Robust Biometric Verification Based On The Eye Region

Algorithms for Visible Spectrum Image Data

Kiran Bylappa Raja

Thesis submitted to
Norwegian University of Science and Technology
for the degree of
Doctor of Philosophy in Computer Science

NTNU
Norwegian University of
Science and Technology

2016

Robust Biometric Verification Based On The Eye Region

Faculty of Computer Science and Media Technology
Norwegian University of Science and Technology

Declaration of Authorship

I, Kiran Bylappa Raja, hereby declare that this thesis and the work presented in it is entirely my own. Where I have consulted the work of others, this is always clearly stated.

Signed:

(Kiran Bylappa Raja)

Date:

Abstract

Popular mobile banking and e-commerce applications like Google Wallet, Apple Pay and Alipay have resulted in using personal devices like smartphones for secure access of services via biometric data captured from embedded sensor. Further, the improved optics on smartphones have been explored for biometric data capture in a contactless manner that can be used for various secure authentication applications. Specifically, the applications are exploring face, periocular, iris and finger photo characteristics for smartphone-based authentication using the embedded camera operating in the visible spectrum.

This thesis is dedicated to explore iris, periocular and face data for authentication applications when captured from smartphones in the visible spectrum. We first present robust algorithms to use iris recognition for the data captured from the smartphones in an unconstrained manner. Further, we propose a new imaging setup to resolve the iris texture for images captured in the visible spectrum for subjects with heavily pigmented iris. As an alternative to mitigate the lower performance of iris recognition in the visible spectrum, we demonstrate the use of periocular characteristics for authentication. We then present a robust algorithm for feature extraction from periocular images for verification purpose. We also present a multi-biometric authentication system fully realized on the smartphone with good verification accuracy for secure access applications. In the end, we present a set of robust algorithms to detect the presentation attacks on the ocular biometric systems working on the smartphones. Additionally, the implementation of most of the proposed algorithms and the databases constructed during the course of this thesis are made available to promote reproducible research in biometrics.

Abstract

Populære mobile banktjenester og e-handel applikasjoner som Google Wallet, Apple Pay og Alipay har resultert i å bruke personlige enheter som smarttelefoner for sikker tilgang på tjenestene via biometriske data som samles inn fra innebygd sensor. Videre har den forbedrede optikken på smarttelefoner blitt utforsket for biometrisk datainnsamling på en kontaktløs måte som da kan brukes til ulike metoder for sikker autentisering i applikasjoner. Nærmere bestemt utforsker applikasjonene egenskaper ved ansikt-, periokulær-, iris- og fingerbilde for smarttelefon basert autentisering ved hjelp av det innebygde kameraet som opererer i det synlige spekteret.

Denne avhandlingen er dedikert til å utforske iris-, periokulær- og ansiktsdata for autentiseringsapplikasjoner når de er samlet inn ved hjelp av en smarttelefon i det synlige spekteret. Vi presenterer robuste algoritmer for å bruke irisgjenkjenning på data samlet inn med smarttelefoner der brukeren ikke ble pålagt noen begrensninger. Videre foreslår vi en ny avbildningsmetode for å ta klarere bilder av iristeksturen for bilder tatt i det synlige spekteret når personen har mørk iris. Som et alternativ for å motvirke lavere ytelse ved irisgjenkjenning i det synlige spekteret, utforsker vi egenskaper ved området rundt øye (periokulærgodkjenning). Deretter presenterer vi en robust algoritme for uthenting av trekk fra de periokulære bildene, som kan brukes til verifisering. Vi presenterer også et multi-biometrisk autentiseringssystem, fullt implementert på smarttelefon, med høy verifiseringsnøyaktighet for sikker tilgang til applikasjoner. Til slutt presenterer vi et sett med robuste algoritmer for å oppdage presentasjonsangrep i okulær biometriske systemer som kjører på smarttelefoner. I tillegg er implementasjoner av de fleste foreslåtte algoritmene, og databasene som er samlet inn i løpet av denne avhandlingen gjort tilgjengelig for å fremme reproduserbar forskning innenfor biometri.

To my parents, my sister and my aunt.
To the love and support they bestowed on me!

Acknowledgements

I am greatly indebted to my advisors, Prof. Dr. Christoph Busch and Dr. Raghavendra Ramachandra for the opportunity. I would like to express my gratitude to 'Safran Identity and Security' for funding this work.

It was a privilege to work under the supervision of Christoph. I wish to thank him for the opportunity and the guidance provided. His words of encouragement have kept me going and helped me focus on the positive aspects of research. He has taught me the importance of being organized in a professional space. His passion and commitment towards mentoring the students has inspired me in many ways. In spite of his busy schedule, he has never compromised on taking the time for discussions in professional and personal front. Christoph has just not been mentoring me in research but also giving me lessons on skiing! Apart from being a professor, Christoph makes a wonderful ski instructor. I extend my special thanks for everything he has offered me including the cabin-trips to take some time off the work. Thank you, Christoph!

My interaction with Raghu started during the final year of my master degree which motivated me to involve myself with the research. He has by far been the best example and inspiration when it comes to dedication and hard-work. He has constantly stood by me in the days of success and failure. His inspirational talks have helped me to carry out research with a positive outlook. Much of the thesis was accomplished through the discussions and arguments with him. He has never settled for anything less than a fine way of conducting research. This thesis would be incomplete without mentioning his special interest in mentoring me on a day-to-day basis. Thank you, Raghu!

My sincere thanks go to Safran Identity and Security (formerly Morpho) for funding this work. I have had the chance to interact with Pierre Gacon, Emine Krichen, Jean Nöel-Braun, Jonathan, Stephane Gentric, Vincent Bouatou and Jean-Christophe Fondeur which gave the opportunity of knowing the industry perspective. I wish to thank Pierre and Emine, who took time out of their busy schedule to review my work and provide the comments to improve the quality of work. A special mention is due for their efforts in reviewing the thesis. Thank you both!

I would also like to thank the members of dissertation committee, Prof. Dr. Arun Ross, Prof. Dr. Raul Sanchez-Reillo, Prof. Dr. Laura Georg and Prof. Dr. Stephen Wolthusen for their time and effort. I would like to extend my gratitude to Nils Kalstad Svendsen for hosting me in NTNU, Gjøvik. Thanks to Laura Georg for supporting the activities of the lab and taking the interest in our research work.

I would also like to acknowledge the support of Kathrine Huke Markengbakken, Hilde Bakke, Jingjing Yang and Rachael McCullum. Special thanks to Katherine and Hilde for helping in many different aspects of administrative procedures. A note of thanks to research administration, student administration and HR department. I would like to acknowledge the efforts of Ingrid von Schantz Bakka for the help poured during the final submission of thesis.

The Norwegian Biometrics Laboratory has provided the best working environment for the research. I have had the opportunity of interacting with the wonderful people during my PhD. I would like to thank Prof. Dr. Patrick Bours, Assoc. Prof. Bian Yang, Guoqiang Li, Martin Stokkenes, Martin Aastrup Olsen, Soumik Mondal, Pankaj Wasnik, Ctirad Sousedik, Edlira Martiri, Andreas Nautsch and Martha Gomez-Barrero for creating such a wonderful environment. I have enjoyed our long nights in the lab as much as the social outings on Friday evenings with most of you. A special thanks to Guoqiang for saving me from the choke!

During the course of my thesis, I have had the opportunity of working with Vinay Vemuri, Jayachander, Martin Stokkenes, Dimitri Abalenkov, Jasmina Pelivani and Morten Krarup Sigaard. I have enjoyed our long sessions of data collection with Vinay. Thank you for the efforts and company. Martin and Morten have helped me immensely, especially in the long coding and debug sessions. I extend my thanks to all of them for their invaluable support during my thesis. Martin, thanks for keeping a watch on me during the ski-trips.

This thesis was possible because of the number of volunteers who have contributed their biometric data for the sake of research. I would like to extend my gratitude for all those volunteers. I would like to thank Peter Nussbaum and Marius Pedersen for letting me use the facilities in Color-Lab for many of my experiments. IT team at NTNU-Gjøvik has supported many times during the course of this thesis, especially for running the large scale experiments. A special note of thanks to them.

During the course of my thesis, I had a wonderful opportunity of working with Frøy Løvåsdal and Cathrine Elisabeth Fari in NID. Working with them was a truly pleasant experience which has enriched me in many aspects of practical considerations for making a biometric system work. A week of ISO meetings in their company was delightful and of course fun-filled. I extend my thanks for the opportunity to work with them.

During my stay in Gjøvik, I have met many people who have made my stay very pleasant. I must acknowledge CIMET group (2011-2013), Thomas Simon, Irina Ciortan, Mihkal Dunfjeld, Rezarta, Gerardo Diego Espinosa De La Riva, Annie Yustina, Vasilisa, Elnaz, Andrii Shalaginov, Kostantin Lenchik, Mariya Chirchenkova, Aditya Sole and Sony George who have kept me occupied socially. I have enjoyed the company of 'Indian' group (Vivek Sharma, Vijay, Priyanka, Aneesh, Gopi, Ramya, Sumanth, Monalisa, Kiran, Venkatesh Kolluru, Venkatesh Chowdary, Praveen, Pankaj Pandey, Vivek Agarwal) who have constantly helped me in all the sessions of data collection. They have taken time out of their schedule number of times even when asked multiple times. Thank you all for making my stay in Gjøvik a memorable one. A special thanks to Sushma Raghavendra for all the food she served to make me feel home away from home!

I would like to thank Raymond and Luuk for the wonderful time during our ski-trip and especially for the long scientific talks while preparing the dinner. Thank you!
I would like to thank Ashitha Binish and Akilandan Solaiappan who have mentored me during the initial days of my career. Their valuable suggestions have helped in making right decisions in my career. Thank you!

Most part of my childhood and adolescence was spent with the most wonderful persons I have met so far. They have walked into my life at the early age of 12 years and stayed with me through every stage of my life so far. Greatest teachers in my schooling have culminated the interest in education and it would not be complete if I do not remember them. I owe the significant part of my success to my teachers - Basavaraj, Gayathri, Gopalakrishna, KC Hubbali, Prema D. Mallapur, Harriet, Manjunath, Hymavathy, Arun, Anasuya (PE), Anasuya (Mess), Steven and Rajshekar Kambar. Specially, Basavaraj, Gopalakrishna and Prema have taught me the value of hard-work and discipline to achieve anything in life, no matter how trivial. I extend my heart-felt thanks to them.

The entire batch of 1996-2003 has just been another family to me. Growing up with all the batch-mates including Mohan, CN Pradeep, Rakesh, Nithin and Jaiswal has taught me the true meaning of friendship. The senior batch of 1995-2002 have shown the true love of siblings throughout the life of Navodaya and beyond, especially, Ashok Kumar HM, Keshav Murthy BN, Vipin Babu V, Bheemesh, KT Srinivas Murthy and BM Shrikanth have shown what it means to have a sibling relation beyond blood-line. It would go incomplete without mentioning Arun Kumar M who has supported me in various spheres. I extend my thanks to everyone from the Navodaya family. Apart from them, two important people to be thanked are Prashanth PB and Anupama, for a great deal of support they have poured in. Thank you!

In the period of thesis, I have had chance to reconnect with Dr. Swetha Kumari. She has poured the help unconditionally for studying the effect of cataract on iris recognition. She took personal interest in setting up the collaboration with KIMS, Hubli and Jayapriya Hospitals, Hubli. A word of thanks would not really suffice her help! It is through her efforts I was able to interact with wonderful surgeons like Dr.Venkatram Katti, Dr.Rajshekar Dyaberi, Dr. YB Bajanthri and Dr. Varsha Bangarshettar. I must also acknowledge the staff of Jayapriya Hospital led by Basavaraj Hugar whose efforts have helped me greatly. A special thanks to everyone.

Special thanks goes to the staff of ophthalmology department, KIMS, Hubli and Dr. Priya Katti for accommodating me to conduct the research. I must acknowledge Dr. Sneha Hegde and Dr. Zain who have walked an extra mile to help me with the research, especially, Dr. Sneha Hegde has given constant company during the stay in Hubli and beyond. Thank you!

I would like to thank my mother, father, my aunt and my extended family who have always supported me in every aspect. A great part of my success should be credited to my sister, Madhushree Bylappa Raju, who has stood by me in the times of need. Thank you for being there for me every time!

Kiran B. Raja
Gjøvik

Contents

List of Figures

List of Tables

List of Algorithms

Acronyms

ANN	Approximate Nearest Neighbours
AHP	Adaptive Hybrid Pattern
BSIF	Binarized Statistical Image Features
CRC	Collaborative Representation Classifier
DET	Detection Error Trade-off
DoG	Difference of Gaussian
EER	Equal Error Rate
EVM	Eulerian Video Magnification
FAR	False Accept Rate
FLANN	Fast Library for Approximate Nearest Neighbours
FMR	False Match Rate
FNMR	False Non-Match Rate
FRR	False Reject Rate
FT	Fourier Transform
GMR	Genuine Match Rate
ICA	Independent Component Analysis
LED	Light Emitting Diode
OSIRIS	Open Source Iris Toolkit
PAD	Presentation Attack Detection
PCA	Principal Component Analysis
PIN	Personal Identification Number
ROC	Receiver Operating Characteristics
SIFT	Scale Invariant Feature Transform
SRDA	Spectral Regression Discriminant Analysis
SRC	Sparse Representation Classifier
STFT	Short Term Fourier Transform
SURF	Speeded Up Robust Features
SVM	Support Vector Machines
USIT	University of Salzburg Iris Toolkit

Chapter 1

Introduction

The need to provide controlled access has increased for applications that range from highly secure border crossing scenario to simple application such as smartphone unlocking. The access control for a secure system can be based on three principles - "what I know", "what I have" and "what I am" [59]. Paradigm of "what I know" follows knowledge based authentication where an user is expected to remember the passwords or passphrases. The passwords can vary from simple 4 digit Personal Identification Number (PIN) to long complex passwords with a combination of alpha-numeric keys along with special symbols. While a simple PIN can be vulnerable to brute force attacks, long and complex passwords are difficult to memorize. The complexity of remembering passwords further increases, when different secure services impose custom rules for choosing passwords which vary from one service to another.

The strategy of token based authentication can be classified under "what I have" authentication modes and it removes the necessity of remembering the passwords by introducing additional authenticating devices. Although, they can be used independently, it is advised to use such tokens for authentication in a complementary mode as they are prone to theft[2, 95].

Alternatively, biometrics introduced a new paradigm of "what I am" approach which is built upon the idea of using the physical or behavioural characteristics any individual possesses or exhibits. The term biometrics refers to "automated recognition of individuals based on their behavioural or biological characteristics" according to the definition given by International Organization for Standardization (ISO) [51]. A biometric system for secure access control establishes/verifies the identity of an user by capturing the measurable physical and behavioural characteristics [56, 51]. The biometric system typically uses fingerprint, face, iris, periocular, vein and voice among the most popular biometric modes in physical characteristics while gait pattern and typing pattern are popularly used among behavioural characteristics.

Biometric systems are becoming ubiquitous mode of secure access control as they remove the necessity of "what I know" or "what I have". Thus, biometric systems present convenience to users while maintaining the security level. Traditional methods of employing numeric, alphabetic or alpha-numeric pin codes to secure the device are limited to fixed length. In consequence, the security of such access control methods is very limited, when expressed as entropy of the pin code. For the most common case of arabic numerals (0-9), the symbol count is 10 and thus, the entropy per digit $H = 3.322$ bits. For a common 4-digit password the entropy is approximately 13 bits. Extending the password in length targeting at higher entropy reduces the usability of the method by creating hassle in managing multiple passwords of longer length [161]. Biometric characteristics for authentication provide higher entropy than limited length passwords [119, 32]. The iris characteristics, for instance provides 249 bits of entropy as compared to simple 4 digit numeric PIN with an entropy of 13 bits [32]. Further, the need of not remembering special passwords or carrying an additional authentication device along with good security level via the use of biometric characteristics has led to prevalent use of biometrics. The success of biometrics is exemplified by the wide range of deployment in various border crossing applications [36, 57, 152, 38] and civilian applications to provide identity in large scale application like AADHAR[44].

The popularity of biometrics has resulted in exploring new generation sensors apart from

traditional sensors. This has led to the use of regular cameras and smartphone embedded cameras that can provide sufficient quality of biometric data [57]. The use of smartphone captured biometric data has been recently explored in commercial applications [148, 48]. Motivated by the robustness and convenience of biometrics, new generation applications such as mobile banking and e-commerce applications have started employing biometric data captured from smartphones [148, 48]. At the same time, smartphones are also providing a platform for authentication in financial transactions via applications like Google Wallet, Apple Pay, Alipay and core banking services [148]. As a result, smartphones are emerging as authentication device through the use of biometric data captured in contactless manner [34, 141, 148, 48, 73, 113]. One can deduce the rise of new paradigm for contact-less biometric authentication using smartphones.

Thus, in this thesis, we explore periocular and iris characteristics for authentication applications via the images captured from smartphones operating in visible spectrum in a contactless manner. This thesis is a step towards the development of robust and reliable algorithms for using iris and periocular images captured from smartphones in visible spectrum for authentication while addressing potential presentation attacks *(a.k.a, spoofing attacks)*.

1.1 Background and Motivation

In this section, we discuss the background of the current thesis listing the major works and then derive the motivation for this thesis by identifying the potential avenues.

1.1.1 Background

The use of smartphones has increased owing to factors such as affordable cost and advanced features meeting the consumers' expectations. Smartphones with good embedded cameras have provided another reason for users to own a smartphone. Further, advanced functionalities in the smartphones have resulted in using them as a device to store personal data. Due to high amount of personal data stored in the current day scenario, the smartphone have to be secured in the first place. Additionally, smartphones can be used as devices for authentication or identification, making them both personal device and a device for authentication.

Smartphone manufacturers like *Apple, Motorola, Samsung, LG and Huawei* have integrated fingerprint sensors to capture contact-based biometric data in the new generation smartphones. To reduce the burden of unnecessary device upgrades to newer generation smartphones with embedded sensors, recent works have illustrated the employability of smartphone cameras for contactless authentication with fingerphoto recognition [113, 128, 143] with good performance for authentication. *Hoyos Labs* has launched a commercial verification system on smartphones which uses four finger photos from a single hand captured in contactless manner for verification purposes [48].

In the same direction of contactless data capture on smartphones using the embedded camera, set of initial works have focused on employing face images captured for biometric authentication [147, 141, 89]. Recent works have progressed from 2D face recognition on smartphones to 3D face recognition where the images were used to reconstruct the 3D face representation from a video recording of face from smartphone [116]. It has to be noted that these earlier works have captured the data from smartphone/mobile-phone and the actual systems were implemented on regular desktop environments [147, 141, 89, 116]. *Thus, the face based biometrics system working on the smartphone itself has not been explored to larger extent except in the very recent work [34].*

The challenges of unconstrained face data capture i.e., changes in pose, illumination and expression (PIE) are also applicable for the data captured from smartphone in real-life. When the face image is captured, the periocular characteristics are acquired inherently.

The periocular region, which is constituted by the region around the eye that includes the eye-lids, creases and folds formed by eye-lids along with the eye-brow was demonstrated as an additional characteristic for biometric recognition [102, 101] and the applicability of periocular region to improve the lower performance of face recognition under degraded data due to factors of PIE was shown earlier [102, 101]. A biometric system can leverage at-least one of the two independent characteristics to provide reliable recognition under the presence of non-uniform pose, illumination or expression in face.

To summarise, periocular region provides two fold benefits:

1. Periocular region inherently captured in iris imaging systems can be used to solve the challenges arising due to non-ideal iris imaging in visible spectrum.

2. It can also be used to complement the performance of biometric systems in unconstrained face recognition scenarios as the face image may present acceptable periocular features.

Even though the benefits provided by the periocular region is motivating for the biometric applications, the need for large scale validation of performance for smartphone captured periocular image remains open.

Another biometric characteristic present in the face region is the iris. Iris recognition has been well explored in Near-Infra-Red (NIR) spectrum due to the robust performance in terms of verification accuracy [29]. The performance of iris recognition and reported error rates with impressive billion comparisons [31] has motivated iris recognition in visible spectrum [108]. The increase in interest for visible spectrum iris recognition in the last decade has resulted in number of works probing the relevance and applicability using regular color capable cameras [108, 107, 100, 35, 106, 16, 124]. The promising performance in visible spectrum has resulted in limited works actively looking to employ various color capable imaging devices such as mobile and smartphones for iris recognition [34, 129, 71, 76, 61].

Even though the promising nature of iris recognition in visible spectrum advocates its applicability, number of challenges arise while using it in practical applications. The key factor for moving to NIR spectrum for iris recognition was due to non-visibility of the complex trabecular mesh of iris pattern in visible spectrum imaging [3, 29]. Larger concentration of melanocytes and melanin in the anterior layer and stroma makes the iris appear dark in color [134]. Alongside melanin, collagen fibrils also contribute to the color appearance of the iris where higher density of collagen results in darker iris color [134]. Further, human melanin pigment is observed to have a peak absorption at $335nm$ [79] and thus, higher density of melanin absorbs light in shorter wavelength (visible spectrum) hiding the complex trabecular mesh of iris. The quality of texture seen is inversely proportional to the pigmentation density implying very low visibility for highly pigmented iris images and vice-versa in visible spectrum. The NIR spectrum usually penetrates beyond the melanin to capture the iris pattern whereas the visible light is absorbed resulting in limited texture visibility. Thus, the challenge of resolving the texture pattern for heavily pigmented iris for images captured in visible spectrum using regular cameras including smartphone embedded cameras is a limiting factor to achieve iris recognition in visible spectrum to full potential [107, 108, 76, 34, 129]. As an alternate solution, smartphones like *Fujitsu NX F-04G, Microsoft Lumia 950, Microsoft Lumia 950 XL and Samsung Note7* have thus advanced to integrate infra-red illumination to capture iris images. However, this limits the use of the smartphone in existing form to capture the iris without a mandatory upgrade to devices with built-in infra-red sensor. *Thus, the need for innovative imaging solutions to obtain iris images with good texture from smartphones in visible spectrum for heavily pigmented iris pattern remains open.*

Further, even the mildly pigmented iris when captured in visible spectrum is impacted by number of factors such as ambient reflection, partial closure of eyes, shadows from eye lashes, out of focus imaging and motion-blur that impact the segmentation and subsequently recognition performance. The complexity of the segmentation increases for dark colored or

heavily pigmented iris as the boundary between the pupil and the iris region cannot be traced easily [76, 61, 60]. Higher segmentation accuracy with robust noise masks is a key to achieve reliable verification performance in iris biometrics [80, 61, 60]. As there is no fixed capture distance between the subject and the camera in unconstrained capture, specifically in a hand held iris acquisition from smartphone, the segmentation can be challenging due to yet another factor resulting in varying pixel spread of iris pattern on the sensor. In the case of standard NIR iris images, the known range of the iris pattern diameter is used to estimate the radius and boundary of the iris in stop-and-stare devices which does not hold good for visible spectrum iris images. The problem of varying iris diameter can also be acknowledged for unconstrained iris capture using NIR sensors embedded on smartphones like *Fujitsu NX F-04G, Microsoft Lumia 950 and Microsoft Lumia 950 XL. Therefore, an approach to estimate the range of iris boundary at-least to coarser range is needed for unconstrained iris capture from smartphone to adapt existing segmentation scheme and obtain good segmentation accuracy.*

Although, iris, periocular or face regions provide a basis for achieving secure biometric systems on smartphones by exploring discriminative information, the systems employing them can be considered reliable if they cannot be attacked at various level of operations [40, 41]. The systems can be attacked starting from the capture level at sensor to decision level in the pipeline of the biometric systems. The possibility of attacking the system can be significantly less under the human supervision at the capture level. However, under the scenarios of capture in smartphones, it is usually assumed there is no human supervision or minimal supervision if any. The risk of attacks on any biometric systems operating via smartphone to capture biometric data is relatively high owing to the ease of artefact creation due to availability of high quality face images in present day social media [103]. The face image obtained from such sources can be used to extract periocular and iris characteristics seen in visible spectrum for preparing the artefact to attack the biometric systems. *Thus, robust presentation attack detection techniques are needed to make biometric systems employing smartphone for capturing the biometric data reliable.*

1.1.2 Motivation

Based on the discussion presented in the previous section, we derive motivation for our research in this thesis which is outlined in the current section.

There are not many works reported on using iris for authentication with the data captured from smartphone (except [34, 129]) even for a considerable population with light or mildly pigmented iris. Hence, this thesis presents a comprehensive work to evaluate the applicability of smartphone captured iris data in visible spectrum. We explore reliable ways of iris recognition for data captured in smartphones by leveraging on discriminant features. Further, we present a image capture set-up using a simple LED to acquire high quality iris data in visible spectrum. The iris images obtained using the set-up demonstrate better texture visibility as compared to images acquired normally in visible spectrum and thereby are explored to achieve good verification performance.

The success of periocular recognition has not been well adopted for biometrics authentication using data captured from smartphone (except in our earlier work [73]). In this thesis, we employ periocular region for reliable authentication on smartphones. We present robust algorithms to obtain good verification performance for unconstrained periocular data captured on smartphones.

The reliability of a biometric system comes with a set of good counter-measures incorporated within the system towards presentation attacks. We present a set of texture based and video based presentation attack detection techniques for addressing attacks on ocular biometric system. In the end, we present a multi-modal system employing face, periocular and iris characteristics for secure authentication on smartphones.

It has to be noted that, a set of algorithms are tailored specifically for the smartphone platform and a couple of algorithms are implemented & tested on the desktop platforms in

this thesis. We thus, formulate the research questions specifically aligned with motivation for this thesis as listed below:

Q1 Can iris images captured using smartphones in visible spectrum be used for reliable iris recognition?

Q2 Is it possible to capture heavily pigmented iris with good texture details in visible spectrum for iris recognition?

