartphones have evolved more into a personal mobile computer than just a communication device and contain sensitive information of the owner of the smartphone. Thus, the need for enhanced security of mobile devices becomes apparent. User identification via biometrics could be one of the promising solutions for the safety of a device. In recent years after the introduction of biometric sensors for fingerprint, iris, and face, in smartphones, there has been tremendous growth in the application of biometric recognition as an essential authentication factor. Moreover, such biometric recognition systems show higher and reliable recognition performance when compared to the traditional authentication mechanisms such as passwords and lock patterns. Furthermore, the addition of dedicated biometric sensors embedded in smartphones can be the critical factor which aids to achieve higher recognition accuracy.

However, biometric recognition systems with the smartphone camera as a biometric sensor have inherent advantages over the systems with dedicated biometric sensors. For example, in the case of fingerphoto recognition, it leaves no latent fingerprints on the camera, that provides a larger recognition area. Similarly, in case of face and iris recognition, it provides a cost-effective biometric recognition system since we can re-use the built-in smartphone cameras to acquire the data from other biometric characteristics. On the other hand, the performance of such a biometric system is often affected by the quality of an input biometric sample. Therefore, it is essential to assess the true quality of a biometric sample that in our case is the facial image of a subject. The thesis aims to provide a unified face recognition system with a FQA framework to determine the quality of a face sample, and the

computed quality can be further used for invoking the recapture pipeline or compensating the artefacts.

In practice, a biometric system is subjected to have systematic and random errors due to the inherent nature of biometric characteristics, and intersession variability. These errors are non-quantifiable and can introduce uncertainty in obtained mated or non-mated comparison scores. In such scenarios, biometric fusion can be employed to improve the performance of a biometric system. By incorporating the uncertainty while performing the fusion of information from multiple sources, we can reduce the effects of systematic and random errors. This thesis aims to provide novel biometric fusion using Subjective Logic to handle the uncertainty correctly. The thesis provides a multi-modal biometric system using face and fingerprint recognition systems.

Recent surveys project that the billions of mobile biometric applications concerning device security, online shopping, data privacy, and online banking will be downloaded by the year 2020. Mobile payment transactions worth billions will be authenticated using mobile biometrics. Ergo, the security threats of biometric systems have exponentially increased. All the widely adopted biometric systems are highly vulnerable to the presentation and indirect attacks. This thesis aims to investigate and propose effective countermeasures to prevent presentation attacks on a biometric system on smartphones. Collectively, this work is concentrating on the formulation of a robust and accurate multi-modal biometric recognition system for smartphones-based online banking applications.

Furthermore, this dissertation is based on the research work carried out under the scope of the SWAN project* at the Norwegian University of Science and Technology (NTNU).

## 1.1 An Overview of the SWAN project

The Research Council of Norway funded SWAN project. The duration of the SWAN project is four years, and it started in November 2015. There are five other partners in the SWAN project: 1. University of Oslo 2. SAFRAN Morpho 3. Institut de Recherche Idiap 4. Association of German Banks 5. Zwipe AS

*For more details about the SWAN project; please visit https://www.ntnu.edu/iik/swan

### 1.1.1 Project Objectives:

From the past few years, there is immense growth in mobile technologies. Today, smartphones come with high power processors, the specially optimized operating systems, broadband connectivity, and high-throughput applications. Nowadays, smartphones have evolved as a personal computer from just a communication device and regularly used for banking and financial applications. The threats like identity theft and impersonation to steal money from someone's bank accounts is one of the most critical issues that directly impact economic development. Consequently, the need for enhanced security of mobile devices becomes apparent in such cases. Thus, user authentication using biometrics could be a promising and convenient solution for the safety of a mobile device. The primary objective of the SWAN project is to research and develop a secure access control platform using multi-modal biometrics on a smartphone. The main research objectives of the SWAN project are divided into four parts and described in subsequent sections:

#### 1.1.1.1 Trustworthy biometrics

There is a massive adoption of smartphones at the consumer level, and it would be fair to say that smartphones have evolved as the personal computer from just a communication device. Due to seamless user experience, convenience and the highest level of protection smartphones are massively being used in the financial transactions. However, having such devices hacked could lead to psychological and financial consequences as it can contain sensitive information about the user. Using biometrics can undoubtedly help to enhance the security of such personal devices. However, most of the biometric systems are vulnerable to the presentation attacks (spoofing or direct attacks). Thus, in the SWAN project, extensive study of vulnerabilities of the 2D face, fingerprint, eye, and voice biometric systems concerning the direct attacks is being carried out. Moreover, a robust system is being developed which will prohibit the presentation attacks by employing novel presentation attack detection schemes.

#### 1.1.1.2 Privacy preserving biometrics

On or off device protection of biometric data is a critical concern since the biometric data is highly sensitive and can be misused in case of loss of theft. Thus, in the SWAN project novel, privacy-preserving techniques are developed by employing advanced template protection methods. This is being achieved by taking privacy into account during the development

process to follow the privacy by design framework. The provided solution will be accurate and robust against the intra-class variability of biometric samples. Finally, the revocable identifier based on the input biometric data will be created which will also support a non-invertibility to protect the biometric data.

#### 1.1.1.3 Trustworthy transaction protocols

Trustworthy transaction authentication protocols are a vital part of a successful online financial transactions. Along with the advancements in web-based technologies, the phishing and malware threats are significantly growing. By spreading malicious software over communication channels, the transactions originated via connected personal computers and smartphones can be hacked and fraudulent financial transactions could be carried out. It has been studied that such malwares are capable of extracting the customers' banking passwords, account numbers, and valid transaction numbers. Such sensitive information can be misused to impersonate a genuine user and can cause financial losses for both banks and the customers. Furthermore, such malwares are not easy to detect. Thus, in the SWAN project, reliable transaction authentication protocols based on biometrics are being explored in order to provide end-to-end security. Besides, biometric-based transaction authentication protocols store a biometric template locally to avoid any possible chance of losing data by removing the need for a centralized database.

#### 1.1.1.4 Information Fusion

Multi-level banking transactions based on the volume of a transaction can be effectively employed using the multi-modal biometrics. For example, for small volume transactions such as 10$,the smartphone itself as token could be sufficient, while for large volumes such as 100$, the presentation of one biometric characteristic such as a face or fingerprint is sufficient. However, for huge volumes such as 1000$ when sending money to another account, a user needs to present two or more biometric modalities as the transaction amount is significant. Thus, in the SWAN project, based on the volume of a transaction a flexible multi-modal system will be employed using more than one biometric characteristic. This will be achieved by research and development of novel techniques to perform the biometric fusion at different levels such as feature, score, or decision level.

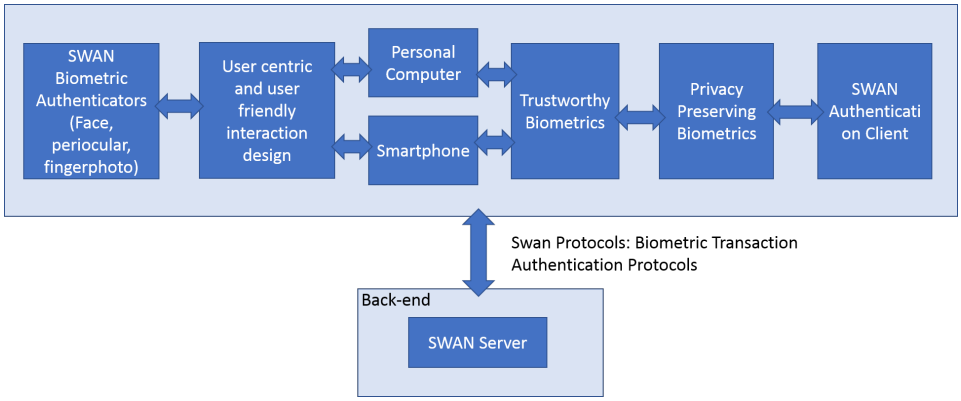



Figure 1.1: Higher-level architecture of the SWAN project

### 1.1.2 An Architecture of the SWAN project

The SWAN Application is a multi-modal biometric system carefully designed as per the objective of the project. It consists of four biometric systems, i.e. 2D face, finger-photo, eye, and voice. Figure 1.1 presents the high-level architecture of the SWAN project. Fundamentally, the architecture of SWAN is a server-client based system where the client software is always running on a smartphone or client pc. Client software, i.e., the SWAN App consists of mainly five modules: 1) Biometric Capture and Quality Assessment Module 2) Presentation Attack Detection Module 3) Template Protection Module, 4) SWAN Authentication Client and 5) Fusion Module. The communication between the SWAN server and client application is carried out through a secure biometric-based transaction protocol.

The data capture module essentially acquires corresponding biometric data as per the volume of an invoked transaction. In-built front or rear cameras are used as sensors to capture the face, eye, and finger-photo biometrics whereas for voice biometrics, an inbuilt microphone is used as a biometric sensor. In theory, a high quality captured sample should retain a high comparison score when compared with the probe sample of the same person. Thus, on a successful data capture, sample quality assessment is performed. Once, a sample with adequate quality is acquired, the PAD module is invoked to analyze the captured data for the presence of any anomalies regarding PAs. On a successful verification from the PAD module, the features are extracted from input sample and a revocable identifier is produced using template protection mechanisms to secure the biometric template. The secured template can further be used to encrypt the shared secret from the

banking server. In the verification process shared secret is decrypted using a template obtained from the probe sample. On successful verification, the transaction is authenticated. Furthermore, multiple decisions or features or scores from the 2D face, eye, fingerphotos, and voice biometrics are fused to obtain an effective multi-modal decision based on the volume of an invoked transaction.

## 1.2 Motivation and Problem Description

Due to the recent developments in mobile technologies nowadays, it is convenient to use mobile phones for many applications where user-specific sensitive information is used. A massive amount of such sensitive information is stored on mobile phones every day. Losing such information could cause mental and financial damage to the user. Thus, a need for enhanced security of mobile device becomes apparent. User identification via biometrics could be one of the most promising solutions for the safety of such devices. In recent years, researchers have investigated a variety of approaches to incorporate biometrics successfully with smartphones. The past decade has seen the rapid development and massive consumer-level adoption of smartphone biometrics in many applications. Many, companies have started manufacturing mobile phones with dedicated biometric sensors such as fingerprint, iris, and face. According to a market report by Acuity Market Intelligence [13], by 2022 the global mobile biometrics market will reach around 60 billion dollars and approximately 1.3 trillion dollars worth financial transactions secured using biometrics will be carried out. Despite its long commercial success, mobile biometric with a dedicated sensor has some problems. For example, in the case of fingerprint recognition system, attackers can obtain the latent fingerprints, 2D printouts of a fingerprint can hack the system [19]. Similarly, the in case of face and iris, the cost of the device is extremely high [26].

Therefore, mobile biometric systems with common smartphone camera as a biometric sensor have inherent advantages over systems with dedicated biometric sensors. For example, in the case of fingerphoto recognition, it leaves no latent fingerprints on the sensor, provides a larger recognition area [192]. Similarly, in case of face and iris recognition, it provides cost-effective biometric recognition system since we can re-use the built-in smartphone cameras to acquire the biometric data. However, the performance of such a biometric system is often affected by the quality of the input biometric sample, and systematic and random errors. Ergo, it is essential to estimate a true quality of biometric samples, in our case, it is a facial image of the sub-

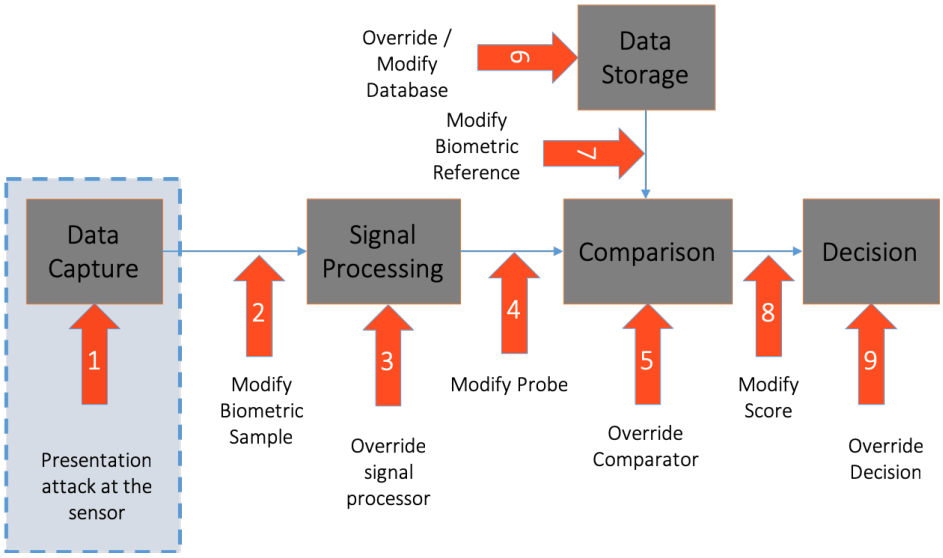



Figure 1.2: Vulnerabilities of Biometric System, inspired by the block diagram given in the International Standard ISO/IEC 30701-1 [73]

ject. Likewise to reduce the effect of authentication errors, one can employ biometric fusion which can improve system performance by incorporating information from multiple sources.

Additionally, a recent survey [149] shows that millions of smartphone-based biometric applications related to phone security, e-commerce, and payment transactions are being downloaded each year. Also, payment transactions worth billions are being authenticated using mobile biometrics. Hence, the security threat associated with the biometric system has increased. Figure 1.2 shows the main vulnerabilities of a typical biometric system described in the standards ISO/IEC 30107-1. One of the critical vulnerabilities of the biometric system is the presentation attack *aka* direct attack that is shown by Red Arrow 1 in the above figure. A presentation attack can be defined as *"an attack carried out by presenting a fake biometric sample (artefact) to a biometric sensor in order to disturb the usual operation of a biometric system"*. Presentation attacks are used to impersonate a genuine as well as to conceal someone's identity. Ergo, this dissertation work primarily concentrates on formulating a robust and accurate BQA, PAD and Biometric Fusion frameworks to address the issues of sample quality assessment, presentation attacks, and multi-modal biometric fusion, thesis aims to develop a robust biometric authentication system on smartphones

for real-world banking applications.

## 1.3 Research Objectives

The research objectives of the thesis are to explores the domains of Trustworthy biometrics and Information Fusion which are critical and important research areas concerning the smartphone biometrics. Thus, the thesis aims to achieve the following five main research objectives:

1. To empirically support the arguments presented against research questions, a database of at least 50 subjects concerning all the modalities that are considered in the SWAN project, i.e., 2D face, fingerphoto, eye, and voice must be constructed.
2. To develop novel methods for quality assessment of acquired biometric data, however, in this thesis a method to analyze the face data is proposed which can easily be adapted for other biometrics.
3. To develop novel algorithms for presentation attack detection for face and fingerphoto recognition systems
4. To develop a novel method for multi-biometric fusion, i.e., score level fusion for face and fingerprint recognition system.
5. The proposed methods should be able to generalize across the databases and modalities. This would help to reduce the computational overhead as the developed methods must run on a smartphone.

## 1.4 Research Questions

Based on the background, motivation, and research objectives, this study intends to address the following research questions. Furthermore, the study also intends to develop an accurate and robust multi-modal biometric system on a smartphone for online banking applications.

**RQ 1: Do the captured data using a smartphone have enough useful features to constitute a reliable biometric recognition decision?** (Related chapter: 6)

In the literature, it has been well studied that the authentication performance of a biometric system is directly proportional to its quality [58, 10]. The biometric recognition system is expected to produce

a high similarity score for mated biometric samples if both involved samples have a high-quality value. Thus, it becomes important for a biometric system to estimate an accurate image quality before a captured sample is stored for reference or before a probe sample is processed for the verification or identification process. Purpose of this research question is to explore a unified quality assessment methodology to quantify the predictive performance of quality assessment algorithms. This could be achieved using the Error Versus Reject Curves (ERC) which determines the performance concerning the rate-of-change of false non-match rate of the system with respect to the percentage of genuine comparison rejected due to qualities of corresponding samples [58, 124].

**RQ 2: How vulnerable is the proposed biometric authentication framework to the existing artefacts?** (Related chapter: 8)

The conventional, as well as smartphone-based biometric systems, are vulnerable to the presentation attacks. The vulnerability of biometric systems needs to be quantified in order to know the attack potential of the presented attack. The purpose of this research question is to quantify the vulnerability of all biometric systems and identify which of them are highly vulnerable to presentation attacks. Also, the vulnerability must be studied quantitatively using metrics defined in ISO standards such as IAPMR (Impostor Attack Presentation Match Rate) [73].

**RQ 3: How can we detect the presentation attacks at sensor effectively to secure smartphone based biometric authentication system?** (Related chapters: 7, 8, 9)

In recent years, researchers have shown that current biometric systems, such as face recognition or fingerprint recognition are prone to presentation attacks [115, 145]. The presentation attack is nothing but an attempt to deceive the biometric system by presenting a fake biometric sample to its sensor. The purpose of this research question is to explore various image processing and machine learning techniques to detect the presentation attacks effectively. This thesis will address the issues of presentation attacks concerning to the 2D face and fingerphoto biometric recognition systems. The presentation attacks considered in this research work are print-photo attacks, display-photo attack and replay video attack.

**RQ 4: How to incorporate confidence of each classifier while fusing the multiple decisions effectively for smartphone-based multi-biometric authentication systems?** (Related chapter: 11)

The unconstrained nature of biometric sample capture leads to random and systematic errors which can not be defined well by any mathematical formulae; however it can be modeled using an uncertainty of the system [70]. In literature, several studies have shown that combining the information from multiple biometric sources which are also known as biometric fusion can improve the accuracy of a biometric system [177, 194]. In order to reduce the effect of random and systematic errors, we have to use biometric fusion since it improves the accuracy of a biometric recognition system. The purpose of this research question is to explore various fusion methods suitable for smartphone biometrics.

## 1.5 Research Methodology

Considering the research questions as our basis, the following research methodologies are designed. These methodologies are used throughout the thesis work to address the research questions and to achieve the research objectives:

- **2D Face, Eye and Fingerphoto database creation using smartphone**
  A database consisting of 2D Face, Eye, Fingerphoto is created using state-of-the-art smartphones as there is no publicly available single database. The database consists of 50 subjects. The database is collected using a state-of-the-art smartphone, i.e., Apple iPhone 6S. Data is captured using both cameras of a smartphone, i.e., front and back camera. The data is collected in 6 sessions, and the gap between each session is approximately two weeks.
- **Presentation attack database creation**
  Since there is no publicly available single database consisting of presentation attacks of 2D Face, and Fingerphoto together, we created a PA database. It contains high-quality bona fide presentations of 2D Face, and Fingerphoto biometrics, and their fake samples created using the recorded bona fide data. This work investigates three presentation attack artefacts and attacks, which are 2D print photo attack, display attack, and video attack. This database is used to develop the

presentation attack detection schemes. The study of existing Presentation Attack Instruments (PAI) and vulnerabilities of the baseline system to these PAIs is carried out to understand the attack potential of PAIs. Performance of the proposed PAD algorithms is evaluated using existing PAIs.

- **Presentation Attack Detection Algorithm**
  Novel methods are investigated to differentiate between a bona fide biometric sample from a genuine subject and an artefact presented by the attackers. Various previous PAD techniques such as [108, 169, 126, 125] are also studied to understand the effect of a print photo, display photo and replay attacks. The study developed robust and accurate PAD frameworks for 2D face and fingerphoto recognition system. In order to generalize the PAD algorithm, a classifier-less approach is also investigated in this work.

- **Fusion scheme**
  The biometric characteristics which are used for fusion are the 2D face and fingerphotos. Biometric fusion can be employed at different levels, e.g., at the feature level, score level, and decision level to make an effective final decision [153]. Furthermore, this thesis work investigates score level fusion which gives the best performance as we can incorporate the confidence of classifier during fusion. Thesis mainly focuses on the camera based systems, i.e., 2D face and fingerphoto recognition systems. Novel methods are developed in order to fuse the unimodal scores effectively.

## 1.6 List of included research publications

Following publications are part of this dissertation:

1. Pankaj Wasnik, Raghavendra Ramachandra, Kiran Raja and Christoph Busch. *"An Empirical Evaluation Of Deep Architectures On Generalization Of Smartphone -Based Face Image Quality Assessment."* In proceedings of $9^{th}$ IEEE International Conference On Biometrics: Theory, Applications, And Systems (BTAS 2018), IEEE, 2018.

2. Pankaj Wasnik, Kiran B. Raja, Ramachandra Raghavendra, and Christoph Busch. *"Presentation Attack Detection In Face Biometric Systems Using Raw Sensor Data From Smartphones."* In the $12^{th}$ International Conference On Signal-Image Technology & Internet-Based Systems (SITIS 2016), Pp. 104-111. IEEE, 2016.

3. Kiran B. Raja, Pankaj Wasnik, Raghavendra Ramachandra and Christoph Busch. *"Robust Face Presentation Attack Detection On Smartphones: An Approach Based On Variable Focus."* In the $3^{rd}$ IEEE International Joint Conference On Biometrics (IJCB 2017), Pp. 651-658. IEEE, 2017.

4. Pankaj Wasnik, Ramachandra Raghavendra, Kiran Raja, and Christoph Busch. *"Presentation Attack Detection for Smartphone Based Fingerphoto Recognition Using Second Order Local Structures"* In the proceedings of $14^{th}$ International Conference on Signal-Image Technology & Internet-Based Systems (SITIS 2018), IEEE, 2018.

5. Pankaj Shivdayal Wasnik, Kiran B. Raja, R. Raghavendra, and Christoph Busch. *"Eye Region Based Multibiometric Fusion To Mitigate The Effects Of Body Weight Variations In Face Recognition."* In Information Fusion (FUSION), 2016 19th International Conference On, Pp. 2007-2014. IEEE, 2016.

6. Pankaj Wasnik, Raghavendra Ramachandra, Kiran Raja, and Christoph Busch. *"Subjective Logic Based Score Level Fusion: Combining Faces and Fingerprints."* In the $21^{st}$ International Conference On Information Fusion (FUSION 2018), pp. 515-520. IEEE, 2018.

## 1.7 List of additional research publications

1. Pankaj Wasnik, Raghavendra Ramachandra, Christoph Busch, and Kiran Raja. *"Improved Fingerphoto Verification System Using Multiscale Second Order Local Structures."* In the 17th International Conference of the Biometrics Special Interest Group (BIOSIG 2018), Darmstadt, Germany, 26.-28.09.2018.

2. Pankaj Wasnik, Mihkal Dunfjeld, Martin Stokkenes, Kiran Raja, Raghavendra Raghavendra and Christoph Busch. *"Baseline Evaluation of Smartphone based Finger-photo Verification System: A Preliminary Study of Technology Readiness."* In The Norwegian Information Security Conference (NISK 2018), Svalbard, Norway. Sept 18-20, 2018.

3. Tommy Thorsen, Pankaj Wasnik, Christoph Busch, R. Raghavendra, and Kiran Raja. *"Assessing Face Image Quality With Lstms."* In The Norwegian Information Security Conference (NISK 2018), Svalbard, Norway. Sept 18-20, 2018.

4. Pankaj Wasnik, Kirstina Schafer, Kiran Raja, Raghavendra Ramachandra, and Christoph Busch. *"Fusing Biometric Scores Using Subjective Logic For Gait Recognition On Smartphone."* In the 16th International Conference of the Biometrics Special Interest Group (BIOSIG), 2017.

## 1.8 Scope of the Thesis

The main scope of the thesis is to investigate various methods for sample quality assessment, presentation attack detection, and biometric fusion. The focus of the thesis is to develop novel techniques to deal with the issues of sample quality, presentation attacks, and effective biometric fusion. The behaviour of system performance is studied corresponding to the quality of input samples. The vulnerabilities of biometric systems are studied quantitatively to analyze the potential of various presentation attack instruments. Furthermore, the effects of uncertainties associated with comparison scores are reduced by employing the biometric fusion; the same approach is also used to provide a volume-based multi-modal biometric system for a smartphone-based banking application. The thesis also presents various algorithms for the 2D face quality assessment, 2D face & fingerphoto PAD, a multi-biometric fusion of face, left and right periocular region, and 2D face & fingerprint score level fusion. The scope of the thesis is limited to two modalities which are 2D face and fingerphoto recognition system. The intended audience of the thesis is biometric security professionals and the researchers from biometrics, image processing, and machine learning domain.

## 1.9 Thesis Outline

This thesis is divided into three parts: Part I presents an overview of the thesis, Part II presents published research articles and appendices are presented in Part III.

In Part I, Chapters 1 discusses an introduction of the thesis by describing details of the SWAN project, motivation and problem description, research objectives, questions and methodology, list of included research articles, list of additional articles followed by the scope of the thesis. Chapter 2 presents the background and related work which presents core concepts and the comprehensive survey of the state-of-the-art, which is an essential factor that helped in the formation of this thesis. Chapter 3 presents a detailed summary of each of the research article included in this thesis.

The research articles are presented in Part II as a reformatted version of the actual publications. Chapter 6 presents the evaluation of various deep

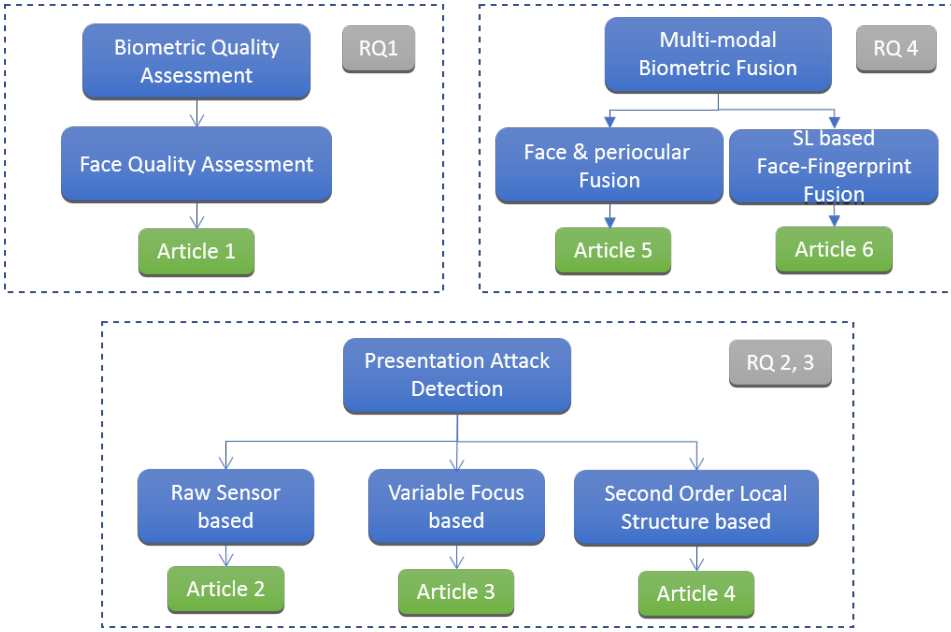



Figure 1.3: Research outline and published articles as per the research questions. Article 1 provides the analysis of FQAA and FQA framework. Article 2,3, and 4 provides the PAD techniques for face and fingerphoto recognition system. Article 5 presents the multi-modal fusion of face and fingerprint based on Subjective Logic.

learning architectures for the task of face image quality assessment along with automatic quality assessment framework for the smartphone-based face recognition system. The presentation attack detection scheme for face recognition systems based on the raw sensor data is discussed in Chapter 7. Chapter 8 presents a presentation attack detection scheme based on the variable focus from a stack of multiple images captured at different focuses. Fingerphoto PAD, based on the second order local structures is given in Chapter 9. The mitigation of effects of weight variations using multi-biometric fusion is presented in Chapter 10. Finally, Chapter 11 presents the novel biometric fusion method based on the subjective logic to fuse the scores from face and fingerprint recognition systems.

The appendices are presented in Part III. Chapter 12 presents Appendix A in which details of the SWAN data capture application are presented. The additional experiments and results are given in Appendix B which is presented in Chapter 13.

# Chapter 2

# Background and Related Work

This chapter provides a brief overview of literature on conventional and smartphone biometrics in Section 2.1 & 2.2. Then moves on to discuss the face quality assessment in Section 2.3. Lastly, Section 2.4 and 2.5 provide literature review on presentation attack detection and biometrics fusion respectively. Throughout this chapter, related background information is provided if necessary.

## 2.1 Biometrics

The international standard ISO/IEC 24741 [74] defines biometrics as *"automated recognition of individuals based on their behavioural and biological characteristics."* A typical biometric system has four main modules: i) a sensing module ii) quality assessment and feature extraction module iii) comparison module and iv) a database module [75]. In the broader spectrum, biometric systems may be divided into two main categories, unimodal and multimodal biometrics [154]. Unimodal biometric systems have several problems of such as sensitivity to the noise in sensed data, intra-class variations, inter-class similarities, non-universality and high vulnerabilities to presentation attacks [153]. Multiple biometric characteristics can be used to address these challenges by fusing the information from each biometric channel to obtain the final similarity or comparison score [154]. Biometrics can further be divided based on the characteristics of the biometric data into physiological and behavioural [74]. Few examples of biometric systems based on physiological characteristics are a face, iris, fingerprint, and

palm-print recognition systems whereas keystrokes, gait, voice recognition systems belong behavioural biometrics.

## 2.2 Smartphone Biometrics

From the past decade, biometrics on a smartphone is widely considered to be an excellent way to secure the mobile device, performing banking transaction, and securing personal data. In contrast to the traditional security measures like passwords and lock patterns which is based on 'what you know', biometrics is based on 'who you are' [76].A typical biometric system on a smartphone is similar to its conventional version except the fact that biometric sensors are embedded is in the mobile device. Despite many similarities, smartphone biometrics seems more challenging due to the unconstrained nature of data capture, limited hardware and lack of standards. Though, smartphone biometrics such as the face, iris, fingerprint, fingerphoto, and voice are a popular medium of security in state-of-the-art mobile phones. One of the earliest works which combines biometric authentication with a mobile phone is proposed in [21], where the authors present a prototype to secure mobile phone with an embedded fingerprint recognition system interfaced with it. However, today's smartphones are embedded with biometric sensors, power-efficient processors, modern operating systems, broadband internet access (e.g., 3G, 4G network connectivity), and productivity-enhancing applications. Billions of units with biometrics have been sold to date, and sales will grow exponentially [150]. The massive consumer level acceptance of smartphone biometrics propelled the development and downloads of millions of mobile biometric applications [149]. With an increase in the number of users, the amount of sensitive information stored on mobile phones has grown immensely.

Several methods are reported in the literature to address the challenges of smartphone biometrics. Many competitions [14, 202, 33, 34, 85, 59, 107] have been held in order to encourage the state-of-the-art research in smartphone biometrics. [111] discusses the comprehensive review of five physiological and six behavioural methods for smartphone biometrics. [22] incorporates high accuracy algorithms for face recognition on smartphones using GPU. In [198], authors review the methods subjected to the crucial aspects of smartphone biometrics. Similarly, [120] presents a survey of various techniques for user authentication for mobile device security. The recent survey [29], provides the literature review from the year 2017-2018.

Following subsections describe the most relevant state-of-the-art concerning to the research work presented in this thesis.

### 2.2.1 Face Recognition:

Face recognition is becoming an active research area in smartphone biometrics. Researchers are trying to incorporate complex methods on smartphones to achieve high recognition accuracy. Many researchers have proposed novel deep learning architectures which show remarkable accuracy on challenging face databases as well as they generalize well across various databases. Two of such architectures are ArcFace and FaceNet. The ArcFace [37] proposes, an additive angular margin loss to extract the highly discriminative features whereas FaceNet [160] proposes a CNN with triplet loss to learn the differences between same and different subject. Further, the OpenFace [5] is a general purpose open source face recognition library which provides support for mobile applications. OpenFace uses the FaceNet for extracting a 128 dimensional feature vector *aka* embedding and trains a linear SVM for classification purpose. A sparse representation of input images is used in [25] to perform the face recognition in an Android smartphone. A. Hadid *et al.* [61] proposed a FRS in mobile phones. The proposed FRS uses Haar-like features with AdaBoost classifier for detection of face and eye. Then, the face is authenticated using intersection distance between probe and gallery LBP histograms. The successful incorporation of holistic face recognition approaches such as PCA, and LDA in a smartphone environment can be seen in [65]. G. Dave *et al.* [31] develop various algorithms for face recognition on mobile phones. To overcome the limited hardware capabilities of the used mobile phone they have selected the lowest computationally complex algorithm to incorporate in a mobile phone. A hybrid classifier using Gabor Wavelet and MLBP for face recognition is proposed in [128]. Further, recent trends in biometrics have led to the successful integration of FRS with a mobile phone and are widely adopted by many consumers [67].

### 2.2.2 Fingerprint and Fingerphoto Recognition:

An early example of research into fingerprint recognition is proposed in [21], where authors presented a fully functional system using a mobile phone and an external fingerprint sensor. Their proposed system achieved EER of the is 4.16%, and FNMR of 5.85% @ 1% FMR. The iPhone 5S [69] is the first mobile phone which successfully integrated the fingerprint recognition system which is called TouchID [182] in September 2013. Further, [175] illustrates a fingerprint recognition scheme based on Minutia Group Matching (MGM) and MGM with selective attention phase (MGM-SA) to unlock or to confirm user actions in a smartphone. Smartphone camera can also be used to capture the photos of fingers of an individual which can further be used for

verification. Preliminary work on user verification using photos of fingerprints is proposed in [63], where SVM is used for classification and trained on the extracted features from the multispace random projections (MRPs). A digital camera was used to capture the fingerphotos. Mohammad *et al.* [39] use such finger photos captured by mobile phones (Nokia N95, HTC Desire) to perform a fingerprint recognition. They used a commercial-off-the-shelf system from Neurotechnology VeriFinger SDK[161] to extract the minutia as well as for verification. Minutia based fingerphotos recognition has been studied by many researchers [116, 172, 39]. Tiwari *et al.* [180] published a paper in which they proposed a similarity measure system based on the extracted SURF features matched between the reference and probe image. Further, recently, smartphone-based fingerphoto recognition is getting adopted as a commercial solution and is being used for secure online payment and mobile device security [186, 47].

## 2.3 Quality Assessment

Performance of a biometric system depends on the quality of input biometric sample [58]. Due to the similarities between conventional and smartphone-based FRS, it is anticipated that facial images captured by smartphones are prone to the challenges of illumination, pose, expression and age variations. Hence, the accurate quality assessment of face images becomes apparent [191]. ISO/IEC TR 29794-5 [72] specifies methodologies for obtaining the objective quality score for face image. Currently, there are no standards like ISO/IEC TR 29794-5 [72] proposed for smartphone-based FRS. However, [191] checks the conformance of images captured using smartphones with the metrics described in ISO/IEC TR 29794-5 [72]. The works [197, 53] propose metrics based on image properties like brightness, contrast, and sharpness for FQA. In [141] an empirical study for quality assessment of facial images is presented for the application of automatic border control systems. Recently, deep learning based approaches are investigated by researchers [201, 135]. Although there are many works in literature proposed for FQA of conventional FRS, there are very few methods which are investigating the FQA for smartphone-based FRS [191, 179]. In [191] a method based on features constructed using ISO/IEC TR 29794-5 metrics and random forest classifier is proposed. Similarly, [179] presents an LSTM based deep learning architecture to assess the facial images captures using iPhone 6 Plus and Samsung Galaxy S7.

## 2.4 Presentation Attack Detection in Smartphone Environment

Biometric systems are vulnerable to two types of attacks: a)presentation or direct attacks; b) indirect attacks [159]. The presentation attacks are carried out by presenting a fake biometric or artefact at the sensors module to deceive the system [159]. On the other hand, indirect attacks are carried out targeting the software components; these attacks may attempt to bypass the feature extractor or the comparator, to manipulate biometric references, or to exploit vulnerabilities in the communications channels [60]. ISO/IEC 30701-1 [73] defines presentation attack as *"the presentation of an artefact or of human characteristics to a biometric capture subsystem in a fashion intended to interfere with system policy is referred to as a presentation attack"*. Several reports have shown that the biometric systems, such as fingerprint or face recognition are more likely susceptible to presentation attacks. Recent work shows, the presentation attack detection schemes can successfully be incorporated on mobile phones to prohibit the presentation attacks. Following subsections are set out to review the PAD techniques for smartphone-based face, fingerprint, and fingerphoto recognition system.