Q3 Can periocular region captured on smartphones be used as a complementary characteristic by exploring robust features to improve the lower performance from iris in visible spectrum?

Q4 Can the ocular biometric systems operating on the smartphones be robust against presentation attacks at the capture level by employing texture and motion based algorithms?

Q5 Can the biometric system be realized to employ multi-biometric characteristics (face, periocular and iris) on smartphone to provide reliable authentication system?

1.2 Contributions

In the due-course of answering the research questions mentioned above, this thesis has resulted in major contributions outlined below.

- **Deep Sparse Filtering for Reliable Iris Verification**

  This contribution corresponds to the first research question [Q1] pertaining to reliable iris recognition.

  As the iris image is influenced by different factors such as ambient light reflection, partially closed eye-lids when captured in visible spectrum, especially on smartphone, the features extracted need to be highly discriminative to achieve reliable recognition performance. The discriminability can be achieved using multi-level features obtained using set of filters that yield multiple meaningful features. This is achieved by learning set of *Sparse Filters* to represent the iris image robustly.

  In particular, the *deep sparse filtered* feature representation achieves higher verification accuracy in unconstrained iris recognition compared to state-of-the-art methods indicating the reliable performance of iris recognition in visible spectrum using the images captured from smartphone.

  - **Related chapter for this contribution:**

    Chapter 3 - Deep Sparse Filtered Features for Iris Recognition in Visible Spectrum.

  - **Related publications for this contribution:**

    – Kiran B. Raja, R. Raghavendra, Vinay K. Vemuri, Christoph Busch, 'Smartphone Based Visible Iris Recognition Using Deep Sparse Filtering', Pattern Recognition Letters (PRL), 57(1), pp. 33-42, 2015.

- **Imaging Heavily Pigmented Iris Using Low Cost White LED Similar to Flash on Smartphones to Resolve Texture in Captured images**

  This contribution relates to the second research question [Q2] on capturing the images in the visible spectrum with good texture details for heavily pigmented iris.

The texture quality of iris is heavily dependent on the pigmentation density of melanin which absorbs the visible light. As the light is not reflected back in heavily pigmented iris, specifically for subjects originating from Asian and African ethnicities, it is challenging to capture the iris images with good texture quality in visible spectrum.

- An imaging set-up with a white Light Emitting Diode (LED) is explored to capture high quality iris images which are unlikely to exhibit trabecular mesh if captured in normal imaging mode in visible spectrum. The LED is employed to mimic the illumination of flash embedded in smartphone cameras to obtain a solution realizable on smartphones.
- The set of images obtained are bench-marked along with images from standard NIR images for the verification performance.

Better texture visibility and at-par verification accuracy in visible iris recognition is demonstrated as compared to standard NIR iris imaging method.

- **Related chapter for this contribution:**

  Chapter 4 - Imaging Heavily Pigmented Iris in Visible Spectrum using White LED.

- **Related publications for this contribution:**

  – Kiran B. Raja, R. Raghavendra, Christoph Busch, 'Iris Imaging in Visible Spectrum using White LED', In proceedings of the 7th IEEE International Conference on Biometrics: Theory, Applications and Systems-BTAS 2015, Arlington-Virginia., USA, Sept. 2015.

- **Deep Sparse Time Frequency Features for Robust Periocular Verification**

  This contribution addresses the research question on improving the lower performance of iris recognition due to unconstrained data in visible spectrum by using the region around eye. A robust feature extraction is presented to obtain the features from periocular region to address the question [Q3].

  As the challenges of iris recognition remain open in unconstrained acquisition on smartphones, we explore robust algorithms for periocular recognition. The idea of making periocular features discriminative leverages on time and frequency features obtained by convolving a set of *Deep Sparse Filters* learnt from natural images.

  - Large scale experiments are conducted on publicly available periocular databases captured on smartphones. In particular, the experiments on heavily degraded periocular data indicate the applicability of *Deep Sparse Time Frequency Features* (*DeSTiFF*) to improve the verification performance.

  - **Related chapter for this contribution:**

    Chapter 5 - Deep Sparse Time Frequency Features for Robust Verification of Periocular Images From Smartphones.

  - **Related publications for this contribution:**

    – Kiran B. Raja, R. Raghavendra, Christoph Busch, 'Collaborative Representation of Deep Sparse Filtered Feature for Robust Verification of Smartphone Periocular Images', In proceedings of 23rd IEEE International Conference in Image Processing (ICIP-2016), Arizona, US, 2016.

    – Kiran B. Raja, R. Raghavendra, Martin Stokkenes, Christoph Busch, 'Smartphone Authentication System using Periocular Biometrics', In proceedings of International Conference on Biometrics - Special Interest Group (BIOSIG-2014), Darmstadt, Germany, 2014.

- **Robust Algorithms for Presentation Attack Detection on Ocular Biometric Systems**

  The algorithms to make the authentication system robust against presentation attacks is presented as a contribution to address the fourth question [Q4].

  In order to ensure the ocular biometric systems to not compromise at the capture level by presenting artefacts, we provide a set of reliable algorithms. The algorithms are based on both texture and video based techniques to detect artefacts presented to smartphone cameras to acquire biometric characteristics.

  - The presentation attack detection algorithms in this thesis address both electronic screen and normal print attacks. The print attacks are reliably detected using texture descriptors based on *Laplacian decomposed time − frequency responses* and *Laplacian decomposed color adaptive hybrid patterns* presented in this thesis.
  - The electronic screen attacks with video replay are detected using the video based attack detection leveraging the phase information from *Eulerian Video Magnification* tailored with a specific decision module for ocular image characteristics.

  Evaluation of the texture and video based techniques on publicly available presentation attack datasets have shown robust performance when compared to state-of-art-techniques.

  - **Related chapter for this contribution:**

    Chapter 6 - Presentation Attack Detection for Ocular Biometrics on Smartphones.

  - **Related publications for this contribution:**

    – Kiran B. Raja, R. Raghavendra, Christoph Busch, 'Presentation Attack Detection using Laplacian Decomposed Frequency Response for Visible Spectrum and Near-Infra-Red Iris Systems', Proceedings of the 7th IEEE International Conference on Biometrics: Theory, Applications and Systems-BTAS 2015, Arlington-Virginia, USA, Sept. 2015.

    – Kiran B. Raja, R. Raghavendra, Christoph Busch, 'Color Adaptive Quantized Patterns for Presentation Attack Detection in Ocular Biometric Systems, In proceedings of 9th International Conference on Security of Information and Networks (SIN 2016), New-Jersy, USA

    – Kiran B. Raja, R. Raghavendra, Christoph Busch, 'Video Presentation Attack Detection in Visible Spectrum Iris Recognition Using Magnified Phase Information', IEEE Transactions on Information Forensics and Security (TIFS), 2015.

- **Multi-Biometric Secure Authentication System on Smartphone**

  This contribution presents the multi-biometric authentication system on smartphones corresponding to the question formulated in [Q5].

  We adapt and apply previously explored state-of-art algorithms to realise an authentication system for smartphones using multi-biometric characteristics. Specifically, we explore face, periocular and iris region to perform authentication using the system on the smartphone. The applicability of the presented system is demonstrated with a high verification performance on a database collected using the authentication system.

- The authentication system is demonstrated systematically to work on smartphones with reliable verification performance for each of the characteristics present in the face region - face, periocular and iris.
- The improved performance with fusion of multi-biometric characteristics at both feature level and comparison score level demonstrates the reliable performance while indicating the efficacy to use it in everyday authentication needs using smartphones.

- **Related chapter for this contribution:**

  Chapter 7 - Multi-biometric Authentication System for Smartphones Using Face, Periocular and Iris.

- **Related papers for this contribution:**
  - Kiran B. Raja, R. Raghavendra, Martin Stokkenes, Christoph Busch, 'Multi-modal authentication system for smartphones using Face, Periocular and Iris', In proceedings of IAPR International Conference on Biometrics (ICB-2015), Phuket-Thailand, 2015.
  - Kiran B. Raja, R. Raghavendra, Martin Stokkenes, Christoph Busch, 'Fusion of Face and Periocular Information for Improved Authentication on Smartphones', In proceedings of 18th International Conference on Information Fusion (FUSION 2015), Washington DC, 2015.

1.2.1 Contributions for Reproducible Research

This section lists the databases and code for the algorithms developed during the course of this thesis along with the details on availing them.

Algorithms and Code

- Matlab implementation of *Deep Sparse Filters* for extracting features from visible spectrum iris images can be availed at:
  www.nislab.no/biometrics_lab/code/deepsparse_iris.
- Matlab implementation of *Deep Sparse Time-Frequency Features (DeSTiFF)* for extracting features from periocular images captured in visible spectrum can be found at:
  www.nislab.no/biometrics_lab/code/destiff_periocular.
- Matlab implementation of two algorithms for presentation attack detection (*Laplacian Pyramid Color Adaptive Quantized Hybrid Patterns* and *Laplacian Pyramid Decomposed Frequency Response Features*) contributed from this thesis can be found at:
  www.nislab.no/biometrics_lab/code/pad_lpfr_lachp.

Databases

- Visible Spectrum Smartphone Iris (VSSIRIS) database consisting of iris images captured in visible spectrum from two different smartphones - iPhone 5S and Nokia Lumia 1020. The database can be availed from:
  www.nislab.no/biometrics_lab/vssiris_db.
- Visible Spectrum Smartphone Periocular (ViSPer) database consists of periocular images collected from three different smartphones (iPhone 5S, Nokia Lumia 1020 and Samsung S5) operating visible spectrum. The details for availing the database can be found at:
  www.nislab.no/biometrics_lab/visper_db.

- Presentation Attack Video Iris Database (PAVID) presents the Bona Fide videos corresponding to normal presentations and Artefact videos corresponding to presentation attacks captured from two different phones (iPhone 5S and Nokia Lumia 1020) in visible spectrum. The attack videos are replayed using high quality display enabled iPad. The details of availing database can be found at: www.nislab.no/biometrics_lab/pavid_db.

1.2.2 Other Contributions

This section lists additional publications on iris and periocular recognition during the course of this thesis work.

- Kiran B. Raja, R. Raghavendra, Christoph Busch, 'Biometric Recognition of Surgically Altered Periocular Region : A Comprehensive Study', In IAPR International Conference on Biometrics (ICB-2016), Sweden, 2016.
- Kiran B. Raja, R. Raghavendra, Christoph Busch, ' Weighted Comparison Score Fusion for Accurate Verification of Surgically Altered Periocular Region', 19th IEEE International Conference on Information Fusion (FUSION 2016), Heidelberg, Germany.
- Kiran B. Raja, R. Raghavendra, Christoph Busch, 'Visible Iris Imaging: A Novel Imaging Solution for Improved Iris Recognition', 12th annual International Conference on Imaging Systems and Techniques (IEEE IST-2015), Macau, China.
- Kiran B. Raja, R. Raghavendra, Christoph Busch, ' Improving Cross-Smartphone Periocular Recognition in Visible Spectrum Using Time-Frequency Features of Laplacian Decomposition', 11th International Conference on Signal-Image Technology & Internet Based Systems (SITIS -2015), Bangkok, Thailand.
- Kiran B. Raja, R. Raghavendra, Christoph Busch, ' Dynamic Scale Selected Laplacian Decomposed Frequency Response for Cross-Smartphone Periocular Verification in Visible Spectrum', 19th IEEE International Conference on Information Fusion (FUSION 2016), Heidelberg, Germany.
- R. Raghavendra*, Kiran B. Raja*, Vinay K. Vemuri, Swetha Kumari, Pierre Gacon, Emine Krichen, Christoph Busch, 'Influence of Cataract Surgery on Iris Recognition: A Preliminary Study', In IAPR International Conference on Biometrics (ICB-2016), Sweden, 2016. (* Equal Contribution).
- Kiran B. Raja, R. Raghavendra, Christoph Busch, 'Binarized Statistical Features for improved Iris and Periocular Recognition in Visible spectrum', 2nd International Workshop on Biometrics and Forensics (IWBF), 2014.
- Kiran B. Raja, R. Raghavendra, Christoph Busch, 'An Empirical Study of Smartphone Based Iris Recognition in Visible Spectrum', 7th International Conference on Security of Information and Networks (SIN 2014), Glasgow, UK, Sep. 2014.
- Kiran B. Raja, R. Raghavendra, Christoph Busch, 'Smartphone Based Robust Iris Recognition in Visible Spectrum using Clustered K-means Features', IEEE Workshop on Biometric measurements and systems for security and medical applications (IEEE-BioMS 2014), Rome, Italy, 2014.
- Kiran B. Raja, R. Raghavendra, Christoph Busch, 'Empirical Evaluation of Visible Spectrum Iris versus Periocular Recognition in Unconstrained Scenario on Smartphones', Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA-ASC 2014), Cambodia, 2014.

1.3 Thesis Outline

This thesis is organized in seven chapters and a summary of each chapter is provided in this section.

In this first chapter, the motivations, objectives and contributions of this work were briefly summarized.

Chapter 2 gives an overview of tools, state-of-art techniques for feature extraction and feature comparison. In addition, this chapter lists a set of metrics, which are used to evaluate presented systems in the rest of thesis.

In Chapter 3, the iris recognition framework for images captured using smartphone in visible spectrum is introduced. Next, an iris radius approximation method is presented along with the experiments indicating the improved verification accuracy as a result of prior processing to existing segmentation scheme. Further, *deep sparse filtering* for iris feature extraction is introduced along with the set of experiments to determine the strength of the method in achieving higher verification accuracy.

Chapter 4 introduces a new imaging set-up using a white LED to resolve the texture pattern from iris for subjects with heavily pigmented iris. The images obtained are benchmarked against the iris images captured using a standard NIR device.

Chapter 5 builds upon the challenges of iris recognition in visible spectrum for unconstrained acquisition from smartphones. An alternative approach using periocular characteristics is explored to address lower verification accuracy due to low quality iris images. Additionally, feature extraction technique leveraging on time and frequency features obtained from *deep sparse filtering* is explored to obtain discriminative features to perform periocular verification along with the large scale experiments on the publicly available periocular databases acquired using smartphones.

Chapter 6 presents the counter-measures against the potential presentation attacks at the capture level on smartphones. In particular, both the texture based and motion based approach are discussed along with the experiments comparing against the well explored state-of-art attack detection schemes. The significance of the methods and the relevance to smartphone ocular images are discussed.

Chapter 7 describes the application of the multi-biometric (face, periocular and iris) characteristics for smartphone based secure authentication systems. The experiments evaluate both unimodal and multi-modal approaches along with the feature level and comparison score level fusion to obtain a decision. In particular, the relevance of the employed methods with respect to smartphone platforms are discussed.

Chapter 8 concludes this thesis by providing a summary of the major contributions and findings. Potential directions for future work are also discussed in this chapter.

Chapter 2

Tools, Techniques and Performance Metrics

This chapter introduces the tools and techniques in current state-of-art methods used throughout the thesis. We also present commonly used metrics for evaluation of biometric systems in this chapter.

2.1 Feature Descriptors for Biometric Data

The image representations in the classification problems including biometrics have a preference to extract meaningful features rather than using raw images. The vast set of features leveraged in face biometrics can be categorized in three levels [77]. *Level* 1 features concentrate on facial characteristics that are easily observable in a face, such as skin color, gender, and the general appearance of the face while *Level* 2 features consist of localized face information that include structure of the face, relations within facial components and the shape of the face. *Level* 3 mainly focus on specific properties such as facial marks, skin discoloration, and moles. The features categorized under *Level* 2 features are highly relevant for biometric recognition of face as the features are locally derived to describe structures and spatial uniqueness [77].

Although, the categorization was mainly based on the face features for biometric recognition, the taxonomy of feature set by itself can be generalized for periocular region which is inherently present in the face region [77]. As periocular region is inherently present in the face region, *Level* 2 features hold good for the periocular recognition. Gabor wavelet features [154], Local Binary Patterns (LBP) [4], Binarized Statitical Image Features (BSIF) [64], Scale Invariant Feature Transform (SIFT) [83] and Speeded-Up Robust Features (SURF) descriptors [9] can be cited as the most prevalently used feature extraction techniques to obtain feature descriptors. The benefit of using *Level* 2 features stem due to the fact that the local features can generally be computed independently of one another in particular spatial coordinates of the image.

In this chapter we present a brief overview of the feature descriptors used in the subsequent chapters throughout this thesis. Specifically, we discuss the approaches of feature extraction with BSIF,SIFT and SURF. The choice of BSIF to obtain texture features is based on the superior performance as compared to Gabor Wavelets or LBP features reported in earlier work [64].

2.1.1 Binarized Statistical Image Features

Any of the texture feature extraction approaches can be assumed to follow the general guidelines of filtering the image with a set of linear filters and quantizing the response of filter. *Binarized Statistical Image Features (BSIF)* [64] obtained using natural image statistics [50] are used to learn the filters instead of hand-crafted filters. The set of filters obtained can be used to convolve the images and extract the texture features [64].

In order to obtain a useful set of filters s_i, the statistical independence is maximized using independent component analysis (ICA). A filter matrix W is decomposed into two parts such as

$$s = Wx = UVx = Uz \tag{2.1}$$

where $z = Vx$, and U is a square matrix of $n \times n$ dimension obtained via ICA. Matrix V is used to perform simultaneous whitening and dimensionality reduction using principal component analysis (PCA). Given whitened data samples z, the independent component analysis is used to estimate an orthogonal matrix U. Finally, given U and V, one obtains the filter matrix $W = UV$ [50, 64].

Generally, a set of filters of patch size $l \times l$ are learnt using natural images and independent component analysis (ICA) [50, 64]. Patch size l is defined as :

$$l = (2 * n + 1)$$

such that n ranges from $\{1, 2...8\}$. Unlike any other binary descriptors such as local binary pattern (LBP) or local phase quantization (LPQ), the filters are learnt using natural image statistics. If an image is represented using $I(x, y)$ and the filter is represented by $H_i(x, y)$ where i represents the basis of the filter, the linear response of the filter s_i can be given as [64]:

$$s_i = \sum_{x,y} I(u, v) H_i(u, v) \tag{2.2}$$

where x, y represents the dimension of image and filter. The response is further binarized based on the obtained response value. If the linear filter response is greater than the threshold, a binarized value of 1 is assigned. This operation can be formulated as [64]:

$$b_i = \begin{cases} 1, & \text{if } s_i > 0 \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

The obtained responses at different basis is used to construct the new gray code for the pixel value.

Since the descriptors are constructed using the filters learnt through set of natural images, the response of the filters achieved are maximally independent in terms of statistical significance [64]. Based on the descriptors being derived from the statistics of the image, the constructed feature set of an image is termed as *Binarized Statistical Image Features* [50, 64].

Figure 2.1 illustrates a sample image and the corresponding filter response obtained using BSIF filters of different patch (scale) size with a basis value of 8. It can be seen that the features obtained in each different patch size varies from the other as shown in Figure 2.1(b)-(d).

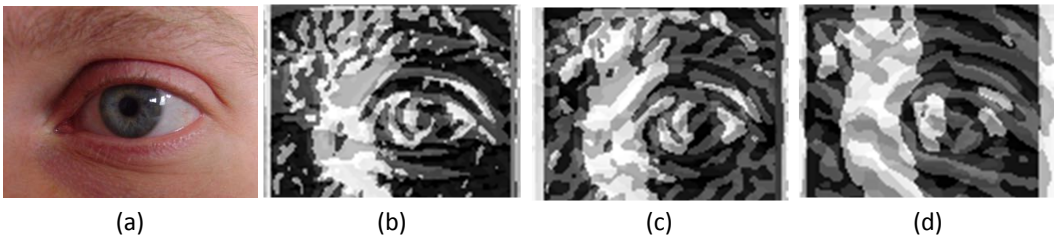


Figure 2.1: Features extracted with various patch sizes (a) Original image; (b)-(d) Features obtained using BSIF filters with patch size 9, 13 and 17

2.1.2 Scale Invariant Feature Transform

Scale Invariant Feature Transforms (SIFT) are based on extracting the key-points and computing the descriptors from the images across the scale space [82]. The images are decomposed using Difference of Gaussian (DoG), which is based on Gaussian blurring of

images. The DoG for an image is first found in different octaves, which are then searched for local extrema in both scale and space [83]. The key-points are localized which is followed by orientation assignment for each key-point. Finally, a neighbourhood of 16×16 is divided into sub-blocks of 4×4 resulting in 128 values which is considered as the descriptor for a particular key-point. The use of SIFT for periocular recognition was well illustrated for biometric applications earlier [13, 101, 158, 73, 75, 74].

2.1.3 Speeded Up Robust Features

Speeded Up Robust Features (SURF) are another class of features based on key-points [9]. As compared to SIFT, SURF was primarily designed to overcome the slower processing time to obtain features at different key-points localized. SURF is based on the Box filters to approximate the Laplacian of Gaussian (LoG) to obtain scale space decomposition of the image. Further, it can be done in parallel for different scales and SURF relies on determinant of Hessian matrix for both scale and location.

SURF also uses wavelet responses in horizontal and vertical direction for a neighbourhood of size 6×6 pixels to obtain the orientation information. The dominant orientation is estimated by calculating the sum of all responses in a sliding orientation window of angle 60 *degrees*. A neighbourhood of size 20×20 pixels is considered for detecting the key-point and the image is divided into 4×4 subregions where horizontal and vertical wavelet responses form feature descriptor with total 64 dimensions. With the lower dimension as compared to SIFT, the speed of computation and comparison is improved for SURF.

SURF was well explored for periocular recognition on regular imaging devices [101, 158] and smartphone platforms in earlier works [73, 75, 74].

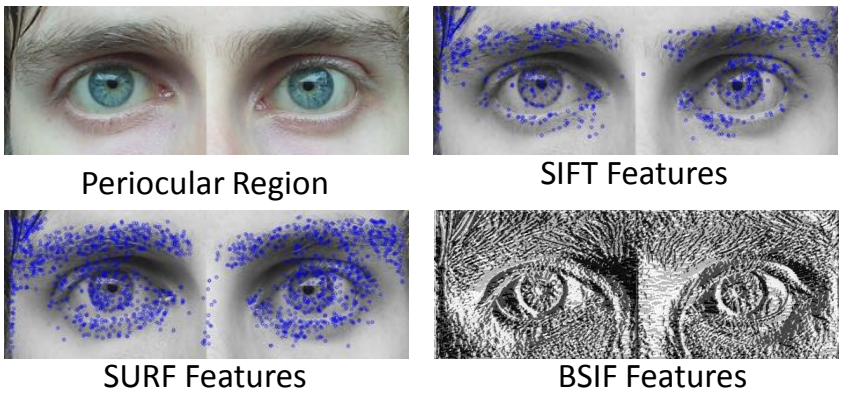


Figure 2.2: Illustration of features extracted using SIFT, SURF and BSIF for a sample periocular region.

Figure 2.2 illustrates the features detected for a sample periocular image using SIFT, SURF and BSIF feature extraction techniques.

2.2 Feature Classification

The obtained features of two images need to be measured for the similarity (or dissimilarity). The scores of similarity (or dissimilarity) are used as the genuine score in biometrics if they stem from the same class or subject for a particular characteristics else it is considered as impostor score. The plausible choice to determine the scores can employ a simple distance metric to measure the similarity (or dissimilarity). As in many cases, simple distance measures do not provide robust scores of similarity (or dissimilarity), advanced

approaches have been devised to leverage the knowledge from machine learning domain which have resulted in better distance computations. The robust classification approaches using dictionary representation of features and classification using sparse representation or collaborative representation have demonstrated better classification accuracy in many applications including biometrics [163, 156]. In this thesis, we have employed both simple distance measures and dictionary classification approaches discussed in this section.

Generally, the computational complexity increases significantly when a large set of features from large sized images are used for computation. A compact representation using histograms of many features, especially in textures have shown similar performance as compared to engaging the set of features in the large feature space [4, 162, 64]. Thus, it can deduced that the features can be represented in compact histogram representation while not loosing significant performance in comparing two images. In this thesis, we have employed the histogram representation of texture features for computing the similarity (or dissimilarity) scores.

2.2.1 Bhattacharya Distance Measure

Histogram features for two different images can be measured using distance measures such as *Euclidean* distance, χ^2 distance, *Bhattacharya* distance among many others. We employ the *Bhattacharya* distance which is related to the *Hellinger* distance to determine the similarity between two histograms. Let the histogram for image i be represented by H_i and histogram for image j be represented by H_j, then the *Bhattacharya* distance between the histograms is given by :

$$B(H_i, H_j) = \sqrt{1 - \frac{1}{\sqrt{\overline{H_i} * \overline{H_j} * N^2}} \sum_k \sqrt{H_i(k) * H_j(k)}} \tag{2.4}$$

where N is the number of histogram bins and $\overline{H_i}$, $\overline{H_j}$ present the mean values of the histogram of image i and j. A score of value 1 is obtained for two histograms which are exactly same. The *Bhattacharya* distance (B) is symmetric between two distributions p and q which can be indicated as in Eqn. 2.5.

$$B(p, q) = B(q, p) \tag{2.5}$$

2.2.2 Approximate Nearest Neighbours (ANN)

The key-points and descriptors obtained from both SIFT and SURF can be compared using distance metrics used in vector spaces. The large size of descriptors can result in heavy computational complexity. An improved way of comparing such large descriptors relies on tree based indexing and matches [93, 18]. Fast Library for Approximate Nearest Neighbours (FLANN) [93, 18] provides a set of optimized feature comparison methods which can automatically select indexing trees for features using Approximate Nearest Neighbours (ANN).

FLANN can be realised using Hierarchical K-means Tree for SIFT and SURF feature comparison with a predefined number of nearest neighbours. Nearest neighbours are discovered by choosing to examine the branch-not-explored nodes along the way in the tree. Specifically, the FLANN with a Best-Bin-First approach to form Approximate Nearest Neighbours (ANN) from Hierarchical K-Means Tree with 2 nearest neighbours are predominantly in many of the object recognition problems [83]. The matched key-point is only counted as a good match when the second nearest neighbour is 0.8 times farther than the first from the set of obtained matches [83].