### 2.4.1 PAD for Face Recognition:

Without effective PAD schemes, most of the state-of-the-art facial biometric systems are vulnerable to presentation attacks. [147] presents a heuristic liveness detection approach by using three different facial images captured at an angle -30°, 0°and +45°for an Android mobile phone. An image distortion analysis (IDA) is used in [195] to perform the PAD. Construction of the IDA feature is done using four different characteristics which are the specular reflection, blurriness, chromatic moment, and color diversity. Multiple SVM classifiers are trained for different presentation attacks (e.g., printed photo and replayed video) to identify the bona fide presentation. Authors used Google Nexus 5 to collect face artefacts and the method is employed to detect three types of attacks, i.e., printed photo, replayed video with iPhone 5S, and replayed video with iPad Air. In [130], K. Patel *et al.* propose a method based on Moiré patterns by comparing fake and bona fide face videos, they observed that Moiré patterns often exist in the entire fake video frame as a distinct texture pattern overlaid on a video frame. The recently conducted competition [14] and OULU-NPU database [16] have boosted the interest of many researchers from academia and industry to explore the domain of face PAD for smartphone environment. In the competition, the approach by the Galician Research and Development Center in Advanced Telecommunications (GRADIANT), Spain has shown consistent

performance across all the four protocols. They fused the information of color [15], texture and motion to extract features from HSV and YCrCb channels, finally SVM based supervised classification is employed to detect the presentation attacks.

### 2.4.2 PAD for Fingerprint and Fingerphoto Recognition:

One of the publicly disclosed successful attacks on the smartphone biometrics is carried out by the Chaos Computer Club. They hacked iPhone 5S, which comes with an in-built capacitive sensor for fingerprint recognition by lifting a fingerprint of the genuine user off a glass surface and then presenting to the fake fingerprint to the sensor [17]. This exposes the vulnerability of the fingerprint recognition system to the presentation attacks despite the claimed security by manufacturers. In [19], authors reported that various types of presentation attacks have not yet been investigated in detail. Further, authors demonstrated the hacking of Android mobile phone with fingerprint sensor by high-quality 2D print of fingerphoto.

Fingerphotos have advantages over fingerprints since data acquisition is in a contact-less manner, which eliminates the problem of latent fingerprints [192]. However, camera-based (here fingerphoto) systems are highly vulnerable to 2D PAs such as print-photo, display photo and replay video attacks [171, 62, 178]. Besides, when we use the camera as a sensor, high-quality attacks can be easily constructed which can be accepted by the system. [173] is one of the earliest work demonstrating the use of fingerphotos for recognition and proposes a PAD algorithm to detect 2D PAs. The proposed method detects liveness based on the reflection from the fingertip area since materials like 2D printouts, and fake fingers do not possess such reflection properties. They have achieved the highest performance of an EER of 1.2-3.0%. Taneja *et al.* [178], proposed a PAD scheme based on SVM classifier and textural and gradient features. Authors investigated the method for print-photos and display photos attacks. Further, their method obtained the best EER of 3.7% for the system based on SVM and LBP features.

## 2.5 Biometric Fusion

Multiple biometric characteristics (e.g., fingerprint, face, voice, palmprints) can be used to perform robust user authentication. In order to have an accurate biometric system, the biometric characteristics should have qualities like universality, distinctiveness and persistence [77]. However, in practice, biometric characteristics do not fully meet these qualities; thus, every single modality biometric system is erroneous. Multimodal biometric systems help

in minimizing some of these limitations [154]. In literature, researchers have proposed various techniques to combine information from multiple biometric systems [156, 12, 46, 181, 154, 189]. Many of these methods show that multimodal biometrics can significantly improve the recognition performance when compared with unimodal biometric systems. In [119, 184] the authors use likelihood ratio (LLR) between the genuine and impostor score distribution to perform the optimal fusion, considering it will reduce the probabilistic errors. [119] uses Gaussian mixture models (GMM) to model the genuine and impostor score distribution. Researchers have also demonstrated the use of auxiliary information such as biometric sample quality [156, 12, 46], and user-specific parameters [181] can be used as weight parameter while fusing the information from multiple unimodal biometric signals. Recently, multimodal biometrics on a smartphone has got the attention of many researchers due to the availability of multiple sensors on the smartphones. [183] is one of the earlier work demonstrating the multimodal biometrics in a smartphone environment, where authors propose a system based on face and voice biometrics. [143] proposes a fusion of face and periocular information employing feature level and score level fusion. In the case of feature level fusion, the features from the face region, left periocular region and right periocular region are concatenated together to obtain the fused feature vector. Similarly, in case of score level fusion, the comparison scores from face region, left periocular region and right periocular region are fused by SUM rule. In [144] proposes a multimodal biometric system using face, iris, and periocular information in a smartphone environment. Recent work proposed in [4] employs a score level fusion method based on logistic regression. The recognition performance of cross-sensor smartphone periocular recognition is improved by mapping the comparison scores to LLRs while performing the fusion. Uncertainties associated with biometric recognition scores can also be utilized to fuse the scores from multiple comparators or biometric channels [82]. [193] uses Subjective Logic [79] to achieve biometric fusion under the influence of uncertainties.

From the reviewed state-of-the-art, it is clear that the smartphone-based multibiometrics and presentation attack detection is an evolving research area. Therefore, this thesis aims to study and investigate the capabilities of a smartphone in this regard. The thesis also aims to formulate an accurate smartphone based multimodal biometric system for real-world mobile banking applications.

# Chapter 3

# Summary of Published Articles

This chapter provides a summary of six publications which are included as a contribution in this thesis. An overview of the research questions and publications is illustrated in Figure 1.3. The summary of these research articles mainly presents an overview of the research problem, motivation, methodology and details of the key findings.

## 3.1 Article 1: An Empirical Evaluation Of Deep Architectures On Generalization Of Smartphone-Based Face Image Quality Assessment

In the literature, it has been well studied that the recognition performance of a biometric system is directly proportional to the input sample quality [58, 10]. Any biometric recognition system is expected to produce a high similarity score for mated biometric samples if both involved samples have a high-quality value. Thus, it becomes apparent need for a biometric system to estimate an accurate image quality before a captured sample is stored for reference or before a probe sample is processed for the verification or identification process. The predictive performance of the quality assessment algorithms is obtained using the Error Versus Reject Curves (ERC) which determines the performance in terms of the rate of change of false non-match rate of the system with respect to the percentage of genuine comparisons rejected due to low qualities [58, 124].

In this article, we formulated a unified framework for smartphone-based

face recognition system with quality assessment. Since many of the Face Quality Assessment Algorithms (FQAA) show poor performance when evaluated against the data samples of unknown origin, we present a robust and accurate quality estimating framework to achieve the generalizability over such data samples. The study presents the comprehensive predictive performance evaluation of 14 FQAAs using ERCs. The presented framework evaluates five well known state-of-the-art Convolutional Neural Networks (CNNs) i.e., AlexNet [94], Vgg16, Vgg19 [164], Inception [176] and Xception [24]. Further, we present evaluation results for 3 state-of-the-art mobile networks [157, 66, 203], and 2 state-of-the-art face quality CNNs [201, 135]. In this paper, we also present results for three blind quality FQAA with one commercial face recognition system [2].

In the case of CNNs, Transfer Learning approach is used to retrain the networks for a task of face image quality assessment. Thus, we first removed the final three layers and added fully connected (FC) layers of size 1024, 512 and number of classes. Each of these FCs is connected to ReLU followed by dropout layer. In order to avoid the over-fitting, we used standard data augmentation techniques. Each CNN is trained for 20 epochs with a batch size of 64 with standard training parameters. The study uses an heterogeneous training database constructed using publicly available face quality databases to achieve generalization across sensors and databases. The used databases are the AR database [109], Extended Yale database [101], FRGC database [133], CAS-PEAL database [52] and NCKU face database [44]. The training database mainly consists of two classes corresponding to good and bad image quality and these images are taken from earlier mentioned databases. In total, it consists of approximately 33000 facial images. In this article, we present our results on five evaluation databases out of which three are camera based databases i.e., ABC [141], Chokepoint [199], SCFace [56] and two mobile databases i.e., Apple iPhone 6 Plus, and Samsung Galaxy S7 Database [191].

Our experimental results show that the inclusion of proposed quality assessment framework based on CNNs like AlexNet or Inception V3 can correctly estimate the quality of input samples. The performance of mobile networks is lower as compared to full CNNs; however, the size and computational time are lesser. One can use such networks in a smartphone environment to develop apps with smaller size but with little lower accuracy. We also observe that the samples which have high quality contributes to the rapid decrements in false non-match rate of the FRS, which in fact is the expected behaviour and can be observed in the obtained ERCs.

## 3.2 Article 2: Presentation Attack Detection In Face Biometric Systems Using Raw Sensor Data From Smartphones

Nowadays it is easy to carry out presentation attack on smartphone biometrics due to the fact of unsupervised data capture, and abundant availability of the face images on social media [23]. Furthermore, recently, researchers have shown that biometric systems such as face recognition or fingerprint recognition are prone to presentation attacks [115, 145]. In general, presentation attack is nothing but an attempt to deceive the biometric system by presenting a fake biometric characteristic to its sensors. Hence, there is a need for countermeasures to detect and deter these attacks to reduce the risk of identity frauds using such biometric systems.

In this article, we propose a robust face presentation attack detection scheme based on the characteristics of the raw data such as a residual noise pattern captured at the sensor. This study uses these noise patterns to classify input images into artefact or bona fide presentations. The method first obtains a residual image to get the corresponding noise component by subtracting the median filtered image from the original raw data. The residue image is then divided into an equal number of blocks. The residual noise pattern is estimated in terms of a summation of block energies. The artefact presentations or fake presentations are then detected based on the computed energy values. The presented approach uses threshold-based classification over learning-based approaches to obtain the robustness along with simplicity which is the intrinsic necessity of a smartphone environment.

The proposed method is evaluated on the database containing bona fide images, paper print attack images, and display attack images. An iOS application is developed to capture the raw data from a smartphone camera. We have used iPhone 6S to collect the data. In total, the database consists of real and fake presentations of 102 subjects. Three types of attack presentations were captured consisting of print attacks, display attacks using Dell UltraSharp 25-inch monitor with QHD display, and Samsung Galaxy Tab 7.0. In total, 510 presentations attack attempts are captured using each type, i.e., two display screens and printed photos. The work presented in this paper processes the data of all three channels along with the combined data of three channels, this gives us a database of 1560 ($390\ images \times (3\ channels\ +\ 1\ RGB)$) live images corresponding to red, green, blue and combined red-blue-green data. Similarly in case of attacks we have 4080 ($2\ devices \times 510\ images \times (3\ channels\ +\ 1\ RGB)$) images display attacks and 2040 ($510\ images \times (3\ channels\ +\ 1\ RGB)$) images

for printed photo attacks. Furthermore, for the experiments, the database is divided into development and testing set. The classification threshold $T$ is determined on the development set, and final results are obtained by applying $T$ on the testing dataset.

This paper evaluates the performance of the proposed PAD in terms of metrics defined in ISO/IEC CD 30107-3 [68]. We used three metrics in this work: (1) Attack Presentation Classification Error Rate (APCER), which is defined as the proportion of misclassified attacks as bona fide presentations (2) Bona Fide Presentation Classification Error Rate (BPCER), which is defined as the proportion of misclassified bona fide presentations as attack presentation (3) Average Classification Error Rate (ACER), which is the average performance from APCER and BPCER.

Our experiments determined the threshold of $T \geq 200000$ and obtained low ACER values indicating the promising performance of the proposed method. The proposed approach successfully demonstrated the high reliability without any especially learned classifier. We have achieved a wide range of thresholds with significantly lower values for APCER, BPCER, and ACER. In the case of print photo attacks, the experiments show an increase in classification error rates as we consider higher values for thresholds indicating a need for further investigation.

## 3.3 Article 3: Robust Face Presentation Attack Detection On Smartphones: An Approach Based On Variable Focus

The smartphone-based face recognition system is widely used in many applications from unlocking the phone to execute an online financial transaction. However, there is increasing concern that face recognition on a smartphone is being disadvantaged due to the successful attacks on the system. The primary challenge faced by such a recognition system is the presentation attacks. A number of methods found to be detecting presentation attacks have been explored in several studies [87, 132, 54, 103, 86, 131, 38, 20, 27]. Most of these earlier studies demonstrate the use of image characteristics such as image quality, textural properties to detect the presentation attacks. In this paper, we have proposed a PAD algorithm based on the total focus difference between the stack of bona fide and artefact images captured at different focal lengths.

The proposed method is mainly divided into three parts (i) Data capture module: it is carried using an Android application developed which uses the intrinsic characteristics of the smartphone camera such as focus variation

to obtain the stack of images; (ii) Stack alignment and its refinement: The stack images generally have parallax motion due to the uncontrolled movement of hands. In order to align the set of stack images the proposed method employs the Inverse Compositional Image Alignment (ICIA) algorithm [8] and refinement is done using Dense Inverse Search based optical flow alignment [95]; (iii) Presentation attack detection: The proposed PAD algorithm first extracts the face region using image at $focus = \infty$, using this face location the face regions from aligned stack are extracted. Secondly, the focus is measured using a focus measure proposed in [96]. Finally, based on the cumulative focus difference between stack images and image at $focus = \infty$ the input image is classified as bona fide or artefact presentation.

This paper evaluates the proposed PAD method on a newly constructed database full-filling the needs of the proposed approach. Our database consists of frontal face images captured at different focal lengths. The dataset is collected using the developed android application and Samsung Galaxy S7. Totally, 50 subjects participated in data collection. The database is further divided into two sets, i.e., development and testing set. Each of them consists of data from 25 subjects. The data collection consists of two sessions: (i) Session 1: consists of 3 images of each subject where the first image is used for vulnerability analysis, the second image is used for enrolment, and the third image used for artefacts (attack) generation. (ii) Session 2: Uses the developed application to collect the stack of images of each subject. Ten different focus values were used to capture the stack of images. For artefact generation, four displays and one printed photo were used to obtain the stack of images. Three recordings per attack instrument were captured, resulting in 1500 attack images per PAI.

In order to assess the proposed scheme, the results are obtained on the testing set by employing the threshold for classifying the images as bona fide or artefact is determined using development set. The paper presents the classification performance in terms of BPCER by fixing the APCER to 5% and 10%. The proposed method achieved very low classification errors indicating the effectiveness of the proposed approach. This paper successfully demonstrated the feature-less and classifier-less approach to detect the presentation attacks. Furthermore, this study discusses the further PAD functionality of the proposed approach in terms of Depth-from-focus, Depth-from-defocus, and Depth based texture analysis.

## 3.4 Article 4: Presentation Attack Detection for Smartphone Based Fingerphoto Recognition Using Second Order Local Structures

Lately, fingerphoto recognition on smartphones is getting attention from many researchers and industry. It provides a good alternative for other security measures. However, due to the fact that fingerphoto recognition systems use the smartphone camera to capture the data, it is vulnerable to the presentation attacks like other biometric systems. So far, the presentation attack detection for smartphone-based fingerphoto recognition systems has not been investigated adequately. There are very few studies which examine vulnerabilities of the system and propose countermeasures for them [171, 178]. This paper proposes a PAD scheme which can detect the printphoto, display and replay attacks. The proposed scheme is based on the second order local structures present in an input image. The paper uses SVM based classification to detect the bona fide and artefact images.

The proposed scheme tries to detect the video display, image display and print photo attacks using the textural features extracted from the maximum response (MFR) images. The very first step in the proposed PAD scheme is to extract the region of interest which is nothing but the finger region. The K-means clustering [170] is used to employ color based segmentation to extract the finger region. MFRs are then obtained from the convolution of the cropped segmented image with second-order derivatives of 2D Gaussian kernel at multiple scales. More specifically, the input image is first convoluted at multiple scales, and the maximum response at every pixel location across the scales is selected to create the MFR image. This response image is then used to extract the textural features. This paper compares three feature, i.e., LBP, BSIF and HOG [122, 83, 28]. The extracted features using each of these methods are then learned using a robust classifier such as support vector machines to classify the image into bona fide or attack presentation.

This paper validates the proposed method on a new database constructed using the developed iOS application. The database consists of 3 sessions: (i) Session 1 & 2: collected indoor with uniform illumination (ii) Session 3: collected outside in daylight conditions. Session 1 data is used for artefact creation. The database consists of 50 subjects out of which randomly selected 33 subjects from Session 2 are used for training. The remaining 17 subjects' data from Session 3 is used for testing purpose. In total, the database has 9900 ($150\ frames \times 33\ subjects\ \times\ 2\ types\ (bona\ fide\ \&\ artefacts)$) images

for training and 5100 (150 $frames \times 17$ $subjects \times 2$ $types$ ($bona\ fide$ & $artefacts$)) images for testing.

In summary, this paper argued that features extracted from MFR images are useful to detect the PAs. The experiments show that the proposed scheme with BSIF feature extractor and SVM has highest performance of BPCER of 1.8%, 0.0% and 0.6% at APCER = 10.0% for the print photo, display photo and replay video attacks respectively. The proposed scheme shows improved performance of an EER of 0.49% for display attacks.

## 3.5 Article 5: Eye Region Based Multibiometric Fusion To Mitigate The Effects Of Body Weight Variations In Face Recognition

The performance of face recognition systems is often affected due to pose, illumination, and expressions. Many researchers have studied these challenges and proposed methods to improves the performance of the system. Furthermore, in recent years many researchers proposed methods to mitigate the effects of various other challenges such as aging, plastic surgery, twin identification, make-up, and hairstyle. However, the effect of weight variations on face recognition has not been studied much.

In this paper, we use the periocular region to reduce the effects of weight variations and improve the performance of a face recognition system. The facial regions such as the cheek or chin area get affected mainly due to the weight variations. However, the periocular region does not get affected much. Motivated by this fact, we propose a robust fusion scheme to reduce the effects of weight variations. The proposed method uses the multi-algorithmic and multimodal fusion strategies to combine information from left and right periocular regions to obtain a robust comparison score.

The proposed method consists of four steps i) ROI extraction: The three ROIs, i.e., face, left periocular and right periocular region is first located using the Viola-Jones algorithm [187], and the ROIS are generated using the cropped region of the input image ii) Feature Extraction: The features are generated using four well-known feature extraction techniques, i.e., local binary patterns, local phase quantization, histogram of gradients, and log-Gabor filters iii) Classification: The extracted features are first learned using the Sparse Representation Classifier (SRC), and the trained classifier is used for the classification step iv) Score level fusion: In this step we fuse the score from multiple sources using proposed fusion strategies.

The proposed fusion scheme is evaluated on the publicly available eWIT [117] database. The database is an extension of the WIT *(WhoIsIt)* database [165]. The eWIT database consists of 2036 images of 200 well-known celebrities downloaded from the Internet. Each subject has at least 10 images of a frontal face with age and weight variations. The average age of the subjects is 34.7 years, and the mean difference between the oldest and youngest age is 28.8 years. To verify the effectiveness of the proposed scheme, we used 50 subjects' data, where each subject consists of 10 frontal face images.

Finally, our experimental results favor the multi-biometric fusion consisting of the fusion of comparison scores based on the face and periocular region. The multiple biometric characteristics and algorithms alleviated the effects of weight variations, and we have obtained the lower EER values when compared against the commercial system. The proposed scheme ameliorated the verification performance by 6.42% in terms of EER. Hence, this paper shows that the fusion of the periocular region with face results in higher performance and reduces the effects of weight variations.

## 3.6 Article 6: Subjective Logic Based Score Level Fusion: Combining Faces And Fingerprints

The unconstrained nature of biometric sample capture leads to random and systematic errors which cannot be defined well by any mathematical formulae however can be modeled in terms of uncertainty of the system [70]. Further, the authentication performance of such a system gets affected due to these types of errors. In literature, several studies have shown that combining the information from multiple biometric sources which is also known as biometric fusion improves the accuracy of a biometric system [177, 194]. This paper considers the case of a multi-modal scenario which combines the comparison scores from a commercial face and fingerprint recognition systems. The use of such a multi-modal approach not only helps to reduce random and systematic errors but also it is robust against Failure-To-Enrol rate (FTE) [154].

The proposed method uses the Subjective Logic which considers uncertainties in the mated or non-mated comparison scores while fusing them. Subjective logic provides a useful fusion framework for combining the information from multiple sources along with their associated uncertainties [79, 78, 80]. The proposed scheme is motivated by the preliminary works introduced in [82, 193], which try to apply the Subjective Logic framework in

the biometric domain. This paper aims to fuse comparison scores obtained from biometric traits such as the face, left and right index finger.

The proposed scheme is mainly divided into two parts: (i) Scores to subjective opinions: The SLF is generally performed on the subjective opinions that are represented by belief, disbelief, and uncertainty associated with input comparison scores. In this paper, we have considered normalized verification scores as belief mass, the uncertainty is estimated at the system level by quantifying the verification errors, and disbelief is computed by Additivity principle [81]. Once we have these values, we can convert the comparison scores to corresponding subjective opinions. (ii) Fusion of Scores: The converted comparison scores, i.e., subjective opinions are then fused using the SL's cumulative fusion operator [81]. Finally, the fused belief is considered as the final fusion output.

We carried our experiments on the well known NIST BSSR1 dataset [30]. The similarity scores from left and right index fingers and two face recognition systems were used in our experiments. In total, the BSSR1 database consists of 517 genuine and 266772 impostor scores. Furthermore, we divided the database as the number of subjects into three parts: (i) Development dataset: it consists of 311 subjects and is used to estimate the uncertainty of the system; (ii) Validation dataset: it consists of 103 subjects and is used to obtain the operating thresholds; (iii) Testing dataset: it consists of 103 subjects and is used for testing the proposed method using the threshold obtained from Validation dataset.

This paper compares the proposed method against the standard fusion rules such as sum, weighted sum, and product rule. These rules were applied on four fusion strategies were i.e., FL, FR, LR, and FLR where F, L, and R stands for face, left index and right index finger respectively. The proposed method shows significant improvement over the baseline performance. However, the obtained results show a slight improvement over the standard fusion techniques. One of the significant advantages of the proposed system is that it correctly performs biometric fusion by incorporating uncertainties associated with the input scores. A benchmarking of the SLF approach with other methods such as fuzzy logic, Bayesian reasoning, and DST would have made a better performance comparison than using the standard fusion rules. However, the paper limits this study to compare SLF with standard fusion rules only.

# Chapter 4

# Conclusions

The thesis aims to improve state-of-the-art research concerning the main research objectives, and questions which are discussed in Section 1.3 & 1.4. The thesis investigated two main parts of the SWAN project, i.e., Trustworthy Biometrics and Information Fusion (see Sec. 1.1.1.1, 1.1.1.4). Extensive research work is done to explore these areas through four research questions and several publications. Six publications are included in the thesis as our main contributions towards the composition of this thesis. Primarily, the thesis emphasizes the issues of sample quality assessment, presentation attacks and multi-modal fusion related to smartphones biometrics. The thesis successfully formulates robust and accurate frameworks for FQA, face & fingerphoto PAD, and multimodal fusion of face and fingerprints for real-world mobile banking applications. Additionally, this thesis investigates the mitigation of weight variations using multi-biometric and multi-algorithmic fusion. Thus, firstly, the thesis proposes a unified FQA framework for smartphone-based FRS. The work studies captured data using smartphones to assess the useful features in order to constitute a reliable biometric recognition system. To this extent, the thesis presents a comprehensive empirical evaluation of various quality assessment algorithms based on state-of-the-art CNNs, quality estimation techniques and also commercial system. Extensive experimentation is conducted to assess and generalize quality algorithms for cross-sensor and cross-database. ERC based formulation is used to quantify the predictive performance of the quality assessment algorithm. The obtained results favour a system based on CNN architectures AlexNet and Inception V3. Interestingly, our experiments show lower performance for the mobile networks MobileNet V2, NASNetMobile and

DenseNet169 compared to full CNNs; however, these networks are smaller in size and their speed of computation higher when tested on evaluation dataset.

Furthermore, this thesis is trying to mitigate the major vulnerability of a biometric recognition system aka presentation attacks. The thesis largely investigates these attacks on a smartphone-based face and fingerphoto recognition systems. It aims to study three types of attacks i) 2D print-photo attack ii) display photo attack and iii) replay video attack. The methods proposed in this thesis successfully detect the presentation attacks with significantly lower error rates. The thesis emphasizes the use of international standards for accurate assessment of PAD algorithms. The face PAD methods proposed in the thesis mainly use intrinsic characteristic of smartphone cameras such as raw sensor noise and camera focus to formulate a classifier-less PAD technique. Similarly, the fingerphoto PAD algorithm mainly exploits the second order characteristics of an input image to classify it into bona fide or artefact presentation. The thesis further proposes a technique for mitigation of effects of weight variations on the face recognition system. Finally, the thesis presents a novel approach to formulate a robust multi-modal biometric system. This work successfully evaluates the application of Subjective Logic to fuse the information/decision from multiple biometric channels. The cumulative fusion operation of subjective logic provided an advantage over SOTA by incorporating the uncertainties present in the mated or non-mated comparison scores. The results, however, do not show significant improvement but could validate correct handling of the probabilistic uncertainties. In order to verify the proposed method, this thesis formulates a multi-modal biometric system comprising the fusion of mated and non-mated comparison scores of commercial face and fingerprint recognition systems. The database used in this study is the NIST BSSR1 Database [30].

Collectively, the thesis concludes that the research objectives are successfully achieved via various presented studies. The thesis empirically addresses all of the research questions in order to meet the research objectives. Furthermore, various novel approaches are developed to constitute a robust multi-modal biometric system for smartphone-based online banking applications.

# Chapter 5

# Future Work

The thesis mainly targets the quality assessment for face biometrics, and it presents the PAD schemes for 2D print-photo, display and video replay attacks and multi-modal fusion using face and fingerprint biometrics. Based on the research work carried under the scope of this thesis, the following limitations and future works are discussed:

## 5.1 Quality Assessment

The current work presented in this thesis is only considering the sample quality assessment for face recognition systems. Future research should propose a quality assessment framework for other involved biometric modalities too since individual QA could very much be the key component in the formulation of a robust multi-modal biometric system. Furthermore, the thesis provides ERC based evaluation for assessing the predictive performance of the FQAA, as ERC is highly dependent of the comparator used to generate the mated comparison scores, potential effects of various comparators on the nature of FQAA could more carefully be observed in the future work. Considering the fact of rapid growth in the 3D face, fingerprint recognition, future studies could fruitfully explore the ERC based quality assessment for 3D face and fingerprint data. Future research should also be devoted to explore the various capabilities of state-of-the-art techniques from deep learning more specifically CNNs. The use of CNNs could elevate generalization performance to assess the data of unknown origin accurately. An academic and industrial effort should be made in order to standardize the quality assessment for smartphone-based biometric recognition.

## 5.2 Presentation Attack Detection

The thesis presents PAD schemes for 2D print-photo, display and video replay attacks. Future research should propose a PAD framework for other evolving presentation attacks such as a 3D mask, wrap attacks, in case of fingerphoto PAD, 3D printed fingers. Future studies could investigate the association between depth information from bona fide and artefact presentations to effectively detect new generation presentation attacks. The results obtained in the presented research work warrant further investigation via the confluence of various image processing and machine learning techniques to improve the results. Future research is needed to generalize the proposed methods for a sample of unknown origin. This thesis provides a good starting point for utilizing the intrinsic characteristics of sensors, and further research can be carried out to explore such approaches to formulate simple and robust PAD scheme. Regardless, future research should consider the hardware capabilities of the smartphone device for supporting more realistic settings. We also believe that apart from looking for effective handcrafted feature extraction methods, it would be promising to use CNNs to learn the characteristics of bona fide and artefact samples naturally.

## 5.3 Multi-modal Fusion

In this thesis, the biometric fusion of face and fingerprint modalities is presented. The thesis uses Subjective Logic to perform the effective biometric fusion by correctly handling the uncertainties present in the mated and non-mated comparison scores. This assumption might be addressed in future studies by modelling the uncertainties with different mathematical assumptions and approaches. Future studies could investigate the association between the comparison score and the belief/disbelief mass, which is a key component in subjective logic fusion. For a robust multi-modal biometric system we recommend that the future studies should investigate combining the individual PAD decision/score with comparison score to formulate a multi-modal decision. This thesis has provided a good starting point for SL based biometric fusion, and further work could be dedicated to exploring various fusion operations and similar frameworks to improve the state-of-the-art. Precise modelling of various masses involved in SL fusion, i.e., uncertainty, belief, base probability mass could be interesting topics for future work.

# Part II

# Published Research Articles

# Chapter 6

# Article 1: An Empirical Evaluation Of Deep Architectures On Generalization Of Smartphone-Based Face Image Quality Assessment

Pankaj Wasnik, Raghavendra Ramachandra, Kiran Raja and Christoph Busch. *"An Empirical Evaluation Of Deep Architectures On Generalization Of Smartphone -Based Face Image Quality Assessment."* In proceedings of $9^{th}$ IEEE International Conference On Biometrics: Theory, Applications, And Systems (BTAS 2018), IEEE, 2018.

## 6.1 Abstract

Often biometric authentication relies on the quality of enrolment and probe sample and it is therefore essential to estimate the image quality before a sample is submitted to the enrolment or verification process. The challenges encountered in estimating the quality is due to generalizability over unknown data samples of different origin. To this extent, we try to evaluate various deep learning networks which in theory, show high-performance in generalization. Due to the massive adoption of biometrics in consumer solu-

tions like smartphones, we have chosen Smartphone based Face Recognition Systems (FRS) to carry out our study. The main factors which impact the operating performance of the FRS are illumination, pose, occlusions and facial expressions. Therefore, it is essential to understand and estimate the quality of a facial image accurately. In this paper we present a robust and accurate quality estimating framework using deep neural networks (DNN). This work leverages the benefits of deep learning by transferring the pre-learned features from already trained DNNs such as AlexNet and Inception to estimate the facial image quality. Furthermore, we present the evaluation results for more than 10 techniques and 5 face image databases to analyze the performance generalization, and our results favor the pre-trained DNN models over the hand-crafted methods.

## 6.2 Introduction

In recent days, face recognition is widely used in smartphones for device security and payment services. Many smartphone manufacturers even provide such biometrics as an inbuilt feature. Smartphone-based face recognition is more challenging than conventional face recognition due to the unconstrained sample capturing. This causes various artifacts unintentionally to the captured sample. The primary artifacts observed in such biometric samples are illumination, pose and expressions [191]. There are well-defined international standards, i.e., ISO/IEC TR 29794-1 for conventional face recognition systems to assess the objective quality of face images and this report also provides the information about face quality assessment algorithms (FQAA) [1]. However, to date, there is no standardized method to assess the quality of face images for smartphone biometrics as well as performance generalization of such FQAAs except for a few recent works [191, 201, 135].

Deep learning in recent days, in particular, Convolution Neural Networks (CNN) are widely used for the tasks of face recognition, especially for achieving high accuracy [129, 160]. CNNs take images as their input and then these images are processed through multiple layers to extract fine features for face classification [64, 84].Due to the requirement of a high number of images to train the deep architectures, an alternative of transfer learning is employed as a viable solution where the learned models are used with suitable adaptations [110]. Further, transfer learning allows the domains, tasks, and distributions used in training to be different from data for testing purpose [127].

Transfer learning can be achieved by mainly two distinct ways *1) Fine-tuning of the network weights using the new database by retraining the last*

*couple of layers of a pre-trained network via back-propagation and 2) Directly utilizing weights from trained network to perform the testing on new database.* The first approach is suitable for large databases with similar data while the second approach is suitable for the small datasets with fewer classes. Based on the properties of the database and similarities between tasks one has to choose lower layer weights as features or higher layer weights [100].

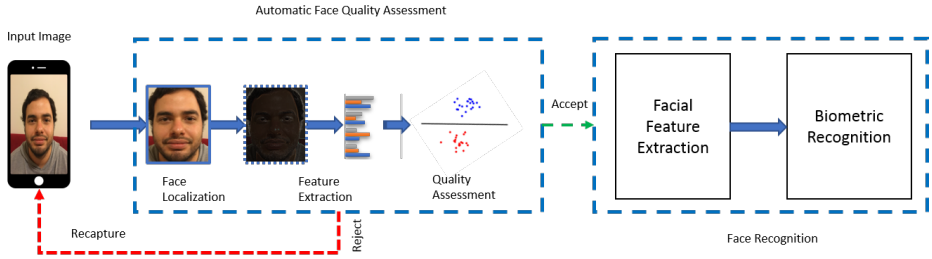



Figure 6.1: Block diagram of a typical face recognition system with automatic quality assessment

This paper presents the comprehensive performance evaluation of five well known CNNs which are AlexNet [94], Vgg16, Vgg19 [164], Inception [176] and Xception [24], three state-of-the-art mobile networks [157, 66, 203], and two state-of-the-art face quality CNNs [201, 135]. We also present results for the classical face quality assessment algorithms (FQAA) [191, 155, 114] and also one commercial off-the-shelf system (COTS) [2]. In particular, to deep learning approaches, we transfer the knowledge acquired from higher layers of a network, which are mainly trained on Imagenet database [36]. We select these networks as they have been very successful in the task of object classification and could be used in facial feature extraction. Also, these models have been the basis for some other face detection and recognition architectures [45, 160]. In order to achieve the performance generalization, we constructed a heterogeneous training database consisting of challenging facial images using the AR database [109], Extended Yale database [101], FRGC database [133], CAS-PEAL database [52] and NCKU face database [44]. Further, we have formulated our problem of quality estimation of facial images as a two-class transfer learning task with deep CNNs.

The rest of the paper is divided as follows: Section 6.3 details the methodology and evaluation criteria for face quality algorithm assessment. Section 6.4 describes the databases used, and Section 6.5 discusses experiments and

results followed by Section 6.6 which gives the concluding remarks.

## 6.3 Methodology

This section describes an overview of the smartphone-based face verification system with inbuilt automatic face quality assessment (FQA) framework. Figure 6.1 shows the block diagram of the typical face recognition flow which is also applicable to this paper. We employed the FQA using mentioned CNNs and other FQAAs. For the task of FQA, the domain $D$ of previously trained task of object recognition, is used to learn the feature vector $X = \{x_1, \ldots, x_n\} \in X$ where $X$ is a feature space which identifies the difference between low quality (bad) and high quality (good) samples by probability distribution $P(X)$. Hence, the face quality assessment can be achieved by the domain $D = \{X, P(X)\}$ and task $T = \{Y, q(.)\}$ where $Y$ is a label space and $q(.)$ is the objective function learned from pair $\{x_i, y_i\}$ where $x_i \in X$ and $y_i \in Y$. Finally, the objective function $q(.)$ can be used to predict the corresponding probability $P(y|x_{in})$ for the given input image. In this paper, we have used the output of the softmax function to obtain the final criteria.

For the practical realization of the system, we developed an iOS application to capture and test the data. The captured image is first processed in order to localize and crop face region. The cropped face is passed as an input to the forward pass of the CNN and quality $q(x)$ is assessed on the input image by applying threshold $t$. The decision to accept the image is based on the computed quality $q(x)$ is given by:

$$Result = \begin{cases} Accept & q(x) > t \\ Reject & q(x) < t \end{cases} \tag{6.1}$$

The effectiveness of the employed FQA is evaluated based on the False Non-Match Rate (FNMR) and percentage of the sample rejection. In this paper we adopt the assertion from [58] that the quality of a sample should be predictive of the recognition performance of the biometric system. Therefore, a good FRS should return a very high genuine comparison score for samples with high quality and vice-versa. In order to benchmark various FQAAs, in this paper we consider the Error versus Reject Curve (ERC) as our main evaluation criteria.

### 6.3.1 Error versus sample rejection

The predictive performance of a quality algorithm can be evaluated in terms of the rate of change of FNMR with respect to the rejection of a biometric

sample due to the low-quality [58]. In [124] authors have adopted equations for the one-dimensional case which is also used in this paper. The minimum of the qualities ($q_i$) in a pair of two samples drives the sample rejection. Hence, we can define a combination function H, as the *min()* function:

$$q_i = H(q_i^{(1)}, q_i^{(2)}) = min(q_i^{(1)}, q_i^{(2)}) \tag{6.2}$$

Therefore, we can derive the set $R(u)$ containing the pairwise minima $< u$ as

$$R(u) = \left\{ j : H(q_i^{(1)}, q_i^{(2)}) < u \right\} \tag{6.3}$$

Consider, $f$ as the FNMR of interest, where $t$ is the corresponding threshold. Exclude the pairs of face samples iteratively which are related specific comparison scores starting with the lowest of the pairwise score minimums up to the threshold $t$. Further, the threshold $t$ is calculated using the empirical cumulative distribution function of the comparison scores given as

$$t = M^{-1}(1 - f) \tag{6.4}$$

where $M$ is the Empirical Cumulative Distribution (ECD) functions computed using genuine comparison scores. Thus, the FNMR of interest can be calculated as:

$$FNMR(t, u) = \frac{\{|\ s_{jj} : s_{jj} \leq t, j \notin R(u)\ |\}}{\{|\ s_{jj} : s_{jj} \leq \infty\ |\}} \tag{6.5}$$

Practically, in each iteration one sample (for example, the sample with lowest quality) is rejected and the corresponding genuine comparison score is excluded, and the FNMR is calculated as the proportion of non-excluded scores below the threshold to total number of remaining samples. In order to quantify the rate of change of FNMR w.r.t % sample rejection we have used two metrics i.e., area under curve and partial area under curve as described in [124] as

$$\eta_{auc}^{erc} = \int_0^1 ERC - \text{area under therotical best} \tag{6.6}$$

where the integral term gives the full area under curve for the input ERC. We then use only 20% of full curve to calculate the second metric as

$$\eta_{pauc20}^{erc} = \int_0^{0.2} ERC - \text{area under therotical best} \tag{6.7}$$

### 6.3.2 Training of CNNs

In case of CNNs, we first removed the final three layers of each of these networks and replaced them with a fully connected (FC) layer of size 1024, 512 and number of classes which is two here. These FCs are then connected to ReLU followed by dropout layers. We used data augmentation techniques such as translation in x and y-direction, mirroring along horizontal and vertical axes and random cropping to have sufficient data for training in order to avoid any over-fitting. The CNNs are trained iteratively via back-propagation for a specified number of epochs, and one epoch is considered as a period where all the samples from the training datasets are used once. We further divide the training dataset into mini-batches of size 64 to achieve batch optimization which resulted in 412 iterations per epochs. In total, the training is carried out for 20 such epochs with total 8240 iterations.

Furthermore, the learning rate for the batch-wise optimization is controlled by Stochastic Gradient Descent with Momentum (SGDM) optimizer [55]. Parameters of SGDM are chosen as Momentum = 0.9000, Initial Learning Rate = $10^{-4}$ and 10% of total training images are used for validation. During validation, the overall loss for each validation batch is computed in the testing mode per epoch. This helps in the early stopping and a regularization to prevent over-fitting of the network. We used the validation frequency of 3 iterations. Finally, the model with the best validation loss is chosen for evaluating the testing datasets. In case of the state-of-the-art method proposed in [201], we used the trained network model provided by the authors for testing of evaluation databases, whereas in case of the method proposed in [135] we tried to recreate the network as defined in their paper and trained it on our training dataset.

## 6.4 Database

### 6.4.1 Training Database

The training dataset mainly contains two classes that correspond to good and bad quality. In order to achieve the performance generalization, we constructed a training database of nearly 33000 facial images. The images corresponding to the bad class consists of various problems such as illumination, pose, expression, occlusion, low resolution, and blur whereas, in the good class, we have the high quality, high-resolution images of frontal faces.

Furthermore, the experiments described in this work were conducted on a

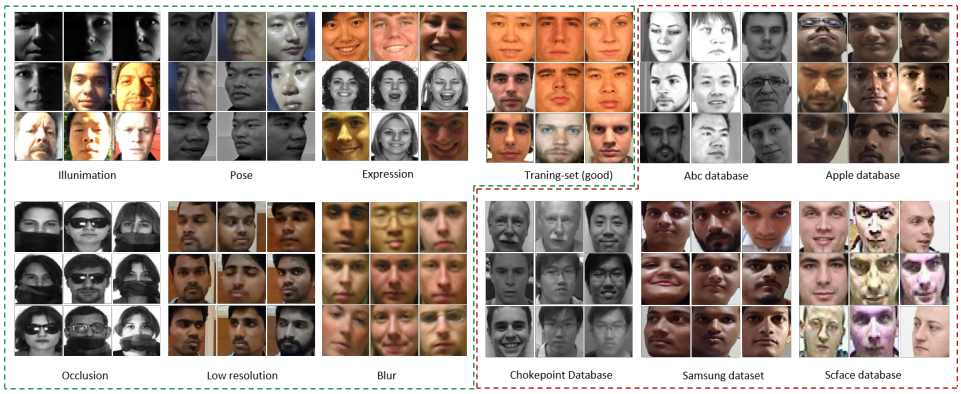



Figure 6.2: Training and evaluation image samples. Images inside green box show the samples from training set while images from red box show images from evaluation datasets.

new and recently constructed dataset. It consists of face video and image samples captured using an application developed for an iPhone 6S. Samples of frontal faces were collected from 201 subjects in 6 sessions over a period of 5 months. However, for this study, we have used the data from Session 1 and Session 3. The Session 1 data consists of 4k high-resolution samples of subjects and Session 3 data consists of different lighting conditions and locations. Bad quality training samples mainly consist of the AR database [109], Extended Yale database [101], FRGC database [133], CAS-PEAL database [52] and NCKU face database [44]. Similarly, the good samples are formed by Session 1 data of the newly constructed database and high-quality samples from the FRGC database. Table 6.1 presents the statistics of the training database.

### 6.4.2 Evaluation Database

This paper reports the results obtained from five different evaluation databases i.e., ABC [141], Chokepoint [199], SCFace [56], Apple iPhone 6 Plus, and Samsung Galaxy S7 Database [191]. In total, we have 16,990 testing images of 419 subjects from all five databases. All the images from evaluation databases were re-sized as per the size of the input layer of CNN under consideration. Further, during the testing phase, the dropouts are replaced by scaling to activate all the neurons from the last fully connected layer. The details of the evaluation databases are given in Table 6.2. One of the major reason we evaluated quality algorithms on the combination of smartphone and non-smartphone databases is to verify the generalizability. In general,

| **Database** | **Bad Images** | **Good Images** | **Data Characteristic** |
|---|---|---|---|
| AR[109] | 2778 | - | Occlusion, expressions, illumination |
| CAS-PEAL[52] | 1250 | - | Pose, illumination |
| Extended Yale[101] | 700 | - | Illumination |
| FRGC[133] | 1580 | 8939 | Blur, expression |
| NCKU face[44] | 4580 | - | Pose |
| Our database | 5605 | 7553 | Illumination, low resolution |

Table 6.1: Statistics of the training database.

| **Database** | **No of Subject** | **No of Images** | **Is smartphone based database?** |
|---|---|---|---|
| ABC [141] | 58 | 8950 | No |
| Apple [191] | 101 | 1010 | Yes |
| Chokepoint [199] | 29 | 2900 | No |
| Samsung [191] | 101 | 1010 | Yes |
| Scface [56] | 130 | 3120 | No |

Table 6.2: Statistics of the Evaluation databases

as we know that CNNs are highly capable of generalizing the feature extraction and further classification, we anticipate seeing higher performance for CNN based methods. To test this, we tried to evaluate these algorithms on facial images with many distortions such as pose, illumination, expression and low resolution. Furthermore, we assume that the smartphone-based data reflect different characteristics than non-smartphone based data due to the difference in sensor type and camera pipeline. Therefore, to achieve generalization of quality assessment we have trained the CNNs with all sorts of distortions mentioned earlier. Hence, we can assume that the CNNs will be able to classify the features more precisely than others.

## 6.5 Results and Discussion

The ERC is plotted to visualize the system's ability to correctly identify good and bad samples as the percentage of sample rejection changes. The ERC visualizes how the FNMR of the system is affected by changing the number of samples rejected. Ideally, when a sample is removed due to its

quality value, it is expected to decrease the FNMR of the system. Therefore, if an algorithm correctly assesses the quality of the sample and it is removed due to poor quality score, we can expect decrements in the FNMR of the system since the ideal quality features can identify the exact samples which were responsible for the FNMR of the system.

Similarly, when the samples which do not contribute to the FNMR of the system are removed due to their quality, it is expected to see an increment in the FNMR of the system. This behavior of the system can be reflected correspondingly in the ERCs. Therefore, the ideal ERC curve is as close to the origin as possible, and area under the curve is as less as possible.

### 6.5.1 ERC calculations

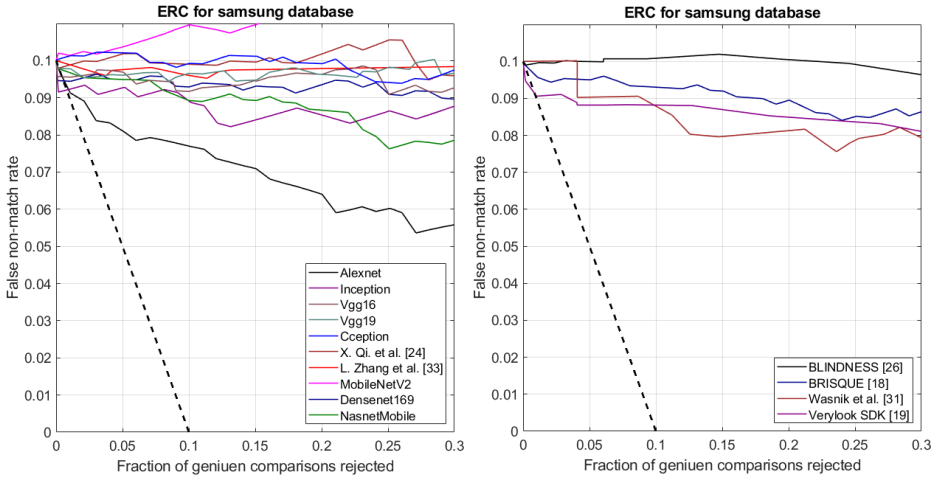
Following are the steps involved in the calculation of the ERC for any given quality assessment algorithm:

1. First, obtain the quality scores for all samples using given FQAA.
2. Sort the quality scores and along with their corresponding input image.
3. Generate the genuine comparison scores considering the first image as enrolment image and all other as probes images.
4. Obtain false non-match rate $f$ of the system using Eq. 6.5, for the % rejection $q_r$.
5. Repeat Step 4 until we reject all samples.

Finally, the ERC is obtained by plotting all the values of false non-match rates against the % of sample rejection. Our experiments resulted in 16,571 genuine scores from 419 testing subjects. We have used the Verilook SDK 5.4 [2] to obtain the comparison scores for a COTS as a baseline. We used $f = 0.1$ to simulate an operational case at the FNMR = 10%.

### 6.5.2 Evaluation of CNN based Methods

For each CNN and dataset we computed the ERC and reported the results in terms of $\eta^{erc}_{auc}$ and $\eta^{erc}_{pauc20}$ as defined in Eq. 6.6 and 6.7. For the sake of simplicity and limited space, we presented ERCs corresponding to Samsung S7 database only (See Figure 6.3a). All the CNN considered providing the wider range of quality values. However, the ERC plot shows that AlexNet outperforms the other approaches. In case of Alexnet, we can see that FNMR of the system is decreasing rapidly as comparisons containing




a ERC for CNN based methods

b ERC for Non-CNN based methods

Figure 6.3: ERC for Samsung database with targeted FNMR of f = 0.1. In figure, the dotted line shows theoretical best.

samples associated with the low-quality level are removed. Except Alexnet, the Inception is showing better results than other networks. Figure 6.3a depicts that MobileNetV2 has the lowest performance in the case of the mobile database; however it also has the smallest file size. Further, in the case of AlexNet, the decrease in FNMR overlaps with the theoretical best around 1.5-2% sample rejection, whereas it drops slightly as we further go on rejecting the samples. At 20% sample rejection the FNMR is approximately decreased by 0.035. Table 6.3, presents the details of $\eta^{erc}_{auc}$ and $\eta^{erc}_{pauc20}$. The AlexNet shows lower values for Apple, Chokepoint, and Samsung databases followed by the Inception which shows better performance for ABC and SCFace Databases. The AlexNet's performance for smartphone databases, i.e., Apple and Samsung show approximately similar behavior with average values of 0.055 and 0.011 for $\eta^{erc}_{auc}$ and $\eta^{erc}_{pauc20}$ respectively. Therefore, from AlexNet's performance, we can say that deep learning can achieve the performance generalization for quality assessment and such algorithm could be robust against the many of the problems such as illumination, pose, expression, occlusion, and low resolution.

### 6.5.3 Evaluation of Non-CNN based Methods

Similarly, for Non-CNN based methods the ERC are given in Figure 6.3b). By carefully analyzing the figure, it is evident that the method proposed in [191] is showing better performance overall. However, until 4% of the

| Algorithm | ABC | | Apple | | Chokepoint | | Samsung | | Scface | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\eta^{erc}_{auc}$ | $\eta^{erc}_{pauc20}$ | $\eta^{erc}_{auc}$ | $\eta^{erc}_{pauc20}$ | $\eta^{erc}_{auc}$ | $\eta^{erc}_{pauc20}$ | $\eta^{erc}_{auc}$ | $\eta^{erc}_{pauc20}$ | $\eta^{erc}_{auc}$ | $\eta^{erc}_{pauc20}$ |
| **Alexnet** | 0.064 | 0.013 | **0.057** | **0.011** | **0.027** | **0.008** | **0.054** | **0.010** | 0.048 | 0.012 |
| **Inception** | **0.035** | **0.009** | 0.084 | 0.012 | 0.045 | 0.012 | 0.072 | 0.012 | **0.027** | **0.008** |
| **Vgg16** | 0.089 | 0.013 | 0.068 | 0.012 | 0.072 | 0.011 | 0.077 | 0.013 | 0.096 | 0.013 |
| **Vgg19** | 0.072 | 0.013 | 0.074 | 0.014 | 0.053 | 0.013 | 0.074 | 0.014 | 0.095 | 0.014 |
| **Xception** | 0.064 | 0.011 | 0.070 | 0.012 | 0.071 | 0.012 | 0.093 | 0.014 | 0.058 | 0.012 |
| **Qi et al.**[135] | 0.062 | 0.012 | 0.089 | 0.014 | 0.070 | 0.012 | 0.094 | 0.014 | 0.084 | 0.013 |
| **Zhang et al.** [201] | 0.083 | 0.013 | 0.083 | 0.014 | 0.088 | 0.014 | 0.091 | 0.012 | 0.083 | 0.012 |
| **MobileNetV2** | 0.060 | 0.012 | 0.104 | 0.015 | 0.071 | 0.012 | 0.108 | 0.015 | 0.053 | 0.011 |
| **Densenet169** | 0.080 | 0.014 | 0.078 | 0.013 | 0.065 | 0.013 | 0.083 | 0.013 | 0.099 | 0.014 |
| **NasnetMobile** | 0.068 | 0.013 | 0.091 | 0.014 | 0.060 | 0.011 | 0.080 | 0.013 | 0.079 | 0.012 |
| **BLINDNESS** [155] | 0.067 | 0.013 | 0.094 | 0.014 | 0.069 | 0.012 | 0.096 | 0.015 | 0.083 | 0.014 |
| **BRISQUE** [114] | 0.077 | 0.012 | 0.067 | 0.013 | 0.094 | 0.012 | 0.099 | 0.014 | 0.087 | 0.013 |
| **Wasnik et al.** [191] | 0.087 | 0.015 | 0.068 | 0.012 | 0.082 | 0.014 | 0.064 | 0.013 | 0.071 | 0.014 |
| **Verylook SDK** [2] | 0.065 | 0.013 | 0.087 | 0.014 | 0.101 | 0.014 | 0.075 | 0.013 | 0.075 | 0.013 |

Table 6.3: Summary of AUC and PAUC for ERC plots of quality algorithms on ABC [141], Chokepoint [199], SCFace [56], Apple iPhone 6 Plus, and Samsung Galaxy S7 Database [191], computed ERC using f = 0.1.

sample rejection, Verilook SDK is showing the best performance. When there is not a sufficient granularity of quality levels, one has to choose to reject the samples corresponding to the entire quality level which results into a step function when plotting the ERC; this phenomenon can be seen in case of [191] where the FNMR of the system remains constant until 4% of the sample rejection, and further it decreases to 8% until 20% sample rejection. Among all of the four hand-crafted methods, the Blindness algorithm [155] is performing worst. From Table 6.3 it is evident that none of the Non-CNN based quality algorithms shows better performance over CNN based methods. However, the method proposed in [191] overall shows better performance over other methods. In case of the smartphone-based database the highest achieved performance is of 0.064 and 0.012 $\eta^{erc}_{auc}$ and $\eta^{erc}_{pauc20}$ respectively.

## 6.6 Conclusion

We successfully present the benchmarking results for various deep architectures for the task of facial quality assessment. We have noticed that our results favor CNN models over the hand-crafted methods indicating that it is possible to learn various types of challenges from facial images to assess precise quality. However, we anticipate that the fine-tuning of CNNs in the context of removing or adding layers, learning more number of layers instead only fully connected layers and hyperparameters optimization could achieve higher performance regarding generalizability when tested with an unknown database.

The experiments indicated that AlexNet performs well for both smartphone and general database. For the Apple and Samsung databases, it achieved approximately the same performance with average values of 0.055 and 0.011 for $\eta^{erc}_{auc}$ and $\eta^{erc}_{pauc20}$ respectively. Moreover, the performance of mobile networks is lower than the AlexNet and Inception, but one can choose to use these mobile networks for faster processing and smaller network file size.

An important point to address in the future research is to improve the results towards the theoretical best. One can apply various fusion methods, or develop a new network architecture, training strategy to achieve this. The assessment of quality algorithms based on ERC heavily depends on the comparator used in obtaining the genuine comparisons, and it would be interesting to see the behavior of CNNs for different comparators.

## Acknowledgement

This work was carried out under the funding from the Research Council of Norway (Grant No. IKTPLUSS 248030/O70).

# Chapter 7

# Article 2: Presentation Attack Detection In Face Biometric Systems Using Raw Sensor Data From Smartphones

Pankaj Wasnik*, Kiran B. Raja*, Ramachandra Raghavendra*, And Christoph Busch. *"Presentation Attack Detection In Face Biometric Systems Using Raw Sensor Data From Smartphones."* In the $12^{th}$ International Conference On Signal-Image Technology & Internet-Based Systems (SITIS 2016), Pp. 104-111. IEEE, 2016.

## 7.1 Abstract

Applicability of the face recognition for smartphone-based authentication applications is increasing for different domains such as banking and e-commerce. The unsupervised data capture of face characteristics in biometric applications on smartphones presents the vulnerability to attack the systems using artefact samples. The threat of presentation attacks *(a.k.a spoofing attacks)* need to be handled to enhance the security of the biometric system. In this work, we present a new approach of using the *raw sensor data*. We first obtain the residual image corresponding to noise by

*All the authors have equally contributed to this article.

subtracting the median filtered version of raw data and then computing simple energy value to detect the artefact based presentations. The presented approach uses simple threshold and thereby overcomes the need for learning complex classifiers which are challenging to work on unseen attacks. The proposed method is evaluated using a newly collected database of 390 live presentation attempts of face characteristics and 1530 attack presentations consisting of electronic screen attacks and printed attacks on the iPhone 6S smartphone. Significantly lower average classification error ($< 3\%$) achieved demonstrates the applicability of proposed approach for detecting the presentation attacks.

## 7.2 Introduction

Face based secure access systems are increasingly becoming popular for many applications such as unlocking the smartphone, e-commerce and e-banking. The ease of unconstrained imaging using a simple color camera has led face recognition to be employed in secure systems, operating in various environments. Recent interest in this direction has led to the use of smartphones for capturing the face characteristics to authenticate the subjects. The key factor to be noted in smartphone based biometric systems is the unsupervised data capture. The use of smartphone-based biometric system for authentication is highly intended to provide the convenience to the user for authentication from any location in an unconstrained manner and thereby is allowed to capture the biometric data in an unsupervised manner.

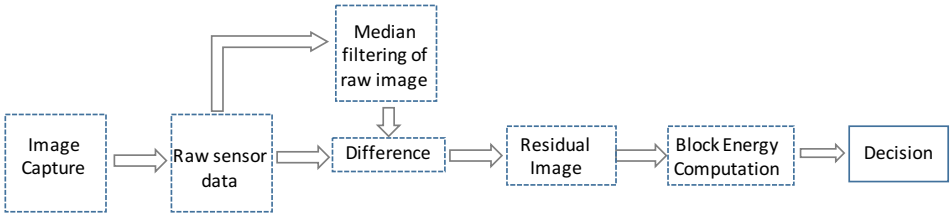



Figure 7.1: Schematic of proposed approach for presentation attack detection. Figure 7.2 shows the complete pipeline of the proposed method.

The freedom of unsupervised data capture, especially in the smartphone based face recognition can be misused by the unauthorized users. The abundant availability of the face pictures on various social media sites can be used for gaining unauthorized access in such unsupervised biometric systems operating on the smartphones[23]. Any attempt to gain the secure access

by presenting the artefact of the genuine subject is classified as presentation attack or spoofing attack. Primitively, an unauthorized user can display the facial image on the electronic screen to gain access to the face-based biometric system on the smartphone. Alternatively, the image can be printed on a paper and presented back to the smartphone data capture system. The failure to detect such attacks on face-based biometric system defeats the purpose of security in the context of face based authentication. Such attacks can be addressed by presentation attack detection (PAD) algorithms incorporated in the biometric system.

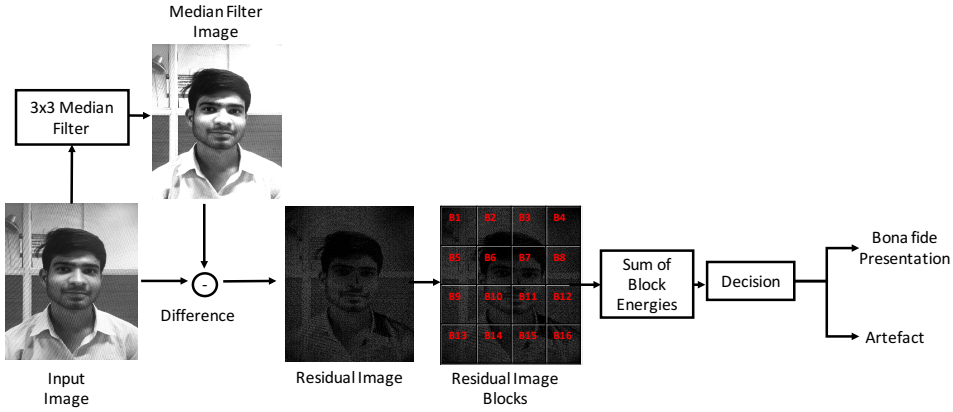



Figure 7.2: Proposed scheme for presentation attack detection.

Many approaches have been proposed to counter such attacks on face-based biometric systems which leverage on the textural characteristics of the live attempt and the presentation attack [142, 89, 136, 137]. Variants of the Local Binary Pattern (LBP) was explored from the images using Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA) was used to differentiate the attacks from live presentation [43, 62]. In a similar way, LBP was used in temporal domain to detect replay attacks using SVM and LDA [32, 62]. Another key factor of difference in the quality of live presentation and attack presentation was fully utilized to detect the presentation attacks by assessing the image quality on both full reference and no-reference quality metrics using LDA and Quadratic Discriminant Analysis (QDA) [50, 49]. Recently, the distortions such as Moiré pattern in the live image and attack images were identified to classify the attacks [195, 130]. Further, applicability of the deep features for detecting the presentation attacks was demonstrated in a recent work [112]. It has to be noted that, all the current state of the art works have focused on the final images obtained from the camera's imaging pipeline to classify the attacks.

In a very different paradigm, earlier works have demonstrated the use of raw sensor data to obtain the unique sensor noise pattern to establish the authenticity of the imaging camera [105, 102]. Another work recently used the raw sensor noise to identify the device and the user in the context of visible spectrum iris recognition [51]. Motivated by the use of raw data for various applications, in this work, we look at the specific characteristics demonstrated in the raw data at the sensor level to determine the presentation attacks against the live (a.k.a bona-fide) attempts. The characteristics are demonstrated at the sensor level is used with simple image analysis and a threshold. The proposed approach thereby removes the necessity for learning classifiers. We evaluate the proposed approach of leveraging on the raw sensor data on a newly collected face image database of 390 live face images using iPhone 6S and 1530 attack images using different electronic screens and printed photos. The key contributions of this work can be summarized as:

1. Presents a new approach to detect presentation attacks on smartphone based face recognition systems by analyzing the raw sensor characteristics of the smartphone camera.
2. Presents a classifier free approach by analyzing the image characteristics at the sensor level data using a simple block-wise energy computation.
3. Extensive set of experiments are carried out on a new and relatively large-scale database collected using the 390 live images and 1530 attack images from iPhone 6S.

In the rest of the paper, Section 7.3 presents the proposed scheme for PAD and Section 7.4 provides the details of newly constructed PAD database. In the Section 7.5, experimental details and protocols are discussed along with the obtained results. Finally, in Section 7.6, the key findings, and remarks are presented along with a possible future direction for this work.

## 7.3 Proposed Scheme for face PAD

Figure 7.1 presents the schematic illustration of the proposed approach for presentation attack detection. As depicted in the Figure 7.1, once the image is captured, we process the raw data of the corresponding image from the smartphone camera sensor. For each of the color channel data such as red, green and blue channel, the image data is extracted from the color filter array (CFA). The raw data is processed for independent color channels such

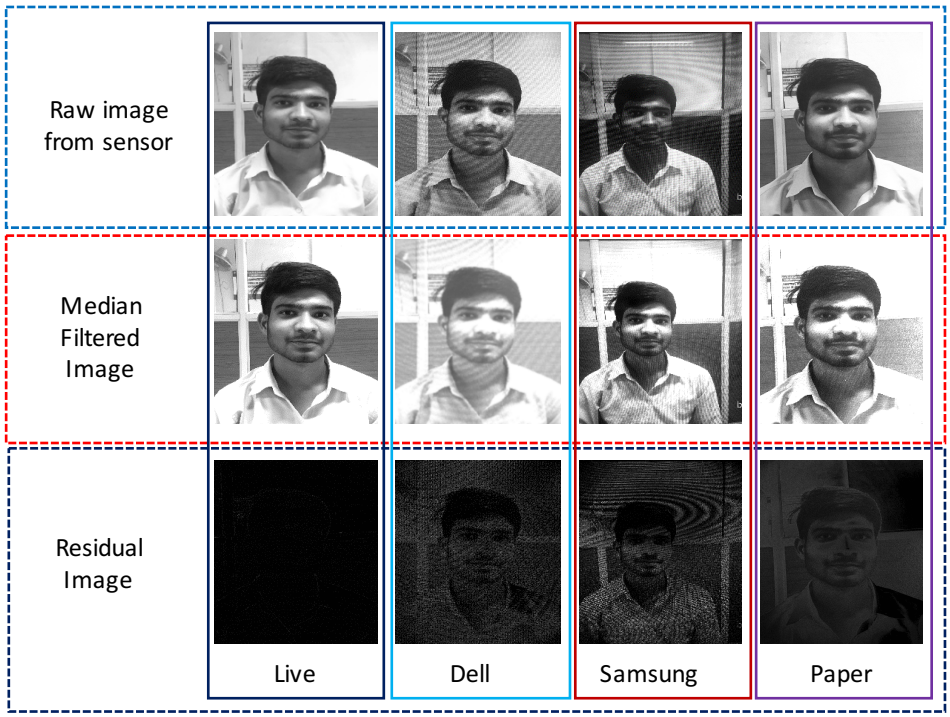



Figure 7.3: Example of residual images obtained using the proposed approach for bona-fide and artefact presentations.

as red, blue and green even before they are processed by imaging pipeline. It is our intuition that the data on the sensor has more noise due to change in reflection properties in live and artefact. Thus, we look for such noise in each of the color channel using a median filter. The detailed explanation of all the steps is given in the section below.

### 7.3.1 Steps in Image Capture to Storage

In general, the process of image capture in smartphone involves following steps:

1. The light from the scene enters through the lens (or set-of-lens) in the camera.

2. The output from a set of lens is further processed by the anti-aliasing filter and later reaches color filter array (CFA) which is capable of storing red, green and blue spectrum information.

3. The information from the sensor (color filter array) is de-mosaiced and post-processed to form the final picture. The number of sequential operations such as color correction, white balancing, gamma correction, enhancing, compression are carried out before storing the picture.

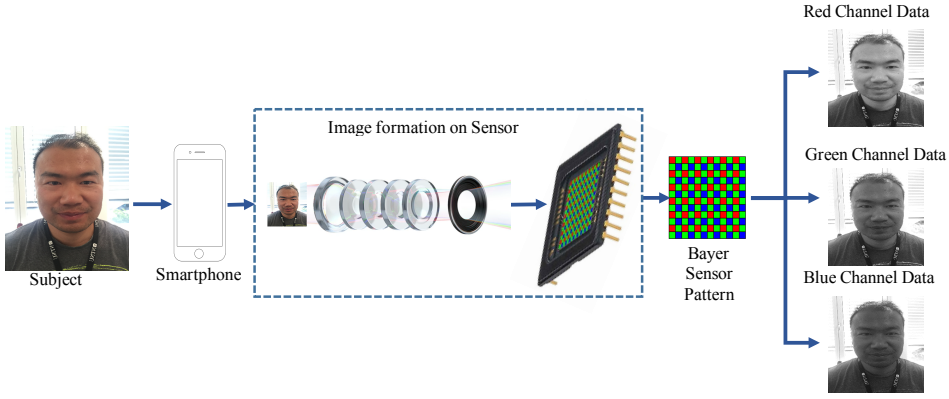



Figure 7.4: Illustration of image data from sensor

In each of these steps such as re-sampling, CFA interpolation or compression, significant loss of the details may happen. Preventing the loss of this information becomes crucial to detect the noise which can be used to separate the live images from the artefact images. Thus, in this work, we propose to use the data directly from the sensor for each different color channel. The information from each color channel is further processed as explained in the subsequent section. Figure 7.4 depicts the general pipeline in the image capture process. When the subject is imaged using the smartphone, the data is captured onto the sensor. The sensor is usually in the form of color filter array which has dedicated cells for red, blue and green spectrum data. In this work, we obtain the image data directly from the CFA and separate them to red, green and blue channel data.

| Data | Red Channel | | Blue Channel | | Green Channel | | RGB Channel | |
|---|---|---|---|---|---|---|---|---|
| | Development | Testing | Development | Testing | Development | Testing | Development | Testing |
| Live | 150 | 240 | 0 | 510 | 0 | 510 | 0 | 510 |
| Attack - Print Attack | 200 | 310 | 0 | 510 | 0 | 510 | 0 | 510 |
| Attack - Dell Monitor | 200 | 310 | 0 | 510 | 0 | 510 | 0 | 510 |
| Attack - Samsung Pad | 200 | 310 | 0 | 510 | 0 | 510 | 0 | 510 |

Table 7.1: Division of the database for experiments.

### 7.3.2 PAD Scheme

Given an image obtained from the sensor, $I_{sr}, I_{sg}$ and $I_{sb}$ for red, green and blue channel respectively, we first filter the images using the median filter. For the sake of simplicity, we henceforth refer all the channel data in sensor using $I_s$ which can represent any channel information. Typically, the median filter removes the repeated noise in a specified window while preserving the useful details of the image. Thus, to remove the external noise introduced due to different factors on the image, we filter the image $I_s$ using a window of $3 \times 3$ size which results in the median-filtered image represented by $I_m$. Further, to obtain the noise data not corresponding to image, we obtain the difference between the image $I_s$ and $I_m$ which we refer as *residual image* and is represented by $I_r$. The residual image characteristics may differ in a different region of the image. In order to obtain the information in a robust manner, a block based approach can be used. The residual image is thus further divided into $k$ blocks of size $n \times n$ resulting in a set of blocks $\{B_1, B_2 \dots B_k\}$. For each block $B$, we compute energy as given by Equation 7.1:

$$E_i = \frac{\sum_{x=1}^{n} \sum_{y=1}^{n} (B_i(x,y))^2}{n \times n} \tag{7.1}$$

Where $i$ represents the $i^{th}$ block. Further, the energies computed for all the blocks are summed to obtain one single value $E_s$ as given by Equation 7.2:

$$E_s = \sum_{i=1}^{k} B_i \tag{7.2}$$

The determined energy value $E_s$ is compared against a threshold $T$ to obtain the decision $D_s$ for the sensor data.

$$D_s = \begin{cases} 1, & \text{if } E_s \leq T \\ 0, & \text{otherwise} \end{cases} \tag{7.3}$$

Where $D_s = 1$ represents the live presentation and the $D_s = 0$ indicates the artefact presentation. Further, to make the decision robust and fully leverage the different channel information, we employ the majority voting for each channel decision. If the decision obtained for red, green and blue channel are represented by $D_r$, $D_g$ and $D_b$ respectively, we take the majority voting as given by:

$$D = \begin{cases} 1, & \text{if } majority\{D_r, D_g, D_b\} = 1 \\ 0, & \text{otherwise} \end{cases} \tag{7.4}$$

Figure 7.3 presents the examples of residual image obtained from sensor data for red color channel for various types of the presentations spanning from live (bona-fide) presentation to artefact presentation.

| Channel | Live | Paper | Dell | Samsung |
|---|---|---|---|---|
| RGB | 390 | 510 | 510 | 510 |
| Red | 390 | 510 | 510 | 510 |
| Green | 390 | 510 | 510 | 510 |
| Blue | 390 | 510 | 510 | 510 |

Table 7.2: Composition of newly created face artefact database in current work.

## 7.4 PAD Database

This section presents the description of the newly created presentation attack database. The database is collected using iPhone 6S using 102 subjects. The database consists of real(bona fide) attempts and presentation attack attempts. The presentation attack subset consists of data corresponding to screen attacks and printed photo attacks. The electronic attacks are carried out using two different electronic displays as below:

1. Dell UltraSharp 25-inch monitor with QHD display.
2. Samsung Galaxy Tab 7.0

A total of 510 attempts are recorded using electronic screen images displayed using Dell monitor and a similar set of 510 images are recorded using the display of Samsung tablet. Further, the attacks are also carried out using printed photos of the subjects resulting in a total of 510 photo attacks.

As the proposed approach is based on employing raw sensor data, the data corresponding to the red, green and blue channel are extracted. Along with the three independent channel data, we also extract combined red-green-blue data without employing any of the correction methods provided by the smartphone. Specifically, we employ "iPhone 6S Plus" to capture the live data and attack data. Thus, the database consists of 1560 ($390\ images \times (3\ channels\ +\ 1\ RGB)$) live images corresponding to red, green, blue and combined red-blue-green data. The total database also consists of 4080 ($2\ devices \times 510\ images \times (3\ channels\ +\ 1\ RGB)$) images for electronic screen attacks and 2040 ($510\ images \times (3\ channels\ +\ 1\ RGB)$) printed photo attacks. Table 7.2 presents the complete composition of the database.

## 7.5 Experiments and Results

This section provides the details of the experimental protocols employed and the obtained results in this work. The performance of the proposed presentation attack detection (PAD) (*a.k.a. spoof detection*) algorithm is given in accordance to ISO/IEC CD 30107-3 [68]. The metrics used in this work are: (1) Attack Presentation Classification Error Rate (APCER), which is defined as a proportion of attack presentation incorrectly classified as Bona Fide presentation (2) Bona Fide Presentation Classification Error Rate (BPCER) which is defined as proportion of Bona Fide presentation incorrectly classified as attack presentation [68]. The average performance derived from both individual performance metrics is provided as an indicative metric which is Average Classification Error Rate (ACER) defined as:

$$ACER = \frac{APCER + BPCER}{2} \tag{7.5}$$

A robust PAD scheme is expected to have lower APCER and BPCER metrics.