The fraction of good matches with respect to the total number of matched points can be treated as the similarity score. Higher number of good matches indicate high similarity of the two samples (images).

2.2.3 Sparse Representation Classification

Any problem in classification which uses labelled training samples from k distinct classes relates to correctly determining the class to which a new test sample belongs. Sparse representation of signals has proven to be highly effective in classification problems [156]. Sparse representation analyzes a signal y over a dictionary Φ such that $y \approx \Phi\alpha$ and α is a sparse vector of features. The sparsity of α can be measured by $l_0 - norm$, which is equivalent to the number of non-zeros in α. An approximation of $l_1 - minimization$, is used as the closest convex function to $l_0 - minimization$ to save the expensive computation and thus solving a sparse representation problem can be measured as $min_\alpha \parallel \alpha \parallel_1 s.t. \parallel y - \Phi\alpha \parallel_2 \epsilon$, where ϵ is a small regularization constant. Different approaches for the $l_1 - minimization$ can be used which include Spectral Gradient Projection, Homotopy, Iterative Shrinkage-Thresholding, Proximal Gradient, and Augmented Lagrange Multiplier (ALM). The steps in the classification approach can be outlined as:

1. Given the set of known samples for a particular class, training representation T is created. Similar training class representation is constructed for different classes in the classification task.
$$T = [T_1, T_2, ..., T_C] \in R^{N \times (n_u . C)} \tag{2.6}$$
where n_u denotes the number of training samples for each class and N indicates the dimension of the features for C different classes.

2. Given the feature vector of the a new sample T_e, one can assume a linear relation between the training class representation (T) and a testing sample (T_e) as:
$$T_e = T \times \alpha \tag{2.7}$$
where, $\alpha = [\alpha_1, ..., \alpha_{1n_u} \mid, \mid \alpha_2, ..., \alpha_{2n_u} \mid, \mid \alpha_C, ..., \alpha_{Cn_u}]$

3. The established relation of linear combination between T and T_e can be modelled as a l_1 minimization problem such that:
$$\hat{\alpha} = \arg \min_{\alpha' \in R^N} \|\alpha'\|_1 Te = T\alpha' \tag{2.8}$$

4. Compute the reconstruction error corresponding to the testing sample and training class representation.
$$err(y) = \|Te - \Pi_C(\alpha')\|_2 \tag{2.9}$$
where y represents the number of test samples for a class and Π represents the combined representation of the training class .

5. The minimum reconstruction error of all the sparsely represented signals is used to map the test sample to the particular class.

The obtained scores provide the closest distance to the number of different classes which can be treated as dissimilarity score. The higher is the score with respect to a particular class, the farther away is the test sample to that training class representation.

2.2.4 Collaborative Representation Classification

A collaborative representation overcomes the deficiencies of sparse representations, when the class representation is imbalanced while significantly reducing the computation time and resource [163]. The features are represented collaboratively from different classes to learn the dictionary Φ. The learnt dictionary Φ is used to classify the features of test sample by imposing regularised l_2 norm [163]. The distance between the testing sample and

the dictionary representing all different classes is then computed by projecting it as Least Square Regression problem [163]. The problem can be represented as:

$$\widehat{f} = arg\ min_{\alpha} \| f - \Phi\alpha \|_2^2 + \lambda \parallel \alpha \parallel_2^2 \tag{2.10}$$

where the Φ is the learned dictionary, α is coefficient vector and λ is the regularization parameter.

The score obtained from the comparison to the dictionary samples is treated as a dissimilarity score which is lower for the samples stemming from the same class in the dictionary representation.

2.2.5 Support Vector Machines (SVM)

Support Vector Machine (SVM) [146, 24] is a classifier which is able to learn a class discriminating hyperplanes separating the set $\chi = \{(x_k, y_k)|k = 1 \dots S\}$ of training samples in $\mathbb{R}^{dx\{-1,1\}}$, while minimizing its generalization error on unseen samples belonging to any of the classes. The minimization of the generalization error is performed via maximization of a decision boundary margin between the classes. SVM can be generalized as maximal margin classifier where the boundary is learnt by the samples from two classes forming the extreme margin such that rest of the samples lie within the boundary and margin boundaries are referred as support vectors.

The training of a SVM is realized by estimating the parameters via real valued linear function $f : \mathbb{R}^d \Rightarrow \mathbb{R}$. The learnt classifier is used to classify the unknown samples or testing samples. For each input sample x_k a score is computed using the linear function $f(x_k)$. This function can be also represented in a dual form, where it is parameterized by the support vectors and a set of Lagrangian multipliers associated with them.

Higher separability between two classes can be achieved by projecting the samples(data) into high dimensionality space which is computationally expensive. However, the dual forms of support vectors also support the computation of score for an input sample x_k through its inner products with the support vectors. The dual form thus allows to bypass the computationally expensive high dimension projection simply by using kernel functions. Among the different kernel functions to learn discriminative margins, linear, Radial Basis Function (RBF), Polynomial, Histogram Intersection and χ^2 are predominantly used.

2.2.6 Spectral Regression Discriminant Analysis (SRDA)

Spectral Regression Discriminant Analysis (SRDA) is another classifier that learns the discriminating boundary by maximizing the between-class covariance and simultaneously minimizing the within-class covariance [22]. SRDA uses spectral graph analysis and casts discriminant analysis into a regression framework which is efficient in computation as compared to Linear Discriminant Analysis(LDA) methods. SRDA performs discriminant analysis on the data projected in space induced by a non-linear mapping. Further, the subspace is analysed using spectral graph analysis followed by regression which facilitates both efficient computation and the use of regularization techniques. Specifically, SRDA only needs to solve a set of regularized least squares problems and there is no eigenvector computation involved, which is a huge save of both time and memory [22]. SRDA solves the classification problem by projecting it as a set of regularized least squares problems which includes no computation of eigenvector [22].

2.3 Iris Recognition Tools

This section presents three iris recognition tools, which are used in this thesis. Of the three systems, two tools are publicly available academic implementation which are distributed as University of Salzburg Iris Toolkit (USIT v1) [122] and Open Source Iris Toolkit (OSIRIS

v4.1) [135]. The third is a commercial-off-the-shelf iris recognition tool from Neurotech Inc.[148].

2.3.1 University of Salzburg Iris Toolkit (USIT v1)

USIT v1 Toolkit [122] provides state-of-the-art reference systems, which serve as adequate starting point for new research in iris biometrics for NIR systems. The toolkit provides two different segmentation techniques tailored for NIR domain which include segmentation schemes with Contrast-adjusted Hough Transform and Weighted Adaptive Hough and Ellipsopolar Transform.

The toolkit also provides seven different reimplemented feature extraction schemes prevalently used in NIR domain such as 1D-LogGabor [91], Complex Gabor features [29], feature extraction based on local variations [85], context based iris features [121], feature extraction based on intensity variations [120] and features based on cumulative-sum-based change analysis [78]. Along with the feature extraction methods, the toolkit provides corresponding comparators to be used for each different feature extraction techniques.

2.3.2 Open Source Iris Toolkit (OSIRIS v4.1)

OSIRIS v4.1 [135] is another freely available toolkit designed to work for iris images captured in NIR domain. The segmentation scheme is based on the robust contour localization based on Daugman's work [29]. Further, the toolkit estimates noise mask for iris pattern robustly using Viterbi search and anisotropic diffusion. The toolkit provides complex 2D Gabor features reimplemented from Daugman's work [29] along with the Hamming Distance (HD) based comparator.

2.3.3 VeriEye - Neurotech

Neurotech Inc.[148] provides a widely used commercial iris recognition tool in secure access control applications. The technical details of the tool are not disclosed to the best of our knowledge.

2.3.4 2D Gabor Features

Gabor filters belonging to the class of bandpass filters are generally used for feature extraction and texture analysis in many other applications and predominantly used in iris recognition [29]. The impulse response of a Gabor filter is formed by multiplying a complex sinusoidal carrier with a Gaussian envelope which can be expressed as $g(x, y)$:

$$g(x, y) = w(x, y) * s(x, y) \tag{2.11}$$

where $s(x, y)$ constitutes complex carrier signal and $w(x, y)$ forms the Gaussian envelope, represented as :

$$w(x, y) = \mathrm{e}^{\frac{-(x^2+y^2)}{\sigma^2}} \tag{2.12}$$

The complex sinusoidal signal is represented mathematically as:

$$s(x, y) = \mathrm{e}^{j(2\pi(u_o x + v_o y) + \psi)} \tag{2.13}$$

with u_o and v_o representing the frequency of the horizontal and vertical component of complex sinusoid. The term ψ represents the phase shift. The complex carrier signal can be separated into real and imaginary parts such that :

$$Re(s(x, y)) = Cos(2\pi(u_o x + v_o y) + \psi) \tag{2.14}$$

$$Im(s(x, y)) = Sin(2\pi(u_o x + v_o y) + \psi) \tag{2.15}$$

The real and imaginary components are the even symmetric and odd symmetric components. The Gabor filtered components of the symmetric and asymmetric components are:

$$g_{sym}(x,y) = \mathrm{e}^{\frac{-(x^2+y^2)}{\sigma^2}} * Cos(2\pi(u_o x + v_o y) + \psi) \tag{2.16}$$

$$g_{asym}(x,y) = \mathrm{e}^{\frac{-(x^2+y^2)}{\sigma^2}} * Sin(2\pi(u_o x + v_o y) + \psi) \tag{2.17}$$

The 2D Gabor filter over the image domain (x, y) is:

$$G(x,y) = \mathrm{e}^{-\pi[(x-x_o/\alpha^2)+(y-y_o/\beta^2)]}\mathrm{e}^{-2\pi i[u_o(x-x_o)+v_o(y-y_o)]} \tag{2.18}$$

where (x_o, y_o) specify position in the image, (α, β) specify the effective width and length, (u_o, v_o) specify modulation[29].

The phase information of Gabor filter response is quantized into four levels corresponding to all quadrants in the complex plane which are represented using two bits of data [29]. The technique originally proposed by Daugman [29] is widely employed in many commercial applications and serves as a baseline for iris recognition systems. In this thesis, we employ the OSIRIS v4.1 implementation of Daugman's 2D Gabor filtering approach.

2.3.5 Hamming Distance

In the context of traditional information theory, Hamming Distance (HD) is the measure of differences corresponding to each position of two strings of equal length. Hamming distance measures the minimum number of substitutions/changes required for converting one string into another. The concept of string matching can be easily adapted to iris template matching where iris templates are typically in the form of binary strings.

Daugman [28] proposed to Hamming distance to measure the similarity between two iris templates which are represented in binary form. Simple boolean Exclusive-OR (XOR) operator is applied to two iris codes. The degree of disagreement with pair of bits are detected using XOR operator. Originally proposed Hamming distance measure also uses noise mask for two iris codes to compensate the errors introduced by eyelashes, eyelids, specular reflections, or other noise. In this case, the iris mask is combined with the iris code using a logical AND operation and the result is used to measure the difference [28].

If the two iris code bit vectors are denoted as $codeA$ and $codeB$ with corresponding mask bit vectors denoted as $maskA$ and $maskB$, Hamming Distance HD is given as:

$$HD = \frac{\parallel (codeA \otimes codeB) \cap maskA \cap maskB \parallel}{\parallel maskA \cap maskB \parallel} \tag{2.19}$$

The resulting HD is a fractional measure of dissimilarity with 0 being a perfect match and higher value corresponding to non-match.

2.4 Performance Metrics

This section presents the commonly used metrics to report the performance of a biometric system. First, general performance metrics is introduced followed by metrics used in presentation attack detection.

2.4.1 Performance Metrics in Biometric Authentication System

In a biometric system, the biometric data such as face, iris, fingerprint, periocular region is collected to enrol a subject into the system. When the subject needs to identified (1 *to many*) or subject claims an identity (1 *to* 1; corresponds to verification), the biometric data collected

is referred as probe data. In both cases of enrolment and probe data, features from the image are extracted such that they are used as templates. The templates from both enrolment sample and probe are used to compare the biometric samples.

If the biometric samples (i.e, templates) are from the same biometric characteristic of the same biometric data subject, the comparison is deemed to be a *mated comparison* and the corresponding comparison result is treated as *genuine* score. When same biometric characteristic of the different biometric data subject are compared, the score is treated as *impostor* score. Two errors corresponding to genuine and impostor scores are [54]:

1. *False-Match-Rate (FMR)*: The proportion of the completed biometric non-mated comparison trials that result in a false match.

2. *False-Non-Match-Rate (FNMR)*:The proportion of the completed biometric mated comparison trials that result in a false non-match.

FMR and FNMR report performance of algorithms employed in the biometric system under the assumption that there is no failure to acquire (FTA) which is related to the interaction of the data subject with the biometric capture device.

The Genuine Match Rate (GMR) is a metric, which is derived from FNMR at a certain FMR [54].

$$GMR = 1 - FNMR \tag{2.20}$$

Higher values of GMR at a specified FMR imply a superior performance of verification accuracy for a particular algorithm. It is common practice to vary the operating points of FMR and present the performance of a system at various operating point in Receiver Operating Characteristic (ROC) curve.

In this thesis, we focus on verifying if the subject claiming his identity is the same as the subject enrolled and thus follow the verification protocol. In a verification scenario, the decision making process consists of comparing a score $c(\chi_{probe} \mid \psi_i)$ threshold with Θ to output a decision of acceptance or rejection, where χ_{probe} represents biometric sample for a probe subject and ψ_i presents the enrolment samples of subject i in the biometric database. When the score c is higher than the pre-determined threshold, it is accepted as the biometric data stemming from probe sample χ_{probe} and is considered to be the same as the one in enrolment samples of ψ_i.

Common errors in the verification system can be classified in two groups:

1. *False Acceptance (FA)* - The biometric system accepting the *impostor* as *genuine* subject.

2. *False Rejection (FR)* - The biometric system rejecting the *genuine* subject as *impostor*.

Thus, two metrics can be used to quantify the verification errors considering FTA (results due to improper interaction of data subjects) is:

1. *False Acceptance Rate (FAR)* - The false accept rate is the expected proportion of zero-effort non-genuine transactions that will be incorrectly accepted. A transaction may consist of one or more non-genuine attempts depending on the decision policy. The proportion of *False Accept (FA)* with respect to number of impostor attempts can be expressed as:

$$FAR(\Theta) = \frac{\mid FA \mid}{\mid impostor\ attempts \mid} \tag{2.21}$$

Taking the FTA into account, FAR can be computed as:

$$FAR = FMR \times (1 - FTA) \tag{2.22}$$

2. *False Rejection Rate (FRR)* - The false reject rate is the proportion of genuine verification transactions that will be incorrectly denied. A transaction may consist of one or more genuine attempts depending on the decision policy. The proportion of *False Reject (FR)* with respect to number of genuine attempts can be expressed as:

$$FRR(\Theta) = \frac{\mid FR \mid}{\mid genuine\ attempts \mid} \tag{2.23}$$

Taking the FTA into account, FRR can be computed as:

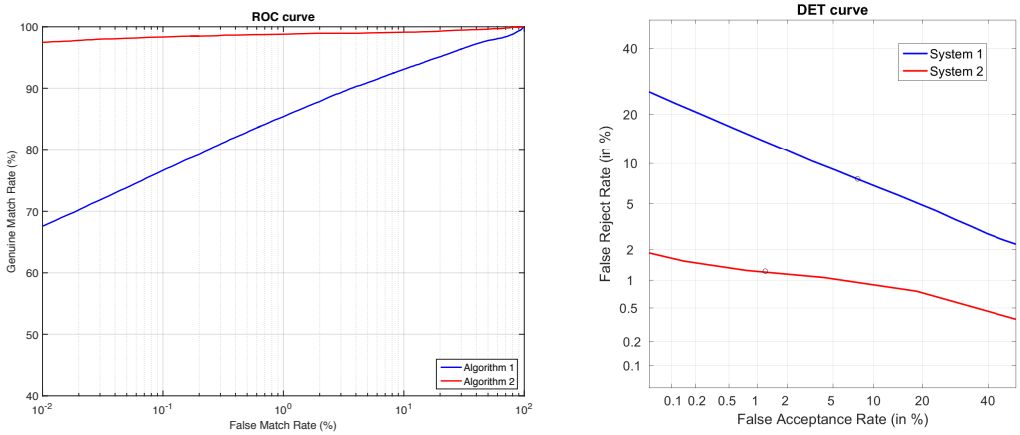
$$FRR = FTA + FNMR \times (1 - FTA) \tag{2.24}$$

where Θ is the decision threshold.

Half-Total-Error-Rate (HTER) is another metric which is the average of FAR and FRR as defined Equation 2.25.

$$HTER(\Theta) = \frac{FAR(\Theta) + FRR(\Theta)}{2} \tag{2.25}$$

Decision threshold Θ is usually selected such that the FAR is equal to FRR. Indicative metric at which the FAR equals FRR is commmonly reported as Equal Error Rate (EER).



(a) Example ROC curve depicting the performance of two algorithms.

(b) Example DET curves presenting the performance of two biometric systems.

Figure 2.3: Illustration of ROC curves and DET curves.

The performance of a system represented in EER is derived for a specific operating point which relates to FAR and FRR being in the same range. Varying the Θ has a direct implication on the error rates where if FAR is decreased, FRR is increased and vice versa. The Detection Error Trade-off (DET) curve presents FAR against the FRR, using axes that are scaled non-linearly by their standard normal deviates [90]. Figure 2.3 presents a sample illustration of ROC and DET curves. The ROC curve is shown in semi-log scale in x-axis to have detailed performance of algorithms at various FMR as illustrated in Figure 2.3(a).

Further, in a scenario evaluation involving the sensor to capture the data, Failure-To-Capture (FTC) needs to measured which indicates number of failed attempts to capture the data by sensor. The captured images can result in Failure-To-eXtract (FTX) biometric data from the captured image. The failure to extract the templates from the captured biometric data to enrol into gallery/reference database results in Failure-To-Enrol(FTE). Thus, the error metrics should account for all these while reporting the performance and thereby present Generalized False Accept Rate (GFAR) along with Generalized False Reject Rate (GFRR).

$$GFAR = FMR \times (1 - FTA) \times (1 - FTE) \tag{2.26}$$

$$GFRR = FTE + (1 - FTE) \times FTA + (1 - FTE) \times (1 - FTA) \times (FNMR) \tag{2.27}$$

Thus, an indicative error for the overall biometric system can be reported using GFAR and GFRR resulting in Generalized Equal Error Rate (GEER) which is defined as a point where GFAR equals GFRR.

2.4.1.1 Interpreting Performance Metrics

This section summarizes interpretation of performance metrics of a biometric system.

1. The lower EER or GEER indicates a better biometric system.
2. The higher GMR indicates superior performance of biometric system at a particular FMR.
3. Analogous to GMR, the lower FNMR indicates better performance.
4. In a practical system, lower FAR is expected with minimal/reasonable FRR.

2.4.2 Performance metrics for Presentation Attack Detection

The performance of a stand-alone presentation attack detection algorithm should be disclosed in terms of Attack Presentation Classification Error Rate (APCER) and Bona fide Presentation Classification Error Rate (BPCER) [53]. APCER is defined as the proportion of attack presentations incorrectly classified as normal/bona fide presentations in a specific scenario while BPCER is defined as the proportion of normal presentations incorrectly classified as attack presentations in a specific scenario [53].

$$APCER = \frac{\mid \text{Attack Presentations classified as Bonafide} \mid}{\mid \text{Total presentations} \mid} \tag{2.28}$$

$$BPCER = \frac{\mid \text{Bonafide Presentations classified as Attack} \mid}{\mid \text{Total presentations} \mid} \tag{2.29}$$

Indicatively, performance of a PAD algorithm can be presented as Average-Classification-Error-Rate (ACER) which is described as the average of APCER and BPCER. ACER is defined by the Equation 2.30 as:

$$ACER = \frac{APCER + BPCER}{2} \tag{2.30}$$

2.4.2.1 Interpreting PAD Metrics

This section presents the interpretation of PAD metrics in the context of biometric system.

1. A lower ACER indicates reliable performance of a PAD mechanism.
2. Typically, the biometric system should exhibit low APCER with low BPCER.

Chapter 3

Deep Sparse Filtered Features for Iris Recognition in Visible Spectrum

Iris recognition in visible spectrum using smartphone cameras is being actively researched for authentication applications. This chapter discusses two challenges of iris recognition, one related to segmentation and another related to feature extraction for using iris data captured in the visible spectrum using smartphone embedded cameras. A new algorithm to estimate iris diameter for the data captured in unconstrained scenario is proposed in this chapter to make an existing open source segmentation scheme robust. Additionally, a new feature extraction technique to obtain robust and discriminant features to achieve higher verification accuracy is presented. Proposed approaches are validated experimentally using publicly available databases consisting of images captured from smartphone embedded cameras in visible spectrum.

3.1 Introduction

The human *iris* is a thin circular muscle structure in the eye, which controls the diameter and size of the *pupil* which is a central opening to regulate the amount of light entering the retina. The observed size of the iris is approximately 12 *mm* in diameter including pupil. The bright light constricts the pupil while it dilates in dark illumination condition. The constriction and dilation of the pupil is primarily controlled by the iris sphincter muscle and iris dilator muscle as illustrated in the Figure 3.2.

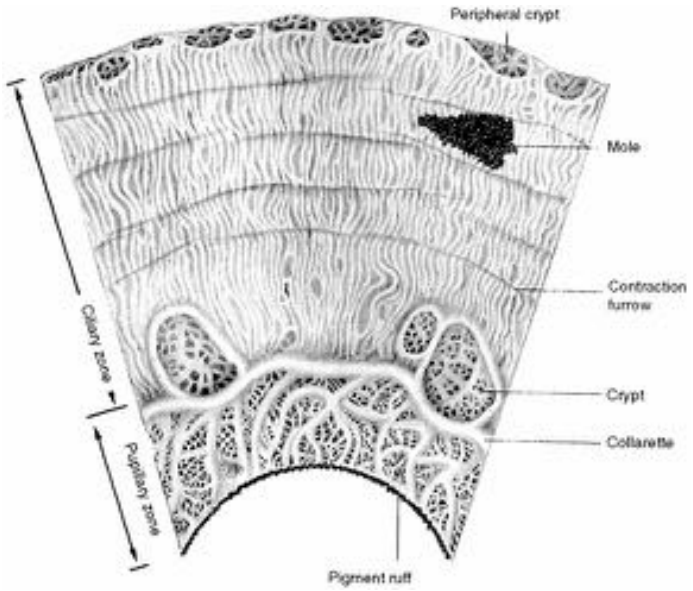


Figure 3.1: Cross section of iris
Source : Text book of Gray's Anatomy, 5th Edition

The components of iris mainly include Fuchs's crypts, nevi, Wolfflin nodules and contraction furrows [134]. The iris can be divided primarily into pupillary zone and ciliary zone as illustrated in Figure 3.1 whose components are well depicted within the picture. The arrangement of these components in different possible ways result in the unique structure of any individual's iris pattern [134]. The complexity, randomness and uniqueness of the iris structure is also heavily contributed by the blood vessels present in the iris region. The cross-sectional view illustrated in the Figure 3.2 shows the contribution of vessels to the structure of complex structure of iris.

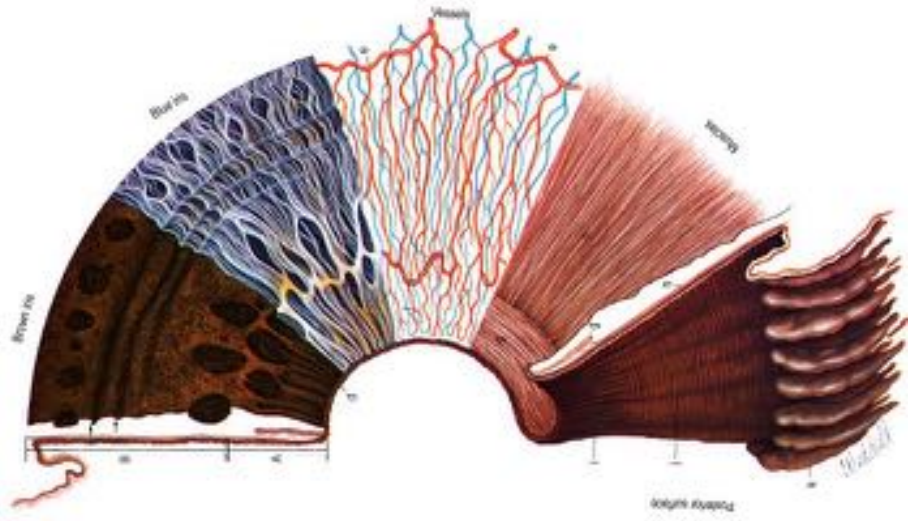


Figure 3.2: Structural anatomy of iris
Source : Text book of Gray's Anatomy, 5th Edition

The color of iris adds another dimension for uniqueness of iris appearance. Abundant presence of the melanocytes and melanin in the anterior layer and stroma makes the iris appear brown. The lower concentration of such pigmentation causes the iris to appear lighter in color as it is the case for blue irises [134]. Along with the melanin, collagen fibrils play a vital role in the appearance of the iris in darker or lighter color. Thus, darker irises have higher density of collagen along with higher density of the melanin [134].