In order to evaluate the proposed approach, we divide the database into development and testing set. The partition to evaluation set is deliberately avoided as there is no training involved in this work which needs the evalu-

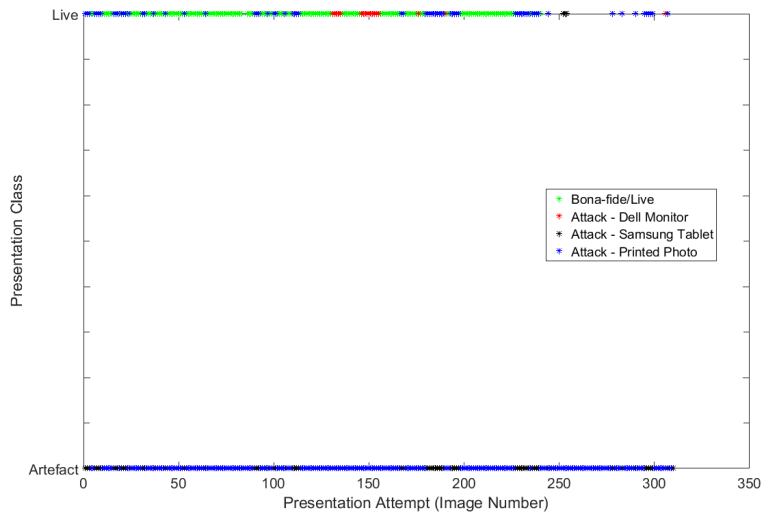



Figure 7.5: Classification of live and artefact presentation with the majority voting.

ation set. The development set is used to determine the decision threshold for classifying the image as live or attack presentation. The division of the database is presented in the Table 7.1.

### 7.5.1 Experiments on development set

In order to determine the values for threshold, we carry out the experiments on the development database. Based on the empirical trials carried out on the development database, we determine the threshold for different values of energy computed. Figure 7.6 presents the classification error rate obtained as a function of threshold for the energy value computed. It can be observed from the Figure 7.6 that $E_s \geq 200000$ reduces the BPCER significantly while retaining the APCER at a very low level. Thus, the threshold of $E_s \geq 200000$ is chosen to make a decision on the presentation using the Equation 7.3.

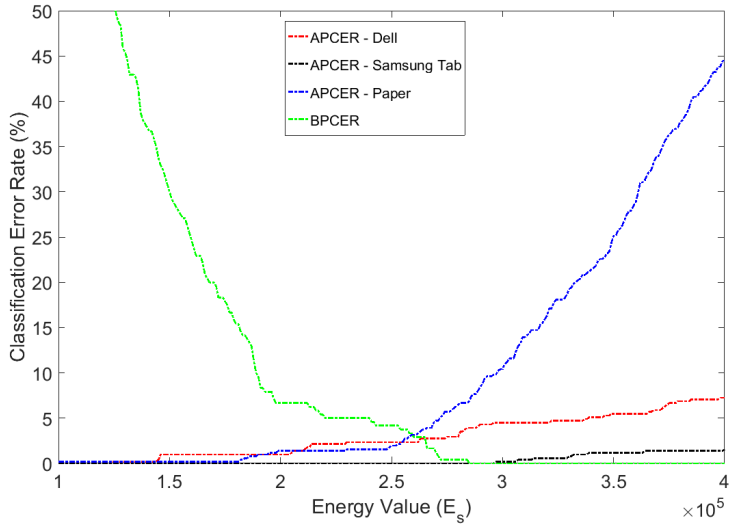



Figure 7.6: Illustration of energy threshold ($E_s$) versus the classification error rate. It can be noted from the image that the $E_s$ after 200000 provides lower classification error rate.

### 7.5.2 Experiments on testing set

The determined threshold of $E_s \geq 200000$ is used in increments on the testing set of images. The Table 7.4 presents the classification error obtained using the data corresponding to green channel. Further, Table 7.3 provides the obtained performance using the majority voting approach. From both

| Threshold | Paper | | | Dell | | | Samsung | | |
|---|---|---|---|---|---|---|---|---|---|
| | BPCER (%) | APCER (%) | ACER (%) | BPCER (%) | APCER (%) | ACER (%) | BPCER (%) | APCER (%) | ACER (%) |
| 200000 | 3.33 | 0.32 | 1.83 | 3.33 | 3.23 | 3.28 | 3.33 | 0.00 | 1.67 |
| 210000 | 3.33 | 0.32 | 1.83 | 3.33 | 3.23 | 3.28 | 3.33 | 0.00 | 1.67 |
| 220000 | 3.33 | 0.32 | 1.83 | 3.33 | 3.23 | 3.28 | 3.33 | 0.00 | 1.67 |
| 230000 | 2.67 | 0.65 | 1.66 | 2.67 | 4.19 | 3.43 | 2.67 | 0.00 | 1.33 |
| 240000 | 2.67 | 0.65 | 1.66 | 2.67 | 4.19 | 3.43 | 2.67 | 0.00 | 1.33 |
| 250000 | 2.00 | 1.29 | 1.65 | 2.00 | 5.48 | 3.74 | 2.00 | 0.00 | 1.00 |
| 260000 | 2.00 | 2.27 | 2.13 | 2.00 | 5.48 | 3.74 | 2.00 | 0.00 | 1.00 |
| 270000 | 2.00 | 3.24 | 2.62 | 2.00 | 5.48 | 3.74 | 2.00 | 0.00 | 1.00 |
| 280000 | 2.00 | 4.21 | 3.10 | 2.00 | 6.13 | 4.06 | 2.00 | 0.00 | 1.00 |
| 290000 | 1.33 | 8.41 | 4.87 | 1.33 | 6.77 | 4.05 | 1.33 | 0.00 | 0.67 |
| 300000 | 1.33 | 9.71 | 5.52 | 1.33 | 6.77 | 4.05 | 1.33 | 0.00 | 0.67 |

Table 7.3: Classification error rates with different threshold of computed energy values with majority voting.

the tables, it can be clearly observed for the low ACER values indicating the reliability of the proposed method. Further, Table 7.7 and Table 7.6 in the Appendix 7.7 present the performance obtained using blue and red channel where the proposed approach has provided very low classification errors.

### 7.5.3 Discussion

The obtained results from the experiments indicate the promising performance of the proposed approach in detecting the presentation attacks by analyzing the sensor level data. The advantage of the proposed method lies in eliminating the need for a specially learnt classifier. It can also be seen that the proposed method has a wide range of threshold with significantly lower classification error rates. An important observation from the set of experiments carried out in this work points to the increase of classification error rate for printed attacks as observed from the Figure 7.6. However, as it can be noted from the Table 7.3 that the chosen range provides very low ACER. In the future works, we intend to investigate more on characteristics of sensor data to separate the print attacks on the biometric systems.

## 7.6 Conclusions

Presentation attacks or what is generally known as spoofing attacks pose a threat to biometric systems. In this work, we have presented a new approach of using raw data directly from the sensor of the smartphone to detect such presentation attacks. The proposed approach employs the computation of the energy values from residual images to detect the amount of artefact noise. The approach has achieved significantly lower classification error in determining live presentations and attack presentations. The note-

| Threshold | Paper | | | Dell | | | Samsung | | |
|---|---|---|---|---|---|---|---|---|---|
| | BPCER (%) | APCER (%) | ACER (%) | BPCER (%) | APCER (%) | ACER (%) | BPCER (%) | APCER (%) | ACER (%) |
| 200000 | 7.50 | 0.32 | 3.91 | 7.50 | 3.23 | 5.36 | 7.50 | 0.00 | 3.75 |
| 210000 | 7.08 | 0.32 | 3.70 | 7.08 | 3.23 | 5.15 | 7.08 | 0.00 | 3.54 |
| 220000 | 6.67 | 0.32 | 3.50 | 6.67 | 3.23 | 4.95 | 6.67 | 0.00 | 3.33 |
| 230000 | 5.83 | 0.97 | 3.40 | 5.83 | 4.19 | 5.01 | 5.83 | 0.00 | 2.92 |
| 240000 | 5.00 | 0.97 | 2.99 | 5.00 | 4.19 | 4.60 | 5.00 | 0.00 | 2.50 |
| 250000 | 3.75 | 2.60 | 3.17 | 3.75 | 5.48 | 4.62 | 3.75 | 0.00 | 1.88 |
| 260000 | 2.08 | 3.57 | 2.83 | 2.08 | 5.48 | 3.78 | 2.08 | 0.00 | 1.04 |
| 270000 | 1.67 | 4.55 | 3.11 | 1.67 | 5.48 | 3.58 | 1.67 | 0.00 | 0.83 |
| 280000 | 1.25 | 6.49 | 3.87 | 1.25 | 6.13 | 3.69 | 1.25 | 0.00 | 0.63 |
| 290000 | 0.83 | 9.42 | 5.12 | 0.83 | 6.77 | 3.80 | 0.83 | 0.00 | 0.42 |
| 300000 | 0.83 | 12.34 | 6.59 | 0.83 | 6.77 | 3.80 | 0.83 | 0.00 | 0.42 |

Table 7.4: Classification error rates for green channel data with different threshold of computed energy values on green channel data alone.

worthy contribution of this work lies in removing the necessity of learning specific classifiers to detect attacks. Extensive experiments carried out using electronic screen attacks and printed photo attacks to demonstrate the applicability of the newly introduced approach. The promising nature of the approach can be extended in the future works to detect the attacks for any natural images using the sensor level raw data.

## 7.7 Appendices

| Threshold | Paper | | | Dell | | | Samsung | | |
|---|---|---|---|---|---|---|---|---|---|
| | BPCER (%) | APCER (%) | ACER (%) | BPCER (%) | APCER (%) | ACER (%) | BPCER (%) | APCER (%) | ACER (%) |
| 200000 | 7.50 | 0.32 | 3.91 | 7.50 | 3.23 | 5.36 | 7.50 | 0.00 | 3.75 |
| 210000 | 7.08 | 0.32 | 3.70 | 7.08 | 3.55 | 5.32 | 7.08 | 0.00 | 3.54 |
| 220000 | 6.67 | 0.32 | 3.50 | 6.67 | 4.84 | 5.75 | 6.67 | 0.00 | 3.33 |
| 230000 | 5.00 | 0.97 | 2.99 | 5.00 | 5.48 | 5.24 | 5.00 | 0.00 | 2.50 |
| 240000 | 4.58 | 0.97 | 2.78 | 4.58 | 5.48 | 5.03 | 4.58 | 0.00 | 2.29 |
| 250000 | 3.33 | 2.27 | 2.80 | 3.33 | 5.48 | 4.41 | 3.33 | 0.00 | 1.67 |
| 260000 | 1.67 | 3.24 | 2.45 | 1.67 | 5.48 | 3.58 | 1.67 | 0.00 | 0.83 |
| 270000 | 1.25 | 4.53 | 2.89 | 1.25 | 5.48 | 3.37 | 1.25 | 0.00 | 0.63 |
| 280000 | 1.25 | 6.80 | 4.02 | 1.25 | 6.45 | 3.85 | 1.25 | 0.00 | 0.63 |
| 290000 | 0.83 | 9.39 | 5.11 | 0.83 | 6.77 | 3.80 | 0.83 | 0.00 | 0.42 |
| 300000 | 0.83 | 11.33 | 6.08 | 0.83 | 6.77 | 3.80 | 0.83 | 0.00 | 0.42 |

Table 7.5: Classification error rates for RGB data with different threshold of computed energy values of RGB data.

| Threshold | Paper | | | Dell | | | Samsung | | |
|---|---|---|---|---|---|---|---|---|---|
| | BPCER (%) | APCER (%) | ACER (%) | BPCER (%) | APCER (%) | ACER (%) | BPCER (%) | APCER (%) | ACER (%) |
| 200000 | 7.50 | 0.32 | 3.91 | 7.50 | 4.84 | 6.17 | 7.50 | 0.00 | 3.75 |
| 210000 | 7.08 | 0.65 | 3.87 | 7.08 | 4.84 | 5.96 | 7.08 | 0.00 | 3.54 |
| 220000 | 5.83 | 0.97 | 3.40 | 5.83 | 5.16 | 5.50 | 5.83 | 0.00 | 2.92 |
| 230000 | 5.83 | 1.30 | 3.57 | 5.83 | 5.48 | 5.66 | 5.83 | 0.00 | 2.92 |
| 240000 | 5.00 | 2.60 | 3.80 | 5.00 | 5.48 | 5.24 | 5.00 | 0.00 | 2.50 |
| 250000 | 4.17 | 3.25 | 3.71 | 4.17 | 5.48 | 4.83 | 4.17 | 0.00 | 2.08 |
| 260000 | 2.50 | 4.87 | 3.69 | 2.50 | 5.48 | 3.99 | 2.50 | 0.00 | 1.25 |
| 270000 | 1.67 | 8.44 | 5.05 | 1.67 | 5.81 | 3.74 | 1.67 | 0.00 | 0.83 |
| 280000 | 1.67 | 10.06 | 5.87 | 1.67 | 6.77 | 4.22 | 1.67 | 0.00 | 0.83 |
| 290000 | 1.67 | 13.31 | 7.49 | 1.67 | 6.77 | 4.22 | 1.67 | 0.00 | 0.83 |
| 300000 | 1.25 | 17.21 | 9.23 | 1.25 | 6.77 | 4.01 | 1.25 | 0.00 | 0.63 |

Table 7.6: Classification error rates for red channel data with different threshold of computed energy values from red channel data.

| Threshold | Paper | | | Dell | | | Samsung | | |
|---|---|---|---|---|---|---|---|---|---|
| | BPCER (%) | APCER (%) | ACER (%) | BPCER (%) | APCER (%) | ACER (%) | BPCER (%) | APCER (%) | ACER (%) |
| 200000 | 6.67 | 0.32 | 3.50 | 6.67 | 1.61 | 4.14 | 6.67 | 0.00 | 3.33 |
| 210000 | 6.67 | 0.32 | 3.50 | 6.67 | 2.26 | 4.46 | 6.67 | 0.00 | 3.33 |
| 220000 | 5.00 | 0.32 | 2.66 | 5.00 | 3.55 | 4.27 | 5.00 | 0.00 | 2.50 |
| 230000 | 5.00 | 0.32 | 2.66 | 5.00 | 3.87 | 4.44 | 5.00 | 0.00 | 2.50 |
| 240000 | 5.00 | 0.32 | 2.66 | 5.00 | 3.87 | 4.44 | 5.00 | 0.00 | 2.50 |
| 250000 | 4.17 | 0.97 | 2.57 | 4.17 | 3.87 | 4.02 | 4.17 | 0.00 | 2.08 |
| 260000 | 2.92 | 1.62 | 2.27 | 2.92 | 3.87 | 3.39 | 2.92 | 0.00 | 1.46 |
| 270000 | 1.25 | 3.25 | 2.25 | 1.25 | 4.19 | 2.72 | 1.25 | 0.00 | 0.63 |
| 280000 | 0.42 | 4.55 | 2.48 | 0.42 | 4.52 | 2.47 | 0.42 | 0.00 | 0.21 |
| 290000 | 0.00 | 6.82 | 3.41 | 0.00 | 6.45 | 3.23 | 0.00 | 0.00 | 0.00 |
| 300000 | 0.00 | 8.44 | 4.22 | 0.00 | 6.77 | 3.39 | 0.00 | 0.32 | 0.16 |

Table 7.7: Classification error rates for blue channel data with different threshold of computed energy values from blue channel data.

## Acknowledgment

This work is carried out under the funding of the Research Council of Norway (Grant No. IKTPLUSS 248030/O70).

# Chapter 8

# Article 3: Robust Face Presentation Attack Detection On Smartphones: An Approach Based On Variable Focus

Kiran B. Raja*, Pankaj Wasnik*, Raghavendra Ramachandra* and Christoph Busch. *"Robust Face Presentation Attack Detection On Smartphones: An Approach Based On Variable Focus."* In the $3^{rd}$ IEEE International Joint Conference On Biometrics (IJCB 2017), Pp. 651-658. IEEE, 2017.

## 8.1 Abstract

Smartphone based facial biometric systems have been well used in many of the security applications starting from simple phone unlocking to secure banking applications. This work presents a new approach of exploring the intrinsic characteristics of the smartphone camera to capture a number of stack images in the depth-of-field. With the set of stack images obtained, we present a new *feature-free* and *classifier-free* approach to provide the presentation attack resistant face biometric system. With the entire system implemented on the smartphone, we demonstrate the applicability of the proposed scheme in obtaining a stack of images with varying focus to effect-

*All the authors have equally contributed to this article.

ively determine the presentation attacks. We create a new database of 13250 images at different focal length to present a detailed analysis of vulnerability together with the evaluation of proposed scheme. An extensive evaluation of the newly created database comprising of 5 different Presentation Attack Instruments (PAI) has demonstrated an outstanding performance on all 5 PAI through proposed approach. With the set of complementary benefits of proposed approach illustrated in this work, we deduce the robustness towards unseen 2D attacks.

## 8.2 Introduction

The recent trends in the smartphone market have demonstrated technological advancement and adaptations to the growing necessity of security through the use of biometrics. The use of smartphone as a biometric sensor is driven by many factors that include convenience at ease for authentication, unconstrained data capture and secure transactions without token based devices. Based on the growing usage of smartphone as a biometric capture device, many works have advocated the use of face biometrics among other possible characteristics due to its usability. The preference towards using face characteristics can be primarily attributed to the ease of data capture with a simple user interaction [132, 91].

With the ease of using face biometrics on smartphone in a unsupervised manner comes the threat of possible presentation attacks [146]. One can easily exploit the advantage of unsupervised data capture to claim else's identity through the use of printed face photos or by displaying the photo from a electronic display of computer, smartphone or tablet [87, 62, 6]. The ease of such attacks is highly attributed to the availability of face pictures across various social media where people publish their high quality pictures. Claiming identity through the illegitimate use of biometric artefacts (such as printed photo or electronic screen display) to gain access to secured systems are popularly termed as *presentation attacks* a.k.a., *spoofing attacks*. Recently Samsung Galaxy S8 was attacked using a simple photo demonstrating the vulnerability of the face recognition system on smartphone towards presentation attacks [†]. Thus, there is a need for stronger counter-measures towards the presentation-attacks, especially from 2D sources such as printed photo and electronic screen displays.

[†] http://tinyurl.com/kxfxmw6

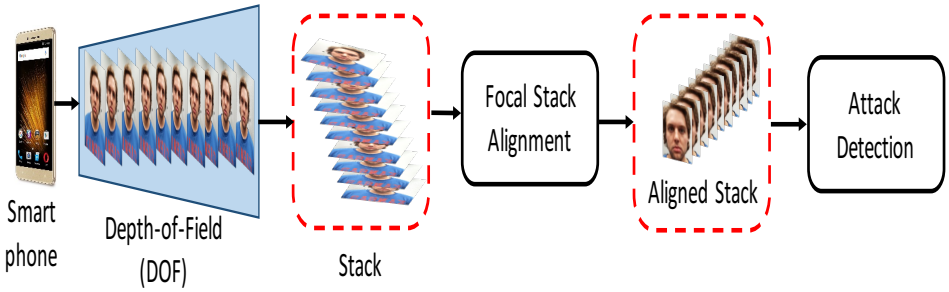



Figure 8.1: Proposed approach for detecting real and attacks using the depth-based approach in smartphone.

### 8.2.1 Related Works

In order to make the smartphone based face biometric system resistant to *presentation-attacks*, a number of works have been carried out [87, 132, 54, 103, 86, 131, 38, 20, 27, 190]. Most of the earlier works have focused on detecting the presentation attacks on smartphones using image quality [27] and textural features such as Gabor features, LBP or scale-space features [27, 6, 6, 62, 90]. A recent work analyzed the Moiré pattern aliasing that commonly appears during the recapture of video or photo replays on a screen in different color channels (R, G, B and grayscale) at holistic level and also at region level [132]. Further, they explored multi-scale LBP and DSIFT features to represent the characteristics of Moiré patterns to detect attacks. Another approach proposed to use the inherent sensors in the smartphones by leveraging the motion and light sensors to defend against 2D face attacks [20]. In similar lines, another approach proposed to use the camera, audio output component (e.g., earpiece), and audio input component (e.g., microphone) to determine the liveness by measuring the signals emitted and reflected which differs from 2D attacks [38], but does not discuss the uniqueness of subject level identification. Yet another kind of works have considered the challenge response of the subjects to perform certain tasks to substantiate the liveness during authentication attempt. Alternative to such approaches, a set of works proposed to use special hardware for addressing the presentation attacks in face recognition using depth-based cameras such as *Microsoft-Kinect, Light-field/Plenoptic* cameras [43, 87, 140]. However, it has to be noted that these works [43, 87, 140], were not related to smartphone based face recognition which is the prime focus of our work in this article.

From the set of works listed above, it can be deduced that most of the earlier works have focused on either software based approaches of analysing the texture or on the hardware based approaches of using additional hardware. On the downside of such hardware based approaches is that they cannot be easily integrated into smartphones without a significant engineering effort [42, 162]. In order to address the vulnerability of the smartphones towards presentation attacks while not using specialized hardware, we formulate a new approach of using the current capabilities of the smartphone camera with a classifier free approach to detect the 2D presentation attacks. We specifically leverage the fundamental approaches of differing focus in an imaging volume (depth-of-field) to obtain a set of images in single capture. The number of images obtained for a particular scene with varying focus results in a stack of images wherein the objects are imaged with different focus. The set of images are obtained for analysing the focus across stack in a cumulative manner to distinguish the real subject attempting to verify using face characteristics against conventional 2D presentation attacks through the use of printed photo or electronic screen. The proposed focus based imaging technique is complemented with a cumulative focus measure to determine the depth in the scene, specifically the face characteristics. To exemplify the proposed approach and provide empirical evaluations, we present a set of experiments on a newly collected database of moderately large number of verification attempts with 500 real (live) attempts and 750 presentation attack attempts which consists of a 5 different attacks. The attacks are carefully engineered to consider most commonly used 2D print attacks and electronic screen attacks with iPad Pro (retina display), iPhone 6S, laptop screen (Macbook with Retina display), high-resolution monitor (Quad-HD resolution) to cover the broad spectrum of possible attacks. The total number of images in the new attack detection database amounts to 13250 images (refer Section 8.4).

Further, we also present a number of possible use-cases of the proposed system to fully utilize and extend for other biometric applications. Of the many advantages, the proposed approach can be used to create a 3D model of the face by utilizing the depth from focus as the set of images are obtained with differing focus. The primary focus of this paper is not to use a classifier based approach and thereby propose a generalized solution towards 2D presentation attacks. The key contributions of this work can be outlined as:

- Presents a new approach of utilizing the differing focus in the imaging volume to capture the face images and determine the liveness. To the best of our knowledge, this is the first work to utilize the varying

focus stack images for presentation attack detection in smartphone based face recognition system.

- Proposed approach presents a learning-free classification approach to determine the presentation attempt through the analysis of a simple cumulative focus measure. The approach is highly generalizable towards 2D presentation attacks, even the unseen kind of attacks.

- Presents an extensive analysis of attack classification by analysing 500 live attempts versus 750 presentation attempts and exemplifies the robustness of detecting 2D attacks.

- Additionally, presents a set of possible use-cases of the proposed approach to provide extended capabilities such as 3D face reconstruction and also, the compatibility of analysing texture to work synchronously with currently deployed texture-based system.

In the rest of this work, we present the proposed approach in Section 8.3 and Section 8.4 presents the details of the newly created database to evaluate the proposed approach. Section 8.5 presents the set of experiments and the results to demonstrate the applicability of proposed approach. Further, Section 8.6 also lists extended capabilities of the proposed approach and in the end, Section 8.7 lists conclusive remarks and also lists limitations of the current work.

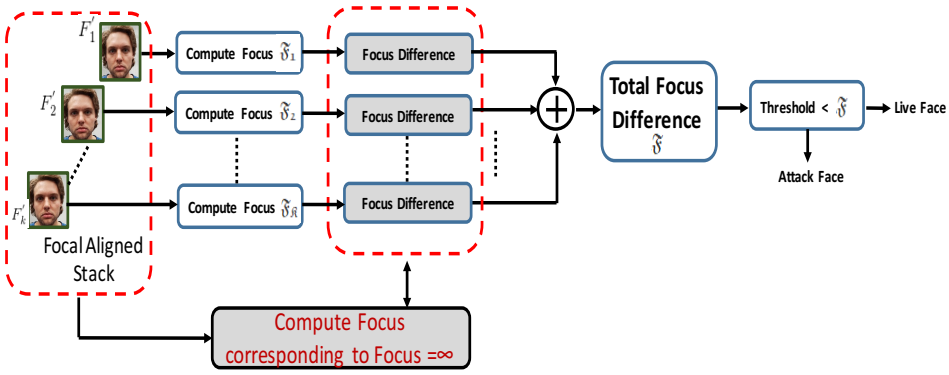



Figure 8.2: Schematic of the proposed attack detection mechanism in the smartphone system.

## 8.3 Proposed Approach

The schematic of the proposed approach is presented in the Figure 8.1. As it can be noted from the illustration, a set of images from the smartphone camera are captured in the imaging volume of a varying depth where the maximum focus is set to infinity. We proceed to sweep the entire focal length in the inverse manner from focus at infinity towards focus of zero and thereby obtain a *focal stack*. However, in the case of smartphone based face recognition, the photos are captured by holding the smartphone in the hand and as a direct impact of this way of capturing the picture, the resulting images are impacted by *motion parallax*. To account for this motion parallax, we first perform *focal stack alignment* such that the images corresponding to different focal length are stabilized. Further, aligned stack images are used for determining the presentation attacks from the presented face (real or artefact). Details of the steps are described in the following sub-sections.

### 8.3.1 Stack Image Alignment

Consider $S = \{S_1, S_2, \ldots S_n\}$ is a set of stack images where $n$ is a number of images in stack $S$ captured at different focal lengths. In this particular scenario of smartphone face recognition system, it has been observed that these stack images have severe parallax motion due to involuntary movement of hands. Further, these set of motions cannot be modelled in a specific manner through mathematical formulations [8]. In order to account for this, we first compute the affine warp using the Inverse Compositional Image Alignment (ICIA) algorithm [8]. Considering the stack image $S_i$ and $S_j$ corresponding to a focal length $i, j$ from the set of $n$ focal images, they can be represented as $S_i = S(x, y)$ and $S_j = S(x + \delta x, y + \delta y)$ where $\delta x$ and $\delta y$ jointly represent the affine transform. Thus, we obtain the correction for the affine transform between $S_i$ and $S_j$ using the ICIA algorithm [8].

### 8.3.2 Stack Alignment Refinement

However, such alignments are not refined to the best accuracy and due to number of factors such as differing motion between the capture time of two subsequent frames. Thus, we adopt optical flow based refinement of the aligned stack to have highly accurate stack-alignment for subsequent operations. Considering the factors such as appearance changes as an impact of environmental conditions such as illumination, chromaticity, blur, deformations due to small pose changes and limited computation resources on smartphone, we use Dense Inverse Search based optical flow alignment

[95]. Specifically, we determine the patch level correspondences using inverse search followed by dense displacement field creation based on the aggregated patches in multiple scales which result in variational refinement [95, 174]. Thus, we obtain the set of well aligned focal stack images represented by $S' = \{S'_1, S'_2, \ldots S'_n\}$ in the stack corresponding to focal length in the range $1, 2, \ldots n$ (Refer Algorithm 1).

**Algorithm 1** Stack Image Alignment

**Require:** $n \geq 1$ , *where n is number of face image in each stack*

**for** $i \neq 0$ *and* $i < n$ **do**
  Capture stack image $S_i$
  $\{S(i)\} \leftarrow S_i \quad \therefore S = \{S_1, S_2, \ldots S_n\}$
  where $S$ is a stack of face images
  $i = i + 1$
**end for**
**for** $j \leq n$ **do**
  Capture stack image $S_i$
  $\{S'_j\} \leftarrow AffineTransform(S_j,\ S_{j+1})$
  where $AffineTransform$ is computed using [8]
  $j = j + 1$
**end for**
**for** $k <= len(S')$ **do**
  Divide each image $S'_k$ into M blocks of equal size
  $\therefore \{B \leftarrow S'_k\} = \{b_1, b_2, ...b_M\}$
  **for** $m = 1$ & $m < M$ **do**
    Then DO inverse search for $b_m$ across stack
    $b_{am}$ = Compute Dense Field Flow($b_m$)
    $m = m + 1$
  **end for**
  $\{S''_k \leftarrow B'\} = \{b_{a1}, b_{a1}, \ldots b_{am}\}$,
  where $m = \{1, 2, ...M\}$
  $k = k + 1$
**end for**
$S' \leftarrow S'' = \{S''_1, S''_2, \ldots S''_n\}$
$OUTPUT \leftarrow S'$

### 8.3.3 Presentation Attack Detection

In order to determine the presentation attacks and considering the facial biometric system, it is essential to obtain the face region alone. This step serves two primary purposes: (1) reduces the computational time by using

the region of interest and (2) accurate depth information can be obtained within facial region. However, processing each of the image for obtaining the face results in redundancy. Thus, considering the fact the images are well aligned through the steps indicated in Algorithm 1, we extract the face region through Haar cascade [188] from the image corresponding to $focus = \infty$ out of all the images in the aligned stack images.

Further, we determine the face region from the image focused at $\infty$ to obtain the bounding box and correspondingly obtain face region from all the stack aligned images which are further indicated by $\{F'_1, F'_2, \ldots F'_n\}$.

In order to determine the presented face image as a normal (bonafide/real) image or artefact (spoof) image, we further measure the detailed focus from each of the face image in $\{F'_1, F'_2, \ldots F'_n\}$. We employ the Tenengrad [96] focus measure to determine the focus of each face region as given by Equation 8.1. Given the set of aligned face images $F' = \{F'_1, F'_2, \ldots F'_n\}$, we compute the focus of each image as given by Equation 8.1.

$$\mathfrak{F}_k = \sum_{(i,j)\in\Omega(x,y)} (F'_{kx}(i,j)^2 + F'_{ky}(i,j)^2) \tag{8.1}$$

where $F'_{kx}$ and $F'_{ky}$ are the horizontal and vertical image gradients computed by convolving the given stack image with Sobel operators. Thus, for a set of stacked images with face region given by $\{F'_1, F'_2, \ldots F'_n\}$, the computed focus is given by $\{\mathfrak{F}'_1, \mathfrak{F}'_2, \ldots \mathfrak{F}'_n\}$. Following the focus measurement of different stack images, we formulate the algorithm for determining the bonafide attempt versus presentation attack attempt as given by Algorithm 2.

As outlined from the Algorithm 2, if the cumulative focus difference obtained ($\mathfrak{F}$) is lesser than $\mathfrak{D}$ (prior determined threshold for focus change using development set), the presentation is considered as a attack attempt and bonafide otherwise.

## 8.4 Database

In this section, we present the details of the newly constructed database by considering the new approach of utilizing intrinsic characteristics of the smartphone camera to obtain the stack of varying focus image which is hereafter referred as *Variable focus Smartphone Face (VaSF)* database. It has to be noted that there exists no such publicly available database as the idea is explored for the first time in this work. The key aim of this database was to collect a set of frontal face images captured at different focal planes from the smartphone camera.

**Algorithm 2** Attack Detection
*Note - The threshold $\mathfrak{D}$ is computed on the development database discussed in Section 8.4

Liveness threshold: $\mathfrak{D}$(Prior determined focus threshold)
$Face_{ROI} \leftarrow \{Face_{ROI} | Face_{ROI}\ has\ focus\ value\ \mathfrak{F}_\infty\}$
$\forall Face_{ROI} \in F\ image\ with\ focus = \infty$ such that the measured focus value is $\mathfrak{F}_\infty$.
**for** stack image (face) in $\{F'_1, F'_2, \ldots F'_n\}$ **do**
    compute focus difference: $\Delta\mathfrak{F} = \mathfrak{F}_\infty - \mathfrak{F}_k$
    where $k = \{1, 2, \ldots n\}$
    Total focus difference: $\mathfrak{F} = \mathfrak{F} + \Delta\mathfrak{F}$
**end for**
**if** $\mathfrak{F} > \mathfrak{D}$ **then**
    $OUTPUT \leftarrow Live$
**else**
    $OUTPUT \leftarrow Spoof$
**end if**

In this work, we have created a new database of 50 subjects which is further divided into two disjoint sets of 25 subjects each corresponding to development and testing set. Note that the proposed scheme does not require the classifier to be trained and hence, there is no training dataset used in this work. The face image database was collected using recent smartphone Samsung Galaxy S7 and the data collection is divided in tow parts as described below.

**Bonafide Database**: The data collection is carried out in 2 sessions. The first session corresponds to enrolment samples where face image of each subject was captured thrice (1) The first sample is used to analyse the vulnerability of the face recognition towards artefacts or presentation attacks. (2) The second sample used as the enrolment image. (3) The third sample is used for creating the artefacts (attacks).

$Session - 2$ corresponds to data captured using the proposed approach where each face image was captured using the focal-stack imaging. Each subject was captured in 10 different instances as the probe attempts. In each capture, focus within the depth-of-field was varied automatically using the *mobile-application developed* in this work to realize the proposed scheme. Table 8.1 tabulates the statistics of the data collected in this session. Each face image was accompanied by 10 stack images with varying focus resulting

| Bonafide Database | | | |
|---|---|---|---|
| **Protocol** | **Subjects** | **No of Recs** | **Samples** |
| **Enrollment** | 50 | 3 | 1500 |
| **Probe** | 50 | 10 | 500 |
| **Attack Database** | | | |
| **Monitor Attack** | 50 | 3 | 1500 |
| **Laptop Attack** | 50 | 3 | 1500 |
| **iPad-Pro Attack** | 50 | 3 | 1500 |
| **iPhone 6S Attack** | 50 | 3 | 1500 |
| **Printed-photo Attack** | 50 | 3 | 1500 |

Table 8.1: Variable focus Smartphone Face (VaSF) Database

in a total of 5000 stack images.

**Attack Database**: To generate this database we have used 5 types of presentation attack instruments (PAI) which are desktop monitor with Quad HD resolution(2560 x 1440), Laptop screen (Macbook Pro), iPad Pro display, iPhone 6S display and printed-photo. Each attack attempt is captured by fixing both PAI and the capture device (Samsung Galaxy S7). The presentation attacks are captured using the same mobile application and device. We have followed same data capture protocol as mentioned in previous section. The complete statistics of the database is presented in the Table 8.1.

## 8.5 Experiments and results

This section presents the details of the experimental evaluation of the proposed approach on the newly collected database. We systematically present the vulnerability analysis followed by the quantitative performance of the proposed scheme in the subsequent sections.

### 8.5.1 Vulnerability analysis

To effectively evaluate the vulnerability of Face Recognition System (FRS) using the newly constructed database for both print photo and display attack, we employ the commercial FRS from Neurotechnology Embedded Face Matcher [2]. The vulnerability results are quantified using the IS0/IEC 30107-3 [68] that defines the metric named: Imposter Attack Presentation Match Rate (IAPMR) which is defined as *the proportion of imposter attack presentations using the same Attack Instrument species (print photo or display) in which the target reference is matched in a full-system evaluation of*

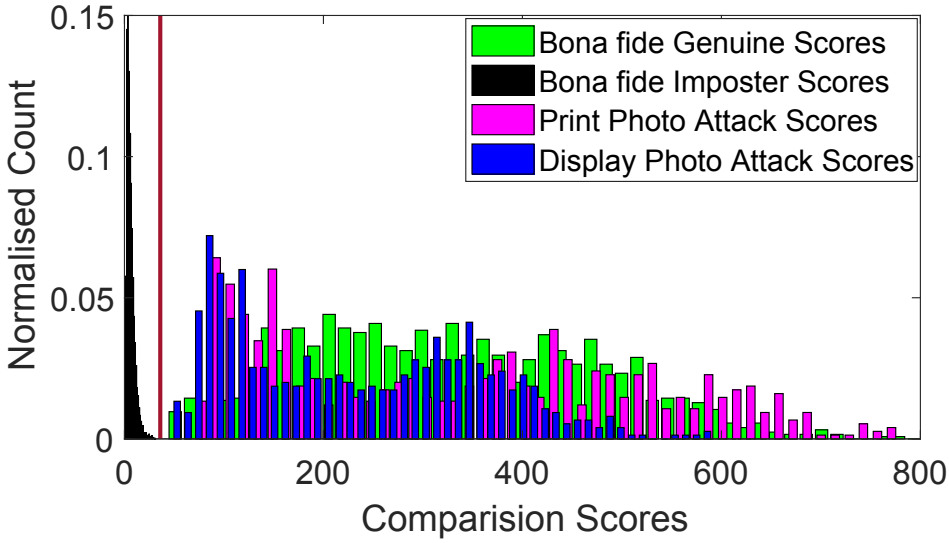



Figure 8.3: Vulnerability analysis of the commercial Neurotech face recognition system

*a verification system.* The higher the value of IAPMR, the higher is the vulnerability of FRS.

Figure 8.3 shows the distribution of the comparison score obtained on both bonafide and presentation attack species such as print photo and display. The comparison scores are computed by operating the FRS in two modes such as: (1) Normal mode: in which the enrolment and probe samples for the subject corresponds to the bona fide presentation. (2) Attack mode: in this case, the enrolment samples for the subject corresponds to the bona fide and probe sample corresponds to the artefact sample (print or display). Since we are evaluating the commercial system, we have used the operating threshold recommended by the vendor [2]. In this work, we have evaluated the vulnerability at two different operating points such as FAR = 0.1% and FAR = 0.01%. For simplicity, we have illustrated the threshold at FAR = 0.1% with a vertical line as shown in the Figure 8.3 and the results are illustrated only on the printed-photo and screen display attacks.

Table 8.2 depicts the quantitative vulnerability of five different kinds of Presentation Attacks Instruments (PAI) employed in this work. It is noted from the experiments that, the commercial FRS employed in this work has demonstrated an vulnerability towards all five PAIs with IAPMR of 100%. This further indicates the near-real quality images of the generated present-

Table 8.2: IAPMR (%) of the commercial FRS

| Method | PAI | IAPMR @ FAR = | |
|---|---|---|---|
| | | 0.1% | 0.01% |
| Neurotech | Display Monitor | 100 | 100 |
| | Laptop | 100 | 100 |
| | iPad-Pro | 100 | 100 |
| | iPhone 6S | 100 | 100 |
| | Printed-Photo | 100 | 100 |

ation attacks.

### 8.5.2 Performance of proposed PAD algorithm

In this section, we present the quantitative performance of the proposed attack detection approach. The evaluation is carried out on the testing set while development set is used to tune the parameters of the proposed scheme and to determine the thresholds for determining the operational points. The key operational points are quantified using *Bonafide Presentation Classification Error Rate (BPCER) and Attack Presentation Classification Error Rate (APCER)* as described in IS0/IEC 30107-3 [68]. **BPCER** is defined as proportion of bonafide presentations incorrectly classified as presentation attacks at the attack detection subsystem in a specific scenario while **APCER** is defined as proportion of attack face images incorrectly classified as bonafide images at the attack detection subsystem in a specific scenario. Besides, we also report the performance of the system by reporting the value of BPCER by fixing the APCER to 5% and 10% corresponding to realistic operating values of commercial face recognition systems.

As observed from the Table 8.3, the proposed approach has demonstrated and outstanding performance on all 5 kinds of presentation attacks evaluated in this work. The applicability of the proposed approach in detecting different PAI further justifies the generalisablity towards unseen 2D attacks. It is interesting to note that the proposed method does not rely on a classifier unlike most of the existing state-of-art attack detection schemes.

The key point to note here is that the proposed work being a new and unique of it's kind, there exists no prior work to compare with. This fact is also justified due to no-use of classifiers or dedicated feature extraction methods (texture based, frequency based so on.).

Table 8.3: Quantitative performance of the proposed approach for detecting the face presentation attack on smartphones.

| PAI | EER (%) | BPCER @ APCER = | |
|---|---|---|---|
| | | 5 % | 10 % |
| Display Monitor | 4.00 | 2.67 | 1.33 |
| Print Photo | 1.33 | 0.00 | 0.00 |
| Laptop Screen | 1.33 | 0.00 | 0.00 |
| iPad-Pro | 1.33 | 0.00 | 0.00 |
| iPhone 6S | 0.00 | 0.00 | 0.00 |

## 8.6 Other Advantages and Limitations

This section presents the advantages of the proposed scheme for other functionalities in-addition to the presentation attack detection. Further, we also list the anticipated drawbacks. To this extent, we present the following cases: (1) Depth-from-focus and Depth-from-defocus, (2) Depth based texture analysis and (3) 3D reconstruction of face.

**Depth-from-focus and Depth-from-defocus:** The basic idea for depth-from-focus (DFF) assumes that the distance of an object to the camera at a certain pixel corresponds to the focal setting at which the pixel is seen sharpest. For a given data set with differently focused images from our proposed approach on smartphone, it is expected to find suitable contrast measure at each pixel and thereby the depth at each pixel can be estimated by determining the focal distance. Thus, with the proposed approach Depth-from-focus and conversely depth-from-defocus, can be used to estimate reliable depth map. One such example is shown in the Figure 8.4 where the depth maps are given in the bottom row corresponding to face images in the top row. It can be noted that the proposed approach extracts robust depth map which distinguishes the real image from 2D attack images.

**Depth based texture analysis:** In line with the earlier works, one can explore the texture based information to detect the presentation attacks together with a classifier [43]. This aspect has not been explored in this specific work as it was mainly focused on the *classier free* approach to detect the presentation attacks. As an another example, this work can be used seamlessly with the existing approaches of using distortions and Moire pattern to

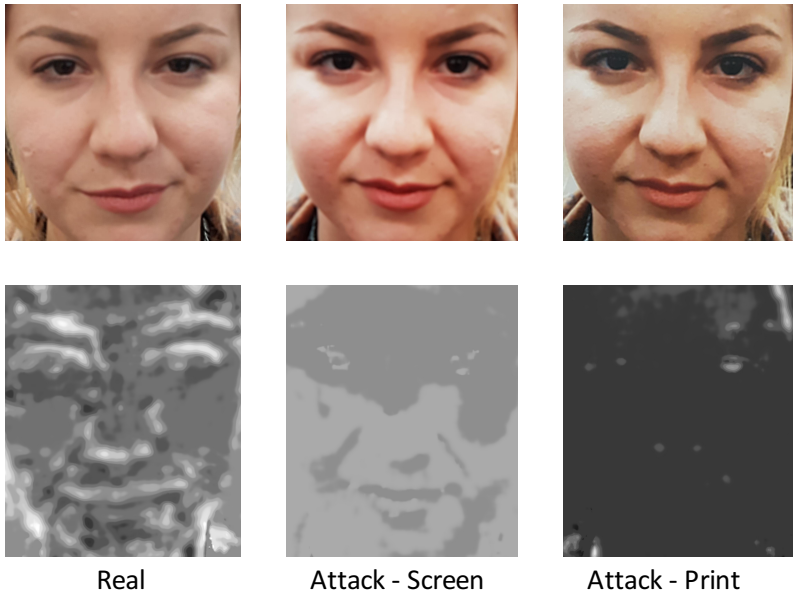



Figure 8.4: Depth map estimation on real and attacks for sample subject.

determine the attacks. Figure 8.5 presents the example of number of stack images captured from the proposed approach where the Moire pattern is visible inherently without additional effort exemplifying the usefulness of the proposed approach.

Figure 8.5: Moire pattern visibility across a sample stack image for a subject.

**3D reconstruction of face:** The proposed approach can be well utilized to construct the 3D image using a number of stack images captured at different depth in a imaging volume. The 3D can further be used to robust verification and also presentation attack detection.

Figure 8.6: 3D reconstruction using the data from proposed approach.

**Limitations of current work:** The current system has been well tested against various 2D presentation attacks emerging from print and screens to prove the usefulness of the proposed approach. However, considering the presence of realistic depth in 3D mask attacks, the proposed approach can be vulnerable. Due to cost factors and non-availability of 3D masks, an analysis of 3D attacks is not included in this work. It should however be noted that 3D attacks can be mitigated with the help of proposed approach along with the texture based counter-measures.

## 8.7 Conclusion

Smartphone based face biometrics is gaining importance due to emerging applications including online banking, mobile authentication and so on. In this work, we have presented a new approach by exploring the intrinsic char-

acteristics of the smartphone camera to capture a number of stack images in the depth-of-field. The proposed method has achieved a *feature-free* and *classifier-free* approach to determine the 2D presentation attacks. Thus, the proposed approach is generalizable for the unseen 2D attacks and the proposed approach is extensively evaluated on the newly created database comprising of 5 different Presentation Attack Instruments (PAI) that include iPad Pro (retina display), iPhone 6S, laptop screen (Macbook with Retina display), high-resolution monitor (Quad-HD resolution) and printed-photo. The obtained results have demonstrated an outstanding performance on all 5 PAI substantiating the applicability of proposed scheme in real-life scenario. Together with the superior performance, we have illustrated a set of complementary use-cases towards achieving robust biometric system on smartphones.

## Acknowledgement

This work is carried out under the funding of the Research Council of Norway (Grant No. IKTPLUSS 248030/O70).

# Chapter 9

# Article 4: Presentation Attack Detection for Smartphone Based Fingerphoto Recognition Using Second Order Local Structures

Pankaj Wasnik, Ramachandra Raghavendra, Kiran Raja, And Christoph Busch. *"Presentation Attack Detection for Smartphone Based Fingerphoto Recognition Using Second Order Local Structures"* In the proceedings of $14^{th}$ International Conference on Signal-Image Technology & Internet-Based Systems (SITIS 2018), IEEE, 2018.

## 9.1 Abstract

Fingerprint recognition on smartphone provides a good alternative over traditional security measures such as lock-patterns and pin. However, such fingerprint systems have some inherent problems such as the fact that user will leave their latent fingerprint on the sensor, and the limited sensor area. Additionally, fingerprint sensors impact on the cost and form factor of the device. Hence, in literature, the camera based approaches such as fingerphoto recognition systems got the attention of many researchers and

manufacturers. However, such systems are highly vulnerable to presentation attacks such as photo-prints, display and replay attacks. To countermeasure these attacks, we propose a robust presentation attack detection scheme based on the features extracted from the maximum response images obtained from the convolution of second order Gaussian derivatives and the input images at multiple scales. The proposed scheme has achieved the detection performance of BPCER of 1.8%, 0.0% and 0.66% at APCER=10% for the presentation attack instrument species i.e., print-photo, display and replay attacks respectively.

## 9.2 Introduction

In the recent years, smartphones are not only used for making the telecommunication calls but also for tasks like internet surfing, accessing emails, various cloud and banking services. Hence, the potential unauthorized access to smartphones increases the risk of exposing sensitive data from the smartphone owner. The smartphone-based biometrics became mainstream in recent years. There has been significant adoption in using biometric authentication as a good alternative to the conventional security measures like passwords. The most common biometric modalities in smartphones include fingerprint, face, voice and iris recognition. However, fingerprint recognition on smartphones is a more reliable, efficient and popular method of user authentication due to its inherent characteristics. For example, touching a fingerprint sensor does not require a user to make any specific posture, unlike face and iris scanner. It also works well in low light conditions where face and iris recognition may fail. Although the dedicated fingerprint sensors can acquire a high-quality biometric sample, which is associated with low error rates, they have some inherent problems such as the user will leave latent fingerprints on the sensor surface and that the sensor area is small. Additionally, these sensors impact on battery usage, the cost and form factor of the device. Hence, authentication of mobile users based on photos of their fingers could be a good alternative since the inherent smartphone camera is used to capture the fingerphotos [158, 39, 138]. Moreover, reusing the smartphone camera could provide an ability to capture more than one sample from a biometric instance with minimal interaction and larger finger area.

In literature, although many researchers have studied the user authentication based on fingerphotos, mostly these methods do not show higher biometric performance. Many works have been proposed using minutia based recognition where different algorithms proposed to extract the minutia from input images [116, 172, 39]. In [180] authors, introduced a smartphone based

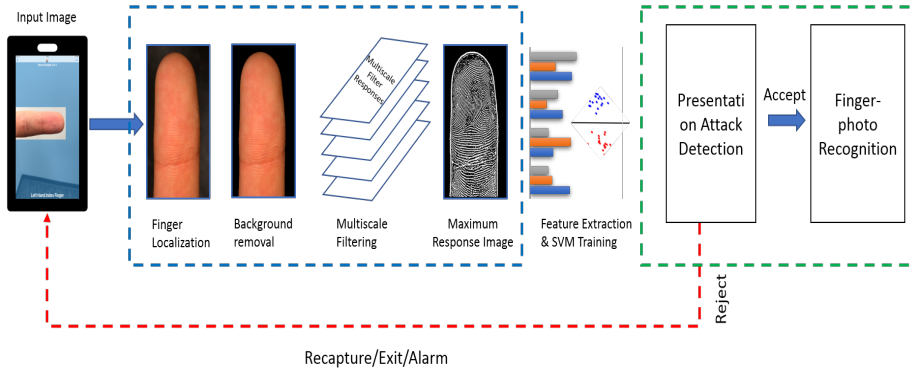



Figure 9.1: Block diagram of a proposed presentation attack detection scheme for fingerphoto recognition system

fingerphoto recognition system, where the number of matched SURF features extracted from fingerphotos used to compute the similarity between the two input images. Furthermore, [63] proposes a support vector machine (SVM) based biometric verification, where SVM is trained on the features extracted from the multispace random projections (MRPs) using the Gabor filters.

Moreover, in [104] authors are trying to analyze the interoperability between the legacy fingerprints and fingerphotos. Their study concludes that when we cross-match the fingerprint images with fingerphotos, the performance of the system is reduced. [97] presents a contactless 3D fingerprint recognition using a 2D image from a single camera, their method uses the Lambertian reflectance based shape from shading technique to develop a 3D fingerprint identification algorithm. [98] presented a novel, entirely touchless 3D fingerprint recognition system based on the 3D models and achieved the best EER of 0.06%.

Despite several advantages, the biometric systems are vulnerable to various attacks namely i) direct and ii) indirect attacks. The direct attacks are also known as presentation attacks where the attack samples are generated using the presentation attack instruments are fed to the biometric systems in order to intervene the normal behavior of the system [62]. Besides, when the camera is used as an acquisition sensor, the difficulty of constructing presentation attacks that will be accepted by the sensor is reduced. As there is no contact required between the sensor and the finger, a simple printout of a person's finger or displaying an image of someone's finger on a display

could be used to fool the biometric system [171, 178]. In [171], authors have proposed a Presentation Attack Detection (PAD) algorithm based on the light reflection properties of bona fide (i.e. real) and attacks images. They used the adaptive threshold, to detect the attacks and achieved an EER of 1.2-3.0% for an in-house database. Furthermore, in [178], authors have proposed a PAD for fingerphoto recognition using SVM and texture and gradient-based features for print-photo and display attacks where they have achieved an EER of 3.7% for LBP based SVM classification.

This paper present a robust PAD scheme as a countermeasure for print, replay and display attacks. Our method illustrates the successful use of the features extracted from the maximum response images obtained from the convolution of input images with the second order Gaussian derivatives at multiple scales. The maximum filter response is mainly obtained by calculating the likelihood of a pixel belonging to the ridges and valleys based on the eigenvalues of the Hessian matrix of convolved images. Considering the inherent distortions introduced in the attack images compared to the real images, we propose a novel presentation attack detection scheme.

The proposed scheme illustrates the successful use of the vesselness filter also known as the *(Frangi filter)* proposed by Frangi et al., in [48]. The *Frangi filter* mainly convolves the input image at multiple scales to produce the Maximum Filter Response image (MFR). MFRs are then processed to get the final decision by learning the the support vector machine (SVM) classifier and features extracted using the local binary patterns (LBP), the binarized statistical image features (BSIF) and the Histogram of Oriented Gradients (HOG) [122, 83, 28].

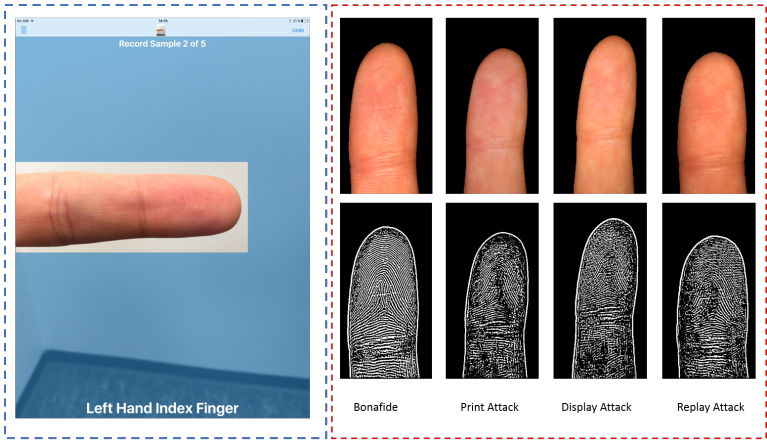
In rest of the paper, Section 2 describes the proposed PAD scheme. Section 3 presents the database used, and Section 4 discusses results and experiments. Finally, Section 5 gives the concluding remarks.

## 9.3 Proposed Scheme

This section describes an overview of the proposed PAD scheme system. Figure 9.1, shows the block diagram of the proposed scheme. The proposed scheme is divided into three main parts.

### 9.3.1 ROI Extraction

The first step in the proposed PAD scheme is the extraction of the region of interest (ROI), i.e., the finger area and background removal. In order to make an effective ROI extraction, we have developed an iOS mobile applic-




(a) Capture Application (b) Input and output of Frangi Filter

Figure 9.2: Data capture screen with the transparent blue mask of the developed iOS application. Fingers in the first row of the red box show the background removed extracted fingers whereas the in the second row the corresponding MFRs are shown.

ation for capturing the samples. We have added a transparent blue overlay mask on the capture screen, where a rectangular region is shown so that users can place their fingers in order to record the sample. The overlay mask has two primary goals 1) to place the finger at the known guided location and thus to increase the usability 2) to extract the ROI with minimalist background information. Once, the region inside the rectangle cropped, it is then processed to mask the background pixels. To remove the background we first convert the cropped image from RGB to L*a*b* color space since here, the colors are represented perceptually linear than other color spaces, which help us to cluster the colors in foreground and background even with finer details. Hence, we then apply K-means clustering [170] to classify these pixels into the foreground and background information. Later the processed image is used as an input to the *Frangi filter* for obtaining the MFR image. In Figure 9.1, the region inside the dotted blue box shows the results at the intermediate and final stages of ROI extraction step. Furthermore, the Figure 9.2.a) presents the capture screen of the data capture application. The images inside the red box show the sample input and output of the *Frangi filter* for bona fide and all three attacks.

### 9.3.2 Maximum Filter Response

This step aims to enhance the fingerprint patterns (i.e., ridges) present in the fingerphotos by suppressing the non-vessel-like structures. In this

approach, the input image is convoluted with the second order derivatives of the Gaussian kernel. The convolution can model the vessel-like structure i.e., the fingerprint structures such as ridges which are similar to thin vascular structures. The convolution is performed at several different scales. The scale which corresponds the most with the underlying fingerprint structures, is assumed to give the highest response.

More specifically, at different scales the image $I(x)$ is convolved with normalized second order Gaussian derivative to form Hessian Matrix ($H$) of the image $I(x)$. The Hessian matrix is matrix that contains the second order derivatives of image. For 2D images, the matrix $H$ is always of size 2x2 at any given point $x$. The properties Hessian matrix based on eigenvalues and eigenvectors are very much useful for the enhancement fingerprint patterns present in the input image $I(x)$.

Let $\lambda_1$ and $\lambda_2$ be the two eigenvalues of $H$ such that $|\lambda_1| \leq |\lambda_2|$ is always true when sorted on their absolute value. Therefore, for any given pixel $x$, the likelihood of the pixel belonging to the fingerprint pattern is given by [48]:

$$fp(x) = \begin{cases} 0 & if\ \lambda_2 > 0 \\ \exp\left(\frac{-\mathcal{D}_\beta^2}{2\beta^2}\right)\left(1 - \exp\left(-\frac{C^2}{2\alpha^2}\right)\right) \end{cases} \tag{9.1}$$

where $fp(x)$* is the probability of pixel $x$ belonging to the vessel-like structure, $\mathcal{D}_\beta = \frac{\lambda_1}{\lambda_2}$ is the anisotropy† of the pixel and $\mathrm{C} = \sqrt{\lambda_1^2 + \lambda_2^2}$ is the second order structureness‡ at the pixel $x$. The constants $\beta$ and $\alpha$ mainly control the sensitivity of $fp(x)$. The Hessian Matrix $H$ is computed at every pixel for different scales($s$) as per the defined range, and it is the standard deviation of the second order derivatives of the Gaussian kernel. As recommended in [48], the value of $\beta$ is set to 0.5 and value of $\alpha$ is computed using the gray pixel values of an input image. Finally, the maximum filter response image is obtained by taking a maximum of the response of the filter across all scales.

*This value is high only if anisotropy and structureness both are high.

†This ratio is mainly distinguishes the plate-like and line-like structures. Its high when there is a high change in the image intensity in one direction and low in other direction in the closed neighbourhood.

‡This value is high only if there is a big change in intensity in closed neighborhood of pixel $x$.

| Modality | PA name | PA source | | | Presentation Attack Instrument (PAI) | Capture Sensor |
|---|---|---|---|---|---|---|
| | | Session | Sensor | Type | | |
| Finger | PA.1 | Session 1 | iPhone 6s rear camera | images | Epson Expression Photo XP-860 Printer Epson Photo Paper Glossy | iPhone 6s back camera (video) |
| | PA.2 | Session 1 | iPhone 6s rear camera | images | iPhone 6s display | iPhone 6s rear camera (video) |
| | PA.3 | Session 1 | iPhone 6s rear camera | videos | iPad-Pro display | iPhone 6s rear camera (video) |

Table 9.1: Details of the PA Sources, Presentation Attack Instruments and capturing sensor

### 9.3.3 Presentation Attack Detection

From Figure 9.2.b), we can see that there is a significant textural difference in the MFR of the live presentation (bona fide) and an attack presentation (artefact). Therefore, we can anticipate that the textural features extracted from the MFRs of bona fide and artefact images can aid into robust and accurate classification. Hence, we first extract the features using techniques such as LBP, BSIF and HOG [122, 83, 28]. These extracted features are then learned by the SVM classifier to categorize the input test images into bona fide and artefacts. The final decision is obtained by applying the threshold $t$ on the predicted scores by the SVM classifier. We can obtain the threshold $t$ corresponding to the equal error rate (EER) or attack presentation classification error rate (APCER) at 5% or 10% acceptance. The final criteria to accept an input image as a bona fide or an artefact can be given by:

$$Result = \begin{cases} Bonafide\ Image & if\ s > t \\ Artefact\ Image & if\ s < t \end{cases} \tag{9.2}$$

where $s$ is the output score predicted by an SVM classifier. Based on the decision from PAD module, the image is further processed for verification or an option for recapture/exit/alarm can be configured in the mobile application.

## 9.4 Databases

As there are no publicly available Presentation Attacks (PA) databases, we have constructed a PA database of 50 subjects consisting of three sessions of bona fide data and three types of attacks. The PA database constructed from the bona fide samples of Session 1 data and generated artefacts from it.

### 9.4.1 Artefacts generation

The artefact generation is carried out using three different Presentation Attack Instruments (PAI).

1. Print artefacts are generated using the printer Epson Expression Photo XP-860 and the high-quality photo paper which A4 size Epson Photo Paper Glossy, whose basis weight is 200 g/m and thickness is 8.1 mm.

2. Electronic replay attack is conducted by displaying the bona fide images and video artefacts on the iPad-Pro screen.

3. Electronic display attack is executed by displaying the bona fide images and video artefacts on the iPhone 6s screen.

### 9.4.2 Data collection

Within the scope of this work, we have developed an application for the iOS environment, i.e., for iPhone and iPad-Pro to capture the data. The data collection consists of mainly 2 phases.

#### 9.4.2.1 Bona fide Data Collection

During the recording of a session, the bona fide samples are captured in one of the two scenarios. Session 1 & 2 are captured in the indoor scenario with uniform illumination, and Session 3 is captured outside during the daylight (sunlight) to have uncontrolled illumination. In each session we collected 5 images of 4K resolution and 5 second slow motion video of 1080p resolution @ 240 fps. The recordings are collected from the left and right-hand index and middle fingers of each subject. However, in this paper, we limit our study to use the data from the left-hand index fingers only.

#### 9.4.2.2 Artefact Data Collection

All the PA are generated using the samples from Session 1 data. In order to have enough variability in artefacts, we have used a different source for each PA and PAI. For example, for type PA.1 images captured using iPhone 6s rear camera is used as PA source and the PAI used is the Epson printer. These photo prints are then presented to the biometric capture device, i.e., iPhone 6s back camera. Table 9.1 summarizes the PA data collection in detail and Table 9.2 provides the particulars of the newly created PA database. Further, in Figure 9.2.b) the binarized images depict the MFRs for the bona fide and attack images.

## 9.5 Experiments and Results

To verify the effectiveness of the proposed PAD scheme we have carried out two experiments 1) PAD with the proposed scheme 2) PAD without the proposed multi-scale feature extraction i.e., PAD with Classical approach. The

| Type | No of Subjects | Total Images | Total Videos | Usage |
|---|---|---|---|---|
| Bona fide Data | | | | |
| Session 1 | 50 | 250 | 50 | Attack Gen |
| Session 2 | 50 | 250 | 50 | Train (SVM) |
| Session 3 | 50 | 250 | 50 | Test |
| Artefact Data | | | | |
| PA.1 | 50 | 0 | 50 | Train (SVM), Test |
| PA.2 | 50 | 0 | 50 | Train (SVM), Test |
| PA.3 | 50 | 0 | 50 | Train (SVM), Test |

Table 9.2: Statistics of newly created PA database.

Table 9.3: Performance of both approaches in terms of EER, BPCER @ APCER = 5% and BPCER @ APCER = 10%. In table the PA.1, PA.2 and PA.3 indicates the Print-Photo Attack, Display Attack and Replay Attack respectively.

| Method | PA.1 | | | PA.2 | | | PA.3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | EER | BPCER@5 | BPCER@10 | EER | BPCER@5 | BPCER@10 | EER | BPCER@5 | BPCER@10 |
| **PAD With Proposed Scheme** | | | | | | | | | |
| LBP | 7.862 | 13.255 | 7.372 | 6.680 | 6.780 | 6.540 | 6.254 | 6.392 | 6.000 |
| BSIF | 6.352 | 18.353 | **1.803** | **0.490** | **0.000** | **0.000** | **4.431** | **3.764** | **0.667** |
| HOG | 8.921 | 17.176 | 7.411 | 6.600 | 7.640 | 4.150 | 6.941 | 28.07 | 5.921 |
| **PAD Without Proposed Scheme (Classical Approach)** | | | | | | | | | |
| LBP | 10.19 | 13.56 | 10.27 | 3.312 | 3.335 | 2.940 | 5.0196 | 5.137 | 2.196 |
| BSIF | **5.902** | **5.882** | 5.882 | 5.882 | 5.882 | 5.882 | 5.882 | 5.882 | 5.882 |
| HOG | 21.41 | 53.52 | 29.33 | 5.459 | 6.784 | 2.940 | 11.09 | 52.07 | 23.60 |

performance of the PAD schemes is evaluated using metrics defined in the International Standard `ISO/IEC 30107-3` [68]. We mainly used two metrics 1) Attack Presentation Classification Error Rate (APCER), defined as the proportion of artefact presentation incorrectly classified as bona fide presentation (2) Bona Fide Presentation Classification Error Rate (BPCER), defined as a proportion of Bona Fide presentation miss-classified as artefact presentation [68]. For a robust PAD scheme, the values of APCER and BPCER metrics should be very low. Furthermore, it is common practice to report the values of BPCER @ 5% APCER or @ 10% APCER. We also report our results in terms of an equal error rate.

In order to evaluate the proposed scheme, we divided our database into training and testing sets based on the number of subjects. We have randomly picked 33 subjects data for training the SVM classifier, and the remaining 17 subjects' data is considered for testing. Furthermore, to have enough variability, we have chosen 150 frames from the bona fide and PA recordings. This resulted in a total number of 4950x2 training and 2550x2 testing samples. In our experiments to train the SVM, we use the bona

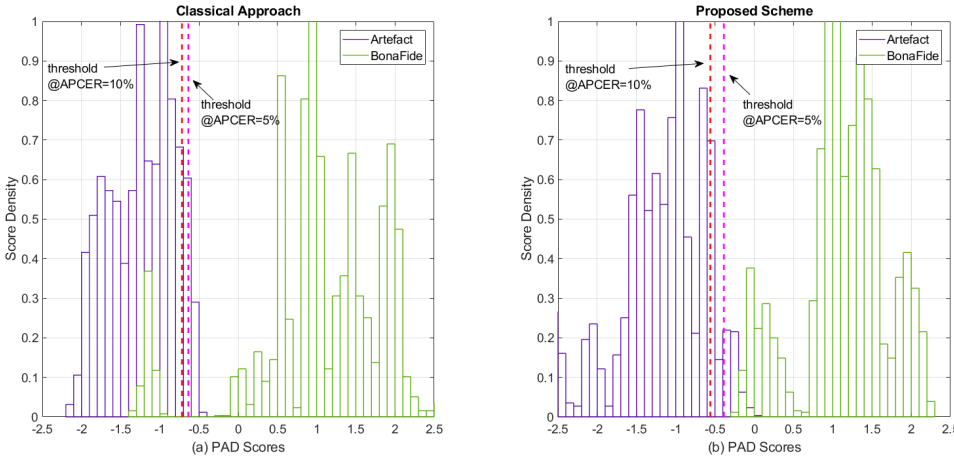



Figure 9.3: PAD score distribution for Classical method and proposed method for BSIF feature extractor and display attack. The magenta dotted lines denotes thresholds @APCER = 5% and red dotted lines indicates thresholds @APCER = 10%.

fide data from Session 2, and the performance is evaluated on the data from Session 3. This arrangement is made in order to verify the generalized performance of the proposed scheme. Further, Figure 9.3 reports the distribution of artefact and bona fide scores for both approaches. From Figure 9.3.b, we can see that there is no overlap between bona fide and artefact scores for both the thresholds, resulting in BPCER of 0.0% whereas, in case of classical approach, there is an overlap for both thresholds (See Figure 9.3.a) resulting in poor performance than proposed scheme.

Furthermore, Figure 9.4 presents the DET curves for both approaches. From figure, we can see that the DET curve of the proposed method for display attack shows the best performance whereas BPCER of classical approaches remains constant around 6.0% for all three attacks. Summary of the obtained results is tabulated in Table 9.3. From table, we can see that the BSIF featured extractor outperforms the LBP and HOG in case of all three attacks.

- For PA.1, the lowest error rate is achieved for classical approach with EER of 5.90%, BPCER @ APCER = 5% of 5.7%. However, for BPCER @ APCER = 10% the proposed method shows best performance with lowest BPCER of 1.80%.
- For PA.2, the proposed method shows best performance in case of

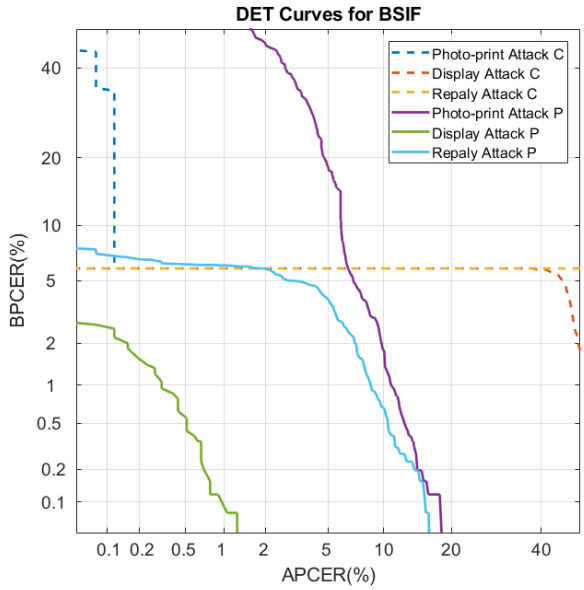



Figure 9.4: DET curves for BSIF feature extractor. In the figure, letter C indicates the Classical Approach whereas P indicates the Proposed Scheme.

all metrics with the values for EER, BPCER @ APCER = 5% and BPCER @ APCER = 10% as 0.49%, 0.0% and 0.0% respectively.

- Similarly, for PA.3, the proposed PAD scheme shows higher performance in-terms of all three performance measures. The achieved performance here is of EER = 4.43%, BPCER = 3.76% and 0.66% @ the accepted APCER of 5 and 10% respectively.

## 9.6 Conclusion

Based on the experiments and obtained results, we conclude that the proposed scheme with BSIF feature extractor shows the highest performance in most of the cases. However, we can see that the classical approach has outperformed the proposed method for the print-photo attack at low tolerated APCER, i.e., of 5.0%. Further, we observe that the proposed scheme shows an improved PAD performance when compared to the state-of-art methods [171, 178], where authors have achieved the best EER of 1.2% and 3.7% respectively. Moreover, if we observe the quality of MFR for bona fide images, the fingerprint patterns are well extracted from the fingerphotos. Hence, one can use these MFRs as an input to any fingerprint recognition system to perform the user verification.

An important point to consider in the future work is to evaluate the proposed scheme for advanced attacks like images of 3D printed fingerprint artefacts. One can also study the application of MFRs for verification purpose. We anticipate that a unified fingerphoto verification system can be formulated, with a single feature extraction technique which not only aids in the PAD mechanism but also for the user verification. Also, it would be interesting to examine the behavior of MFRs with advance features extraction techniques such as transfer learning or deep learning.

## Acknowledgment

This work was carried out under the funding from the Research Council of Norway (Grant No. IKTPLUSS 248030/O70).

# Chapter 10

# Article 5: Eye Region Based Multibiometric Fusion To Mitigate The Effects Of Body Weight Variations In Face Recognition

Pankaj Shivdayal Wasnik*, Kiran B. Raja*, R. Raghavendra* and Christoph Busch. *""Eye Region Based Multibiometric Fusion To Mitigate The Effects Of Body Weight Variations In Face Recognition."* In the $19^{th}$ International Conference On Information Fusion (FUSION 2016), pp. 2007-2014. IEEE, 2016.

## 10.1 Abstract

Face recognition has certain impediments due to alignment, illumination, facial expressions. Several techniques have been proposed to rectify these challenges. In recent years, many researchers have addressed challenges due to ageing, plastic surgery, twin identification, make-up and hairstyle. But, the impact of weight variation on face recognition has not been explored much. In contrary to other facial regions such as the cheek or chin area, the region near the human eye is not much affected due to the body weight changes. In this paper, we explore the use of eye region information to mit-

*All three authorshave contributed equally to this article.

igate the effects and stabilize the performance of the biometric recognition system. To this extent, we propose a multi-algorithmic and multimodal fusion strategies to combine the information from eye region (left and right). The experiments carried out on the publicly available eWIT database indicates the improved recognition performance by 6.42% when benchmarked with commercial face recognition system.

## 10.2 Introduction

Face Recognition Systems are known to provide better recognition performance under constrained environment than unconstrained environment. However, unconstrained or uncontrolled conditions can cause huge variations in the captured data, which can significantly reduce the recognition performance [134] of the face recognition system. The biometric performance of face systems has pushed even smartphones based system to explore 3D face recognition [139]. At the same time, the challenges due to illumination conditions, facial expressions and pose variations have attracted the attention of the research community resulting in various approaches proposed to mitigate these problems [75, 57, 151, 140]. Further, use of glasses, hairstyle, make-up, weight effects are gaining the attention in face recognition research.

The variations due to physiological changes which occur over time such as ageing, weight (gain or loss) can directly affect the recognition performance. The effects of ageing on the performance of the face recognition systems was illustrated in [151]. In [99] authors have proposed the simulation of ageing in facial images to understand it's effects on face recognition. Furthermore, the facial geometry changes significantly after plastic surgeries [167] and the impact of this on recognition performance was demonstrated in [166].

Many works have experimentally shown how various surgeries affect different conventional face recognition techniques. In [35] authors have evaluated various face recognition techniques such as FARO *(Face Recognition Against Occlusions)* and FACE *(Face Analysis for Commercial Entities)* to address the challenges due to plastic surgery procedures by dividing the facial image at different granule level. In similar terms, [11] proposed a model for face recognition to overcome the non-linear variations due to plastic surgery.