Human melanin pigment is observed to have a peak absorption at $335nm$ [79]. It was thus well advised to employ Near-Infra-Red (NIR) illumination in the wavelength of $700nm - 900nm$ to obtain optimal iris texture in the images captured [33, 29, 20]. The reflectance of the iris is observed to be relatively constant over the wavelengths $700nm - 900nm$ [29]. Further, the NIR light in the range of 780 nm to 840 nm is highly effective in resolving the iris pattern as the light in this range can be scattered by collagen fibrils, melanin pigments in the anterior layer and stroma which typically absorb the illumination in shorter wavelengths corresponding to visible spectrum [29]. Higher density of fibrils causes the light to be absorbed heavily and lower density absorbs the light partially. Therefore, using visible spectrum light to capture iris does not result in superior quality image with textural details as observed in NIR illumination for heavily pigmented iris [29]. Thus, traditional iris biometric systems employ NIR light to illuminate the iris to capture the unique and random pattern [33, 29].

Iris recognition has been well explored in Near-Infra-Red (NIR) spectrum due to the robust performance in terms of verification accuracy [29]. The success of iris recognition in NIR spectrum with impressive error rates in billion comparisons [31] has led to a set of earlier works to investigate the feasibility of iris recognition in visible spectrum using regular color capable cameras [107, 57, 139, 137, 16]. Motivated by the rich

texture available in NIR domain, an earlier work has investigated the visibility of iris texture information across wide range of spectrum using multi-spectral illumination [17]. It was well demonstrated with a limited dataset that the information from RGB channel can achieve a good Genuine Accept Rate (GAR) at lower False Accept Rates (FAR) [17]. Based on the biometric performance demonstrated in earlier works, visible spectrum iris recognition is gaining importance as an alternative/viable option with considerable performance [108, 57, 139, 137, 16, 15, 72, 117, 76, 118, 71] owing to fact that a regular RGB camera can be used to capture the iris image without the necessity of NIR illumination. Although the visible spectrum iris systems have not been benchmarked yet against the NIR spectrum iris systems to the size of billion subjects [31], the visible spectrum iris systems are known to serve good purpose for on-the-move, at-a-distance and unconstrained iris imaging scenarios[107, 108]. Further, as the camera on smartphones are similar to regular color capable RGB cameras used for iris recognition, there is a new interest in the direction of smartphone based authentication using iris biometric characteristics [34, 61, 76].

The advantages of constrained iris acquisition or contact based iris acquisition within a fixed imaging volume in NIR imaging set-up is that the range of the iris diameter does not vary largely. This *apriori* information of a particular range determined with the help of set of images captured in fixed capture volume results in accurate segmentation in scenarios involving fixed stand-off distance [29, 80]. In an unconstrained imaging condition such as iris image captured from smartphones, the range of the iris and pupil diameter vary largely. Some of the factors influencing the images include the intensity of the illumination (i.e. incoming light on the eye) which results in dilation and constriction of the pupil. The size of pupil varies on the basis of amount of light entering the eyes as it cannot be controlled in an unconstrained outdoor illumination condition and thereby the ratio of iris to pupil diameter varies which is usually regarded as $0.7 : 0.3$ in NIR images [135]. Further, the optical resolution of the imaging device results in image pattern with pixels which are not bound to certain limited range as the smartphones provide cameras with different resolution. The focal length and the distance of the camera from the capture subject also presents another variable factor which results in non-ideal iris images. Further, different ways of interaction of subjects with the smartphones while capturing the image and the placement of camera on smartphones also influence the iris images. Thus, the *apriori* knowledge of iris diameter cannot be assured due to factors mentioned above restricting the adaptation of segmentation schemes designed for NIR iris images. In this chapter, we present an approach for approximating the iris diameter range for images captured in unconstrained scenarios, specifically for the data captured from smartphones. The approximated diameter range is further provided to OSIRIS v4.1 [135] to adapt it for visible spectrum iris recognition. The proposed approach is experimentally validated as discussed in upcoming sections.

Another key challenge in the visible spectrum iris recognition is the low visibility of trabecular mesh due to various factors which result in degraded image quality, lesser usable area from iris due to partial closure of eyes and ambient light reflection [108]. For the iris images captured in NIR spectrum, features from 2D Gabor wavelets have proven to perform very well [29] whereas similar performance cannot be expected when the images are of sub-optimal quality in visible spectrum [107]. In this chapter, we present a new iris feature extraction method based on *deep sparse filtering* to obtain robust features. The features are further experimentally validated for the performance as described in upcoming sections.

This chapter, therefore, focuses on the aforementioned two aspects of the iris recognition framework for the data captured using smartphones operating in visible spectrum. In rest of the chapter, Section 3.2 presents the general pipeline of iris recognition in visible spectrum. It also discusses two of the many challenges related to segmentation and robust feature extraction in adapting iris recognition for data captured from smartphone. A new approach for improving the segmentation accuracy is presented in the same section. Further, a brief description of databases used in this work are presented in Section 3.4. The databases

are used to validate improved segmentation approach presented in this chapter. Later, in the Section 3.3, we present a new feature extraction scheme for extracting robust features from visible spectrum iris images. The set of experiments and results for demonstrating the superiority of new feature extraction scheme are presented in Section 3.5. Finally, Section 3.6 discusses the key observations from the experimental validation carried out in this chapter.

3.1.1 Contributions

The key contributions of this chapter can be listed as below:

1. Presents a new approach to adapt the open source segmentation scheme - OSIRIS v4.1 to iris images captured in unconstrained scenario on smartphones operating in visible spectrum.

2. Presents a new feature extraction scheme to obtain robust and discriminant features from iris images using *deep sparse filtering* when the iris texture is not fully utilizable due to number of factors such as partial closure of eyes, ambient light reflection, shadow from eye-lashes and low texture visibility.

3. The code corresponding to the newly proposed feature extraction is freely distributed for non-profitable scientific and academic research purposes [76].

4. Further, this chapter presents a new iris image database of 28 subjects that is collected during the course of this thesis. The database is distributed freely for the purpose of non-profitable scientific and academic research.

3.2 Iris Recognition for Data Captured Using Smartphone Cameras

Most of the iris recognition systems consist of components that are indicated in the Figure 3.3. Given the eye image captured using any imaging device, it is pre-processed. Preprocessing involves in localizing the eye region followed by segmenting the iris in which the boundary between sclera-iris and iris-pupil is determined. Due to the unconstrained nature of the image capture process and large unrestricted field-of-view, an image acquired from a camera does not necessarily contain just the eye region but also background. Thus, in order to avoid segmentation failure of the iris region, it is essential to localize the eye region first. The iris-sclera and iris-pupil boundary is then localized that corresponds to the texture of iris. Once the boundaries are well-identified, the iris is unrolled to represent it in polar-coordinates for the convenience of processing using the approach of Daugman's rubber sheet expansion[29].

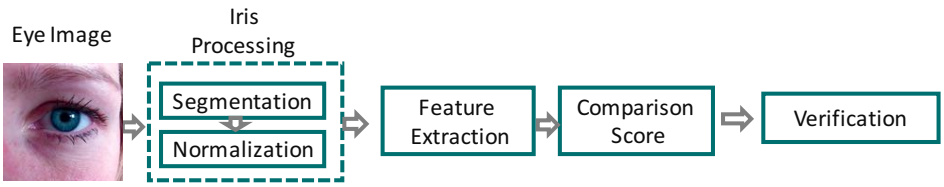


Figure 3.3: Block diagram of the general iris recognition framework.

The normalized iris image is processed further to extract the features using approaches of texture localization such as 2D Gabor features [29] or the variant 1D Log-Gabor representation [91] among many other approaches. The extracted features can be used as reference template in an enrolment transaction and the features extracted from an identity claim transaction (verification) are used as probe template. The distance between the

reference template and probe template is used to determine if the claimed identity is accepted as genuine subject or rejected as impostor subject.

Further, Figure 3.4 illustrates a simplified version of the components in this chapter specifically. The eye image is preprocessed to segment the iris and pupil boundary. The segmented iris region is normalized to a fixed dimension of 512 $\times$ 64 pixels which is further used to extract robust iris features by employing the $deep - sparse - filter$ responses and the feature vector is generated as outlined in the upcoming sections. The generated feature vectors are used for comparison with the Sparse Representation Classification (SRC) method [156, 117].

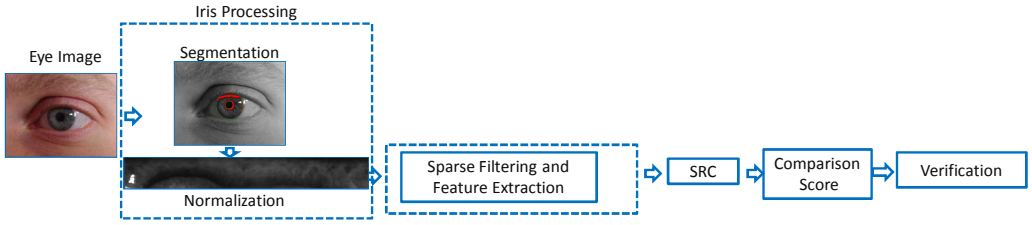


Figure 3.4: Schematic for deep sparse filtering based iris recognition. The segmented iris boundary is marked with green color tracing the boundary of pupil in inner circle and another green color in outer boundary. The texture pattern masked in red corresponds to noise detected in iris region which can correspond to ambient light reflection and shadow from eye-lashes.

3.2.1 Estimating Iris Diameter Range to Improve Segmentation

A key to achieve reliable result in terms of recognition accuracy is driven by well-segmented iris images. Incorrect segmentation eventually results in the degraded recognition accuracy for the biometric system [80, 108]. As a prior step, it is essential to localize the eye region alone to obtain better estimation of iris boundaries. Based on the success of Haar cascade based object detectors, we employ an eye detector trained using Haar Cascade network to detect the eye region [149]. Figure 3.5(a) illustrates the input image and Figure 3.5(b) depicts the detected eye region using the employed Haar cascade eye detector.

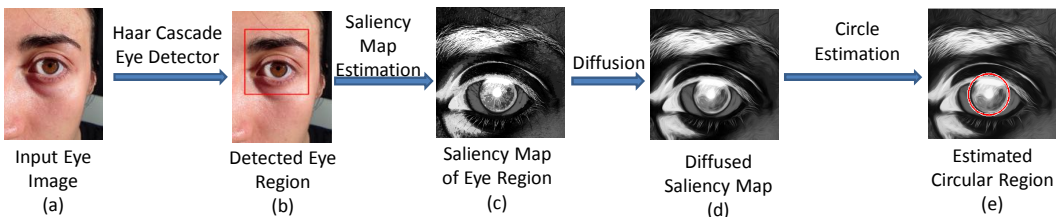


Figure 3.5: Estimation of approximate radius of iris region using saliency maps.

Following the localization of eye, the iris and pupil boundary are detected as a part of segmenting iris pattern. Locating the boundary is a challenge as the iris samples are captured in the visible spectrum in unconstrained manner as discussed in Section 3.1. OSIRIS v4.1 [135] provides an open source iris recognition toolkit which can be adapted for visible spectrum iris recognition [65, 71, 117]. However, the OSIRIS v4.1 works with a known range of iris diameter for successful segmentation [76] which becomes a key aspect in adapting the OSIRIS v4.1 for segmenting images captured from smartphone embedded cameras in visible spectrum.

The problem can be systematically addressed by estimating the iris diameter in coarse manner to reduce the search space for locating the fine boundary. Recent works have explored different methods to achieve coarse-localization of the iris boundary [1, 39]. In a similar paradigm, this chapter exploits supplementary information obtained from saliency maps of the localized eye image. Saliency maps provide sharp responses along the edges mimicking the human visual system and thereby helping to trace the iris boundary in a coarse manner.

Robust approach to estimate the saliency map for a given image in a manner close to human visual system is by measuring the variation in contrast [55, 25]. When the image of eye is captured, edge information constituted by the iris-sclera boundary presents high contrast response corresponding to near iris region boundary. Further, the intensity change in sclera and iris region gives an indication for the approximate location of the iris. Thus, we estimate the saliency map to coarsely localize iris from the image of eye region. The high intensity edge responses along the boundary of iris region provides a rough area to constrain the search space. Figure 3.5(c) presents the saliency map providing rich information consisting of strong edges in iris-sclera boundary [25]. The Figure 3.5(c) depicts the changes in intensity of image across regions.

The change in contrast appears in many regions including the boundary of iris-sclera and also sclera-skin which occasionally results in a falsely estimated iris region. To mitigate the falsely estimated iris regions, the saliency map is anisotropically diffused such that the low contrast changes are minimized and the high contrast regions along with strong boundaries are retained. Rotation invariant anisotropic diffusion is employed to retain the boundaries and edges [153]. As compared to other strategies of diffusion, rotation invariant approach employed in this chapter does not affect the stronger boundaries while merging the regions with smaller intensity boundaries. Figure 3.5(d) shows the diffused saliency map with strong edge transitions retained while the local intensity changes are smoothened.

Algorithm 1 Approximating Iris Radius

1: Reference image set : Ref
2: Detect eye region using Haar cascade eye detector (Refer Figure 3.5(b))
3: Extract Saliency map for the detected region. (Refer Figure 3.5(c))
4: Perform anisotropic diffusion on the saliency map. (Refer Figure 3.5(d))
5: Estimate the circles using the circular Hough transform.
6: **for** `<i circles detected (`C_{iD}`)>` **do**
7: Obtain corresponding region
8: Compute 2D correlation coefficient for estimated iris region using the formula below:

$$CR_i = \frac{\sum_m \sum_n (Ref - \bar{Ref})(C_{iD} - \bar{C}_{iD})}{\sqrt{\left(\sum_m \sum_n (Ref - \bar{Ref})^2\right)\left(\sum_m \sum_n (C_{iD} - \bar{C}_{iD})^2\right)}}$$

where $\bar{C}_{iD}$ and $\bar{Ref}$ represent mean values.
9: $Approx\_Radius \leftarrow (radius \mapsto max(CR_i))$ (Refer Figure 3.5(e))

Further, to increase the robustness and mitigate any false detections of iris boundary, the estimated iris region is correlated to a set of reference iris images from an independent and disjoint set consisting of 20 cropped iris images obtained from UBIRIS v2 dataset [108]. The reference images from this dataset serve as prototypes for various iris diameters. Each estimated region is correlated to the prototype set of images to obtain the correlation factor. The radius of the reference iris image having the highest correlation is used to provide the approximated iris radius. Figure 3.5(e) presents the located circular region in the image. The diameter of the approximated iris region is provided to the OSIRIS v4.1 [135] segmentation technique for further processing and locating the exact iris and pupil boundary. OSIRIS v4.1

internally performs anisotropic diffusion on high resolution images to estimate the iris-pupil boundary. The diffused image is used to detect the coarse boundaries and estimates robust noise mask by employing Viterbi search algorithm [37].

The set of all steps in approximating the iris boundary are outlined in Algorithm 1. The reference image set is indicated by Ref and the i number of detected circles are indicated by C_{iD}.

With the help of the preprocessing method for locating the coarse iris boundary, the necessity for *apriori* knowledge of an iris diameter range for OSIRIS v4.1 is eliminated and making the OSIRIS segmentation scheme robust. As compared to estimating the iris radius manually for every single image in unconstrained scenarios under varying resolutions and focus of different cameras, the technique discussed to approximate the radius has recorded the segmentation accuracy as given in Table 3.2. Note that the percentage is indicating the number of cases for which the automated segmentation operated with no error. The accuracy of the proposed improvement to segmentation technique is computed by manual inspection. The segmented iris is further normalized using the Daugman's rubber sheet model [29]. The normalized iris is then processed to extract the features for verification as discussed in Section 3.3.

3.3 Feature Extraction Scheme for Iris

This section provides the details of the new feature extraction technique for visible spectrum iris images using *Deep Sparse Filtering* approach. A brief summary of *sparse filtering* is outlined first and the feature extraction scheme based on *deep sparse filtering* for robust iris recognition is discussed in detail later.

3.3.1 Deep Sparse Filtering

Sparse Filtering is a recent paradigm of unsupervised algorithm to learn the number of specified features which does not explicitly attempt to model the distribution of data [94]. *Sparse filtering* optimizes a simple cost function of sparsity using $l_2 - normalized$ features. The key aspect in using *sparse filtering* is that unlike other algorithms in machine learning, *sparse filtering* does not necessarily include hyper-parameter tuning and typically converges to optimal solution easily. The only parameter required in learning *sparse filters* is the number of features, as the *sparse filters* are learnt by optimizing sparsity in feature distribution. The set of filters learnt can be used to extract features from any image using simple convolutional approach.

Different number of layers form the building blocks in learning *deep sparse filters*. In order to achieve deep sparse filtering, one has to employ more than one layer for learning. We learn the deep sparse filter using two layers such that *layer* 1 is trained using $200,000$ random patches of size 16×16 pixels from 4212 natural images [145] and the *layer* 2 is trained using the output of *layer* 1. The patches of natural images are first normalized to obtain the absolute values of feature data. In learning techniques involving multi-layered approaches, output from one layer is provided as input to the subsequent layer. The number of layers can be varied to n number of layers for various tasks. The choice of 2 layered *deep sparse filtering* through canonical greedy layer-wise approach [10, 47] is to learn more robust features with low computations [94]. The sparse features obtained as output from the *layer* 1 are normalized and provided to *layer* 2 using the feed-forward network employing the soft-absolute function provided in Equation 3.1. Generally, the same soft-absolute function is used across multiple layers [94].

$$f_j^{(i)} = \sqrt{\epsilon + (W_j^T . X^{(i)})^2} \qquad \text{where } \epsilon = 10^{-8} \tag{3.1}$$

where $f_j^{(i)}$ represents the j^{th} feature value corresponding to rows in the i^{th} column. X represents the input vector and W represents the weights in the training network.

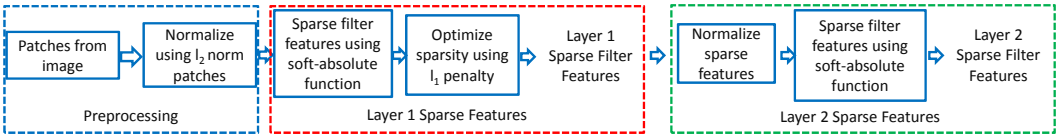


Figure 3.6: Schematic of sparse feature based filter learning

Figure 3.6 presents the schematic processing of the sparse filter learning in this chapter. As described in Figure 3.6, the image patches are preprocessed and used to train the *layer* 1 and subsequently *layer* 2. In this chapter, the sparse filter is trained to have *layer* 2 with 256 sparse filters of 16 $\times$ 16 features. Figure 3.7 illustrates some sample sparse features obtained in *layer* 2. We use the *sparse features* obtained from *layer* 2 as filters to extract features from the iris images as explained in the next section.

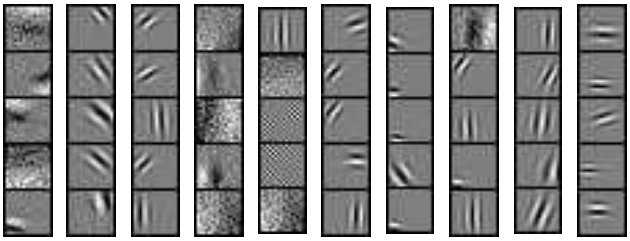


Figure 3.7: Sample sparse filters obtained from layer 2.

3.3.2 Deep Sparse Filtered Features for Iris Recognition

As the iris pattern is known to consist of unique texture pattern that is randomly formed [134, 29], we employ *deep sparse filters* to extract robust and unique features. Robust feature extraction from the iris texture is important for the performance of an iris biometric system, particularly when the data is heavily degraded as in the case of visible spectrum data. The visibility of the clear texture under NIR illumination contributes to obtain reliable features for recognition using techniques like 2D Gabor wavelets [29]. In the case of visible spectrum iris recognition on smartphones, due to the unconstrained nature of the imaging process, the captured iris pattern may be influenced by a number of factors such as imaging device, the ability to resolve the texture based on the color of iris among and many other environmental factors causing heavy reflections as discussed earlier [107]. Thus, it becomes essential to obtain robust features from the captured iris images to perform reliable recognition. Considering the unconstrained nature of iris data along with the high dimension of data, *deep sparse filters* can be successfully used to obtain more meaningful and reliable features.

The key motivation to extract features using the *sparse filters* learnt from natural images is that they contain different kind of information. The varying set of statistical properties in the natural images cover the broad spectrum relating to different ways human primary visual cortex receives the information in space and time when seeing an image [145]. The set of basis functions learnt in different approaches on natural images have resulted in filters similar to Gabor-like filters which are necessary to detect the edge responses in different direction and orientation [145]. The resulting filters are localized, oriented and band-pass, resembling the spatial receptive fields of simple cells in the primary visual cortex that are deemed to obtain features from images [145]. The set of basis functions learnt using natural images cover broad range of responses as compared to limited Gabor filters motivating us to employ the natural images to learn sparse representations to obtain highly varying filter responses. Thus, we use natural images to learn different basis functions using

sparse filtering which serve as set of independent filters to obtain the response on iris images. These set of sparse filters are used to extract discriminant and robust features from iris as detailed below.

In this section, we discuss forming feature vector using *deep sparse filters* for iris recognition. Given the segmented iris boundaries, the normalization technique unwraps the circular iris region into a rectangular image using Daugman's rubber sheet model [29]. Predominantly used dimensions of normalized iris image vary from 2048 $\times$ 1024 pixels to 128 $\times$ 32 pixels. The larger dimension of normalized iris image results in higher processing time while the smallest size may result in loss of many features due to the effect of resizing. Based on the popular dimension used for iris recognition in many works [29, 122, 135], in this chapter, we have employed a normalized iris dimension of 512 $\times$ 64 pixels. As the iris image captured in visible spectrum records three different color channels, i.e., red, green and blue channel, we average the information across color channels.

Further, we have employed 256 filters obtained from the output of *layer* 2 in the learning framework as discussed in section 3.3.1 where each of the filters have a dimension of 16 $\times$ 16 pixels. When the iris image is convolved with 256 filters of *layer* 2, a total of 256 response images are obtained. If the gray level iris image is represented by I and a sparse filter is represented by S, the sparse filter responses can be denoted as:

$$R = I * S \tag{3.2}$$

where $*$ represents the convolution operation. Considering 256 filters employed in this chapter, we can adjust the Equation 3.2 as:

$$R_{i=1:256} = I * S_{i=1:256} \tag{3.3}$$

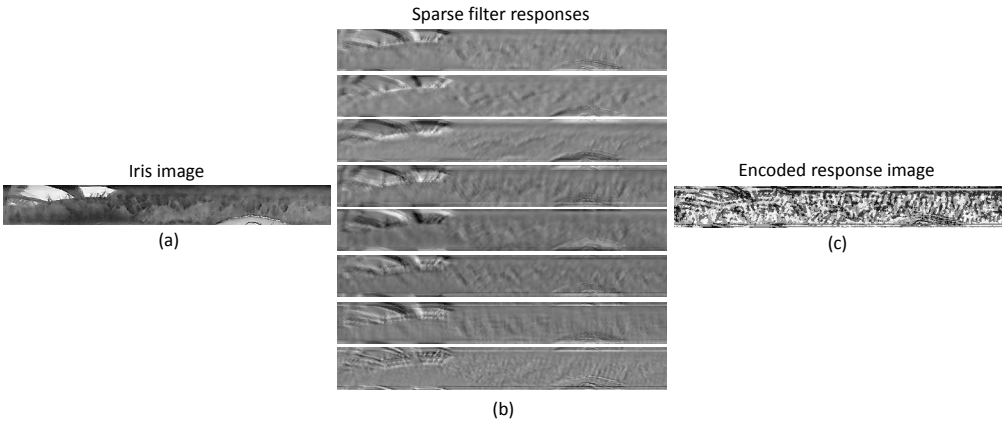


Figure 3.8: Sample responses to sparse filters and encoded image

Figure 3.8 illustrates the sample response to eight different sparse filters. Figure 3.8(a) provides the normalized iris image and Figure 3.8(b) presents eight different responses. Since processing 256 responses represented by $R_{i=1:256}$ at the feature level becomes tedious, hence we employ a simple way of binning the data to reduce the process time while still retaining the features. Each of the response image is thresholded and binarized based on the pixel value. For a pixel at position (x, y) in the response image, the thresholded value is represented by $T(x, y)$.

$$T(x,y) = \begin{cases} 1, & \text{if } R(x,y) > 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.4}$$

Each of the binarized feature image $T(x, y)$ are grouped in 8 images (i.e. the binarized

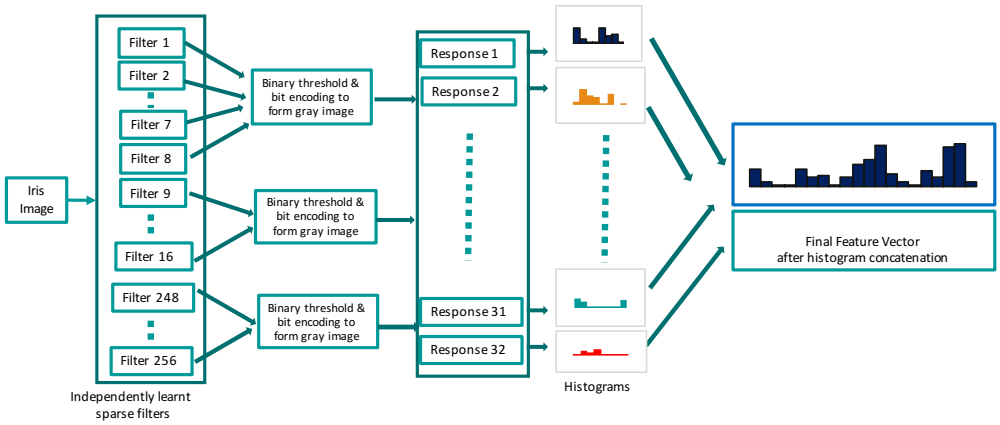


Figure 3.9: Proposed feature vector construction scheme

feature matrices) to transform into a new feature domain [1]. Considering a single pixel at a particular position (x, y), a set of binary values from images in the pool consisting of 8 matrices can be used to construct a binary code of 8 bits. Formulating it mathematically, for a pixel at position (x, y) in a set of 8 binary images, the pooled pixel feature encoded as gray value is presented as $P(x, y)$:

$$P(x,y) = \sum_{j=1}^{8} T_j(x,y) \times (2^{(j-1)}); \tag{3.5}$$

Figure 3.8(c) presents the pooled image formed by 8 responses of sparse filtering. Similarly, pooling of 256 response images in groups of 8 images result in 32 gray level response images and are represented by $P_{l=1:32}$. The obtained 32 gray level images result in high number of features and thus cause the overhead for computation at the second level. In order to avoid this, we employ histogram representation of each of the 32 response images to form a feature vector. The histogram for a single image is given by :

$$H_l = \sum_{m=0}^{255} (P_l)_m \qquad \text{for l = 1, 2 ... 32} \tag{3.6}$$

The final feature vector denoted by F is formed by concatenating the histograms of all the 32 gray level response images. The final feature vector can be represented as F given by :

$$F = [H_1, H_2, ...H_{32}] \tag{3.7}$$

Thus, each of the iris image is represented using $32 \times 256 = 8192$ features obtained using the sparse filtering and histogram vector of grouped binary images. Further in this chapter, we normalize the histogram in order to obtain the final feature vector. Figure 3.9 presents the proposed iris feature extraction using *deep sparse filtering*. The final feature vector is used in conjunction with the sparse representation classification (SRC) [156, 117, 76] to obtain the comparison scores.