Even though many of the face recognition challenges were predominantly studied in the literature, there is still a limited work addressing the issues due to body weight variations. As the facial appearance indicates a gain or reduction of weight, it changes some of its geometrical properties and

Figure 10.1: Sample images from the eWIT database and details of left and right periocular regions.

may acquire new attributes such as big wrinkles or furrows in the facial region. These changes may have an equivalent impact as the above described changes due to plastic surgery. There are only very few works [117, 118, 196], which have attempted to address this challenge. In [117] authors have used deep learning techniques to learn about good features using existing regularization approaches like $l_1$ *norm*, $l_2$ *norm* and dropout to train the classifier. Whereas, in [118] they have used three different neural networks to train the main *RDF* classifier. Wen et al. [196] proposed a face recognition technique using well-known methods i.e., LBP [122] and SIFT [75] feature descriptors on a synthetic images database created using Photoshop. They used the Partial Least Squares (PLS) method to reduce the effects of weight changes and improve the accuracy of face recognition.

The state-of-art as mentioned above mainly concentrated towards the machine learning techniques and are designed to work with face characteristic only. In this paper, we explore the possibility of utilizing the eye region

Figure 10.2: eWIT sample images with age and weight variations

information along with holistic face information to address the problem of weight variation effectively using multi-algorithmic and multimodal fusion strategies. The region near the human eyes does not change as much as other areas (e.g. cheek and chin) in the case of weight gain or loss [41]. Motivated by this, we utilize the eye region information along with the holistic face. Thus the use of multi-biometrics is anticipated to improve the biometric performance as the information provided by multiple characteristics provides complementary information. To best of our knowledge, there are now works in this direction to use fusion approaches to mitigate the effects of body weight variations on face recognition systems. Further, the use of distinctive features from ocular characteristics to improve the biometric performance has been explored in different context [88].

In this work, we propose a novel fusion scheme based on score level fusion of comparison scores obtained from different feature extraction algorithms using multi-biometrics. Thus, given the captured face image, the proposed method will first perform face and eye detection using Haar Cascade Classifiers [187] to get the corresponding features. In the second stage, we extract the feature vectors using four different algorithms. We then obtain comparison scores using the *(Sparse Representation Classification)* (SRC) [200]. In the last step, we propose a novel fusion scheme which computes the final

comparison score. Extensive experiments are carried out on the images from eWIT database.

The outline of the paper is as follows: Section 10.3 presents details of the used dataset. Section 10.4 introduces the proposed multi-biometric, multi-algorithmic score level fusion scheme, Section 10.5 presents the experimental protocols and results. Section 6.6 draws the conclusion.

## 10.3 Database

In this paper, we have used the publicly available eWIT [117] database. This database is the extension of the WIT *(WhoIsIt)* database, which is also accessible to public [165], WIT contains publicly available images from the internet with 110 subjects and has 1109 images whereas, eWIT database is an extension of WIT. This database has 200 subjects with 2036 images and it mainly contains the images of well-known celebrities which are available on the internet. Each subject has at least 10 images with a clearly visible frontal face with age and weight variations. The following Table 10.1 gives particulars of the both databases.

For this paper we have used eWIT database and all images are marked up with one of the three labels: Thin, Moderate and Heavy. There are 437 images labelled as thin, 1309 as moderate and 290 as heavy. These images also contain age variation from 1 to 96 years. The average is 34.7 years, Further, the average difference between youngest and oldest image is 28.8 years. In this paper, we have used only 50 subjects from the database to obtain the results as we could not generate the whole database with 200 subjects because many of the provided web-links were not working or broken, since the authors [117] have provided only the set of web-links and not the actual images from which many of the links were broken. Thus, we had to consider only 50 subjects with 10 samples each. The average age of considered subjects is 43.8 years with 23 and 73 as minimum and maximum age. Fig. 10.2 shows the frontal face images of subjects with weight and age variations.

## 10.4 Proposed scheme

In this section, we present the proposed multi-biometric score level fusion scheme based on the fusion of comparison scores obtained using four different feature extraction techniques on multiple biometric characteristics. Fig.10.4 shows the block diagram of the proposed scheme. It mainly consists of following steps:

| Database | No of Subjects | Labeling | | | Total no of images |
|---|---|---|---|---|---|
| | | Thin | Moderate | Heavy | |
| WIT | 110 | 527 | 448 | 124 | 1109 |
| eWIT | 200 | 437 | 1309 | 290 | 2036 |

Table 10.1: Details of WIT and eWIT database

1. ROI Detection and Segmentation
2. Feature Extraction
3. Sparse Representation Classifier
4. Score Level Fusion Scheme.

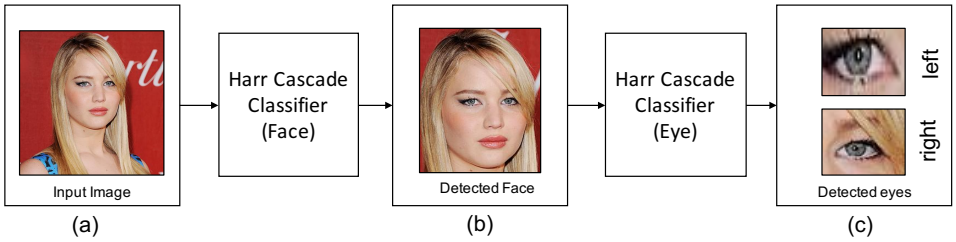



Figure 10.3: ROI Extraction: (a) illustrates the input image (b) is the detected face image (c) detected left and right eye images.

The following subsections describe each component of the proposed scheme. The input image is preprocessed to get the region of interest, and then four different feature extraction algorithms: Histogram of Oriented Gradients (HOG), Local Binary Patterns (LPB), Local Phase Quantization (LPQ) and Log-Gabor Filters (LGF) are used to generate the respective feature vectors. The generated feature vectors are then given as input to SRC *(Sparse Representation Classifier)* for obtaining the comparison score for the recognition process.

### 10.4.1 ROI Detection And Extraction

The segmentation of the images is crucial part of this scheme as the images are taken from the internet and captured in an unconstrained environment. The proposed scheme extracts three segmented parts from the input image:

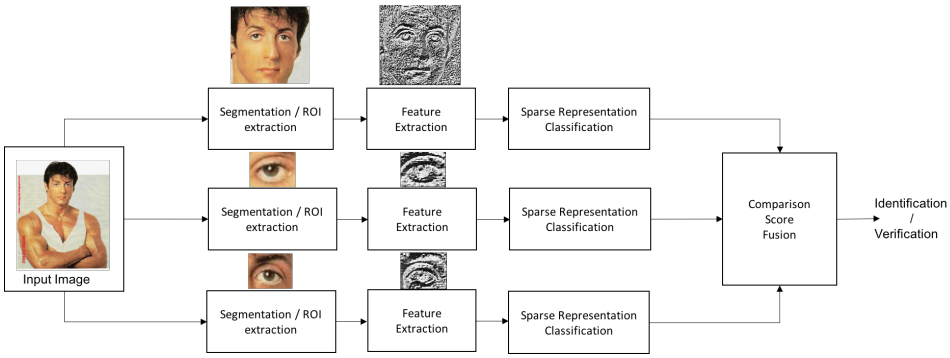



Figure 10.4: Block diagram of proposed multibiometric score level fusion approach

(1) Face Region (2) Left Eye Region and (3) Right Eye Region. Given the input image $I$, we first perform the face detection and segmentation using the Viola-Jones algorithm [187], which is extremely efficient for rapid object detection and it is widely used. The segmented face image is further given as input to the Viola-Jones algorithm trained for eye detection to localize the left and right eye region. Fig. 10.3 gives the overview of the segmentation process. Fig. 10.3(a) illustrates the input image, Fig. 10.3(b) is the detected face image and the Fig. 10.3(c) shows the detected eye regions.

### 10.4.2 Feature Extraction

Feature extraction is employed using four techniques to evaluate the best performing feature extraction algorithm to work with proposed fusion scheme. Further, the best algorithm is utilized in the final fusion process. These techniques are discussed in the subsequent sections:

#### 10.4.2.1 Histogram of Oriented Gradients (HOG)

The appearance and shape of the local object can be well described by the distribution of local intensity variations i.e. intensity gradients or edge directions. This can be achieved by dividing the image into small spatial regions (cells) in rectangles or radial. Where, each cell is accumulating a weighted local 1-D histogram of gradient directions over the pixels of the cell.

#### 10.4.2.2 Local Binary Patterns (LBP)

The LBP operator is best suited for texture description. The image is divided into several regions from which the LBP feature distributions are extracted and accumulated into an enhanced feature vector, which can be

used an object descriptor. The operator assigns a binary 0 or 1 by comparing the center pixel with neighboring pixels. If the center pixel's value is greater than the neighboring pixel value, it assigns 0 otherwise 1. Finally, the histogram of the accumulated labels can be used as a texture descriptor [122].

#### 10.4.2.3 Local Phase Quantization (LPQ)

The Local Phase Quantization (LPQ) method is mainly based on the blur invariance property of the Fourier phase spectrum of the image. The LPQ operator has to be applied to every pixel of the image can be used to identify the textures. The resultant codes are accumulated in a histogram. Codes and histograms are generated in same manner as for the LBP [122] operation [123].

#### 10.4.2.4 Log-Gabor Filters (LGF)

The Log-Gabor filter represents a signal in terms of the local frequency responses. Feature vectors derived using Log-Gabor filters can be very useful in these type of application as it contains the high information of the local frequencies. In [152], the author has successfully used Log-Gabor filters for facial expression recognition.

### 10.4.3 Sparse Representation Classifier

We consider each image as $I(x, y)$, where all images are labeled to their respective $p$ class. The labeled training samples from $P|_{\mathrm{c}}$ distinct classes will be used to determine the class of given test sample correctly. We have considered 50 subjects with 9 samples of each for training purpose. We arrange the features of given training samples from $i^{\mathrm{th}}$ class as the columns of matrix A and can be defined as:
$A_i = [v_{i,1}, v_{i,2}, \ldots, v_{i,n_i}]$
where, $\mathrm{n_i} = 9$ i.e no. of samples, A is the dictionary of atoms and $\mathrm{A} \in \mathrm{R}^{\mathrm{mxn_i}}$, $v_i$ is feature vector or atoms in the sparse space of $\mathrm{i}^{\mathrm{th}}$ training sample, $\mathrm{v} \in \mathrm{R}^{\mathrm{m}}$ and m = size of the feature vector. Thus, in general we can determine the matrix A as:
$A = [A_1, A_2, \ldots, A_n] = [v_{1,1}, v_{1,2}, \ldots, v_{k,n_k}]$ for all the subjects. The size of a feature vector varies with feature extraction technique but one can consider whole the image as a feature vector by stacking its columns together. In [9] the authors have described that the images of the same class lie in a linear span of the training samples. Then, any test sample $\mathrm{y} \in \mathrm{R}^{\mathrm{m}}$ from the class $i$

can be represented as linear combination of the feature vectors of class i as:

$$y = a_{i,1}v_{i,1} + a_{i,2}v_{i,2} + \cdots + a_{i,n_i}v_{i,n_i} \tag{10.1}$$

From the above equation, $y$ can be defined as the linear system given by Eq. (10.2)

$$y = Ax_r \in \mathrm{R} \tag{10.2}$$

Where, $x_r = [0, \ldots, 0, a_{i,1}, a_{i,2}, \ldots, a_{i,n_i}, 0, \ldots, 0]^{\mathrm{T}} \in R^{\mathrm{n}}$ is a coefficient vector whose entries are zero all over except the ones associated with the $i^{\mathrm{th}}$ class. Hence for $i^{\mathrm{th}}$ class $x_r$ becomes:

$$x_r = [a_1, a_2, \ldots, a_n]^T \tag{10.3}$$

As the entries of the vector $x_r$ contain the identity of the test sample $y$, it can be obtained by solving the linear system of equations $y = Ax_r$. The conventional techniques to solve the linear system will not be helpful as the number of features $m$ need not to be equal to number of samples $n$. In case of $m > n$, the system will be overdetermined and in case of $m < n$ it is underdetermined, in both cases the system will not have a unique solution. Conventionally, this type of problem is resolved by choosing the minimization $\ell$-norm solution. Using the proposed technique in [40] we can consider the system as a $\ell_1$ optimization problem to obtain the sparse vector $x_r$. Eq. (10.4) describes the $\ell_1$ optimization problem:

$$x_{\ell_1} = \arg_x \min \|x\|_1, \ \textit{subject to}\ Ax = y \tag{10.4}$$

Where, $\|\cdot\|_{\ell_1}$ is the $\ell_1$-norm indicating the sum of non-zero coefficients of the operand vector. For a given test sample $y$ the using Eq. (10.2) and (10.4), the identity can be found as :

$$S(y) = \arg\max_i \frac{\|\delta_i(\hat{x}_{\ell_1})\|_1}{\|x_{\ell_1}\|_1} \tag{10.5}$$

Where, $S(y)$ is the classified label for the given sample $y$ and $\delta_i(\hat{x}_{\ell_1})$ is characteristic function that selects non zero coefficient from solution vector $x_r$ of the i$^{\mathrm{th}}$ class.

### 10.4.4 Score Level Fusion Scheme

As defined in ISO/IEC TR 24722[71] a fusion of different biometric information can be performed at various stages such as feature level, score level, and decision level fusion on multimodal and other multi-biometric fusion. This section of the paper describes the proposed score level fusion technique to improve the recognition performance. The paper proposes three fusion schemes:

1. Multi-biometric Fusion
2. Multi-algorithmic Fusion
3. Multi-algorithmic and Multi-biometric Fusion

Following subsections describe each of the proposed fusion scheme in details.

#### 10.4.4.1 Multibiometric Fusion

In this scheme, we extract the features from any given sample using four different feature extraction techniques which are explained in Subsection *Feature Extraction*. The extracted features are given as input to the Sparse Representation Classifier (SRC) to get the comparison scores. Further, comparison scores of face, left eye, and right eye images are fused to formulate the final decision. Eq. (10.6) represents the multi-biometric fusion score:

$$S_{fuse} = f(S_{face}, S_{left}, S_{right}) \tag{10.6}$$

Where $f$ is the fusion rule, $S_{fuse}$ represents the fused score, $S_{face}$, $S_{left}$, and $S_{right}$ are the comparison scores obtained using one of the above mentioned feature extraction technique.

#### 10.4.4.2 Multi-algorithmic Fusion

Here, the fusion score is determined using the comparison scores obtained by application of different feature extraction algorithms on multiple biometrics. Eq. (10.7) illustrates the multi-algorithm score level fusion:

$$S_{fuse} = f(S_{fe_1(face)}, S_{fe_2(left)}, S_{fe_3(right)}) \tag{10.7}$$

Where, $f$ is the fusion rule, $S_{fuse}$ represents the fused score, $S_{fe_1(face)}$, $S_{fe_2(face)}$ and $S_{fe_3(face)}$ are the comparison scores obtained using $fe_1$, $fe_2$ and $fe_3$ feature extraction algorithms.

#### 10.4.4.3 Multi-algorithmic and Multi-biometric Fusion

The feature extractors are applied on all biometric characteristics i.e., holistic face sample, left eye sample and right eye sample. Then the comparison scores are obtained using SRC classifier. The scores from the best-performing algorithms for face, left and right eye samples are fused together to get the final similarity scores. Eq. (10.8) illustrates the multi-algorithm and multi-biometric score level fusion:

$$S_{Fuse} = f((S_{Face})_{best}, (S_{left})_{best}, (S_{right})_{best}) \tag{10.8}$$

Where $f$ is the fusion rule, $S_{fuse}$ represents the fused score, $(S_{face})_{best}$, $(S_{left})_{best}$ and $(S_{right})_{best}$ are the comparison scores obtained using the best performing feature extraction technique for the respective biometric instance.

#### 10.4.4.4 Fusion Rules

Following are the fusion rules applied on the comparison scores:

1. Weighted Summation:
   Fusion score is obtained using weighted summation of the input comparison scores

$$S_{fuse} = \frac{\omega_1 * (S_{face}) + \omega_1 * (S_{left}) + \omega_1 * (S_{right})}{3} \tag{10.9}$$

   Where $\omega_1, \omega_2$ and $\omega_3$ are the weights, such that $\omega_1 > \omega_2 > \omega_3$ or $\omega_1 > \omega_3 > \omega_2$

2. Summation:
   Fusion score is obtained using summation of the input comparison scores

$$S_{fuse} = \frac{S_{face} + S_{left} + S_{right}}{3} \tag{10.10}$$

3. Min:
   Fusion score is obtained using minimum of input comparison scores

$$S_{fuse} = min(S_{face}, S_{left}, S_{right}) \tag{10.11}$$

4. Max:
   Fusion score is obtained using maximum of input comparison scores

$$S_{fuse} = max(S_{face}, S_{left}, S_{right}) \tag{10.12}$$

5. Product:
   Fuse score is obtained by multiplication of input comparison scores

$$S_{fuse} = \frac{S_{face} * S_{left} * S_{right}}{3} \tag{10.13}$$

Where $S_{fuse}$ represents the fused score, $S_{face}, S_{left}$ and $S_{right}$ are the comparison scores for face, left eye and right eye biometric.

| **Instances** | **EER %** | | | | |
|---|---|---|---|---|---|
| | **HOG** | **LBP** | **LPQ** | **LG** | **Neurotech** |
| Face | 10.02 | 18.08 | 22.00 | 16.00 | 13.91 |
| Left Eye | 17.81 | 28.02 | 26.02 | 19.83 | NA |
| Right Eye | 20.02 | 24.00 | 24.00 | 16.28 | NA |

Table 10.2: EER values of all feature extraction algorithms for individual biometric instances

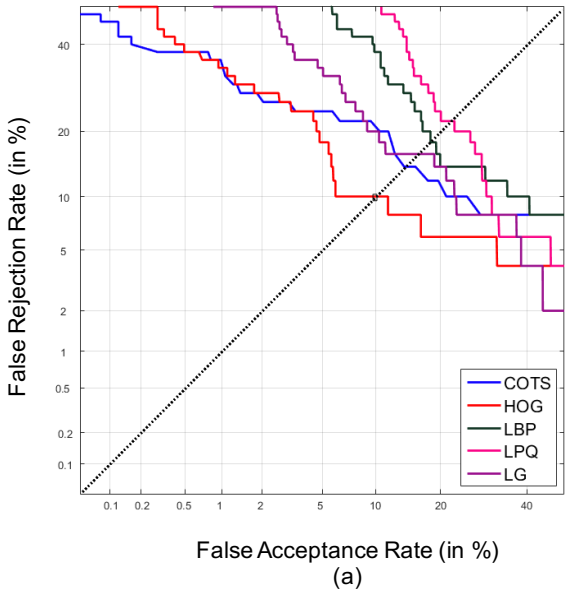



Figure 10.5: DET curves for face with different FE techniques and COTS

From Table 10.3 it is clear that the unimodal multi-algorithmic fusion does not give better results, this motivates us to perform the multi-modal fusion scheme to improve the results. In Table 10.3 and Table 11.3 we consider, F = face image, L = left eye image, R = right eye image, M-ALGO = Multi-algorithmic and M-BIO = Multi-biometric.

## 10.5 Experiments & Results

This section details the experiments and quantitative results obtained using proposed fusion scheme on the eWIT database images. The experiments

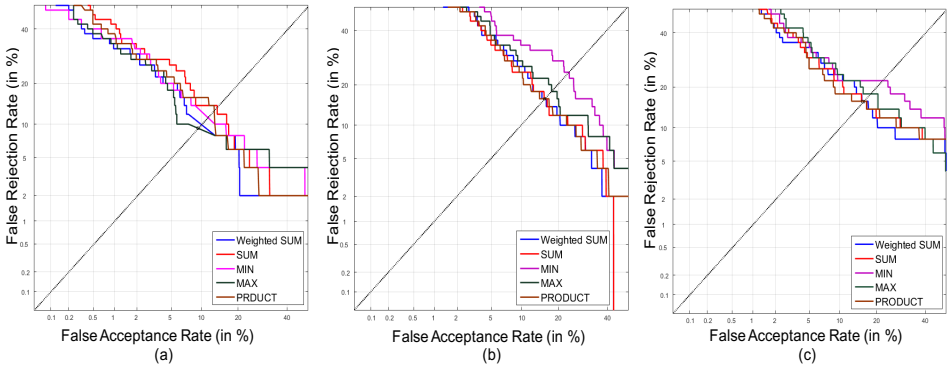



Figure 10.6: DET curves for unimodal multi-algorithmic fusion scheme: a) DET for face b) DET for left eye and c) DET for right eye

| **FE Algorithm and Biometric Fusion** | **Fusion Rule (EER % Values)** | | | | |
|---|---|---|---|---|---|
| | **Weighted Sum** | **Sum** | **Min** | **Max** | **Product** |
| M-ALGO F | 10.02 | 13.87 | 12.00 | 10.00 | 13.46 |
| M-ALGO L | 15.97 | 16.00 | 23.91 | 18.00 | 16.00 |
| M-ALGO R | 16.00 | 16.00 | 22.00 | 18.00 | 15.97 |

Table 10.3: Unimodal score level fusion

were conducted on 50 subjects with 10 samples each. For all 50 subjects face and eye, detection was performed using Viola-Jones object detection algorithm [187]. Feature extraction was carried out using four different techniques, and the classification is accomplished using sparse representation classifier (SRC). The experiments resulted in 50 genuine comparisons and 2450 impostor comparisons. This paper presents the results in terms of Equal Error Rate (ERR), False Acceptance Rate (FAR) and False Rejection Rate (FRR). The lower values of EER signifies better performance of the biometric system. Further, we present the Detection Error Trade-off (DET) Curves as a performance report plotted as FRR versus FAR values.

Table 10.2 tabulates the quantitative performance of individual biometric with different feature extraction algorithms benchmarked with a commercial-off-the-shelf-system (COTS) from Neurotech VeryLook 5.7 SDK. Please refer here [2] for more information about the VeryLook SDK. From Table 10.2, it can be perceived that the HOG feature extraction algorithm with sparse representation classifier for face biometric outperforms the commercial FRS with an EER = 10.02% where commercial FRS has EER = 13.91%.

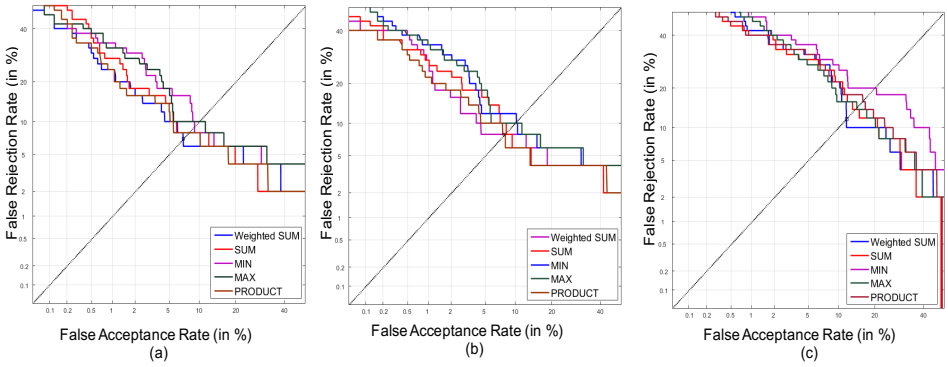



Figure 10.7: DET curves for multi-modal and multi-algorithmic fusion scheme: a)DET for face, left and right eye score fusion using HOG and Log-Gabor as feature extractors (multi-algorithms) b)DET for face, left and right eye score fusion using HOG as feature extractors (single algorithm) c)DET for left and right eye score fusion using Log-Gabor as feature extractors(single algorithm)

Further, from the table it can be understood that EER values for left and right eye are not so competent since the images were taken in unconstrained environment and in many of the images eyes are occluded due subject's hair (Ref. Fig. 10.8) illustrates this.

Table 11.3 indicates the quantitative performance of the proposed scheme along with different score level fusion case-studies. Further, the overall performance and results of the proposed system can be summarized in following points:

- Multi-biometric and multi-algorithmic fusion scheme with weighted summation rule, gave the best result with EER = 7.48%.
- From Table 11.3, it is evident that the performance of the proposed system is improved by 6.43% when compared with COTS.
- Another interesting thing can be observed from Table 11.3, the fusion of comparison score of the left and right eye with Log-Gabor feature extractor gave EER = 11.91% which is lesser that the COTS.
- The overall performance of the system was improved when HOG or Log-Gabor filters were used as a feature extractor. We have obtained 5.94% and 3.94% improvement when used HOG and Log-Gabor feature extractors respectively.

Figure 10.8: Occlusion due to hair

| FE Algorithm and Fused Biometrics | Fusion Rule (EER % Values) | | | | |
|---|---|---|---|---|---|
| | Weighted Sum | Sum | Min | Max | Product |
| (F,L,R) HOG | 8.00 | 8.12 | 10.14 | 10.0 | 7.97 |
| (F,L,R) LBP | 15.97 | 16.0 | 23.87 | 16.0 | 17.81 |
| (F,L,R) LPQ | 18.44 | 16.38 | 20.00 | 19.57 | 15.97 |
| (F,L,R) LGAB | 10.00 | 12.00 | 17.91 | 15.89 | 12.00 |
| (L,R) LGAB | 11.91 | 14.00 | 20.02 | 14.26 | 16.12 |
| M-BIO &M-ALGO | 7.48 | 8.00 | 9.53 | 10.00 | 8.00 |

Table 10.4: Multimodal score level fusion

- The average EER obtained using HOG and Log-Gabor with all case studies for weighted summation rule is 9.34%.

- The average EER obtained using LBP and LPQ with multiple biometrics for weighted summation rule is 17.20% which is not good as HOG and Log-Gabor.

The following figures show the Detection Error Trade-off (DET) Curves as a performance evaluation report for the proposed fusion scheme. From Fig. 10.6.b and 10.6.c, it can be interpreted that the proposed fusion scheme does not perform well with for unimodal (with only left or right eye ) biometric. Where the Fig. 10.7.a has the lowest FAR and FRR near the origin hence it is the best system. Further, the Fig. 10.5 represents the comparison of VeriLook 5.4 SDK COTS [2] with different feature extraction algorithms for face biometric also we can see that HOG feature extraction algorithm with sparse representation classifier outperforms COTS.

## 10.6 Conclusion

The performance of biometric recognition drops when the subject gains weight over a period of the time. In order to mitigate the effect due to the variations in body weight, in this work, we explore multi-biometric characteristics by employing both the holistic face and the periocular region. This work shows that a multi-biometric score level fusion approach can significantly improve the performance of face recognition systems for subjects with body weight variations. Use of multiple biometric characteristics and algorithms results in lower EER when compared with the COTS face recognition system. The proposed scheme improved the recognition performance by 6.42%. Thus, this study shows that the use of the eye region coupled with face can mitigate the effects of the body weight variations for face based biometrics. Future works shall include fusion of different biometric information at various stages such as data level, feature level and, decision level to improve the robustness of system to adapt it in real life scenarios.

## Acknowledgment

This work was carried out under the funding of the Research Council of Norway (Grant No. IKTPLUSS 248030/O70).

# Chapter 11

# Article 6: Subjective Logic Based Score Level Fusion: Combining Faces And Fingerprints

Pankaj Wasnik, Raghavendra Ramachandra, Kiran Raja, and Christoph Busch. *"Subjective Logic Based Score Level Fusion: Combining Faces and Fingerprints."* In the $21^{st}$ International Conference On Information Fusion (FUSION 2018), pp. 515-520. IEEE, 2018.

## 11.1 Abstract

Biometric systems are prone to random and systematic errors which are typically attributed to the variations in terms of inter-session data capture and intra-session variability. Furthermore, these errors cannot be defined and modeled mathematically in many cases, but we can associate them with uncertainty based on certain conditions. In such cases, one of the possible approach to improve biometric system performance is to employ multi-biometric fusion by incorporating the uncertainties. In the literature, researchers have proposed many fusion techniques, but most of these techniques do not take uncertainty into account while performing fusion. Since the decision made by uni-modal biometric comparators do not consider the uncertainty involved in such decisions, it is essential first to model the uncertainty before combining the decision from multiple uni-modal biometric

systems efficiently. To this end, we propose a score level multi-biometric fusion scheme using Subjective Logic which incorporates the uncertainty of the system's information channels while fusing the scores. Extensive experiments are carried out on the multi-biometric NIST BSSR1, and the proposed scheme has indicated a superior performance with a genuine match rate of 99.02% at a false match rate fixed to 0.01%.

## 11.2 Introduction

The biometric systems are often affected due to random and systematic errors [70]. Usually, it is difficult to quantify such type of errors; this introduces an uncertainty in the output recognition score. In literature, researchers have shown that biometric fusion can resolve such problems and improve the recognition performance. Further, many of the limitations of unimodal biometric systems are addressed by multi-modal systems and even shows improved robustness against Failure-To-Enrol rate (FTE) [154].

In the earlier work, many biometric fusion methods have shown promising performance, but very few of them consider an uncertainty while performing the fusion [121]. Many researchers have proposed methods based on Dempster Shafer Theory (DST) to incorporate factors like uncertainty by combining the evidence from multiple sources however they do not model the uncertainty appropriately [185, 168, 7]. Furthermore, in a few other approaches, authors do not propose a method to model the uncertainty, and few of them even consider the uncertainty equal to zero [168, 185, 106, 92, 93]. In [148, 3] authors have considered the uncertainty mass equal to the complement of the genuine scores, while in [7, 113] authors consider the uncertainty as some constant value. However, in [121] authors have proposed a method to model the uncertainty in terms of quality and EER. They have used DST for combining the scores from multiple biometric information channels. However, the DST has only one fusion rule to combine the evidence [163]. This motivates us to explore the theories like Subjective Logic which formalizes the probabilities and uncertainties associated with them as well as provides different fusion rules [79, 81].

Subjective logic (SL) overcomes the limits of probabilistic calculus and binary logic by incorporating subjectivity and uncertainty. It provides an opinion about a proposition by combining the belief, uncertainty mass and prior knowledge about the proposition [79]. Subjective logic fusion (SLF) operators mainly operate on the subjective opinions which are the function of belief mass supporting the proposition, associated uncertainty mass and the prior knowledge about the proposition [81]. The application of SL in

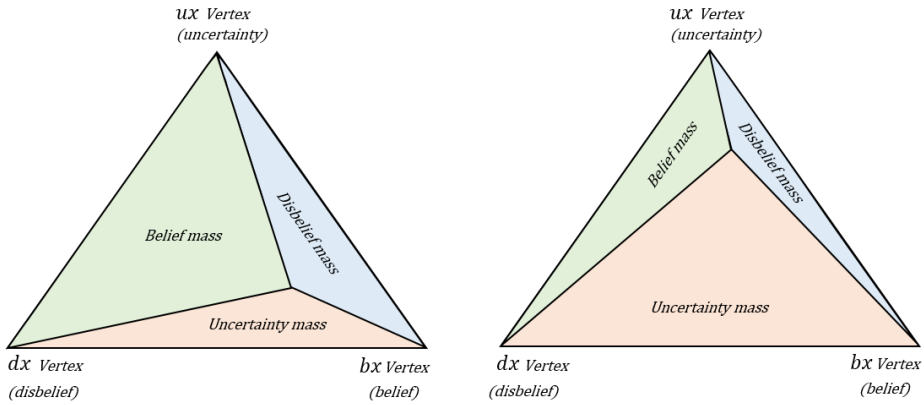



Figure 11.1: The Barycentric representation of input biometric score as a subjective opinion with low and high uncertainty masses.

the domain of biometrics was first introduced in [82], where an overview of the underlying theory of different SLF operators is explained concerning biometric fusion. Further, in [193], authors have proposed a method to combine the outputs of the various classifiers using subjective logic for smartphone-based gait recognition.

In this paper, we propose a fusion scheme based on subjective logic which fuses the comparison scores from different biometric modalities such as the face, with left and right index fingers using cumulative SLF operator. The extensive experiments are carried out on the publicly available NIST BSSR1 database* to evaluate the performance of the proposed scheme. The rest of the paper is divided into the sections describing the proposed fusion scheme, database, and experiments, results, and a discussion followed by the conclusions.

## 11.3 Proposed Fusion Scheme

This section describes an overview of the proposed fusion scheme using subjective logic. Figure 11.2 illustrates the block-diagram of the steps involved in the proposed scheme. The binomial subjective opinions can well define the verification systems since they usually give two output states, i.e., either the subject is genuine or an impostor. Therefore consider a random variable

*The BSSR1 database can be availed from: https://www.nist.gov/itl/iad/image-group/nist-biometric-scores-set-bssr1

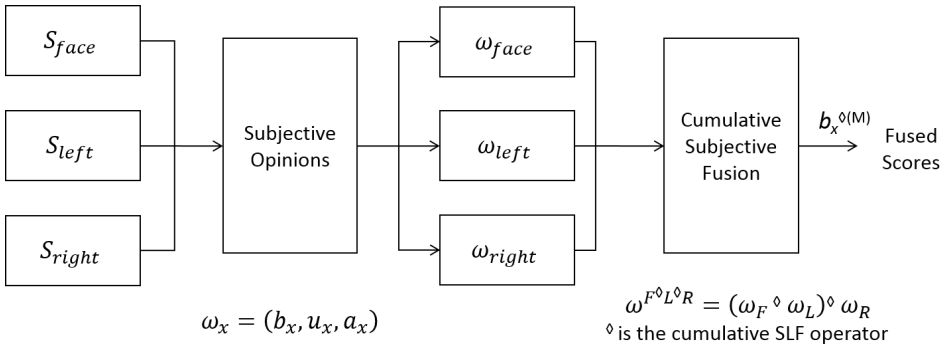



Figure 11.2: Proposed fusion scheme based on cumulative subjective fusion operation.

$X$ over the domain $\mathbb{X} = \{x_1, x_2\} \in \mathbb{R}$ where $\mathscr{R}(\mathbb{X})$ is reduced powerset of $\mathbb{X}$ such that all proper subsets of $\mathbb{X}$ are the element of $\mathscr{R}(\mathbb{X})$.

More specifically, let $X = \{x, \bar{x}\}$ be the binomial opinion which represents the verification scores for the input sample, where $x$ represents a user being genuine and $\bar{x}$ that user being an impostor. Thus for each biometric instance $M$, we can represent a subjective opinion $\omega^M$ by a tuple $\omega_X^M = (b_X^M, d_X^M, u_X^M, a_X^M)$, where $b_X^M, d_X^M, u_X^M, a_X^M$ are the belief, disbelief, uncertainty and base rate mass for the input scores. And $M$ denotes the biometric instance, in this case it is face, left finger or right finger. For any given subjective opinion, Equation 11.1 is always true [81] under the following condition:

$$b_x + d_x + u_x = 1 \tag{11.1}$$

For binomial opinions, the projected probability of $x$ can be expressed as defined by Equation 11.2 which is the expected output value for any given input verification score.

$$P(x) = b_x + u_x a_x \tag{11.2}$$

The normalized verification scores are considered as belief mass $b_x$, as is almost equal to the probability of the sample belonging to the genuine or impostor class. The uncertainty of the system is obtained by quantifying the errors of the verification system. We can define Equation 11.3 using all the mislabeled samples to get the uncertainty of the system.

$$u_x = \frac{\sum_{i=1}^{n} x_i^2 + \hat{p}^2 + \sum_{i=1}^{n} y_i^2}{\frac{n-1}{n} \sum_{i=1}^{n} y_i}, \qquad (11.3)$$

$$where\ \hat{p} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i}$$

where $n$ is the number of test subjects, $y_i$ is the number of attempts by $i^{th}$ subject, $x_i$ is the number of false-non matches for $i^{th}$ subject and $\hat{p}$ is the observed false-non match rate. The term $a_x$ is nothing but the base probability and can be expressed as the classification accuracy of the verification system. Thus, every verification score will produce a subjective opinion i.e. $S_i \rightarrow \omega_i$. In this paper, we have used three biometric instances (i.e. information channels) the face with left and right index finger.

We then convert comparison scores of these information channels (i.e. sources) to individual subjective opinions based on the theory as mentioned above (See Equation 11.1-11.3). In subjective logic, there are several fusion operators based on the scenario one can choose which operator to be used. In this paper, we have focused on only one SLF operator called Cumulative Subjective Fusion operator.

### 11.3.1 Cumulative Fusion:

The cumulative belief fusion is used when we want to increase an evidence for any hypothesis by adding the evidences from different sources. In our case, we have three different sources such as Face, Left, and Right Finger. The discriminant frame of sources can be defined by $\mathbb{M} = \{F, L, R\}$. Let $M \in \mathbb{M}$ denote be the specific biometric instance, and let $\omega_X^M$ denote its opinion about the variable $X$, hence we can assume that all of the three sources produce independent opinions about the variable $X$. Consider the cumulative fusion operator be the symbol $'\diamond'$ and the cumulative fused opinion is denoted as $\diamond(M) = \omega_X^{\diamond(M)}$ and can be expressed by the following equation:

$$\begin{cases} \boldsymbol{b}_X^{\diamond(\mathbb{M})}(x) = \dfrac{\sum\limits_{M\in\mathbb{M}}\left(\boldsymbol{b}_X^M(x)\prod\limits_{M_j\neq M} u_X^{M_j}\right)}{\sum\limits_{M\in\mathbb{M}}\left(\prod\limits_{M_j\neq M} u_X^{M_j}\right)-(N-1)\prod\limits_{M\in\mathbb{M}} u_X^M}, \\ u_X^{\diamond(\mathbb{M})} = \dfrac{\prod\limits_{M\in\mathbb{M}} u_X^M}{\sum\limits_{M\in\mathbb{M}}\left(\prod\limits_{M_j\neq M} u_X^{M_j}\right)-(N-1)\prod\limits_{c\in\mathbb{M}} u_X^M}, \end{cases} \tag{11.4}$$

where $\exists u_X^M \neq 0$ and for $\exists u_X^M = 0$, the fusion can be obtained as given below,

$$\begin{cases} \boldsymbol{b}_X^{\diamond(\mathbb{M})}(x) = \sum\limits_{M\in\mathbb{M}} \gamma_X^M \boldsymbol{b}_X^M(x), \\ u_X^{\diamond(\mathbb{M})} = 0 \end{cases}$$

$$\text{where } \gamma_X^M = \lim_{u_X^{\mathbb{M}}\to 0} \frac{u_X^M}{\sum\limits_{M_j\in\mathbb{M}} u_X^{M_j}}$$

where, $\boldsymbol{b}_X^{\diamond(\mathbb{M})}, u_X^{\diamond(\mathbb{M})}$ are the fused belief and uncertainty which defines the fused subjective opinion $\omega_X^{\diamond(M)}$. We further obtain the final result by applying threshold $t$ to the $\boldsymbol{b}_X^{\diamond(\mathbb{M})}$ as:

$$Result = \begin{cases} Accept & \boldsymbol{b}_X^{\diamond(\mathbb{M})} > t \\ Reject & \boldsymbol{b}_X^{\diamond(\mathbb{M})} < t \end{cases} \tag{11.5}$$

The threshold $t$ in the Equation 11.5, is obtained from the validation dataset (See Section III) and applied on the testing dataset.

Additionally, we perform the biometric fusion using standard fusion rules such as sum rule, weighted sum rule and product rule (with veto-power of each contributing information channel) to evaluated the performance on the proposed fusion scheme. The standard fusion rules are given by following equation:

$$Fused\ Sum\ Score = \sum_{i=1}^{n} S_i/n \tag{11.6}$$

$$Fused\ Weighted\ Sum\ Score = \sum_{i=1}^{n} w_i S_i / n \tag{11.7}$$

where $w_i$ is the weight given to the $i^{th}$ biometric instance. We further assign higher weights to the better performing biometric instance as per the baseline EER calculation i.e., $w_1 > w_2 \cdots > w_n$.

$$Fused\ Product\ Score = \prod_{i=1}^{n} S_i / n \tag{11.8}$$

A benchmark of the SLF approach with other methods such as fuzzy logic, Bayesian reasoning, and DST will make a better comparison than standard fusion rules. However, in this paper we limit our work to compare the SLF with standard fusion methods only.

## 11.4 Database and Experiments

This section describes the database in details along with the overview of the experiments that we carried out. The performance of the proposed scheme is validated using a subset of the well-known benchmark and publicly available NIST Biometric Scores Set (BSSR1). The database mainly consists of the verification scores of the face, left and right index finger obtained from 517 individuals. The face scores consist of two sets obtained using commercial comparators C and G whereas the fingerprint comparison scores were generated using only one commercial fingerprint recognition system. Table 11.1 tabulates the statistics of NIST BSSR1 score database. In total, our score database consists of four such sets of 517 genuine and 266772 impostor scores. Based on the number of subjects, we divide the database into three sets as development, validation, and testing set. The development set consists of 311 subjects whereas validation, and testing set consists of 103 subjects each. The development dataset is mainly used to estimate the uncertainty associated with the scores whereas validation dataset is used to obtain the operating threshold of the system. The paper presents the results obtained by applying the operating threshold on the testing dataset.

In the next section, we compare the results of both experiments to test the effectiveness of the proposed fusion scheme. We have carefully designed the fusion strategies to analyze the behavior of the fusion operations (See Table 11.2). Each strategy shows the modalities and instances used in the

| **Datasets (BSSR1 Scores)** | **No of subjects** | **Genuine Scores** | **Impostor Scores** |
|---|---|---|---|
| Face Comp C | 517 | 517 | 266772 |
| Face Comp G | 517 | 517 | 266772 |
| Left Finger | 517 | 517 | 266772 |
| Right Finger | 517 | 517 | 266772 |

Table 11.1: Statistics of the NIST BSSR1 scores database

| **Strategy** | **Biometric modalities** |
|---|---|
| S1 | Face and Left Finger(FL) |
| S2 | Face and Right Finger(FR) |
| S3 | Left and Right Finger(LR) |
| S4 | Face, Left and Right Finger(FLR) |

Table 11.2: Fusion strategies applied in case of both standard and subjective logic fusion approaches

fusion operation. For both fusion techniques, i.e., standard fusion rules and subjective logic fusion, we have used all of the four strategies given in Table 11.2.

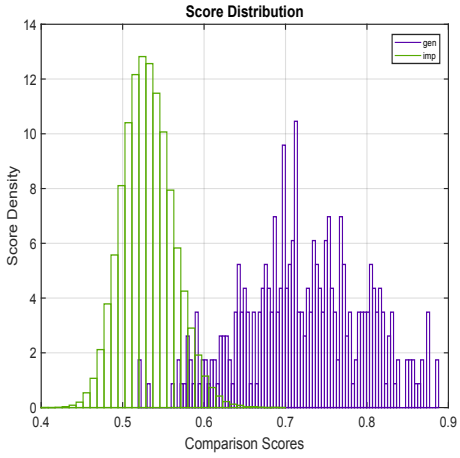
| **Dataset** | **% EER** | **% GMR** |
|---|---|---|
| Face Comparator C | 4.36 | 83.60 |
| Face Comparator G | 5.80 | 77.49 |
| Left Index Finger | 8.51 | 85.53 |
| Right Index Finger | 5.04 | 90.35 |

Table 11.3: Baseline performance for all four biometric instances. Here, GMR is calculated @ FMR = $10^{-3}$

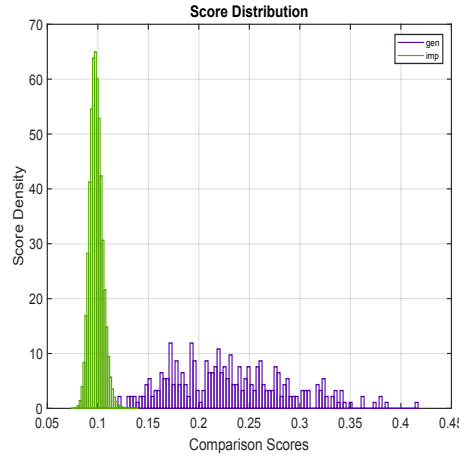
## 11.5 Results and Discussion

This section presents the obtained results from the experiments that we carried out. We present our results in terms of Detection Error Trade-off (DET) curve, Receiver Operating Curves (ROC) and Equal Error Rate (EER%), and Genuine Match Rate. We also analyze the score distributions of baselines, standard and subjective logic fusion. Figure 11.3, presents the score distribution for face comparator C and fusion strategy S4. From the Figure 11.3, it is apparent that the genuine and impostor score distributions

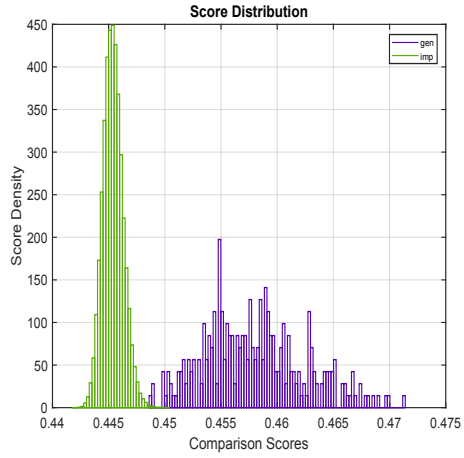
for subjective logic and weighted sum fusion rule show low overlapping area, whereas the baseline system has a significant overlapping area.




a Baseline systems



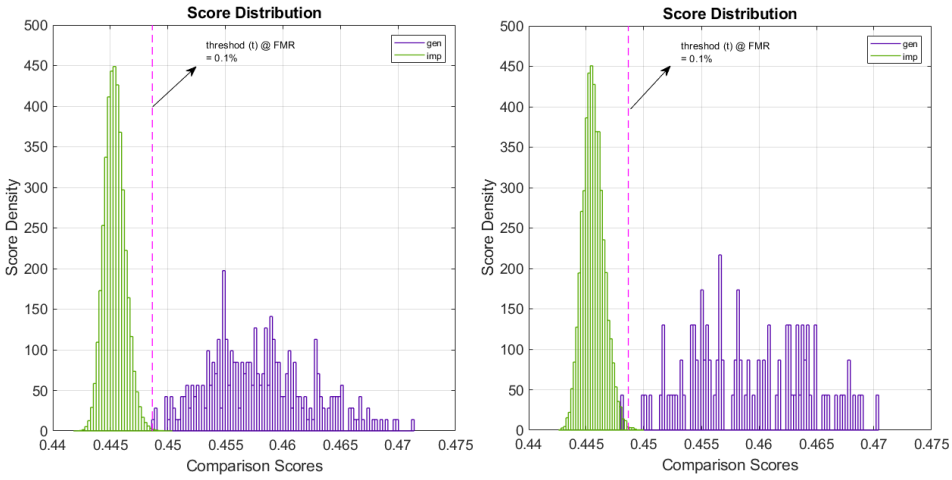

b Weighted sum fusion




c Subjective logic fusion

Figure 11.3: Score distribution for face comparator C and fusion strategy S4 (See Table 11.2).

The Figure 11.4, shows the score distributions for validation and testing datasets for face comparator C and fusion strategy S4. The pink dotted line in the figure shows the threshold obtained @ FMR $= 10^{-3}$ for validation dataset and the same threshold is applied whenever there is a new testing sample. From the Figure 11.4b, we can see the score distributions after application of threshold $t$ on testing dataset. It is evident that the $t$ is




a Validation Scores

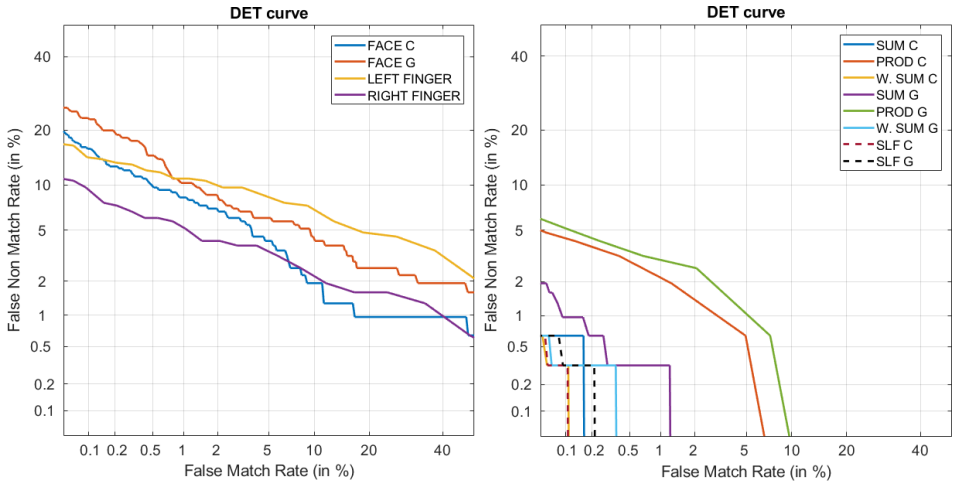
b Testing Scores

Figure 11.4: Score distributions for validation and testing datasets for face comparator C and fusion strategy S4, here the threshold (t) is determined on validation dataset where FMR = $10^{-3}$

a well-generalized threshold since there are very few genuine scores miss-classified as impostor scores.

Table 11.3 presents the performance of the baseline systems in terms of EER and GMR @ FMR = $10^{-3}$. The EER of face recognition system with comparator C gives the best baseline result of 4.46%, and similar observation can be drawn from the Figure 11.5a. However, in terms of GMR, the fingerprint baseline system with right index finger data shows the best performance of GMR of 90.35%.

The operating threshold for classifying the subject as genuine or impostor is obtained on the validation dataset. Hence it is imperative to analyze the validation performance of the system. The DET curve for both of the fusion approaches for fusion strategy S4 is given in the Figure 11.5b. The DET plots of proposed and weighted summation fusion are highly overlapping resulting approximately similar performance in terms of an EER. Furthermore, by comparing Figure 11.5a and 11.5b, we can say that the performance of the system has improved significantly in terms of EER. The ROC plot is given in Figure 11.6, it also validates this behavior of the system. From the figure, it is evident that the GMR for proposed system with the face comparator C achieved the highest performance in terms of GMR of 99.05% at FMR = $10^{-3}$.




a Baseline systems

b Fusion approaches

Figure 11.5: DET curves for baselines systems , SLF and standard fusion techniques for validation dataset

The evaluation performance of our proposed fusion scheme is presented in the Table 11.4. The GMR values in Table 11.4 are obtained by applying the operating threshold on the testing dataset. In case of standard fusion methods, for both face comparators, the weighted averaging of scores of the face, left and right index fingers achieved the highest performance of GMR of 99.02%. Further, it is evident that not only fusion strategy S4 where all biometric modalities are used shows better performance but also the fusion of face and left index finger (S1) has shown the higher performance in both fusion cases. Furthermore, the fusion strategy S3 has shown lowest performance with the average GMR of 92.95%. Also, the product fusion rule has achieved the lowest performance than other fusion rules.

## 11.6 Conclusion

From the experiments and results reported in this paper, we conclude that the proposed fusion scheme based on subjective logic outperforms all of the baselines, standard fusion techniques except for fusion strategy S2 for face comparator G. Though the weighted averaging of scores has shown nearly equal performance as that of the proposed system. However, these methods do not take care of the irregularities of the system, moreover finding the appropriate weights is also a challenging task. On the other hand, our proposed system provides mathematical formulations to deal with such errors

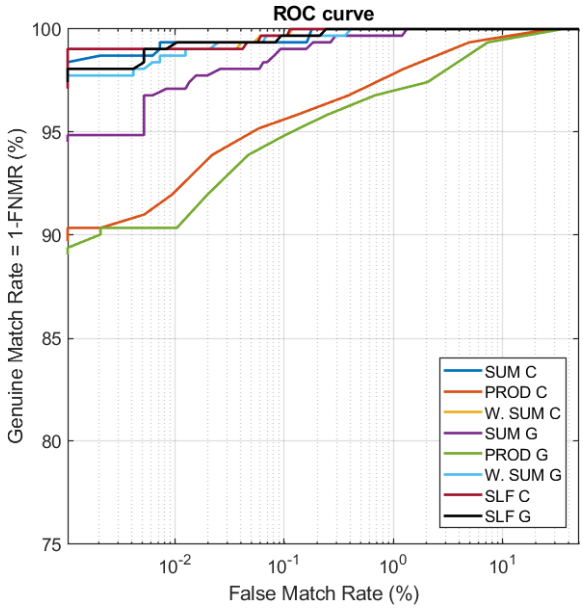



Figure 11.6: ROC curves for SLF and standard fusion techniques for validation dataset

| **Fusion Rule** | **FLR GMR(%)** | **FL GMR(%)** | **FR GMR(%)** | **LR GMR(%)** |
|---|---|---|---|---|
| Face Comparator C | | | | |
| Sum | 99.02 | 99.02 | 98.05 | 93.20 |
| W. Sum | 99.02 | 98.05 | 98.05 | 92.23 |
| Product | 94.17 | 85.43 | 90.02 | **94.17** |
| **SLF** | **99.02** | **99.02** | **98.05** | 92.23 |
| Face Comparator G | | | | |
| Sum | 98.05 | 96.11 | **98.05** | 93.20 |
| W. Sum | 99.02 | 90.29 | 97.08 | 92.23 |
| Product | 94.17 | 83.49 | 89.32 | **94.17** |
| **SLF** | **99.02** | **97.08** | 97.08 | 92.23 |

Table 11.4: GMR calculated @ FMR = $10^{-3}$ for SLF and standard fusion techniques with all four fusion strategies

logically, and it correctly handles the irregularities of the system. Further, we conclude the SLF correctly handles the random and systematic errors which are not taken into account by standard fusion techniques. Finally, this work not only explores the application of subjective logic fusion but also provides the experimental evidence supporting the SL as a better fu-

sion technique to deal with uncertain probabilities.

For the future work, we would like to study the behavior of different SLF operators in comparison with other methods such as fuzzy logic, Bayesian reasoning, and DST. Approaches to define belief mass and uncertainty can be explored to understand the biometric fusion with the subjective logic point of view.

## Acknowledgment

This work was carried out under the funding from the Research Council of Norway (Grant No. IKTPLUSS 248030/O70).

# Bibliography


[1] *ISO/IEC IS 29794-1 Information technology - Biometric sample quality - Part 1: Framework*. ISO/IEC.

[2] NeuroTechnology VeriLook 5.4. "http://www.neurotechnology.com/face-biometrics.html".

[3] Ahmed Al-Ani and Mohamed Deriche. A new technique for combining multiple classifiers using the dempster-shafer theory of evidence. *Journal of Artificial Intelligence Research*, 17:333–361, 2002.

[4] Fernando Alonso-Fernandez, Kiran B Raja, Christoph Busch, and Josef Bigun. Log-likelihood score level fusion for improved cross-sensor smartphone periocular recognition. In *Signal Processing Conference (EUSIPCO), 2017 25th European*, pages 271–275. IEEE, 2017.

[5] Brandon Amos, Bartosz Ludwiczuk, Mahadev Satyanarayanan, et al. Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, 2016.

[6] André Anjos, Ivana Chingovska, and Sébastien Marcel. Anti-spoofing, face databases. *Encyclopedia of Biometrics*, pages 55–66, 2015.

[7] Muhammad Arif, Thierry Brouard, and Nicole Vincent. A fusion methodology based on dempster-shafer evidence theory for two biometric applications. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 4, pages 590–593. IEEE, 2006.

[8] Simon Baker and Iain Matthews. Equivalence and efficiency of image alignment algorithms. In *Computer Vision and Pattern Recognition,*

*2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2001.

[9] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[10] Samarth Bharadwaj, Mayank Vatsa, and Richa Singh. Biometric quality: a review of fingerprint, iris, and face. *EURASIP journal on Image and Video Processing*, 2014(1):34, 2014.

[11] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa. Recognizing surgically altered face images using multiobjective evolutionary algorithm. *IEEE Transactions on Information Forensics and Security*, 8(1):89–100, 2013.

[12] Josef Bigun, Julian Fiérrez-Aguilar, Javier Ortega-Garcia, and Joaquin Gonzalez-Rodriguez. Multimodal biometric authentication using quality signals in mobile communications. In *null*, page 2. IEEE, 2003.

[13] The Global Biometrics and Mobility Report. "http://www.acuity-mi.com/GBMR_Report.php", 2018.

[14] Zinelabdine Boulkenafet, Jukka Komulainen, Zahid Akhtar, Azeddine Benlamoudi, Djamel Samai, Salah Eddine Bekhouche, Abdelkrim Ouafi, Fadi Dornaika, Abdelmalik Taleb-Ahmed, Le Qin, et al. A competition on generalized software-based face presentation attack detection in mobile scenarios. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 688–696. IEEE, 2017.

[15] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid. Face anti-spoofing based on color texture analysis. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 2636–2640. IEEE, 2015.

[16] Zinelabinde Boulkenafet, Jukka Komulainen, Lei Li, Xiaoyi Feng, and Abdenour Hadid. Oulu-npu: A mobile face presentation attack database with real-world variations. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*, pages 612–618. IEEE, 2017.

[17] Chaos Computer Club breaks Apple TouchID. "https://www.ccc.de/en/updates/2013/ccc-breaks-apple-touchid", 2013.

[18] Sijia Cai, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. A probabilistic collaborative representation based approach for pattern classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2950–2959, 2016.

[19] Kai Cao and Anil K Jain. Hacking mobile phones using 2d printed fingerprints. 2016.

[20] Shaxun Chen, Amit Pande, and Prasant Mohapatra. Sensor-assisted facial recognition: An enhanced biometric authentication system for smartphones. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*, pages 109–122. ACM, 2014.

[21] Xinjian Chen, Jie Tian, Qi Su, Xin Yang, and Fei Yue Wang. A secured mobile phone based on embedded fingerprint recognition systems. In *International Conference on Intelligence and Security Informatics*, pages 549–553. 2005.

[22] K. T. Cheng and Y. C. Wang. Using mobile gpu for general-purpose computing 2013; a case study of face recognition on smartphones. In *VLSI Design, Automation and Test (VLSI-DAT), 2011 International Symposium on*, pages 1–4, 2011.

[23] Ivana Chingovska, André Anjos, and Sébastien Marcel. On the effectiveness of local binary patterns in face anti-spoofing. In *Biometrics Special Interest Group (BIOSIG), 2012 BIOSIG-Proceedings of the International Conference of the*, pages 1–7. IEEE, 2012.

[24] François Chollet. Xception: Deep learning with depthwise separable convolutions. *arXiv preprint*, 2016.

[25] K. Y. Chou, G. M. Huang, H. C. Tseng, and Y. P. Chen. Face recognition based on sparse representation applied to mobile device. In *Automatic Control Conference (CACS), 2014 CACS International*, pages 81–86, Nov 2014.

[26] GSMArena Price Compare. "https://www.gsmarena.com/compare.php3?idPhone1=9343&idPhone2=8858&idPhone3=8966", 2018.

[27] Artur Costa-Pazo, Sushil Bhattacharjee, Esteban Vazquez-Fernandez, and Sebastien Marcel. The replay-mobile face presentation-attack database. In *Biometrics Special Interest Group (BIOSIG), 2016 International Conference of the*, pages 1–7. IEEE, 2016.

[28] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1, June 2005.

[29] Abhijit Das, Chiara Galdi, Hu Han, Raghavendra Ramachandra, Jean-Luc Dugelay, and Antitza Dantcheva. Recent advances in biometric technology for mobile devices. In *BTAS'18, 9th IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2018.

[30] NIST BSSR1 Database. https://www.nist.gov/itl/iad/image-group/nist-biometric-scores-set-bssr1, 2010.

[31] Guillaume Dave, Xing Chao, and Kishore Sriadibhatla. Face recognition in mobile phones. *Department of Electrical Engineering Stanford University, USA*, 2010.

[32] Tiago de Freitas Pereira, Jukka Komulainen, André Anjos, José Mario De Martino, Abdenour Hadid, Matti Pietikäinen, and Sébastien Marcel. Face liveness detection using dynamic texture. *EURASIP Journal on Image and Video Processing*, 2014(1):1–15, 2014.

[33] Maria De Marsico, Michele Nappi, Fabio Narducci, and Hugo Proença. Insights into the results of miche i-mobile iris challenge evaluation. *Pattern Recognition*, 74:286–304, 2018.

[34] Maria De Marsico, Michele Nappi, and Hugo Proença. Results from miche ii–mobile iris challenge evaluation ii. *Pattern Recognition Letters*, 91:3–10, 2017.

[35] Maria De Marsico, Michele Nappi, Daniel Riccio, and Harry Wechsler. Robust face recognition after plastic surgery using local region analysis. In *Image Analysis and Recognition*, pages 191–200. Springer, 2011.

[36] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR 2009*, pages 248–255. IEEE, 2009.

[37] Jiankang Deng, Jia Guo, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. *arXiv prepr,it arXiv:1801.07698*, 2018.

[38] Reza R Derakhshani and Joel Teply. Systems and methods for spoof detection and liveness analysis, May 31 2016. US Patent App. 15/169,107.

[39] Mohammad Omar Derawi, Bian Yang, and Christoph Busch. *Fingerprint Recognition with Embedded Cameras on Mobile Phones*, pages 136–147. Springer Berlin Heidelberg, 2012.

[40] David L Donoho. For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution. *Communications on pure and applied mathematics*, 59(6):797–829, 2006.

[41] Ted Dunstone and Neil Yager. *Biometric System and Data Analysis: Design, Evaluation, and Data Mining*. Springer Publishing Company, Incorporated, 1st edition, 2008.

[42] Jacques Duparre. Fabrication process for mastering imaging lens arrays, July 31 2012. US Patent 8,231,814.

[43] Nesli Erdogmus and Sebastien Marcel. Spoofing face recognition with 3d masks. *IEEE Transactions on Information Forensics and Security*, 9(7):1084–1097, 2014.

[44] The National Cheng Kung University face database. "http://www.datatang.com/data/14866".

[45] Sachin Sudhakar Farfade, Mohammad J Saberian, and Li-Jia Li. Multi-view face detection using deep convolutional neural networks. In *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pages 643–650. ACM, 2015.

[46] Julian Fierrez-Aguilar, Javier Ortega-Garcia, Joaquin Gonzalez-Rodriguez, and Josef Bigun. Discriminative multimodal biometric authentication based on quality measures. *Pattern recognition*, 38(5):777–779, 2005.

[47] Onyx Touchless fingerphoto. "http://www.diamondfortress.com/resources/the-white-paper-3", 2018.

[48] Alejandro F Frangi, Wiro J Niessen, Koen L Vincken, and Max A Viergever. Multiscale vessel enhancement filtering. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 130–137. Springer, 1998.

[49] Javier Galbally and Sébastien Marcel. Face anti-spoofing based on general image quality assessment. In *ICPR*, pages 1173–1178, 2014.

[50] Javier Galbally, Sébastien Marcel, and Julian Fierrez. Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. *Image Processing, IEEE Transactions on*, 23(2):710–724, 2014.

[51] Chiara Galdi, Michele Nappi, and Jean-Luc Dugelay. Multimodal authentication on smartphones: Combining iris and sensor recognition for a double check of user identity. *Pattern Recognition Letters*, 2015.

[52] Wen Gao, Bo Cao, Shiguang Shan, Xilin Chen, Delong Zhou, Xiaohua Zhang, and Debin Zhao. The cas-peal large-scale chinese face database and baseline evaluations. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 38(1):149–161, 2008.

[53] Xiufeng Gao, Stan Z Li, Rong Liu, and Peiren Zhang. Standardization of face image sample quality. Springer, 2007.

[54] Diogo Caetano Garcia and Ricardo L de Queiroz. Face-spoofing 2d-detection based on moiré-pattern analysis. *IEEE transactions on information forensics and security*, 10(4):778–786, 2015.

[55] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[56] Mislav Grgic, Kresimir Delac, and Sonja Grgic. Scface–surveillance cameras face database. *Multimedia tools and applications*, 2011.

[57] Ralph Gross, Simon Baker, Iain Matthews, and Takeo Kanade. Face recognition across pose and illumination. In *Handbook of Face Recognition*, pages 193–216. Springer, 2005.

[58] Patrick Grother and Elham Tabassi. Performance of biometric quality measures. 2007.

[59] Manuel Günther, Artur Costa-Pazo, Changxing Ding, Elhocine Boutellaa, Giovani Chiachia, Honglei Zhang, Marcus de Assis Angeloni, V Štruc, Elie Khoury, Esteban Vazquez-Fernandez, et al. The 2013 face recognition evaluation in mobile environment. In *Biometrics (ICB), 2013 International Conference on*, pages 1–7. IEEE, 2013.

[60] A. Hadid, N. Evans, S. Marcel, and J. Fierrez. Biometrics systems under spoofing attack: An evaluation methodology and lessons learned. *IEEE Signal Processing Magazine*, 32(5):20–30, 2015.

[61] A. Hadid, J. Y. Heikkila, O. Silven, and M. Pietikainen. Face and eye detection for person authentication in mobile phones. In *Distributed Smart Cameras, 2007. ICDSC '07. First ACM/IEEE International Conference on*, pages 101–108, Sept 2007.

[62] Abdenour Hadid, Nicholas Evans, Sébastien Marcel, and Julian Fierrez. Biometrics systems under spoofing attack: an evaluation methodology and lessons learned. *IEEE Signal Processing Magazine*, 32(5):20–30, 2015.

[63] Bee Yan Hiew, Andrew Beng Jin Teoh, and Ooi Shih Yin. A secure digital camera based fingerprint verification system. *Journal of Visual Communication and Image Representation*, 21(3):219 – 231, 2010.

[64] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[65] J. Hu, L. Peng, and L. Zheng. Xface: A face recognition system for android mobile phones. In *Cyber-Physical Systems, Networks, and Applications (CPSNA), 2015 IEEE 3rd International Conference on*, pages 13–18, Aug 2015.

[66] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *CVPR*, 2017.

[67] Apple Face ID. https://en.wikipedia.org/wiki/Face_ID, 2017.

[68] International Organization for Standardization and International Electrotechnical Committee. *ISO/IEC CD 30107-3. Information Technology – Biometric Performance Testing and Reporting – Part 3: Testing and reporting*, 2015.

[69] Apple iPhone 5S. "https://en.wikipedia.org/wiki/IPhone_5S", 2013.

[70] ISO/IEC. *ISO/IEC IS 19795-1 Information Technology - Biometric performance testing and reporting- Part 1: Principles and framework*. ISO/IEC, 2006.

[71] ISO/IEC. *ISO/IEC TR 24722:2007. Information technology – Biometrics– Multimodal and other multibiometric fusion.* ISO/IEC, 2007.

[72] ISO/IEC. *ISO/IEC TR 29794-5 Information technology - Biometric sample quality - Part 5: Face image data.* ISO/IEC, 2010.

[73] ISO/IEC. *ISO/IEC CD 30107-3.2 - Biometric presentation attack detection.* ISO/IEC, 2016.

[74] ISO/IEC TR 24741:2018. Information technology - Biometrics - Overview and application. Standard, International Organization for Standardization, February 2018.

[75] Anil Jain, Patrick Flynn, and Arun A Ross. *Handbook of biometrics.* Springer Science & Business Media, 2007.

[76] Anil K Jain. Technology: biometric recognition. *Nature*, 449(7158):38, 2007.

[77] Anil K Jain, Arun Ross, and Salil Prabhakar. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1):4–20, 2004.

[78] Peter A Johnson, Bozhao Tan, and Stephanie Schuckers. Multimodal f, pages = 1–7,usion vulnerability to non-zero effort (spoof) imposters. In *Information Forensics and Security (WIFS), 2010 IEEE International Workshop on*, pages 1–5. IEEE, 2010.

[79] Audun Jøsang. Artificial reasoning with subjective logic. In *Proceedings of the second Australian workshop on commonsense reasoning*, volume 48, page 34. Perth:[sn], 1997.

[80] Audun Jøsang. A logic for uncertain probabilities. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(03):279–311, 2001.

[81] Audun Jøsang. Subjective logic. *Book draft*, 2011.

[82] Audun Jøsang and Thorvald H Munch-Møller. Biometric data fusion based on subjective logic. In *Information Fusion (FUSION), 2014 17th International Conference on*, pages 1–8. IEEE, 2014.

[83] J. Kannala and E. Rahtu. Binarized statistical image features. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 1363–1366, Nov 2012.

[84] Koray Kavukcuoglu, Pierre Sermanet, Y-Lan Boureau, Karol Gregor, Michaël Mathieu, and Yann L Cun. Learning convolutional feature hierarchies for visual recognition. In *Advances in neural information processing systems*, pages 1090–1098, 2010.

[85] Elie Khoury, Bostjan Vesnicer, Javier Franco-Pedroso, Ricardo Violato, Z Boulkcnafet, LM Mazaira Fernández, Mireia Diez, Justina Kosmala, Houssemeddine Khemiri, Tomas Cipr, et al. The 2013 speaker recognition evaluation in mobile environment. In *Biometrics (ICB), 2013 International Conference on*, pages 1–8. IEEE, 2013.

[86] Inhan Kim, Juhyun Ahn, and Daijin Kim. Face spoofing detection with highlight removal effect and distortions. In *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*, pages 004299–004304. IEEE, 2016.

[87] Wonjun Kim, Sungjoo Suh, and Jae-Joon Han. Face liveness detection from a single image via diffusion speed model. *IEEE transactions on Image processing*, 24(8):2456–2465, 2015.

[88] Kiran B. Raja, R Raghavendra, and Christoph Busch. Binarized Statistical Image Features for Robust Iris and Periocular Recognition in Visible Spectrum. In *In proceedings of IEEE conference on International Workshop on Forensics and Biometrics (IWBF), Malta*. IEEE, 2014.

[89] Kiran B. Raja, R Raghavendra, and Christoph Busch. Presentation attack detection using laplacian decomposed frequency response for visible spectrum and near-infra-red iris systems. In *The 7th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, September 2015.

[90] Kiran B Raja, R Raghavendra, and Christoph Busch. Presentation attack detection using laplacian decomposed frequency response for visible spectrum and near-infra-red iris systems. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–8. IEEE, 2015.

[91] Kiran B. Raja, R Raghavendra, Martin Stokkenes, and Christoph Busch. Multi-modal authentication system for smartphones using face,

iris and periocular. In *IEEE International Conf. Biometrics (ICB), Phuket, Thailand*, 2015.

[92] Dakshina R Kisku, Massimo Tistarelli, Jamuna Kanta Sing, and Phalguni Gupta. Face recognition by fusion of local and global matching scores using ds theory: An evaluation with uni-classifier and multi-classifier paradigm. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 60–65. IEEE, 2009.

[93] Dakshina Ranjan Kisku, Phalguni Gupta, Hunny Mehrotra, and Jamuna Kanta Sing. Multimodal belief fusion for face and ear biometrics. *Intelligent Information Management*, 1(03):166, 2009.

[94] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[95] Till Kroeger, Radu Timofte, Dengxin Dai, and Luc Van Gool. Fast optical flow using dense inverse search. In *European Conference on Computer Vision*, pages 471–488. Springer, 2016.

[96] Eric Krotkov and J-P Martin. Range from focus. In *Robotics and Automation. Proceedings. 1986 IEEE International Conference on*, volume 3, pages 1093–1098. IEEE, 1986.

[97] A. Kumar and C. Kwong. Towards contactless, low-cost and accurate 3d fingerprint identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015.

[98] R. Donida Labati, A. Genovese, V. Piuri, and F. Scotti. Toward unconstrained fingerprint recognition: A fully touchless 3-d system based on two views on the move. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2016.

[99] A. Lanitis, C. J. Taylor, and T. F. Cootes. Toward automatic simulation of aging effects on face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):442–455, 2002.

[100] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[101] K.C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 27(5):684–698, 2005.

[102] Chang-Tsun Li. Source camera identification using enhanced sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, 5(2):280–287, 2010.

[103] Yan Li, Yingjiu Li, Qiang Yan, Hancong Kong, and Robert H Deng. Seeing your face is not enough: An inertial sensor-based liveness detection for face authentication. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 1558–1569. ACM, 2015.

[104] C. Lin and A. Kumar. Matching contactless and contact-based conventional fingerprint images for biometrics identification. *IEEE Transactions on Image Processing*, 2018.

[105] Jan Lukas, Jessica Fridrich, and Miroslav Goljan. Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, 1(2):205–214, 2006.

[106] Mohammad H Mahoor and Mohamed Abdel-Mottaleb. A multimodal approach for face modeling and recognition. *IEEE Transactions on Information Forensics and Security*, 3(3):431–440, 2008.

[107] Sébastien Marcel, Chris McCool, Pavel Matějka, Timo Ahonen, Jan Černockỳ, Shayok Chakraborty, Vineeth Balasubramanian, Sethuraman Panchanathan, Chi Ho Chan, Josef Kittler, et al. On the results of the first mobile biometry (mobio) face and speaker verification evaluation. In *Recognizing Patterns in Signals, Speech, Images and Videos*, pages 210–225. Springer, 2010.