3.3.3 Feature Classification

The obtained features are further classified to obtain comparison scores. In this chapter, we employ Sparse Representation Classification (SRC) to improve the recognition accuracy

[1] The number of response images to be binned can be chosen specific to any applications. In this chapter, the number was set to 8 on the basis of empirical trials. Similar results as comparable to binning size of 8 can be observed when the binning size was varied up-to 12.

inspired by the success of SRC in other biometric applications [156, 117]. The distinct features obtained using *deep sparse filtering* scheme are classified by projecting them on l_1 *norm* $-$ *minimization via* $SPGL_1$ solver based on a spectral gradient projection [156, 117].

Given the feature vector from *deep sparse filtering* of the image as in Equation 3.7, we further enhance the uniqueness of the histogram signature by representing it sparsely. The classification is performed according to the steps mentioned in Section 2.2.3 in Chapter 2 to obtain the residual error. The residual errors are treated as dissimilarity scores to classify them as genuine and imposter scores. These scores are used to compute the final False Match Rate (FMR) and False Non-Match Rate (FNMR) to obtain the biometric performance.

3.4 Smartphone Iris Databases

Owing to the difficulty in pattern visibility of the iris in visible spectrum light and lower resolution of cameras on smartphone, there are limited number of works on smartphone based iris recognition and thereby limited datasets are publicly available for research. Of the few datasets from smartphones captured specifically for iris recognition, MICHE I dataset [34] and Visible Spectrum Smartphone Iris (VSSIRIS) database [76] (collected during the course of this thesis) are publicly available. Using the two different databases, one can measure the robustness of various stages in an iris recognition pipeline such as segmentation and feature extraction. A brief description of the databases are presented in this section.

3.4.1 MICHE I Database

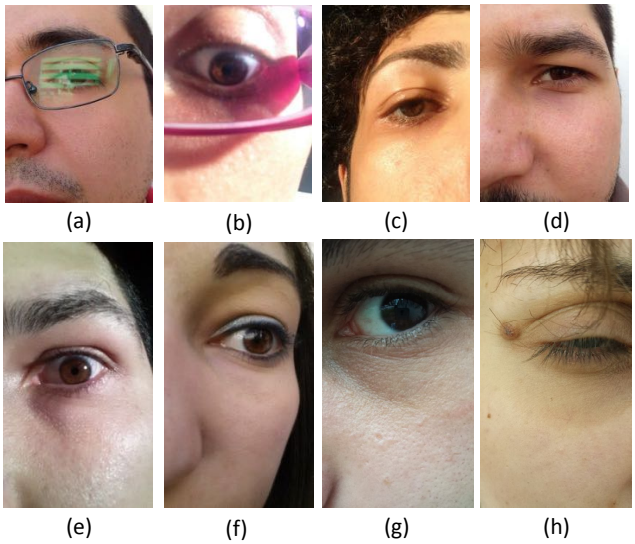


Figure 3.10: Sample images from the MICHE I database acquired using two different phones. (a)-(d) correspond to images captured using iPhone 5 and (e)-(h) correspond to images captured using Samsung Galaxy S4.

MICHE-I [14, 34] provides iris image database collected using two different smartphones - *iPhone 5* and *Samsung Galaxy S4*. The dataset presents the iris images mostly from southern European ethnicities. The images are captured using both frontal and rear camera in indoor and outdoor illumination conditions. MICHE-I consists of images obtained from 75 unique iris instances with more than 3 samples for each. In this chapter, we employ a subset of 50 iris instances such that every iris instance has at-least 4 samples. Equal number of samples for each subject is chosen for the experimental protocols discussed later.

Figure 3.10 presents the sample images from the MICHE-I database. It can be observed from these examples that the iris images do not correspond to iris images acquired in ideal conditions and thereby exhibit number of degradation factors. Figure 3.10(a) exhibits impact of illumination. Figure 3.10(b) illustrates off-angle iris. Figure 3.10(c) and Figure 3.10(d) exemplify occluded iris. Figure 3.10(e) presents a good quality image. Figure 3.10(f) and (g) again demonstrate off-angle iris images while Figure 3.10(h) presents completely occluded iris image.

3.4.2 Visible Spectrum Smartphone Iris (VSSIRIS) Database

VSSIRIS database constructed during the course of this thesis consists of images acquired from volunteers originating mostly from northern European countries. The VSSIRIS database has been acquired with two recent phones - *iPhone 5S and Nokia Lumia 1020*. The iris images in the VSSIRIS database were captured using the rear camera of both smartphones. The specifications of the camera and operating environments are provided in Table 3.1. The database consists of images acquired under the influence of mixed illumination constituted by artificial indoor illumination and natural daylight illumination. This database thus provides an opportunity to explore the challenges presented by the mixed illumination for iris recognition in visible spectrum. The VSSIRIS database consists of iris instances obtained from 28 subjects with 5 samples for each acquisition.

Figure 3.11: Sample images from the VSSIRIS database acquired using two different phones.

Figure 3.11 depicts sample images from the VSSIRIS iris image database. The VSSIRIS database consists of images acquired from 28 subjects in a single session which constitutes to a total of 56 unique iris instances. Each unique iris instance is captured in 5 different presentations per device in a single session under semi-cooperation from the subjects and under unconstrained conditions. A total of 560 images are present in the database. The participants in the VSSIRIS database consist of various nationalities originating from eastern, northern and southern European countries.

Table 3.1: Camera parameters of smartphones

Parameters	iPhone 5S	Nokia Lumia 1020
Resolution	3264 x 2448 Pixels	7712 x 5360 Pixels
Color Representation	sRGB	sRGB
Bit Depth	24	24
F-Stop	f/2.2	f/2.2
White Balance	Auto	Auto
Flash	No Flash	No Flash
Metering Mode	Average	Average
File Format	JPEG	JPEG
Focus	Auto	Auto
Illumination	Mixed Illumination	Mixed Illumination

3.5 Experiments & Results

This section reports the experiments and the obtained results for improved segmentation accuracy comparing with standard OSIRIS v4.1. Further, this section also presents the verification results using the proposed *deep sparse filtering* approach on two publicly available smartphone iris database - MICHE-I [34] and VSSIRIS database [76]. All results on both databases are reported using the Equal Error Rate (EER), which is a metric defined as a point for which the False Match Rate (FMR) equals the False Non-Match Rate (FNMR) [54].

Further, the performance of seven well known feature extraction techniques are presented along-with. These feature extraction techniques are predominantly used for iris recognition for images captured in NIR spectrum [29, 91, 84, 78, 120, 121, 117]. We compare the results of the *deep sparse filtering* approach against the state-of-art feature extraction techniques. Most of the mentioned techniques [91, 84, 78, 120, 121] are evaluated on both databases using the implementation obtained from USIT - University of Salzburg Iris Toolkit v1.0 [122] and OSIRIS v4.1[135].

3.5.1 Evaluation of Improved Segmentation Scheme

This section presents the results of segmentation by estimating the iris diameter range and compares it against segmentation of standard OSIRIS v4.1 i.e, without iris diameter estimation. Results are reported on the experiments conducted on MICHE-I [34] and VSSIRIS databases [76].

Figure 3.12 presents an illustration of the segmentation accuracy when the radius is estimated as compared to segmentation with OSIRIS v4.1 alone for MICHE-I. Figure 3.12(a) illustrates the robustness of OSIRIS v4.1 which has resulted in accurate segmentation even obstruction of hairs. Figure 3.12(b) presents the clear segmentation with both OSIRIS v4.1 and OSIRIS v4.1 with estimated iris diameter. It can be observed from the image that approximation of iris diameter has resulted in fine iris boundary localization and good noise mask. Figure 3.12(c) presents highly pigmented image where the OSIRIS v4.1 works reasonably well to localize boundaries, however fails to localize the exact boundary. The same iris is segmented well when the iris diameter is approximated. Figure 3.12(d) depicts a highly challenging image both in-terms of pigmentation and off-angle iris representation. The image in the second row is segmented using OSIRIS v4.1, the segmentation has failed to localize the clear boundary while the bottom row indicates successful segmentation when the search space for iris boundary is constrained using coarse iris radius estimated using saliency maps.

The images from each illumination condition corresponding to single smartphone camera are treated as one subset of database for MICHE-I database following the protocols reported

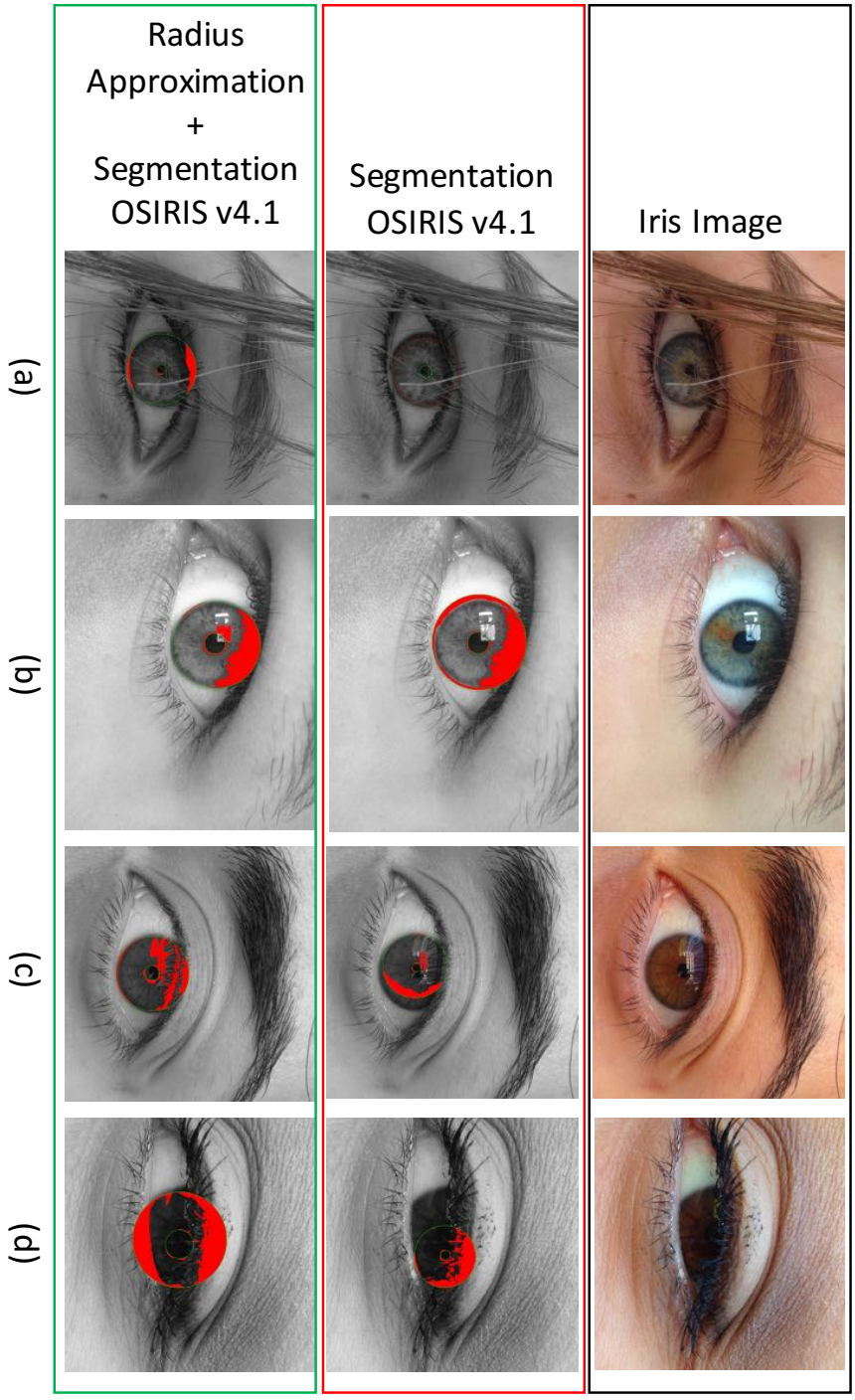


Figure 3.12: Illustration of segmentation with the radius approximation as compared to segmentation with OSIRIS v4.1 alone for MICHE-I database. The segmented iris boundary is marked with green color tracing the boundary of pupil in inner circle and another green color in outer boundary. The texture pattern masked in red corresponds to noise detected in iris region which can correspond to ambient light reflection.

earlier in our work [76]. A subset consisting of 50 unique iris has been employed in this chapter to maintain the same number of iris samples in all different acquisition conditions and provide unbiased comparison. The detailed description of database can be obtained from Section 3.4.

Table 3.2 lists the obtained improvement in segmentation accuracy as compared to using OSIRIS v4.1 alone. The accuracy reported in the Table 3.2 is verified using visual inspection of the segmented images by manually correlating the boundaries of iris and pupil to the boundaries localized by automated segmentation.

Table 3.2: Segmentation accuracy with the iris diameter approximation for OSIRIS v4.1

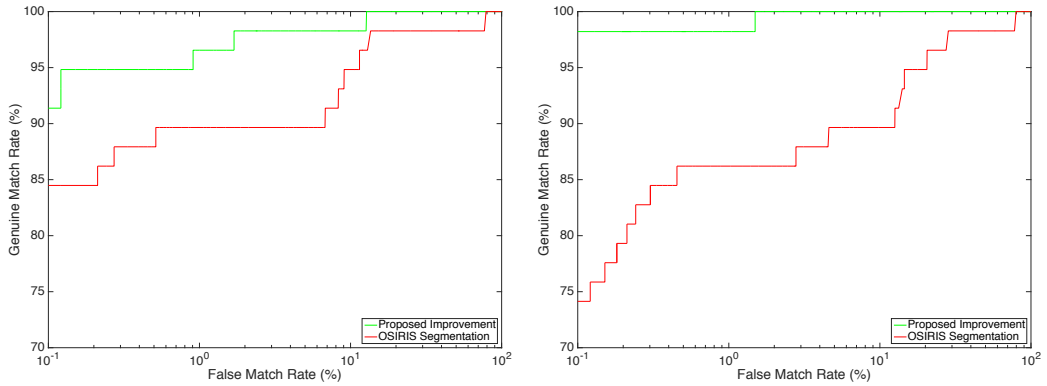
Smartphone	Illumination	Camera	Segmentation Accuracy (%)
Database : MICHE-I [34]			
iPhone	Outdoor	Rear	81
		Frontal	64
	Indoor	Rear	64.5
		Frontal	76.5
Samsung	Outdoor	Rear	74.5
		Frontal	62
	Indoor	Rear	65
		Frontal	77
Database : VSSIRIS [76]			
iPhone 5S	Mixed	Rear	85
Nokia Lumia 1020	Mixed	Rear	78.5

3.5.2 Improvement of Verification Accuracy Due to Improved Segmentation

To benchmark the impact of the radius approximation for segmentation technique on the verification performance, we evaluate the verification accuracy and report the Equal Error Rate (EER). The verification accuracy is obtained using the proposed approach of new feature extraction based on *Deep Sparse Filtering*. The Table 3.3 provides the EER of the segmented iris images using standard OSIRIS v4.1 and OSIRIS v4.1 with approximated iris diameter for the unconstrained iris images in both the databases employed in this chapter. Figure 3.13, Figure 3.14 and Figure 3.15 provide the plots of Receiver Operating Characteristics curves (ROC) for images obtained from different phones. The improvement in Genuine Match rate at lower False Match Rates (FMR) for the verification performance indicates the suitability of approximating radius to achieve a robust biometric system in unconstrained iris acquisition systems.

Table 3.3: EER(%) with the standard OSIRIS v4.1 system and the improvement due to the iris diameter estimation.

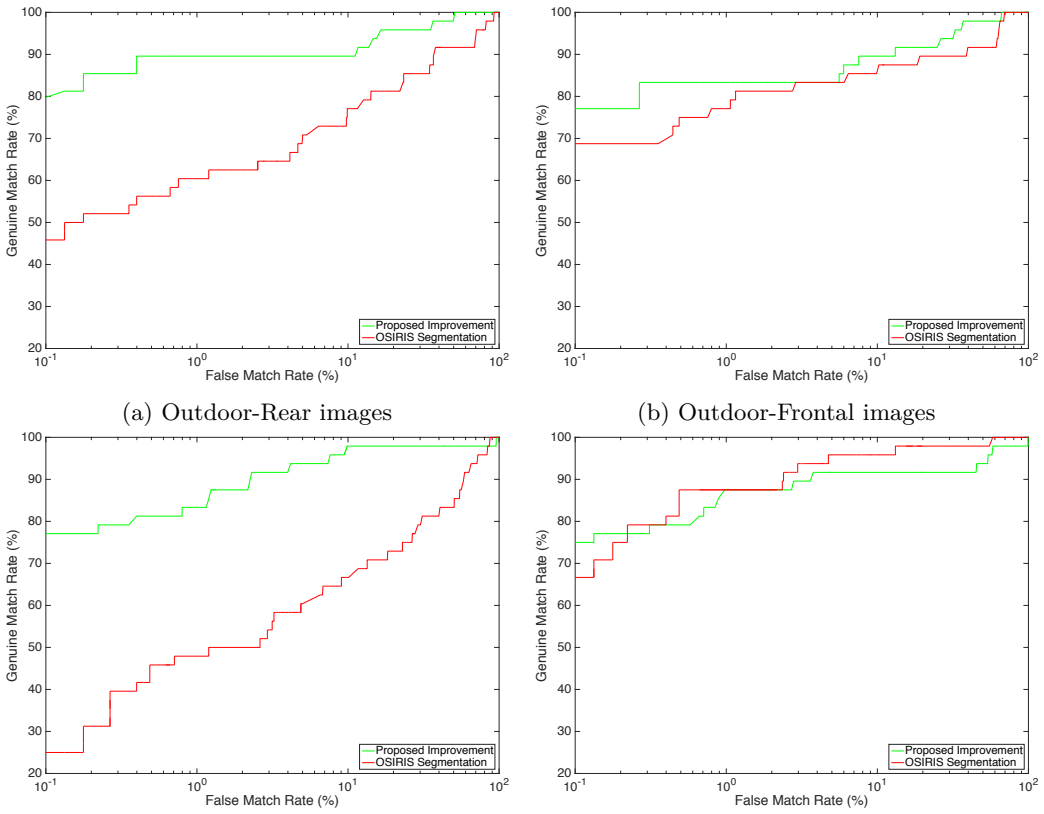
Smartphone	Illumination	Camera	EER (%)	
			OSIRIS v4.1	Diameter Approximation + OSIRIS v4.1
Database : MICHE-I [34]				
iPhone	Outdoor	Rear	14.62	10.40
		Frontal	11.14	6.78
	Indoor	Rear	8.35	8.35
		Frontal	3.74	4.16
Samsung	Outdoor	Rear	18.86	10.52
		Frontal	12.45	10.26
	Indoor	Rear	25.06	6.16
		Frontal	8.40	4.47
Database : VSSIRIS [76]				
iPhone 5S	Mixed	Rear	14.62	10.40
Nokia Lumia 1020	Mixed	Rear	7.98	2.01



(a) VSSIRIS Nokia images (b) VSSIRIS iPhone images

Figure 3.13: ROC curves depicting performance with the native OSIRIS segmentation and the improvement based on radius approximation scheme for VSSIRIS database

Further, the impact of the iris diameter approximation on the verification performance is presented in Figure 3.13 for the VSSIRIS database. Figure 3.13(a) and Figure 3.13(b) presents the improvement in verification accuracy with the *Deep Sparse Filtering* approach[76] for data captured in Nokia Lumia 1020 and iPhone 5S.



(a) Outdoor-Rear images

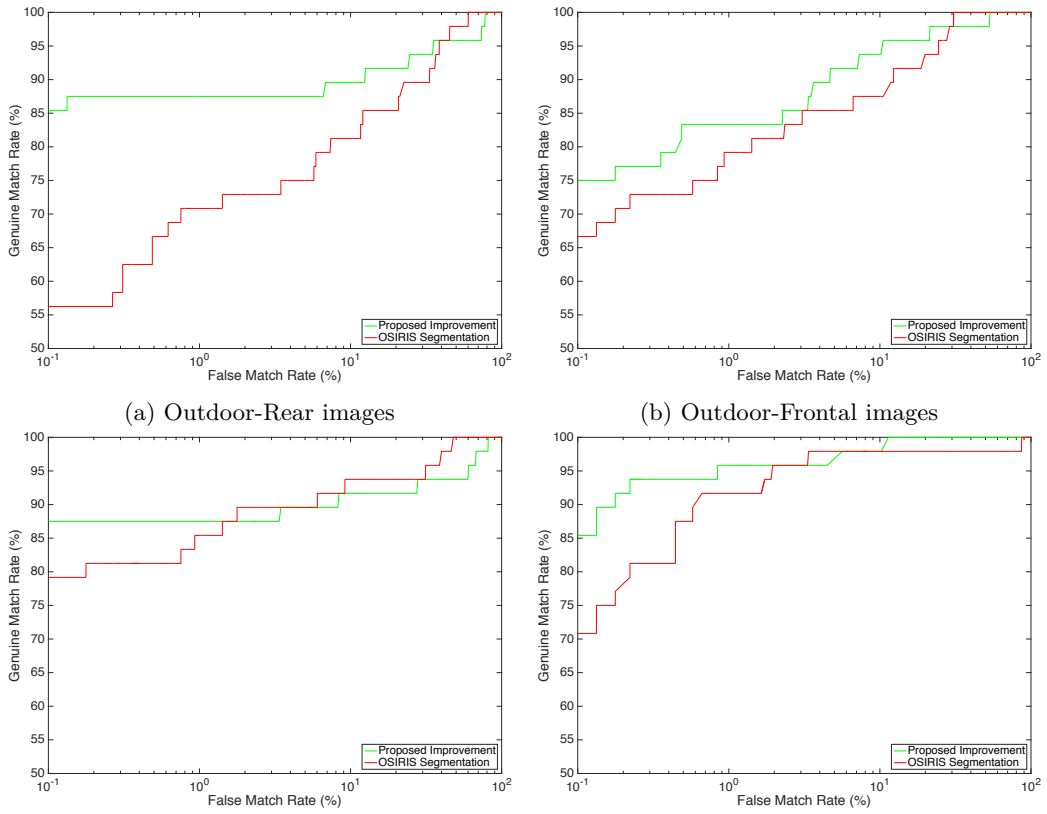
(b) Outdoor-Frontal images

(c) Indoor-Rear images

(d) Indoor-Frontal images

Figure 3.14: ROC curves depicting performance of improvised segmentation scheme for Samsung images of MICHE-I subset.

Similarly, Figure 3.14 and Figure 3.15 presents the verification performance for MICHE-I database. The improvement in the verification can be clearly marked with the large difference in red and green curves. The verification accuracy can be observed to improve in all cases of MICHE-I subsets as the green curves are closer to $GMR = 100\%$ indicating higher performance.



(a) Outdoor-Rear images

(b) Outdoor-Frontal images

(c) Indoor-Rear images

(d) Indoor-Frontal images

Figure 3.15: ROC curves depicting performance of improvised segmentation scheme for iPhone 5S images of MICHE-I subset [34].

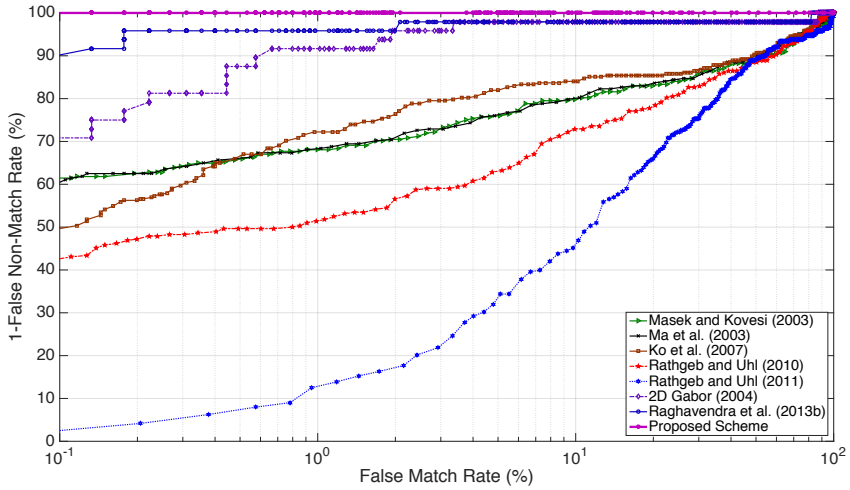
3.5.3 Evaluation of Deep Sparse Filtering on MICHE-I Database

This section presents the results of proposed feature extraction obtained on the MICHE-I database and compares it standard state-of-art schemes mentioned above. In the subset of MICHE-I database employed in this work, each unique iris instance has 4 samples and thus, we adopt the leave-one-out approach by dividing the data in $3:1$ ratio with 3 samples as reference and 1 as probe sample. The minimum score from three comparisons is used as the comparison score for the pair of probe and reference. The partition is continuously swapped to make each iris sample reference and probe at different times. Further, the reference and probe partition is repeatedly changed m times with $m = 10$ under the leave-one-out cross-validation strategy. The final results are obtained by averaging the results obtained from all iterations of the leave-one-out approach. The results thus represent the mean value of all the 10 different trials taking care of statistical variations.

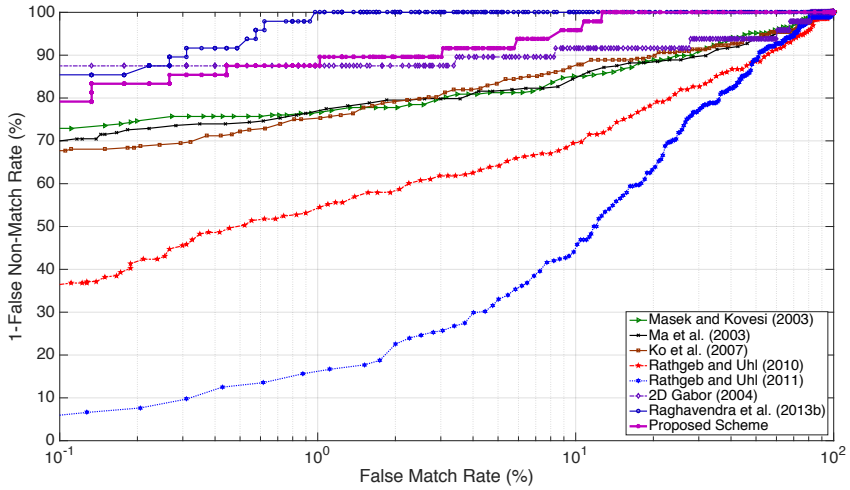
Table 3.4 provides the EER scores obtained for the images on MICHE-I database [34]. It can be observed from Table 3.4 that the *deep sparse filtering* feature extraction technique performs well on the publicly available database validating the robustness of method. On average, the obtained gain in performance is around 2% on all different cameras and illuminations. The obtained error rate in terms of GMR and FMR is presented in the plot provided by Figure 3.16 and Figure 3.17 for iPhone data in MICHE-I dataset. Similarly, the obtained error rates is presented in the plots in Figure 3.18 and Figure 3.19 for Samsung data in MICHE-I dataset.

Table 3.4: EER (%) obtained for various schemes on MICHE-I database [34]
(* Scores obtained using USIT v1.0 [122]; $ Scores obtained using OSIRIS v4.1 [135])

Schemes	EER (%)							
	Outdoor				Indoor			
	iPhone		Samsung		iPhone		Samsung	
	Rear	Front	Rear	Front	Rear	Front	Rear	Front
2D Gabor [29] $	10.41	6.78	10.52	12.45	8.35	3.74	6.16	8.399
1D Log Gabor [91]*	24.01	21.86	20.83	21.22	13.90	17.01	17.26	18.83
Li Ma [84]*	29.98	22.01	20.63	20.60	12.84	17.01	18.68	17.10
Ko et. al [78]*	21.78	18.11	18.05	17.70	11.29	14.58	14.68	14.06
Rathgeb & Uhl [120]*	28.12	27.43	24.39	27.30	20.83	21.05	25.68	21.13
Rathgeb & Uhl [121]*	30.39	26.76	26.04	31.09	26.31	26.84	33.17	30.43
Raghavendra et al. [117]	8.57	7.82	6.29	8.11	**0.48**	2.07	4.18	4.23
Deep Sparse Filtering [76]	**6.25**	**4.18**	**2.06**	**6.27**	6.25	**0.02**	**3.96**	**2.50**

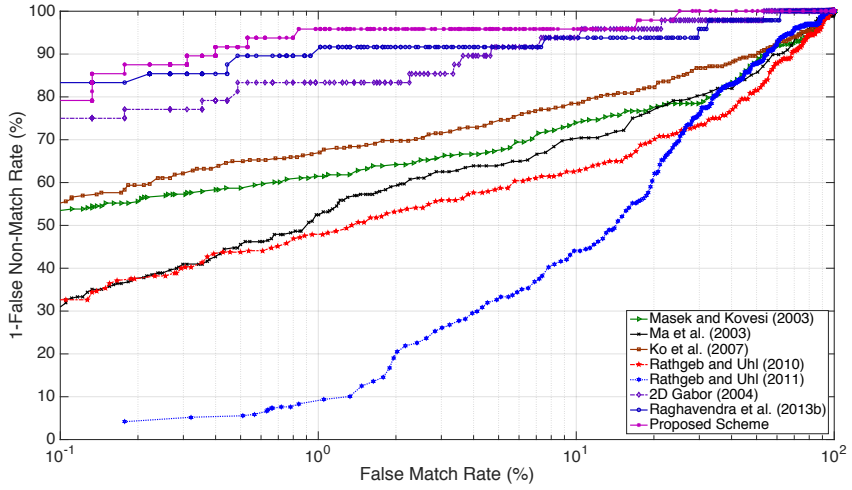


(a) iPhone-Indoor-Frontal camera

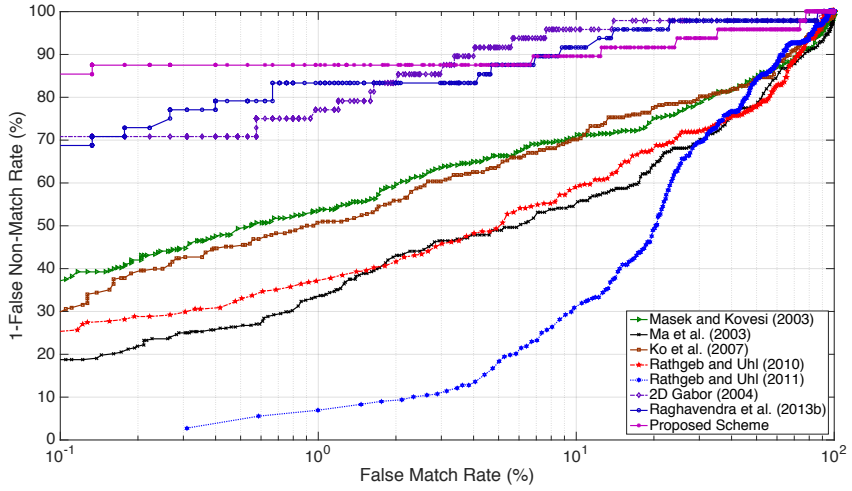


(b) iPhone-Indoor-Rear camera

Figure 3.16: ROC curves obtained for various schemes applied on the iPhone images from MICHE-I database (Indoor illumination) [34].



(a) iPhone-Outdoor-Frontal camera

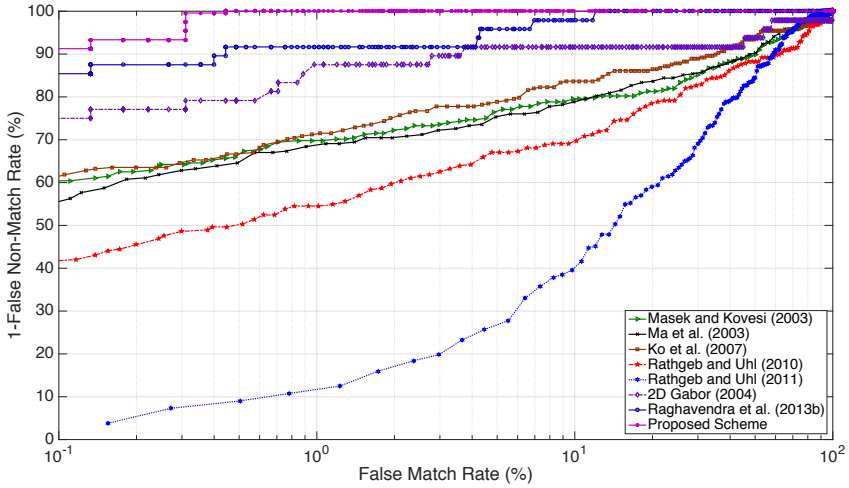


(b) iPhone-Outdoor-Rear camera

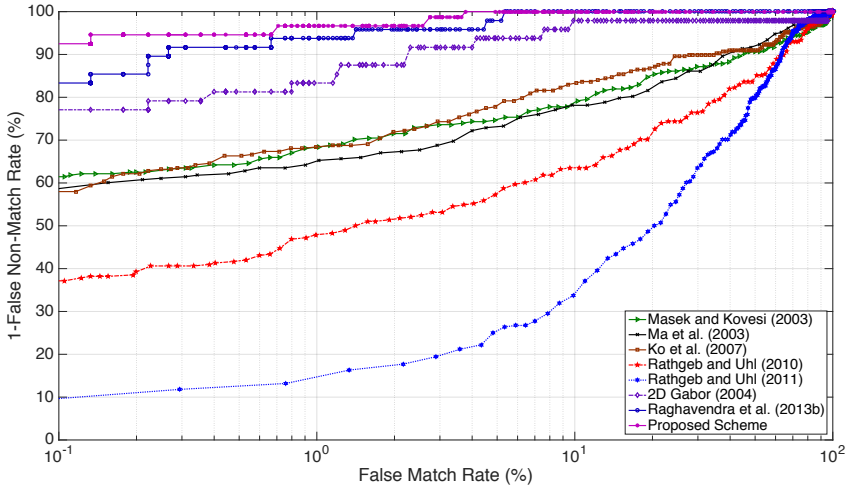
Figure 3.17: ROC curves obtained for various schemes applied on the iPhone images from MICHE-I database (Outdoor illumination) [34].

3.5.4 Evaluation of Deep Sparse Filtering on VSSIRIS Database

In this section, we experimentally validate the *deep sparse filtering* feature extraction technique on the VSSIRIS database [76] and compare against other state-of-art feature extraction techniques.



(a) Indoor-Frontal camera



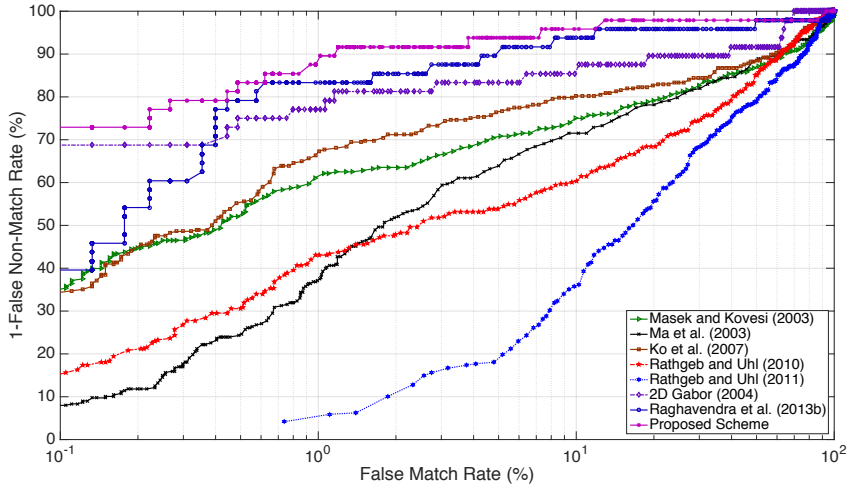
(b) Indoor-Rear camera

Figure 3.18: ROC curves obtained for various schemes applied on the Samsung images from MICHE-I database (Indoor illumination)[34].

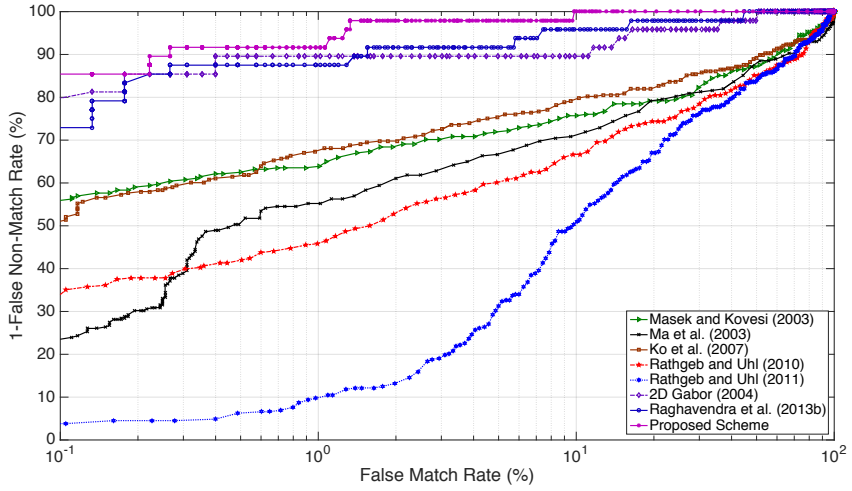
Table 3.5: Biometric performance obtained for various schemes on the VSSIRIS database
(* Scores obtained using USIT v1.0 [122]; $ Scores obtained using OSIRIS v4.1 [135])

Schemes	EER (%)	
	iPhone 5S	Nokia 1020
2D Gabor [29]$	3.62	3.52
1D Log Gabor [91]*	5.73	11.66
Li Ma et al. [84]*	7.89	13.88
Ko et. al [78]*	7.88	11.79
Rathgeb & Uhl [120]*	16.26	24.16
Rathgeb & Uhl [121]*	19.45	27.54
Raghavendra et al. [117]	8.31	10.59
Deep Sparse Filtering [76]	**1.62**	**1.78**

Each unique iris instance has 5 samples in VSSIRIS database and thus, this work adopts



(a) Outdoor-Frontal camera

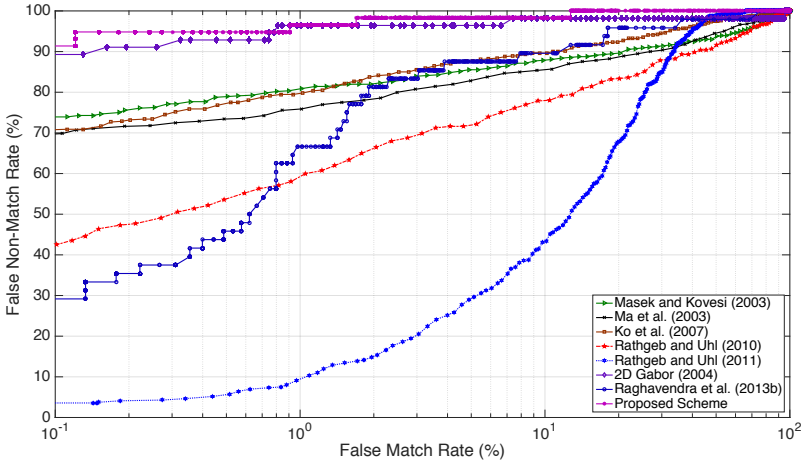


(b) Outdoor-Rear camera

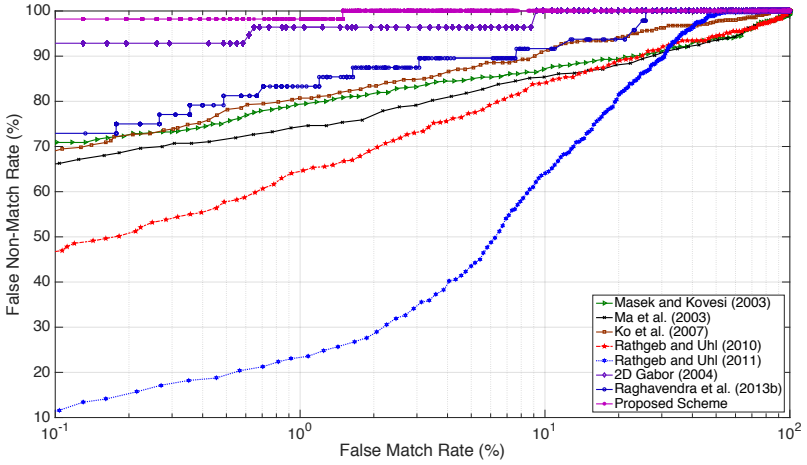
Figure 3.19: ROC curves obtained for various schemes applied on the Samsung images from MICHE-I database (Outdoor illumination) [34].

the leave-one-out approach by dividing the data in 4 : 1 ratio with 4 samples as reference and 1 other sample as probe. The minimum score from 4 comparisons is used as the comparison score for the pair of probe and reference. The reference and the probe samples are swapped continuously such that each of the image is reference in a particular trial. The reference and probe partition is repeatedly changed m times with $m = 10$ under the leave-one-out cross-validation strategy. The final results are obtained by averaging the results obtained from all iterations. The results obtained represent the averaged value of all the 10 trials providing statistically meaningful result.

Table 3.5 presents the results of all different schemes on the VSSIRIS database. It can be observed from the table that the *deep sparse filter* based feature extraction technique has outperformed the rest of the state-of-art techniques by providing the best EER of 1.62% for iPhone 5S images and 1.78% for Nokia images. Figure 3.20 presents the plots of the obtained GMR at different FMR for various feature extraction schemes.



(a) iPhone 5S images from the VSSIRIS database



(b) Nokia Lumia 1020 images from the VSSIRIS database

Figure 3.20: ROC curves obtained for various schemes applied on the VSSIRIS database

3.5.5 Discussion

The challenges of iris recognition in the visible spectrum stem from many factors such as unconstrained imaging, low visibility of texture pattern, ambient reflection from external illumination and non-standard iris-pupil dilation ratio. The verification accuracy can be influenced by any or all of these factors. A careful inspection of the scores and the corresponding images indicate the drop in verification performance due to contribution towards both false matches and false non-matches. Significantly, lower verification is attributed to false matches and a key factor resulting in high false matches is the low visibility of unique texture information when the images are captured in unconstrained conditions, especially for the heavily pigmented iris patterns. The trend of higher false match rates can be observed in MICHE-I database which consists of southern European ethnicities presenting mild to heavily pigmented iris patterns as observed in Figure 3.19, Figure 3.18, Figure 3.17 and Figure 3.16. As it happens, the lower EER in VSSIRIS database can be attributed to mildly pigmented iris which are light in color and thereby providing better texture visibility. With the evident observations from the set of experiments presented in this chapter, it can be noted that the iris recognition on smartphones remains a challenge which needs considerable amount of effort in achieving better imaging sensors that will provide detailed texture visibility, especially for subjects with higher pigmentation density. Iris pattern as an

authentication mode from smartphone captured data in visible spectrum can be adapted for larger population of world with heavily pigmented iris if the images can be captured with better texture visibility. The problem of obtaining better texture resolution in image needs considerable research efforts in the future works.

3.6 Conclusions

The rise of iris recognition in visible spectrum has seen growing interest for numerous applications. Specifically, many recent works have investigated smartphones as a biometric sensor for iris recognition. The performance of the iris recognition depends on the accuracy of segmentation taking care of noise via robust masks. The challenge of segmenting iris images increases for visible spectrum due to unconstrained nature of image capture which result in strong ambient reflection, shadow of eye lashes and partial closure of the eyes. Thus, it is essential to devise robust segmentation algorithms specifically tailored for visible spectrum data in unconstrained capture scenarios.

Further, to adapt the existing segmentation schemes like OSIRIS v4.1 [135] commonly used for NIR spectrum data, the range of iris diameter needs to be estimated. This chapter has contributed in improving open source segmentation scheme OSIRIS v4.1 by eliminating the need of *apriori* knowledge of iris diameter. The robustness of the OSIRIS v4.1 segmentation scheme is further enhanced by the presented iris diameter approximation as demonstrated in this chapter. Although the proposed improvement is specifically tested on images captured in visible spectrum, the solution can be adapted for images captured in unconstrained iris capture systems operating in NIR spectrum.

The improvement in segmentation accuracy due to radius approximation has been evaluated both manually and also in terms of the verification performance on two publicly available iris databases - MICHE-I [34] and VSSIRIS database [76]. The approximated iris radius is used as *apriori* information for standard OSIRIS v4.1 segmentation scheme [135]. The method has resulted in an accuracy of 75% in average with the best performance of 85% accuracy for iPhone 5S images.

Further, the need for robust feature extraction techniques in visible spectrum iris recognition is demonstrated through a set of experiments using state-of-art techniques on two public iris datasets. *Deep sparse filtering* based feature extraction for iris images is evaluated with respect to the verification performance in this chapter. The feature extraction has resulted in good accuracy for both the publicly available databases. Obtained results on VSSIRIS database report an EER of 1.62% for iPhone 5S and EER of 1.78% for Nokia Lumia 1020 phone. The robustness of the feature extraction technique can also be observed from the average gain of around 2% on the EER for all different smartphone images over the state-of-art techniques for MICHE-I database.

As a remark, the need for better imaging ability, robust segmentation approaches and reliable feature extraction are needed to employ iris recognition to full potential, especially for heavily pigmented iris patterns. The higher false match rates and false non-match rates as observed from Figure 3.19, Figure 3.18, Figure 3.17 and Figure 3.16, specially for the heavily pigmented iris in MICHE-I need to be addressed before this can be adapted in real-life verification scenarios. Alternatively, innovative imaging techniques should be designed to illuminate the iris such that the light is scattered back to resolve better texture.

Chapter 4

Imaging Heavily Pigmented Iris in Visible Spectrum using White LED

Low visibility of iris texture for the images captured in visible spectrum limits the employability of iris recognition for large population of subjects with heavily pigmented iris. This chapter is dedicated to present a new imaging set-up to obtain iris images with good texture visibility, especially for heavily pigmented iris from smartphone visible spectrum. The proposed approach is validated by set of experiments and compared against the images captured in the NIR spectrum for the set of corresponding subjects.

4.1 Introduction

Two specific challenges of iris recognition in visible spectrum arises due to low visibility of texture and challenges in segmentation due to quality of iris image which are discussed in Chapter 3. In order to resolve the texture for dark iris, NIR spectrum illumination in the range of 780 *nm* to 840 *nm* is used as the light in this range penetrates melanin and is scattered by collagen fibrils as discussed in Section 3.1 in Chapter 3.

However, most of the everyday-imaging sensors such as off-the-shelve cameras or smartphone embedded cameras, do not have inbuilt NIR illumination unit to support the capture process in iris. Given the interest in using smartphones with embedded cameras to capture iris images for various biometric applications, the absence of NIR illumination is limiting factor for heavily pigmented iris. Recent smartphones *Microsoft - Nokia 950, Nokia 950 XL, Fujitsu NX F-04G and Samsung Galaxy Note7* are designed with inbuilt NIR illumination for capturing iris data. Although the solution is appealing, it does not allow the current smartphones (without NIR illumination) to capture iris data. Thus, if one intends to use visible spectrum light in the range of 380 *nm* to 720 *nm* to capture iris patterns, the success is limited and restricted to only those iris instances that have light colors and that are captured in a controlled scenario. It is therefore important to address such a problem of low texture visibility for images captured in visible spectrum to support the increasing popularity of smartphone based biometrics employing iris recognition[34, 129, 76, 142].

In this chapter, we present an imaging set-up with a white light-emitting-diode (LED) to obtain iris images with detailed texture visibility [66]. The idea of using LED is to mimic the flash illumination present in the smartphone cameras and thereby to find alternative ways of engaging smartphones without NIR illumination in an optimal way. The LED used in this thesis is highly similar to the flash embedded in the present day smartphones. The presented prototype captures good quality iris images by placing a white LED inclined at an acute angle to the iris position. The main advantage of the presented approach is that it can capture good quality iris samples with minimal additional hardware (like LED) on a smartphone platform and demonstrating the ability using the flash on the smartphone itself. A recent smartphone like Oppo N1 [1] has the flash LED and camera installed on a rotating unit which follows a similar paradigm of presented setup. Therefore, the new approach can be integrated into the smartphones with minimal engineering effort while manufacturing following the design approach of Oppo N1.

[1]Specifications of Oppo N1 can be availed at: http://www.oppo.com/en/smartphone-n1

We present significance of the imaging set-up by capturing the iris images of heavily pigmented subjects to create a new database with 62 unique iris instances with 10 samples each. The database is evaluated using different state-of-art techniques to determine the applicability of the new imaging set-up. This chapter also presents a benchmark evaluation of the proposed method with conventional Near-Infra-Red (NIR) images for a subset of the subjects in the database. Extensive experiments are carried out using five different well-established iris recognition algorithms and one commercial-of-the-shelf algorithm.

In the rest of this chapter, Section 4.2 presents the iris recognition pipeline with the imaging set-up based on LED. Section 4.3 provides a detailed description of the database captured using the imaging setup. Section 4.4 provides the discussion on the experimental protocols and results. In the Section 4.5, the conclusive remarks and observations are presented for this chapter.

4.1.1 Contributions

The key contributions of this chapter can be outlined as below:

1. Presents an image capturing setup based on normal white LED mimicking the flash illumination embedded on the smartphones to capture heavily pigmented iris.
2. Presents extensive experiments to demonstrate the applicability by analysing the database of images from 62 heavily pigmented unique iris instances captured using the proposed set-up.
3. Additionally, this chapter presents evaluation of five different iris recognition algorithms and one commercial-off-the-shelf (COTS) algorithm to render a significance of LED set-up along with comparative performance analysis on three different smartphone samples.
4. Also benchmarks the images obtained from LED imaging setup with images from conventional NIR iris capture setup using commercial iris capture device *MorphoTrust Mobile-Eyes* [92].