[108] Sébastien Marcel, Mark S Nixon, and Stan Z Li. *Handbook of Biometric Anti-Spoofing*. Springer, 2014.

[109] Aleix M Martinez. The ar face database. *CVC technical report*, 24, 1998.

[110] Mostafa Mehdipour Ghazi and Hazim Kemal Ekenel. A comprehensive analysis of deep learning based representation for face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 34–41, 2016.

[111] Weizhi Meng, Duncan S Wong, Steven Furnell, and Jianying Zhou. Surveying the development of biometric user authentication on mobile phones. *IEEE Communications Surveys & Tutorials*, 17(3):1268–1293, 2015.

[112] David Menotti, Giovani Chiachia, Allan Pinto, William Robson Schwartz, Helio Pedrini, Alexandre Xavier Falcao, and Anderson Rocha. Deep representations for iris, face, and fingerprint spoofing detection. *IEEE Transactions on Information Forensics and Security*, 10(4):864–879, 2015.

[113] Lamia Mezai, Fella Hachouf, and Messaoud Bengherabi. Score fusion of face and voice using dempster-shafer theory for person authentication. In *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*, pages 894–899. IEEE, 2011.

[114] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012.

[115] Amir Mohammadi, Sushil Bhattacharjee, and Sébastien Marcel. Deeply vulnerable: a study of the robustness of face recognition to presentation attacks. *IET Biometrics*, 7(1):15–26, 2017.

[116] R. Mueller and R. Sanchez-Reillo. An approach to biometric identity management using low cost equipment. In *2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pages 1096–1100, Sept 2009.

[117] S. Nagpal, M. Singh, R. Singh, and M. Vatsa. Regularized deep learning for face recognition with weight variations. *IEEE Access*, 3:3010–3018, 2015.

[118] S. Nagpal, M. Singh, M. Vatsa, and R. Singh. Regularizing deep learning architecture for face recognition with weight variations. In *Biometrics Theory, Applications and Systems (BTAS), 2015 IEEE 7th International Conference on*, pages 1–6, 2015.

[119] Karthik Nandakumar, Yi Chen, Sarat C Dass, and Anil K Jain. Likelihood ratio-based biometric score fusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(2):342–347, 2008.

[120] Tempestt J Neal and Damon L Woodard. Surveying biometric authentication for mobile device security. *Journal of Pattern Recognition Research*, 1:74–110, 2016.

[121] Kien Nguyen, Simon Denman, Sridha Sridharan, and Clinton Fookes. Score-level multibiometric fusion based on dempster–shafer theory

incorporating uncertainty factors. *IEEE Transactions on Human-Machine Systems*, 45(1):132–140, 2015.

[122] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.

[123] Ville Ojansivu and Janne Heikkilä. Blur insensitive texture classification using local phase quantization. In *Image and signal processing*, pages 236–243. Springer, 2008.

[124] Martin Aastrup Olsen, Vladimír Šmida, and Christoph Busch. Finger image quality assessment features–definitions and evaluation. *IET Biometrics*, 2015.

[125] Andrzej Pacut and Adam Czajka. Aliveness detection for iris biometrics. In *Carnahan Conferences Security Technology, Proceedings 2006 40th Annual IEEE International*, pages 122–129. IEEE, 2006.

[126] Gang Pan, Lin Sun, Zhaohui Wu, and Shihong Lao. Eyeblink-based anti-spoofing in face recognition from a generic webcamera. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.

[127] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

[128] Junghun Park, Bowon Jung, and Okkyung Choi. Two-factor authentication methodology using hybrid face recognition. *biometrics*, 2016.

[129] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. In *BMVC*, volume 1, page 6, 2015.

[130] K. Patel, H. Han, A. K. Jain, and G. Ott. Live face video vs. spoof face video: Use of moir patterns to detect replay video attacks. In *Biometrics (ICB), 2015 International Conference on*, pages 98–105, 2015.

[131] Keyurkumar Patel, Hu Han, and Anil K Jain. Cross-database face antispoofing with robust feature representation. In *Chinese Conference on Biometric Recognition*, pages 611–619. Springer, 2016.

[132] Keyurkumar Patel, Hu Han, and Anil K Jain. Secure face unlock: Spoof detection on smartphones. *IEEE Transactions on Information Forensics and Security*, 11(10):2268–2283, 2016.

[133] P Jonathon Phillips, Patrick J Flynn, Todd Scruggs, Kevin W Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek. Overview of the face recognition grand challenge. In *CVPR 2005*, volume 1, pages 947–954. IEEE, 2005.

[134] P Jonathon Phillips, W Todd Scruggs, Alice J O'Toole, Patrick J Flynn, Kevin W Bowyer, Cathy L Schott, and Matthew Sharpe. Frvt 2006 and ice 2006 large-scale results. *National Institute of Standards and Technology, NISTIR*, 7408:1, 2007.

[135] Xuan Qi, Chen Liu, and Stephanie Schuckers. Boosting face in video recognition via cnn based key frame extraction.

[136] R Raghavendra and Christoph Busch. Presentation attack detection algorithm for face and iris biometrics. In *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, pages 1387–1391. IEEE, 2014.

[137] R Raghavendra and Christoph Busch. Robust scheme for iris presentation attack detection using multiscale binarized statistical image features. *IEEE Transactions on Information Forensics and Security*, 10(4):703–715, 2015.

[138] R Raghavendra, Christoph Busch, and Bian Yang. Scaling-robust fingerprint verification with smartphone camera in real-life scenarios. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pages 1–8. IEEE, 2013.

[139] R Raghavendra, Kiran B. Raja, Anika Pflug, Bian Yang, and Christoph Busch. 3d face reconstruction and multimodal person identification from video captured using smartphone camera. In *Technologies for Homeland Security (HST), 2013 IEEE International Conference on*, pages 552–557. IEEE, 2013.

[140] Ramachandra Raghavendra, Kiran B. Raja, and Christoph Busch. Presentation attack detection for face recognition using light field camera. *IEEE Transactions on Image Processing*, 24(3):1060–1075, 2015.

[141] Ramachandra Raghavendra, Kiran B Raja, Bian Yang, and Christoph Busch. Automatic face quality assessment from video using gray level

co-occurrence matrix: An empirical study on automatic border control system. pages 438–443, 2014.

[142] K. B. Raja, R. Raghavendra, and C. Busch. Video presentation attack detection in visible spectrum iris recognition using magnified phase information. *IEEE Transactions on Information Forensics and Security*, 2015.

[143] K. B. Raja, R. Raghavendra, M. Stokkenes, and C. Busch. Fusion of face and periocular information for improved authentication on smartphones. In *Information Fusion (Fusion), 2015 18th International Conference on*, pages 2115–2120, 2015.

[144] K. B. Raja, R. Raghavendra, M. Stokkenes, and C. Busch. Multimodal authentication system for smartphones using face, iris and periocular. In *Biometrics (ICB), 2015 International Conference on*, pages 143–150, 2015.

[145] Kiran B Raja, Pankaj Wasnik, R Raghavendra, and Christoph Busch. Robust face presentation attack detection on smartphones: An approach based on variable focus. In *Biometrics (IJCB), 2017 IEEE International Joint Conference on*, pages 651–658. IEEE, 2017.

[146] Raghavendra Ramachandra and Christoph Busch. Presentation attack detection methods for face recognition systems: a comprehensive survey. *ACM Computing Surveys (CSUR)*, 50(1):8, 2017.

[147] V. Ravibabu and N. Krishnan. A vary approach to face recognition veritable mechanisms for android mobile against spoofing. In *Computational Intelligence and Computing Research (ICCIC), 2014 IEEE International Conference on*, pages 1–4, Dec 2014.

[148] Xiaohui Ren, Jinfeng Yang, Henghui Li, and Renbiao Wu. Multifingerprint information fusion for personal identification based on improved dempster-shafer evidence theory. In *Electronic Computer Technology, 2009 International Conference on*, pages 281–285. IEEE, 2009.

[149] Juniper Research. "http://www.juniperresearch.com/press/press-releases/biometric-authentication-app-downloads-to-reach-77", 2015.

[150] Janessa Rivera and Rob Van der Meulen. Gartner says worldwide pc, tablet and mobile phone combined shipments to reach 2.4 billion units in 2013. *Tablet*, 116(197,202):265–731, 2013.

[151] Sami Romdhani, Jeffrey Ho, Thomas Vetter, and David J Kriegman. Face recognition using 3-d models: Pose and illumination. *Proceedings of the IEEE*, 94(11):1977–1999, 2006.

[152] Nectarios Rose. Facial expression classification using gabor and log-gabor filters. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 346–350. IEEE, 2006.

[153] A. Ross and A.K. Jain. Multimodal biometrics: An overview. In *Signal Processing Conference, 2004 12th European*, pages 1221–1224, 2004.

[154] Arun Ross and Anil Jain. Information fusion in biometrics. *Pattern recognition letters*, 24(13):2115–2125, 2003.

[155] Michele A Saad, Alan C Bovik, and Christophe Charrier. Blind image quality assessment: A natural scene statistics approach in the dct domain. *IEEE transactions on Image Processing*, 21(8):3339–3352, 2012.

[156] Conrad Sanderson and Kuldip K Paliwal. Identity verification using speech and face information. *Digital Signal Processing*, 14(5), = 2004.

[157] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *CVPR*, 2018.

[158] A. Sankaran, A. Malhotra, A. Mittal, M. Vatsa, and R. Singh. On smartphone camera based fingerphoto authentication. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7, Sept 2015.

[159] Mark S. NixonStan Z. Li Sébastien Marcel, editor. *Handbook of Biometric Anti-Spoofing*. Springer, Springer London Heidelberg New York Dordrecht, 2014.

[160] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823, 2015.

[161] Neurotechnology Verifinger SDK. "http://www.neurotechnology.com/verifinger.html".

[162] Takehiko Senba and Takeshi Misawa. Imaging device and digital camera, June 8 2006. US Patent App. 11/448,676.

[163] Glenn Shafer. *A mathematical theory of evidence*, volume 42. Princeton university press, 1976.

[164] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[165] M. Singh, S. Nagpal, R. Singh, and M. Vatsa. On recognizing face images with weight and age variations. *IEEE Access*, 2:822–830, 2014.

[166] R. Singh, M. Vatsa, H. S. Bhatt, S. Bharadwaj, A. Noore, and S. S. Nooreyezdan. Plastic surgery: A new dimension to face recognition. *IEEE Transactions on Information Forensics and Security*, 5(3):441–448, 2010.

[167] R. Singh, M. Vatsa, and A. Noore. Effect of plastic surgery on face recognition: A preliminary study. In *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*, pages 72–77, 2009.

[168] Richa Singh, Mayank Vatsa, Afzel Noore, and Sanjay K Singh. Dempster-shafer theory based classifier fusion for improved fingerprint verification performance. In *Computer vision, graphics and image processing*, pages 941–949. Springer, 2006.

[169] Ctirad Sousedik and Christoph Busch. Presentation attack detection methods for fingerprint recognition systems: a survey. *IET Biometrics*, 3(4):219–233, 2014.

[170] Helmuth Spath. *The cluster dissection and analysis theory fortran programs examples*. Prentice-Hall, Inc., 1985.

[171] C. Stein, V. Bouatou, and C. Busch. Video-based fingerphoto recognition with anti-spoofing techniques with smartphone cameras. In *2013 International Conference of the BIOSIG Special Interest Group (BIOSIG)*, pages 1–12, Sept 2013.

[172] C. Stein, C. Nickel, and C. Busch. Fingerphoto recognition with smartphone cameras. In *BIOSIG*, 2012.

[173] Chris Stein, Vincent Bouatou, and Christoph Busch. Video-based fingerphoto recognition with anti-spoofing techniques with smartphone cameras. In *BIOSIG*, 2013.

[174] Supasorn Suwajanakorn, Carlos Hernandez, and Steven M Seitz. Depth from focus with your mobile phone. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3497–3506, 2015.

[175] Michał Szczepanik, Ireneusz J Jóźwiak, Tomasz Jamka, and Karol Stasiński. Security lock system for mobile devices based on fingerprint recognition algorithm. In *Information Systems Architecture and Technology: Proceedings of 36th International Conference on Information Systems Architecture and Technology–ISAT 2015–Part III*, pages 25–35. Springer, 2016.

[176] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–9, 2015.

[177] Elham Tabassi, George W Quinn, and Patrick Grother. Whe = n to fuse two biometrics. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Workshop on Biometrics*, 2006.

[178] Archit Taneja, Aakriti Tayal, Aakarsh Malhorta, Anush Sankaran, Mayank Vatsa, and Rieha Singh. Fingerphoto spoofing in mobile devices: A preliminary study. In *Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on*, pages 1–7. IEEE, 2016.

[179] Tommy Thorsen, Pankaj Wasnik, Christoph Busch, R Raghavendra, and Kiran Raja. Assessing face image quality with lstms. In *Proceedings of the 11th Norwegian Information Security Conference*, 2018.

[180] K. Tiwari and P. Gupta. A touch-less fingerphoto recognition system for mobile hand-held devices. In *2015 International Conference on Biometrics (ICB)*, pages 151–156, May 2015.

[181] K-A Toh, Xudong Jiang, and Wei-Yun Yau. Exploiting global and local decisions for multimodal biometrics verification. *IEEE Transactions on Signal Processing*, 52(10):3059–3072, 2004.

[182] Apple TouchID. "https://en.wikipedia.org/wiki/Touch_ID", 2013.

[183] Philip Tresadern, Chris McCool, Norman Poh, Pavel Matejka, Abdenour Hadid, Christophe Levy, Tim Cootes, and Sebastien Marcel. Mobile biometrics (mobio): Joint face and voice verification for a mobile platform. *IEEE pervasive computing*, 2012.

[184] Sergey Tulyakov and Venu Govindaraju. Classifier combination types for biometric applications. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 58–58. IEEE, 2006.

[185] Mayank Vatsa, Richa Singh, Afzel Noore, and Arun Ross. On the dynamic selection of biometric fusion algorithms. *IEEE Transactions on Information Forensics and Security*, 5(3):470–479, 2010.

[186] Veridium. "https://www.veridiumid.com/biometric-authentication/fingerprint-recognition/", 2018.

[187] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 1, pages I–511. IEEE, 2001.

[188] Paul Viola and Michael Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57:137–154, 2004.

[189] Yunhong Wang, Tieniu Tan, and Anil K Jain. Combining face and iris biometrics for identity verification. In *International Conference on Audio-and Video-Based Biometric Person Authentication*, pages 805–813. Springer, 2003.

[190] Pankaj Wasnik, Kiran B. Raja, Ramachandra Raghavendra, and Christoph Busch. Presentation attack detection in face biometric systems using raw sensor data from smartphones. In *Signal-Image Technology & Internet-Based Systems (SITIS), 2016 12th International Conference on*, pages 104–111. IEEE, 2016.

[191] Pankaj Wasnik, Kiran B Raja, Raghavendra Ramachandra, and Christoph Busch. Assessing face image quality for smartphone based face recognition system. In *Biometrics and Forensics (IWBF), 2017 5th International Workshop on*, pages 1–6. IEEE, 2017.

[192] Pankaj Wasnik, R Ramachandra, Martin Stokkenes, Kiran Raja, and Christoph Busch. Improved fingerphoto verification system using multi-scale second order local structures. In *2018 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pages 1–5. IEEE, 2018.

[193] Pankaj Wasnik, Kirstina Schäfer, Kiran Raja, Raghavendra Ramachandra, and Christoph Busch. Fusing biometric scores using subjective logic for gait recognition on smartphone. *BIOSIG 2017*, 2017.

[194] Craig I Watson, Brad Ulery, RA Hicklin, William Fellner, and P Halinan. Studies of biometric fusion executive summary. Technical report, 2004.

[195] D. Wen, H. Han, and A. K. Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, 2015.

[196] Lingyun Wen, Guodong Guo, and Xin Li. A study on the influence of body weight changes on face recognition. In *Biometrics (IJCB), 2014 IEEE International Joint Conference on*, pages 1–6. IEEE, 2014.

[197] M. Werner and M. Brauckmann. Quality values for face recognition. In NIST Biometric Quality Workshop, 2006.

[198] Agata Wojciechowska, Michał Choraś, and Rafał Kozik. The overview of trends and challenges in mobile biometrics. *Journal of Applied Mathematics and Computational Mechanics*, 16, 2017.

[199] Yongkang Wong, Shaokang Chen, Sandra Mau, Conrad Sanderson, and Brian C Lovell. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. 2011.

[200] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210 – 227, 2009.

[201] Lijun Zhang, Lin Zhang, and Lida Li. Illumination quality assessment for face images: A benchmark and a convolutional neural networks based model. In *Neural Information Processing*. Springer, 2017.

[202] Man Zhang, Qi Zhang, Zhenan Sun, Shujuan Zhou, and Nasir Uddin Ahmed. The btas competition on mobile iris recognition. In *Biometrics Theory, Applications and Systems (BTAS), 2016 IEEE 8th International Conference on*, pages 1–7. IEEE, 2016.

[203] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V Le. Learning transferable architectures for scalable image recognition. *arXiv preprint arXiv:1707.07012*, 2017.

# Part III

# Appendix

# Chapter 12

# Appendix A

This section provides details of the SWAN data capture application which developed for the collection of subjects' data participated in the SWAN data collection.

## 12.1 Implementation Details

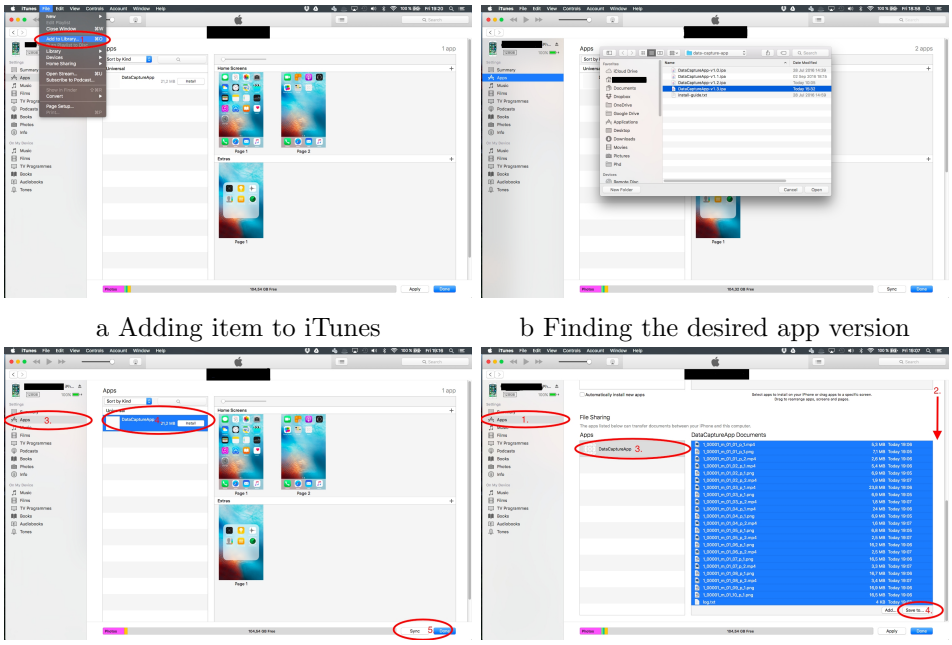
This application is developed on the iOS platform version 9 and intended to use for the devices iPhone 6S and iPad Pro. These devices are used throughout the SWAN project.

### 12.1.1 Setup

The application can be distributed in the form of a compiled installable (.ipa) file for the iOS environment. The name of the application is DataCaptureApp-vX.X.ipa, where X represents the digit used for specifying the version of the application. For now, the data capture application can only be installed on devices which are registered with the NTNU Apple developer license. Figure 12.1 shows the steps involved in the installation data capture application and Figure 12.1d provides the details of transfer of captured data from application.

### 12.1.2 Capture Settings

Once the user opens the SWAN data capture application, it is advised to configure some settings related to a number of image and video recordings, capture site where data is captured, and the length of videos. However, these settings can be reset using the "RESET TO DEFAULT VALUES".



a Adding item to iTunes

b Finding the desired app version

c Installing app on device

d Transferring data from device

Figure 12.1: Application setup using iTunes and data transfer from mobile device to computer

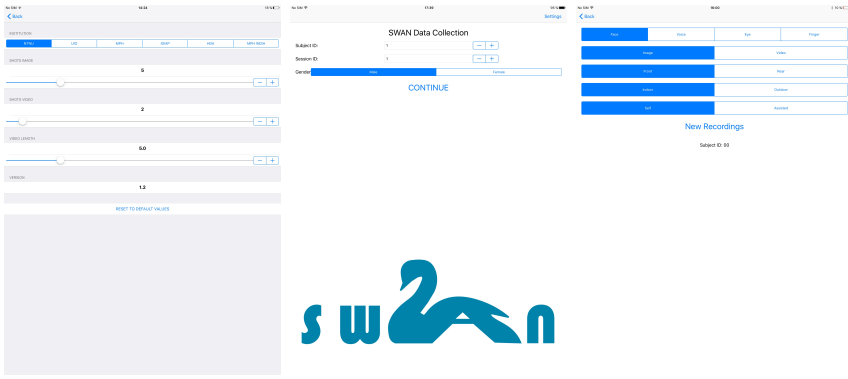
### 12.1.3 Exporting Data Files

The acquired data can be exported by connecting the capture device to a computer with iTunes installed. See Figure 12.1d. Following are the steps involved:

1. Select the mobile/tablet device in iTunes menu bar and then select the Apps tab shown on the left side.
2. Scroll down to the File Sharing option.
3. Select ”DataCaptureApp” under the Apps list.
4. Select all the files and click on ”Save to...” for exporting the selected files to the computer.

### 12.1.4 Subject and Session Selection

The first step in the data capture process is to select the proper subject and session IDs with the gender of the current subject. Using the -/+ buttons




a Settings menu b Home menu c Capture setup menu

Figure 12.2: Main setup screens

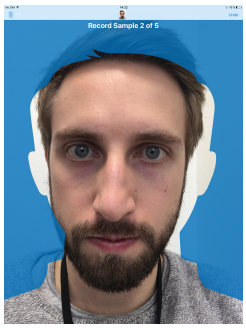
or by tapping on the text field user can set the current subject ID and session ID in use. Once the configuration is done, a user needs to click on "CONTINUE" button to proceed. Figure 12.2b shows the setup for a male subject with Subject ID 1 and Session 1.

### 12.1.5 Capture selection

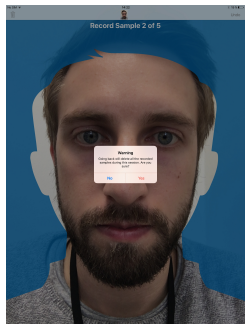
This menu provides the capture setup where a user can select biometric modality for which data will be collected. There are four options to capture the data of face, voice, periocular and fingerphoto. Figure 12.2c shows the setup for capturing the facial data of a subject in self-capture mode, i.e., using the front camera and indoors. One has to use the respective options, as described in the SWAN project data collection protocols. Once the capture setup is done, the data acquisition process is started by tapping the "New Recordings" button.

## 12.2 Capture GUI

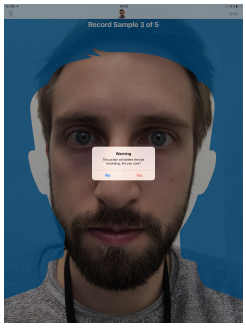
Figure 12.3a shows the different elements in the recording mode. In case of capturing of the face, eye, and finger an acquisition is started by double tapping the screen as the single tap will focus the rear camera. For voice, a single tap or pressing the "START" button will start the data acquisition. A progress bar is displayed at the top of the screen to show the recording progress. The current recording session can be deleted by tapping the trash bin icon in the top left corner. This will delete the recorded files for the current acquisition. Figure 12.3b, shows the displayed alert message in case of cancellation. Recapture of the sample is possible by tapping the "UNDO"




a GUI for recording samples




b Abort current recording session alert message




c Undo alert message

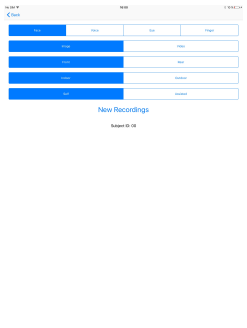
d Preview of last recorded sample

Figure 12.3: Capture GUI

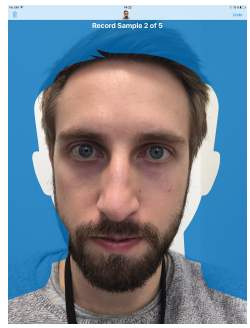
button in the top right corner. To verify the captured recording, a preview option is provided by tapping the icon in the middle of the top bar.

### 12.2.1 Face Capture

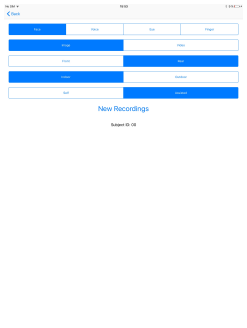
Figure 12.4a and 12.4c shows the capture setup used to acquire a face data using front and rear camera. An overlay face mask is displayed on the screen to guide the user for placing the face at the recommended location. The users need to double tap the screen to start the recording. In the case of data capture using rear camera, a single tap is used to focus the camera, and double tap is used for recording the data. The user needs to tap the DONE button when all samples are recorded.



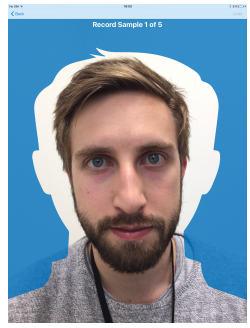

a Capture setup for front camera face capture




b Front camera face capture screen




c Capture setup for front camera face capture

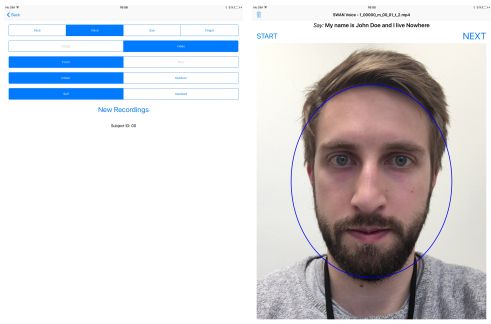



d Rear camera face capture screen

Figure 12.4: Face capture setup and screens

### 12.2.2 Voice Capture

Figure 12.5a shows the setup which is used to capture voice data. The blue circle is displayed on the screen to guide the user to place the face at the appropriate location. Eight sentences are then displayed on the screen, and the user needs to read out loud in order to record the data. There are two parts in this recording, firstly, the 8 sentences are displayed in the English language, and later they are displayed in the local language. Appropriate labels and instructions are provided for users in order to simplify the data capture.

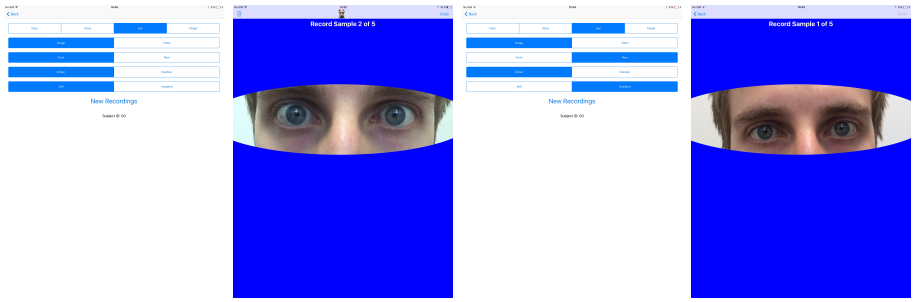



a b

Figure 12.5: Voice capture setup and screen. In figure, from left to right the captions are: a) Capture setup for voice capture and b) Voice capture screen

### 12.2.3 Eye Capture

Figure 12.6a and 12.6c shows the selection to capture a periocular data using the front and rear camera. The user needs to place the eyes in the visible area of the blue screen. The actions required to capture the eye data are similar to that of face capture mode.




a b c d

Figure 12.6: Eye capture setup and capture screens. In figure, from left to right the captions are: a) Front camera setup b) Capture screen c) Rear camera setup and d) Rear camera screen

### 12.2.4 Finger Capture

Figure 12.7a shows the option selected to capture a fingerphoto data. Capturing of the fingerphotos is only available with the rear camera. At the bottom of the screen, a label for correct finger is displayed. The user needs

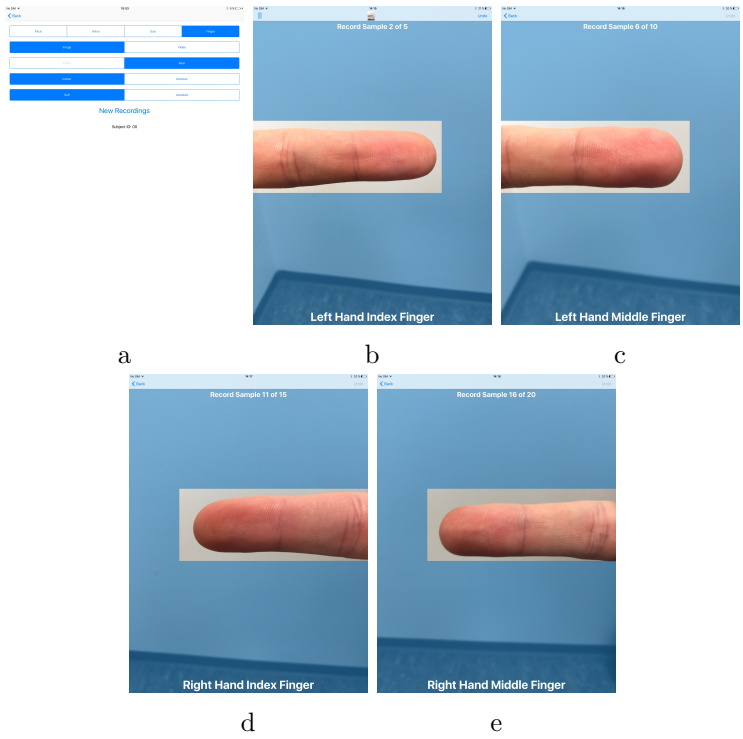



Figure 12.7: Finger capture setup and capture screens. In figure, from left to right the captions are: a) Capture setup screen for finger b) Capturing left index finger c) Capturing left middle finger e) Capturing right index finger and e) Capturing right middle finger

to place the finger in the guided area by transparent overlay mask. As similar to the face and periocular data capture mode, the single tap is autofocus the finger and double tap to start the recording. Figures 12.7b, 12.7c, 12.7d and 12.7e shows capture of different fingers.

# Chapter 13

# Appendix B

This section provides the details of the additional experiments and results obtained during the course of this thesis. These results further strengthen the motivation and effectiveness of proposed methods in Chapter 6 and 9.

## 13.1 Real-time on-device results

This section gives the details of the real-time on-device results of the best performing framework proposed in Chapter 6. The experimental results show that the FQAA based on the fine-tuning of Inception V3 network.

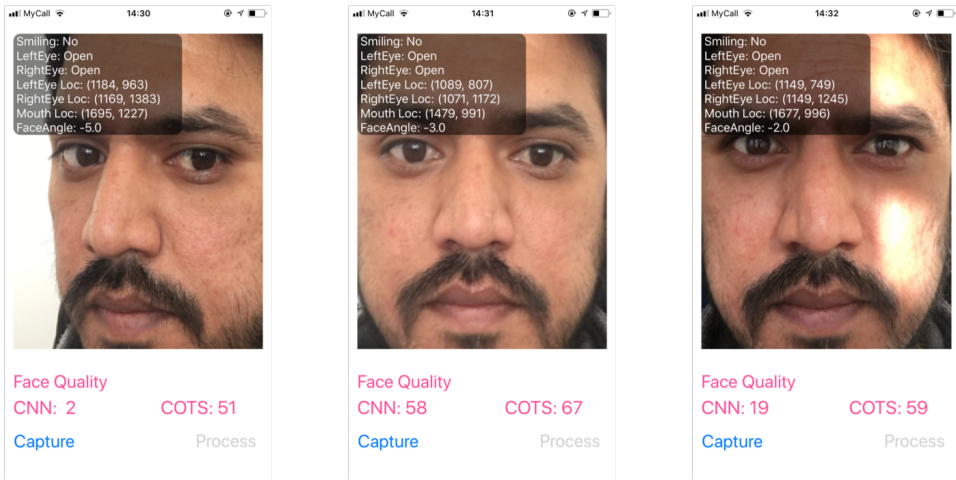



Figure 13.1: Results of the on-device experiments using Inception V3 [176] and COTS [2]

achieved high performance. Hence, in order to test the effectiveness of the method, we implemented the proposed framework on a smartphone. Figure 13.1 shows the output for three face inputs. The obtained results are compared with the commercial mobile SDK from Neurotechnology which was also executed alongside the proposed framework. From the figure, we can observe that the performance of the proposed framework is better than the COTS. This can be validated by observing the first figure where we can see that there is a pose in the input image. Thus, the CNN outputs low score whereas of COTS shows high-quality score which is not the expected behaviour. Similarly, in the case of the rightmost figure, the output of the CNN is low-quality value as there is an illumination present in the input face image. However, COTS still outputs high-quality value which is not an anticipated behaviour. In the end, we can also observe that in the case of an input image with decent quality both of the algorithms show approximately similar behaviour.

## 13.2 Vulnerability analysis of fingerphoto recognition system

In order to evaluate the vulnerability, we first formulated the fingerphoto recognition system using the probabilistic collaborative representation based approach for pattern classification (Pro-CRC) proposed in the recent work [18]. We have used Histogram of Gradient (HoG) features to train the (Pro-CRC) classifier. The database consisting of 48 subjects is used to obtain the verification results. For training, we used ten randomly selected frames from the videos of Session 1 data. Mainly, the dataset consists of data from the left-hand index finger of a particular subject which is treated as a unique identity. The gallery template features are extracted using HoG feature extractor with a block size of $16 \times 16$ pixels which resulted in approximately 26000 features per frame. Finally, these 480 templates are used to train the Pro-CRC classifier. The data from Session 2 is used as bona fide data. The presentation attacks are generated using the Session 1 data. The mated and non-mated comparison scores are generated using the trained classifier and the Session 2 data.

Similarly, in order to get the mated comparison scores for the presentation attacks randomly picked seven frames from the PA sample and compared against the gallery samples. The vulnerability analysis of the fingerphoto recognition system is shown in Figure 13.2. The green score distribution depicts the mated comparison scores for the bona-fide samples. The blue score distribution shows the zero-effort impostor, i.e., non-mated comparison score. Similarly, the gray score distribution represents the mated com-

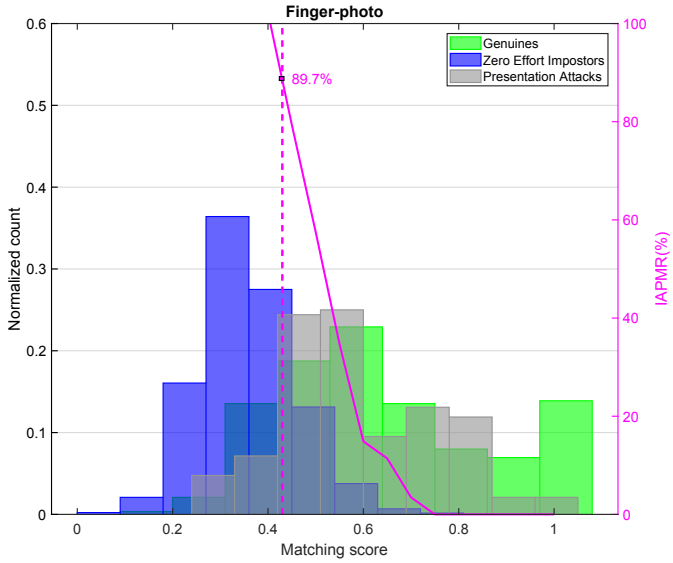



Figure 13.2: Vulnerability analysis of the fingerphoto recognition system. The reported IAPMR value corresponds to the equal error rate threshold.

parison score distribution of attempted presentation attack. Ideally, for a robust fingerphoto recognition system, the zero-effort impostor and presentation attack scores should be lower compared to the mated comparisons. However, from the Figure 13.2, we can observe that the score distribution of presentation attacks largely overlaps with the mated score distribution indicating the high vulnerability towards presentation attacks. Furthermore, our results show that the system is 89.7% vulnerable to the presented attacks in terms of IAPMR.