4.2 Iris Recognition Framework

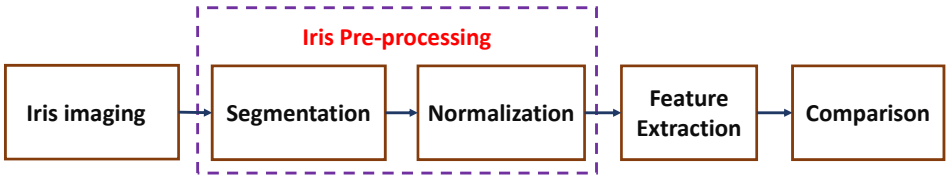


Figure 4.1: Block diagram of the iris recognition framework

Figure 4.1 shows the block diagram of the regular iris recognition framework. After acquiring the eye image, the iris region is localized from the captured sample. The localization of the eye is important for two main reasons - (i) The field- of-view of smartphone camera is generally wider, and hence details from background have to be discarded. (2) The segmentation errors in localizing the iris boundary can be significantly reduced as the region is small. Thus, we employ the Haar cascade based eye detector for locating the specific region of eye [150]. Once the eye is localized, the iris has to be segmented. The details of each component are given in the section below.

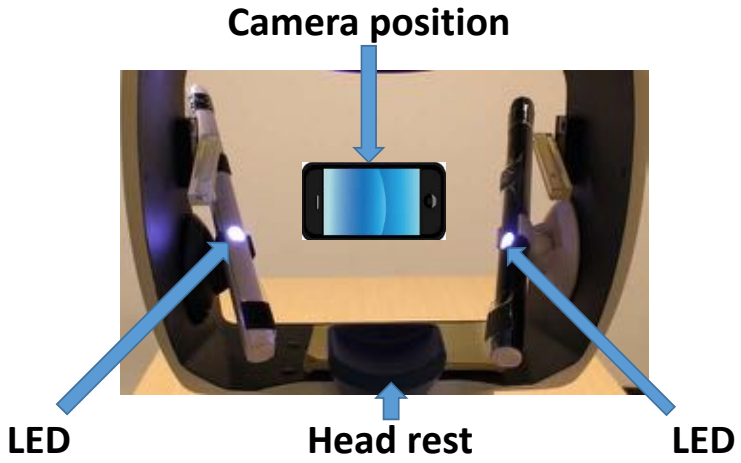


Figure 4.2: Prototype of the visible iris image acquisition setup

4.2.1 Approach for Imaging Heavily Pigmented (Dark) Iris

To capture the iris pattern in the visible spectrum, we employ white LED at an acute angle. Figure 4.2 illustrates the prototype of the visible iris imaging setup. Placing the light at an acute angle illuminates the iris such that complete light is not absorbed by human melanin pigmentation. Due to non-absorption of light, a maximal portion is reflected back resulting in a visibility of iris structure. Although the iris texture information obtained does not fully correspond to information obtained in NIR spectrum, iris recognition can still be achieved with the obtained information as demonstrated later. Further, the placement of LED is carefully designed to make sure there is no strain for the eyes of the subject while the texture of iris is well observed. As the white LED is medically accepted for the fact that it does not impact or damage the vision and moreover, similar white LEDs are also used in smartphones, we have employed white LED in this chapter to serve two purposes -

1. Study the feasibility of using the LED on smartphone for iris recognition with minimal engineering effort to place it optimally.

2. Provide an alternative imaging set-up for capturing iris in visible spectrum.

Figure 4.3 provides the detailed illustration of images obtained from three different smartphones under the presented setup. Along with the images obtained from three different smartphones, an image of the same eye obtained in the conventional NIR setup is presented. The key factor to note from the illustration in Figure 4.3 is the visibility of the iris information obtained due to the set-up. It has to be further noted that, on the left part of the Figure 4.3, the iris pattern is not visible either with or without flash. However, for the samples in the third column that are captured using LED based approach, the pattern of the iris becomes visible to a greater degree. It has to be noted that iris texture information obtained using the proposed approach significantly differs from iris texture information obtained in NIR spectrum. Indeed, the proposed lightning does improve the visibility of dark eyes but the melanin still partly hides the texture which would have been fully visible with NIR lighting. Nonetheless, the tests show the observed texture has enough entropy to be used for biometrics purposes. Further tests with longer span of time between acquisition sessions would be required to assess the stability.

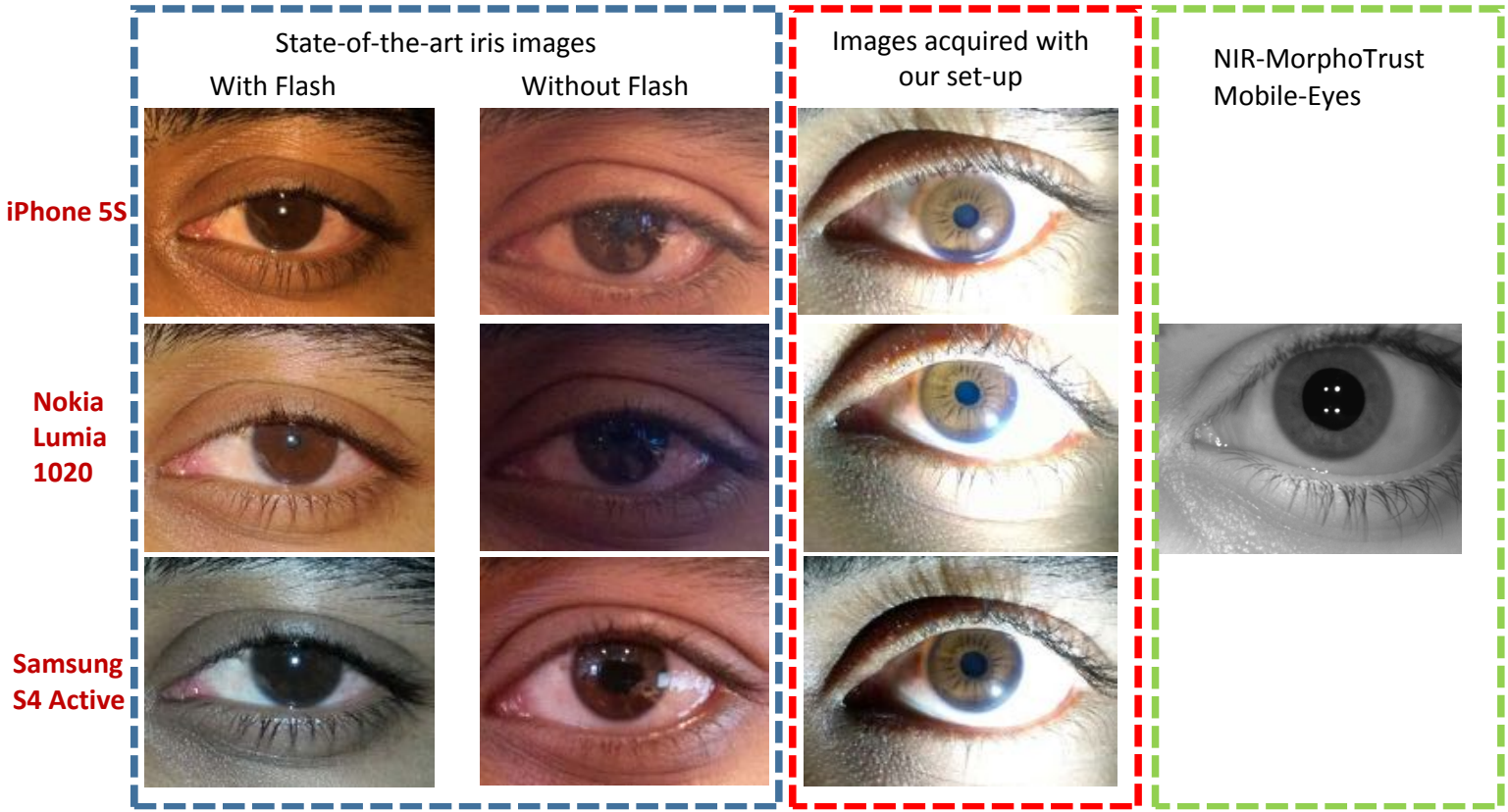


Figure 4.3: Comparison of improved visibility of iris pattern obtained using various smartphones and a conventional NIR capture device as baseline. It can be observed from the images that the proposed approach provides better texture visibility but does not necessarily correspond to textural information obtained in NIR spectrum using NIR-MorphoTrust device.

4.2.2 Segmentation and Normalization

The captured iris images are segmented prior to processing the texture information. Based on the proven robustness for segmentation of iris in both NIR domain [135] and also in the visible spectrum domain [76], in this chapter, OSIRIS V 4.1 [135] is employed. As the segmentation algorithm is complemented by Viterbi search for robust mask creation, the iris and pupil boundary are well localized in OSIRIS V 4.1 [135]. In the case of visible spectrum iris recognition, factors such as non-uniform illumination require a robust noise mask and OSIRIS v4.1 performs localization of noise to a better degree [135]. Followed by the segmentation, the iris texture is normalized using Daugman's rubber sheet expansion technique [29]. The dimension of a normalized iris image in this chapter is fixed to 512×64 pixels.

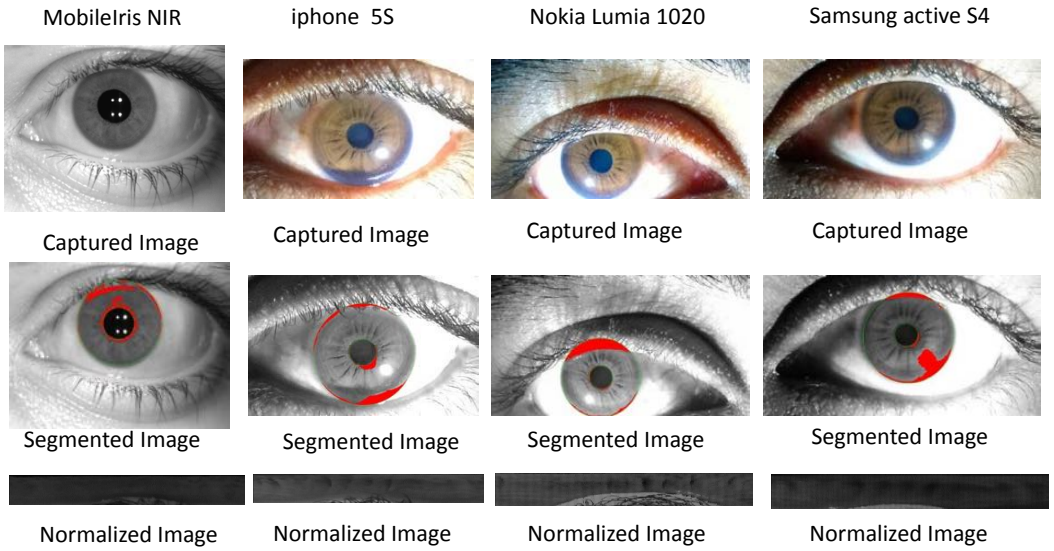


Figure 4.4: Illustration of segmentation and normalization of iris images.

4.2.3 Feature extraction and comparison

The features are extracted from the normalized iris images. The texture features are obtained using five state-of-the-art techniques. We have employed texture feature extraction based on well-established algorithms - 2D Gabor features [29], 1D Log Gabor features [91], multichannel spatially filtered features [84], cumulative sum of gray value features [78] and 1D Log Gabor features [91] with sparse representation [76].

In order to compare the feature vector from reference and probe iris, we employ Hamming Distance (HD) score in comparison subsystem for most of the techniques [29, 91, 78, 84] and residual scores obtained from comparison of sparse representation of 1D Log Gabor feature [76].

4.3 Database

In order to evaluate the applicability of the imaging set-up based on LED, a database has been constructed. The database consists of images obtained from 31 unique subjects amounting to 62 unique iris instances. Each unique iris instance is captured using three different smartphone cameras. For each subject, images are captured in 10 different attempts with a duration of approximately 5-15 minutes between each acquisition. As the LED based setup is intended to improve the iris texture visibility of dark irises, the data is collected

from individuals of Eastern ethnicity, who typically exhibit dark iris pattern due to high melanin pigmentation and higher density of collagen fibrils. Typically, the texture pattern and trabecular mesh is not visible fully or it is visible to the minimal level with regular imaging method using the smartphones. Figure 4.5 presents an illustration of the exact challenge. It can be observed that Figure 4.5(a) captured with smartphone, which has almost nil texture visibility and Figure 4.5(b) presents the same iris instance captured using LED based imaging set-up with clear texture visibility. The crypts, furrows and annular rings in the iris can be observed clearly in the image captured using LED based set-up.

Figure 4.5: Illustration of texture visibility with LED based set-up; (a) Iris instance captured in smartphone demonstrating poor texture visibility; (b) Iris instance with superior texture visibility when captured using LED based set-up.

Three smartphones are employed in this work - Apple iPhone 5S, Nokia Lumia 1020 and Samsung Active S4. Each unique eye is captured with all three smartphones. A total of 10 samples are captured for each iris instance with each smartphone and thus a total of 620 images are obtained in one set. The complete overview of a number of iris images in the database is presented in the Table 4.1. The database consists of 1860 images in total.

Table 4.1: Details of the database for iris recognition in our current work

Smartphone	Subjects	Unique Eyes	Samples	Total images
Apple iPhone 5S	31	62	10	620
Nokia Lumia 1020	31	62	10	620
Samsung Active S4	31	62	10	620
NIR - *MorphoTrust Mobile-Eyes* [92]	12	24	10	240

4.3.1 Near Infra Red Database

In order to benchmark the LED based illumination method in the visible spectrum iris recognition on smartphones, we also use a complementary iris image database acquired using a conventional Near Infra Red (NIR) iris imaging device *MorphoTrust Mobile-Eyes* [92].

A smaller subset of 24 unique iris was acquired based on the willingness of the participants to participate in both capture processes. Thus, the images are obtained from subjects corresponding to the volunteers in the visible spectrum iris database obtained using smartphones. 10 samples were acquired in 10 different attempts with a duration of $5 - 15$ minutes between each acquisition. Table 4.1 also provides details on the number of images present in the NIR database employed in this chapter.

4.4 Experiments and Results

This section provides the details of the experimental protocol to evaluate the applicability of images obtained using LED based set up. As described in Section 4.3, the database consists of two different subsets. The first subset of data stems from the capture of the iris in the visible spectrum using the LED based setup. The second subset of data originates from using the standard NIR iris acquisition device. Thus, we have two subsets and corresponding experiments outlined in this chapter.

4.4.1 Experiments on Visible Spectrum Iris Recognition

In this set of experiments, we evaluate the performance of the iris recognition algorithms for the iris captured using white LED. As the data is collected using three different smartphones, we have three different sets of evaluation. Exploiting the availability of 10 samples for each unique eye instance, we employ one sample as the reference image and the rest of the 9 samples as probe images. We continuously change the reference image 10 times such that all the images corresponding to one unique eye become reference image at least once. All scores obtained from the comparison are accumulated to generate the final set of genuine and impostor scores. For each set of evaluation, we have obtained 2790 genuine scores and 189100 impostor scores for iPhone, Nokia and Samsung data.

4.4.1.1 Results on Visible Spectrum Iris Recognition

Table 4.2 presents the results obtained on the visible spectrum iris dataset acquired using LED. It can be observed that the verification rate is consistently good when following the LED based approach. It has to be noted that the iris pattern would not be visible without using LED for the involved dark eyed capture subjects. Figure 4.5 has provided an illustrative example of the low texture visibility.

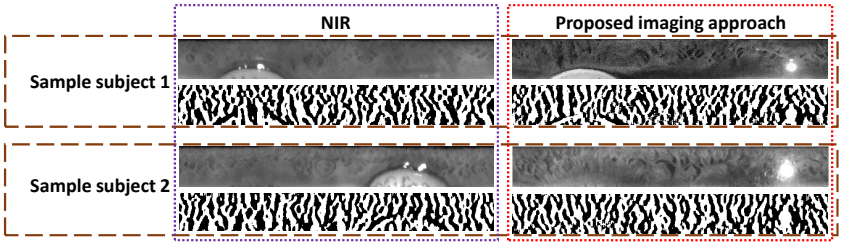


Figure 4.6: Illustration of iris texture visibility in normalized iris pattern using LED based approach for sample subjects. Both the images from NIR and visible spectrum capture using LED approach present good texture visibility but different information.

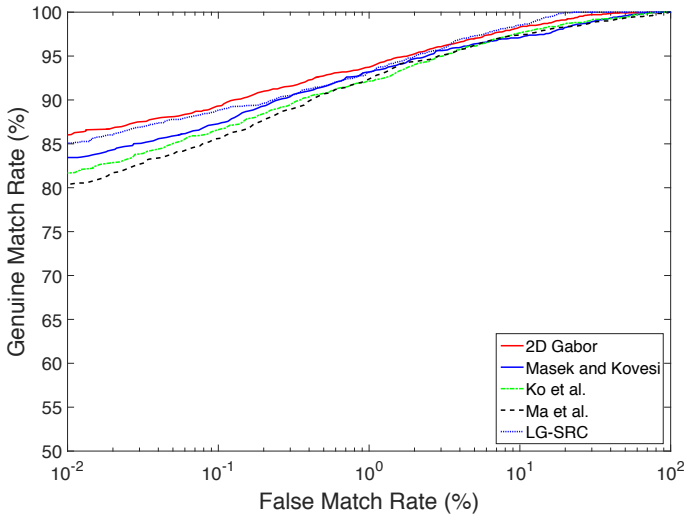
Figure 4.6 provides a sample illustration of the iris texture visibility using the LED based setup. It can be noted that the texture is highly visible even for a dark-eyed subject. For the simplicity of illustration, we have also provided the iris images of the same subject captured in the NIR domain. Furthermore, each iris is accompanied by the 2D Gabor features corresponding to one single scale as illustrated in Figure 4.6.

Table 4.2: LED based iris recognition verification rate
*OSIRIS v4.1 implementation of algorithms †University of Salzburg Iris-Toolkit v1.0 implementation of algorithms

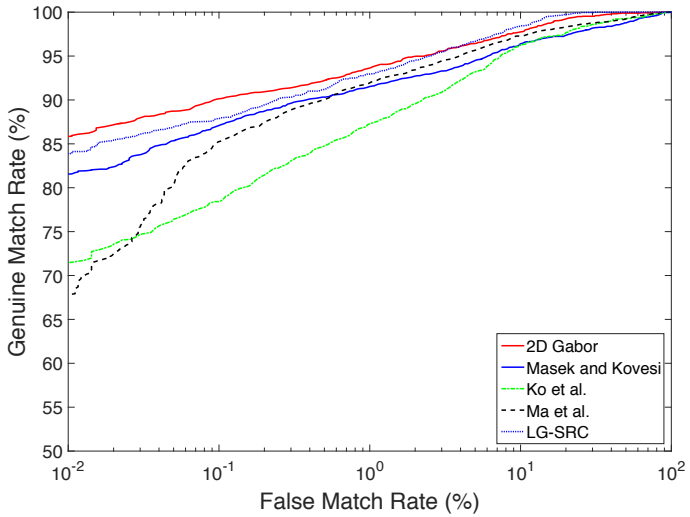
Phone	Verification Accuracy					
	iPhone 5S		Nokia Lumia 1020		Samsung Active S4	
Feature Extraction	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)
2D Gabor *[29]	85.98	3.61	85.8	3.94	78.04	4.55
Masek & Kovesi †[91]	83.44	4.01	81.57	5.38	66.81	5.71
Ko et al. †[78]	81.72	4.15	71.57	6.09	63.33	4.54
Ma et al. †[84]	80.35	4.18	67.84	4.53	53.44	5.49
LG-SRC †[91]	85.08	3.54	83.87	3.94	80.03	4.52

The best GMR of 85.98% is obtained from the data acquired in the LED setup with iPhone 5S as indicated in the Table 4.2. Further, the GMR of 85.8% and 80.03% are obtained for Nokia and Samsung respectively, again at the FMR of 0.01% indicating the robust performance of the LED based setup. It can be correlated to the superior texture visibility as illustrated in the Figure 4.6.

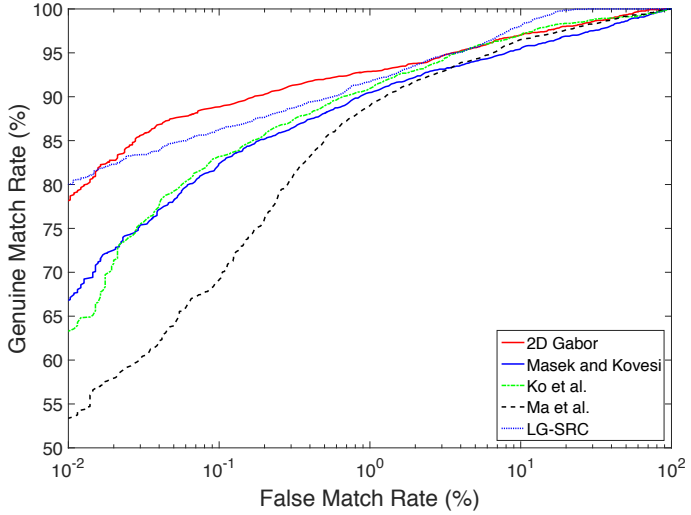
The analysis of the iris data captured indicates that the LED based setup can be supported by any of the employed smartphone hardware and can easily be extended to other smartphones. It is interesting to note that the GMR obtained from different smartphones under the LED based setup is consistently higher than 80% at $FMR = 0.1\%$, which indicate the applicability in a real life verification scenario. Figure 4.7 presents the Receiver Operating Characteristics (ROC) curves for iPhone, Nokia, and Samsung phone. The consistent performance of the LED based setup can be confirmed at various FMR by observing the ROC curves.



(a) iPhone Iris Dataset



(b) Nokia Iris Dataset



(c) Samsung Iris Dataset

Figure 4.7: ROC curves for iris recognition on various smartphones using the LED based approach

4.4.2 Experiments on NIR spectrum iris recognition

As discussed in the earlier section, in order to benchmark the performance of images from LED based set-up versus a conventional NIR performance, we use a subset of iris images in the NIR domain corresponding to involved volunteers, that were also enrolled in the visible spectrum database. Similar to the protocol mentioned earlier, due to the availability of 10 samples per eye instance, we consider one sample as a reference and the rest of 9 samples as a probe. The reference sample is iteratively swapped such that each sample becomes reference at least once. All the scores obtained from the comparisons are combined to form final genuine and impostor scores. The number of genuine and impostor scores generated in this set of experiments is detailed in the Table 4.3.

Table 4.3: Distribution of genuine and impostor scores in NIR-Visible Spectrum iris database

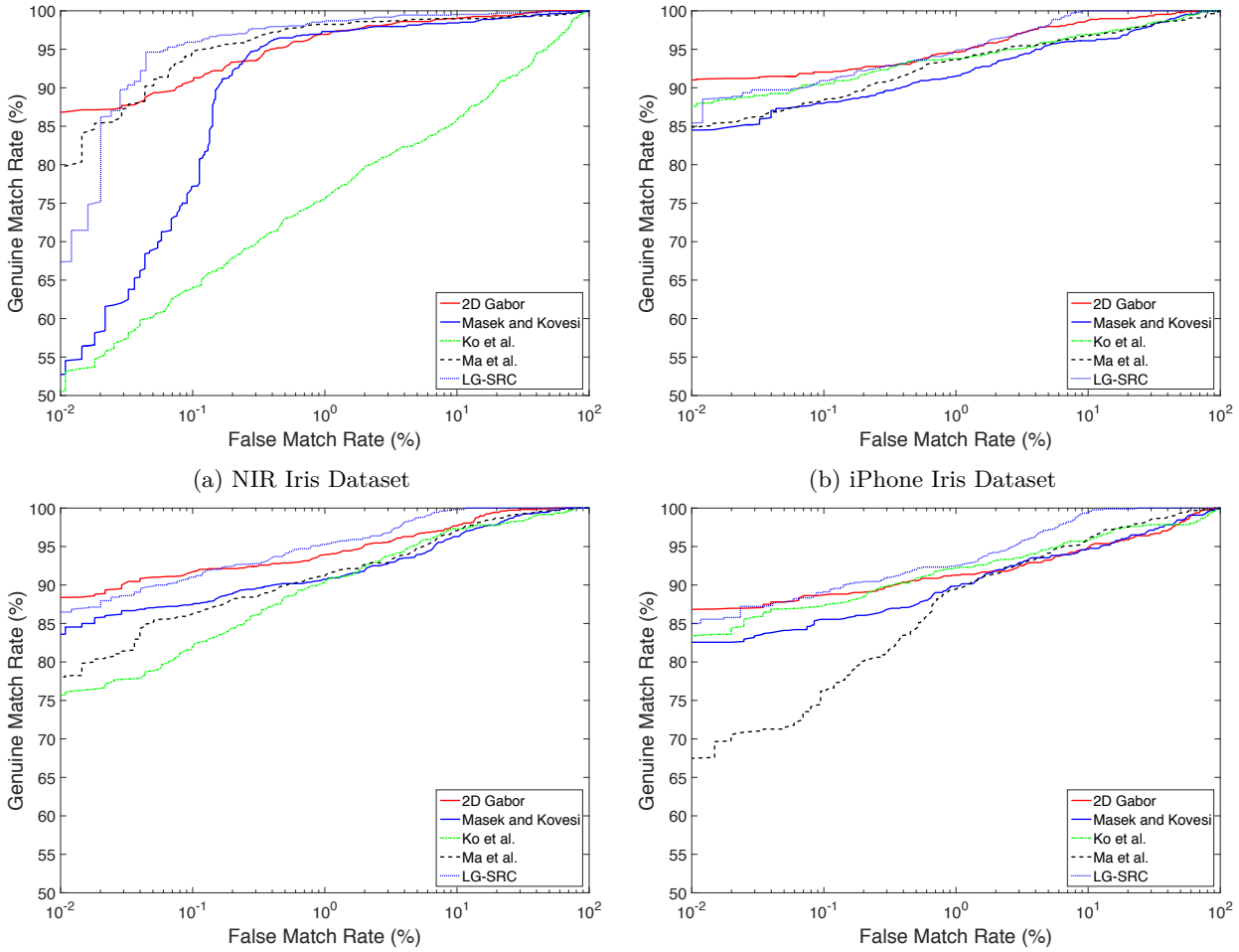
Phone	Number of unique eyes	Samples per eye	Genuine score	Impostor score
iPhone 5S	24	10	1080	27600
Nokia Lumia 1020	24	10	1080	27600
Samsung Active S4	24	10	1080	27600
NIR -*MorphoTrust Mobile-Eyes* [92]	24	10	1080	27600

4.4.2.1 Results on NIR spectrum iris recognition

The benchmark performance of iris recognition obtained using NIR versus the visible spectrum using LED based approach for the subset of data as mentioned in Section 4.4.2 is provided in Table 4.3. The corresponding ROC curves are presented in Figure 4.8. The obtained performance for 24 unique eyes instances in NIR data in terms of GMR is 91.01% at FMR of 0.01% that corresponds to images captured with iPhone 5S. Comparing the GMR obtained for the same set of data with the NIR device, LED based set-up indicates an equivalent performance. It can also be observed from the ROC curves that the obtained GMR is consistently higher than 80% even at lower FMR indicating the superior or equivalent performance as compared to NIR images.

Table 4.4: Benchmark performance of LED versus NIR dataset. Reported accuracy for individual feature extraction methods on smartphone differs from results in Table 4.2, as the dataset is significantly smaller.
*OSIRIS v4.1 implementation of algorithms
†University of Salzburg Iris-Toolkit v1.0 implementation of algorithms

Feature Extraction	GMR @ specified FMR (%)							
	iPhone 5S		Nokia Lumia 1020		Samsung Active S4		NIR - *MorphoTrust Mobile-Eyes* [92]	
	FMR=0.01%	EER (%)	FMR=0.01%	EER (%)	FMR=0.01%	EER (%)	FMR=0.01%	EER (%)
2D Gabor *[29]	91.01	2.96	88.33	3.88	86.87	6.15	86.85	2.05
Masek & Kovesi †[91]	84.53	4.62	83.61	5.76	82.56	6.11	52.68	2.15
Ko et al. †[78]	87.59	4.26	75.74	4.83	83.58	5.15	50.55	12.49
Ma et al. †[84]	84.9	4.38	78.05	5.18	67.58	5.43	79.81	1.56
LG-SRC †[91]	85.37	2.96	86.48	3.14	85.89	3.86	67.4	1.3



(a) NIR Iris Dataset

(b) iPhone Iris Dataset

(c) Nokia Iris Dataset

(d) Samsung Iris Dataset

Figure 4.8: ROC curves for various dataset acquired using LED based approach. (a) Benchmark performance obtained using NIR images; (b)-(d) Performance obtained using LED based approach for various smartphones

4.4.3 Evaluation of Commercial Iris Recognition System

Most of the iris recognition based authentication system in real-life verification scenarios employ commercial-off-the-shelf (COTS) algorithms. To measure the applicability of LED based set-up, we have evaluated VeriEye commercial algorithm [148]. To provide the comparison with respect to an open-source algorithms, we have retained the experimental protocols as discussed in previous sections.

VeriEye SDK is highly tuned to work with NIR iris images and has been proven to work well even in NIST IREX performance evaluation [148]. As the VeriEye SDK fails to extract the features/template for images obtained using the smartphone in the visible spectrum, the performance metrics such as EER, FMR and FMNR do not hold good [54]. All the images for which the template extraction fails must be treated as Failure-to-Enroll (FTE) as discussed in Chapter 2. As the acquisition of images are continued until the images of satisfactory quality are obtained, Failure-to-Acquire (FTA) equals to zero. The final performance is provided in Genuine Match Rate (GMR) and Generalized Equal Error Rate (GEER) considering the FTA and FTE.

The performance obtained from the VeriEye SDK is provided in the Table 4.5. Significant FTE can be observed from the Table 4.5 for iPhone data and moderately low amount of FTE for Nokia and Samsung data. However, the FTE equals to zero for NIR data as VeriEye successfully extracts template for NIR iris images. Further, from the Table 4.5, it can be noted that GMR of 91.86% is obtained for data acquired using Nokia Lumia 1020 and GMR of 100% is observed for NIR data acquired from MorphoEyes. Along the same lines, one can observe GEER obtained for NIR data is 0% and 6.74% for data obtained using Nokia smartphone. Figure 4.9 presents the ROC for the COTS evaluation of NIR images and images acquired using proposed set-up.

Table 4.5: Performance for images captured from NIR sensor and smartphones in visible spectrum using VeriEye SDK

Sensor / Smartphone	FTE(%)	GMR (%) @ FMR = 0.01%	GEER (%)
NIR - MorphoTrust Mobile-Eyes	0	100	0
iPhone 5S	29.09	78.98	14.15
Nokia Lumia 1020	4.64	91.86	6.74
Samsung Active S4	6.70	89.53	8.46

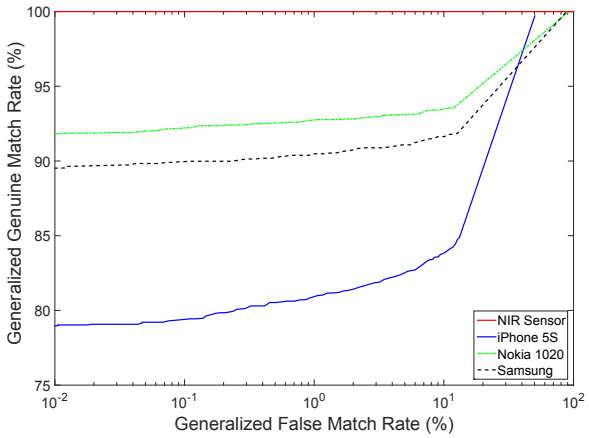


Figure 4.9: Performance curves obtained using VeriEye commercial SDK for NIR iris images and images acquired using proposed setup in visible spectrum

The performance obtained using COTS algorithm supports the applicability of the LED based imaging set-up for the real-life verification system.

4.5 Conclusions

The concentration of melanin pigments and the density of the collagen fibrils in iris structure governs the color of the iris. A lower concentration of melanin and lower density of collagen fibrils result in the light colored eye. A light colored eye can easily be captured in the visible spectrum as the shorter wavelength light is scattered. However, to image dark colored iris, NIR has been used as the light in the shorter wavelength range is easily absorbed.

In this chapter, we have explored an imaging set-up specifically deigned to image dark colored iris using white LED illuminated in a acute angle. The specifications of the LED used in this chapter is close to the specification of flash illumination used in smartphones and thereby presented an alternative way to engage the smartphones effectively to capture iris in visible spectrum, even heavily pigmented iris. Furthermore, to validate the robustness of the LED based setup for imaging the dark iris, we have employed a database captured using three new smartphones. The best GMR of 91.01% at $FMR = 0.01\%$ obtained for iPhone data validates the applicability of presented approach in everyday authentication in low security applications for dark colored iris.

The presented approach provides an alternative to the NIR light on smartphones by making use of the existing LED flash illumination. The realization of the proposed illumination needs engineering effort to integrate the flash such that it can illuminate the iris at an inclined angle in a similar manner to the design of recent smartphone Oppo N1 as discussed in this chapter. Benchmark performance evaluation of iris recognition in the visible spectrum using LED based approach versus NIR illumination has indicated good performance that is comparable to NIR data. The LED based approach of imaging is promising to adopt in real life verification scenarios for everyday authentication. Further, the large scale experiments with bigger dataset of images needs to be carried out in future works of this direction and compared against the NIR database to validate the initial observations from this chapter.

Chapter 5

Deep Sparse Time Frequency Features for Robust Verification of Periocular Images From Smartphones

The periocular region is used for authentication in the recent days under unconstrained acquisition in biometrics. This chapter presents a new approach to perform biometric authentication using periocular region captured in smartphone embedded camera operating in visible spectrum. A new feature extraction is presented to achieve robust verification performance and is experimentally validated using three publicly available databases.

5.1 Introduction

The use of iris recognition on smartphones in visible spectrum is limited by number of factors detailed in the Chapter 3 and Chapter 4 . The low visibility of texture due to multiple factors plays key role in segmenting iris pattern, which in turn influences the performance of biometric system. As an alternative, many works [100, 108, 107, 102, 101, 73, 117] have employed the region around the eye for biometric recognition of an individual with considerable biometric performance.

The periocular region *a.k.a ocular region* can be defined as the area of the face that includes the eyelids, eyelashes, eyebrow, and the skin surrounding the eye [102]. Ocular characteristics can be obtained either while capturing the face image or the iris image alone [102, 11]. Many of the earlier works have advocated periocular region as a supplement information for iris systems [155, 43, 136] or face recognition systems [102]. In the context of smartphone captured data, it was also shown to supplement/improve the performance in a multi-modal authentication scenario [129]. Further, in our earlier work [73], periocular region was used as a stand-alone characteristics on smartphone based authentication systems operating in a semi-cooperative environment. The usefulness of periocular region in multi-modal system was also iterated in our works [75, 74]. The rise of interest in using ocular biometrics in the visible spectrum is partially due to ease of imaging with a simple RGB camera without additional hardware. Thus, one can use any existing cameras including cameras embedded on smartphones operating in visible spectrum to capture ocular characteristics.

In this chapter, we explore periocular region as a stand-alone biometric characteristic for authentication applications using smartphone captured data in both semi-cooperative and unconstrained capture conditions. We first present a new feature extraction scheme based on time-frequency features of *deep sparse filtering* which we refer hereafter as **De***ep* **S***parse* **Ti***me* **F***requency* **F***eatures* (*DeSTiFF*) and employ it for feature extraction of periocular images. We evaluate the strengths of the new feature extraction scheme on publicly available periocular databases captured using smartphones to gauge the applicability. The features from new scheme are evaluated on both semi-cooperative and unconstrained data captured from smartphones in visible spectrum. It has to be noted that the results on the preliminary version of the newly proposed scheme (i.e., *Deep Sparse Filtering*) has been reported on a large scale database (VISOB) in our earlier

work [69, 123].[1]

In the rest of the chapter, Section 5.3.1 presents the newly proposed feature extraction scheme followed by the Section 5.4 which presents the set of databases employed in this work. Further, the evaluation and results obtained on the databases are presented in Section 5.5. Finally, the observations and remarks are detailed in Section 5.6.

5.1.1 Contributions

The main contributions of this chapter are:

1. Presents a new feature extraction technique to obtain robust and discriminant features from periocular images captured from smartphone embedded camera operating in visible spectrum.

2. Presents a new periocular database hereafter referred as Visible Spectrum Smartphone Periocular (ViSPer) database consisting of 152 unique periocular instances collected using two different smartphones (iPhone 5S & Nokia Lumia 1020) and 100 unique periocular instances from Samsung S5. The ViSPer database is freely distributed for the non-profitable research purpose.

3. Presents extensive set of experiments to demonstrate the applicability of proposed technique through evaluation on publicly available periocular databases along with the newly captured database from smartphones operating in visible spectrum.

5.2 Framework for Periocular Recognition

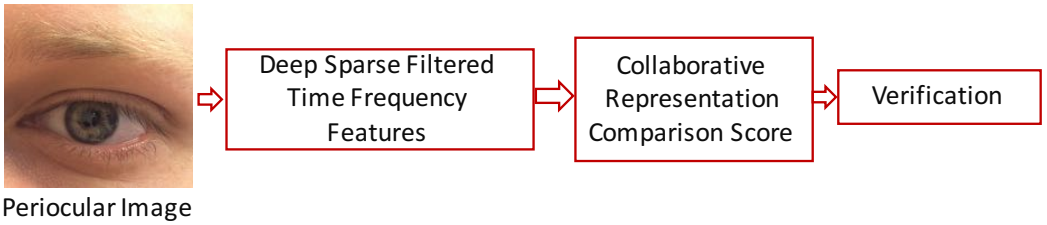


Figure 5.1: Illustration of the proposed framework for verification.

Compliant to a general biometric verification system, the enrolment images (captured from smartphone) are used to extract the features and are stored as templates in the enrolment database. When the probe image (captured from smartphone) is given, the features are extracted in a similar manner as for enrolment images. Features from both enrolment and probe image are compared to make a decision to accept the genuine user or reject the impostor. The framework for periocular based verification is depicted in the Figure 5.1.

From the periocular image, $DeSTiFF$ features are extracted using newly proposed scheme. The features from probe image are compared against the enrolment features using collaborative representation classifier [163] as detailed in Section 2.2.4 of Chapter 2. The score obtained from the collaborative representation classification is used to verify or reject the subject.

[1] A detailed ROC on the performance obtained on large scale database is presented in the Appendix C.

5.3 Feature Extraction Schemes for Periocular Images

This section presents the new feature extraction scheme based on *Deep Sparse Filtering* followed by time frequency feature localization employed in this chapter to obtain robust features from ocular images.

5.3.1 Deep Sparse Time Frequency Features (DeSTiFF) for Periocular Images

As discussed in Section 3.3.1 of Chapter 3, *Sparse Filtering* can be used to learn a set of filters to extract features from periocular images. Similar to the approach described earlier in Section 3.3.1 of Chapter 3, we employ two layered architecture to learn the filters where the input to learn *layer* 1 is the patches obtained from natural images [50] and the output from *layer* 1 is used to learn the *layer* 2. The set of filters obtained from the *layer* 2 are used to obtain the features from the images. Following the discussions in Chapter 3, Section 3.3.1, *layer* 1 is optimized using l_1 norm and the *layer* 2 using l_2 norm. Based on the performance obtained in earlier work [76], we limit the framework to learn 256 filters of size 16×16 pixels using natural scene data. We convolve each of the gray level periocular image with 256 filters learnt.

The responses are further binarized using a simple threshold as given in Equation 5.1.

$$b_i = \begin{cases} 1, & \text{if } s_i > 0 \\ 0, & \text{otherwise} \end{cases} \tag{5.1}$$

where s_i is the convolved response for i^{th} pixel obtained using a deep sparse filter. As the there are 256 binary responses available for a pixel, it is critical to reduce the dimensionality while preserving the uniqueness of the responses. Thus, we combine a set of 8 binarized responses in a sequential format using Equation 5.2.

$$g(x,y) = g(x,y) + (b(x,y)_i) * (2^{(i-1)}) \; for \; i = 1:8 \tag{5.2}$$

where $g(x,y)$ is the final encoded response for the pixel at x,y location. Thus the final encoded gray-level responses amount to 32 images. The set of gray-level response images can be represented as $G \in \{G_1, G_2 \ldots G_{32}\}$ which is a group of 32 images.

As discussed in the earlier chapter 3, the image characteristics change based on different environmental conditions while capturing the periocular data. The non-uniform illumination limits the visibility of local texture from periocular region. Further, unconstrained capture of periocular region results in out-of-focus and motion blurred image as observed commonly. Thus, the descriptor needs to be blur invariant to obtain the texture information optimally from periocular image. Techniques like Local Phase Quantization (LPQ) [97] have proposed to explore the phase information to be invariant to illumination changes. The technique also advocates the use of low-frequency phase components which are invariant to the centrally symmetric blur [97]. Motivated by such approach, we obtain the localized frequency responses to obtain the features in illumination and motion invariant manner.

Thus, in this chapter, we extract localized time and frequency features in large windowed region by using Short Term Fourier Transform (STFT) with a window of size 60×60 based on empirical trials on the MICHE-I database. Given an image G_k from the set of gray-level response images where k ranges from $1 - 32$, the STFT of the image can be represented as F_k which is the image resulting to response of frequency components in four different orientations such that $\Phi = \{0°, 45°, 90°, 135°\}$.

The filter response obtained from each orientation are separated for real and complex values subsequently. Each of the responses denoted by b is finally encoded to form the final response image with localized time and frequency as given by FR_k where i corresponds to

different orientation angles given by $\phi = \{0°, 45°, 90°, 135°\}$.

$$FR_k = \mathrm{Re}(\sum_{i=1}^{4}(b_i) * (2^{(i-1)})) + \mathrm{Im}(\sum_{i=1}^{4}(b_i) * (2^{(i-1)})) \tag{5.3}$$

The image denoted by FR corresponds to the final features which is here after referred as *Deep Sparse Time Frequency Features* (*DeSTiFF*). The features from each of the 32 images result in very large dimension if processed. Thus, each of the response image is represented as histogram in this chapter. The final *DeSTiFF* feature vector $\mathfrak{F}$ for each image in the response image set $k = 1 : 32$ is given by:

$$\mathfrak{F}_k = \sum_{i=0}^{255}\{FR_k\}_i \quad \forall\ k = 1 : 32 \tag{5.4}$$

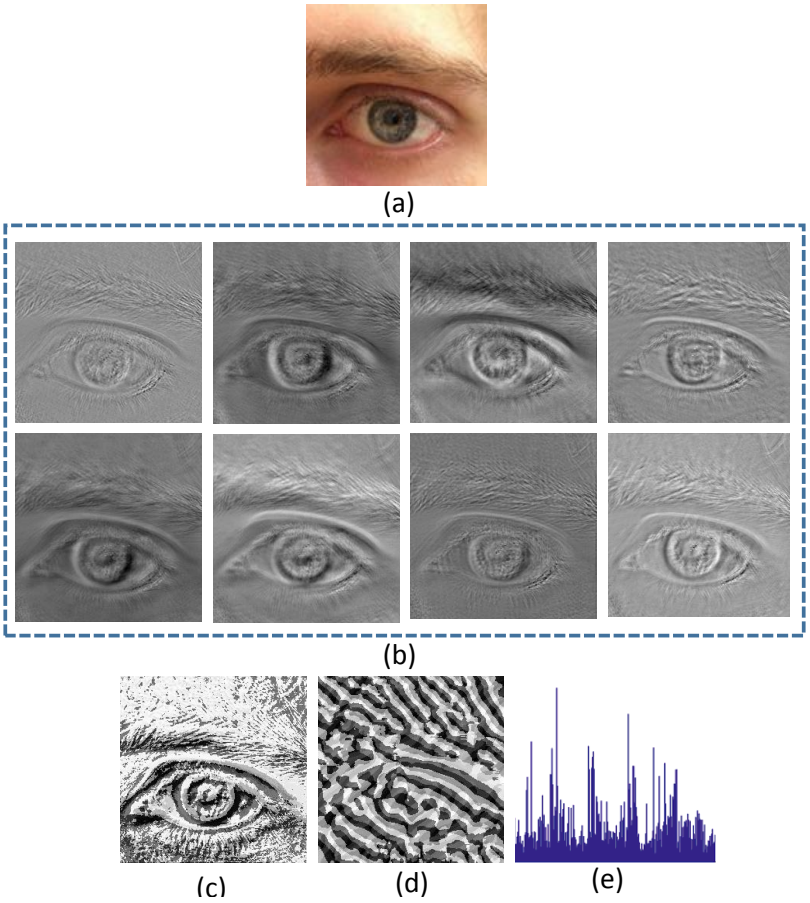


Figure 5.2: Illustration of *DesTiFF* features. (a) Sample periocular image; (b) Deep Sparse response for a set of 8 filters out of 256 filters; (c) Grouped response of 8 sequential response from (b); (d) STFT response for (c); (e) Final histogram features (*DeSTiFF*) for set of 8 filters.

Figure 5.2 presents the illustration of all the steps involved in extracting the *DeSTiFF* features. Figure 5.2(a) depicts a sample periocular image and Figure 5.2(b) shows a set of 8 responses corresponding to set of 8 filters in deep sparse filters. Figure 5.2(c) is the grouped response of all the 8 filters shown in Figure 5.2(b). Finally, the time-frequency features are

extracted using STFT as shown in Figure 5.2(d). The $DeSTiFF$ features correspond to histogram features depicted in Figure 5.2(e). In the similar fashion, the histogram for all 32 responses are obtained to form the final features as given by Equation 5.4.

5.4 Periocular Databases

The ocular biometric databases collected using the smartphones in visible spectrum are very limited, amongst which few are publicly available for research. The limited number of databases are further posed with a problem of limited number of subjects. In this thesis, we explore three publicly available periocular databases collected using various smartphones which can be listed as - **Vi**sible spectrum **S**martphone **Per**iocular (ViSPer) database, MICHE-I database [34] and VISOB database [69, 112].

5.4.1 Visible Spectrum Smartphone Periocular (ViSPer) Database

Visible spectrum **S**martphone **Per**iocular (ViSPer) database was constructed in the due course of this thesis. It presents a set of periocular images captured using three different smartphones - iPhone 5S, Nokia Lumia 1020 and Samsung S5. The ocular images are captured in a mixed illumination environment using the rear camera of the smartphones in a semi-cooperative manner. A total of 152 unique periocular instances are captured from 76 unique subjects from iPhone 5S and Nokia Lumia 1020 smartphones while 100 unique periocular instances are captured from Samsung S5. Each unique periocular image has 10 samples captured in different instances. The total distribution of the images in the database is presented in the Table 5.1.

Table 5.1: Distribution of ViSPer database

Details	Smartphone		
	iPhone 5S	Nokia Lumia 1020	Samsung S5
Capture Scenario	Mixed Illumination	Mixed Illumination	Mixed Illumination
Resolution	12 Mp	41 Mp	16 Mp
Number of subjects	76	76	50
Unique periocular instances	152	152	100
Samples per unique periocular instance	10	10	10
Total images	1520	1520	1000

The ocular images are captured in the uncontrolled environment where the influence of both the external illumination from sunlight and the illumination from artificial room light are present. This database is constructed to have the ocular image, which consists of eye region and the eye-brow region. The sample images from the periocular database are illustrated in Figure 5.3. The images in the database also present everyday appearances that include the make-up in the case of female subjects and non-uniform illumination. Further, as the images correspond to the real life acquisition scenario, the images present various forms of degradation due to motion blur and blinking of eye.

Figure 5.3: Sample images from ViSPer database

It can be observed from Figure 5.3 that the images vary in terms of appearance in different smartphones and the factors of degradation are not constant across the phones/subjects.

5.4.2 MICHE-I Periocular Database

MICHE-I database provides periocular images collected using two different smartphones - *iPhone 5* and *Samsung Galaxy S4* [34]. The images in the database are acquired from 75 subjects using the frontal and rear camera in both indoor and outdoor illumination conditions. During the indoor acquisition mode various sources of artificial light are combined with natural light sources while during the outdoor acquisition mode data capture takes place using natural light only. For each subject only one of the two ocular region was acquired in this dataset. Each unique periocular instance is captured in four different sessions resulting in 4 samples. Figure 5.4 presents illustrative images from the MICHE-I database. Table 5.2 presents the composition of the database detailing the images in all different conditions.

Table 5.2: Distribution of MICHE-I Periocular database.
* Not uniform across all subjects

Details	Smartphone	
	iPhone 5	Samsung S4
Rear Camera Resolution	8 Mp	13 Mp
Front Camera Resolution	1.2 Mp	2 Mp
Number of subjects	75	75
Unique periocular instances	75	75
Samples per unique periocular instance	4*	4*
Frontal Camera		
Capture scenario	Indoor Illumination	Indoor Illumination
Total images	300	300
Capture scenario	Outdoor Illumination	Outdoor Illumination
Total images	300	300
Rear Camera		
capture scenario	Indoor Illumination	Indoor Illumination
Total images	300	300
Capture scenario	Outdoor Illumination	Outdoor Illumination
Total images	300	300

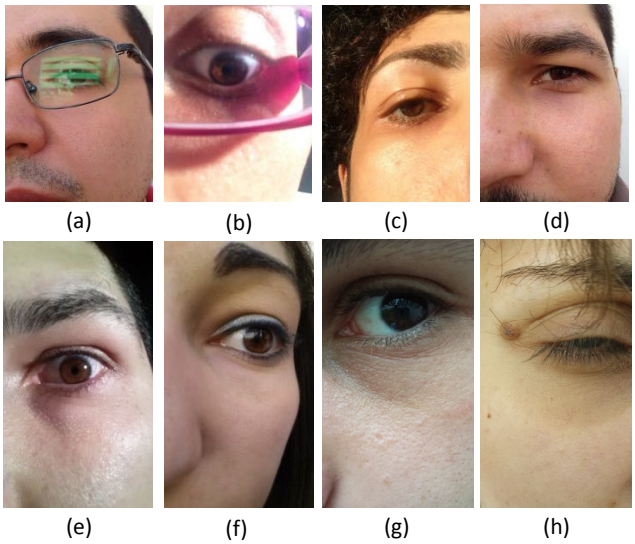


Figure 5.4: Sample periocular images from the MICHE-I database acquired using two different phones. (a)-(d) images captured from iPhone 5 and (e)-(h) images captures from Samsung S4.

5.4.3 Visible Light Mobile Ocular Biometric Database (VISOB) Database

A recent large scale database for ocular images captured using smartphones in visible spectrum is Visible Light Mobile Ocular Biometric Database (VISOB) [123, 69, 112]. Unlike other existing databases, this database consists of 580 subjects with variable number of images in disjoint set of enrolment and probes. The images are collected using three different smartphones - iPhone 5S, Samsung and Oppo. The key relevance of this database are listed below:

1. Visible Light Mobile Ocular Biometric Database (VISOB) is the largest ocular database collected from more than 550 subjects using front facing cameras of three different mobile devices: Oppo (13 MP), Samsung and iPhone 5S.
2. The key difference of the VISOB database from others is that it presents intra-class variations due to the characteristics of front facing cameras which reflect everyday mobile biometric use cases which result in out-of-focus images, occlusions due to prescription glasses, different illumination conditions, gaze deviations, eye-makeup (i.e., eye liner and mascara), specular reflections, and motion blur.

Figure 5.5: Illustration of sample periocular images from VISOB database.

The degraded quality and challenging nature of the images can be observed from the Figure 5.5. It can also be noted that the images in VISOB database do not have the eyebrows in the ocular images. The statistics of the database is provided in the Table 5.3. As it can be noted from the Table 5.3, there are no fixed number of samples in either enrolment or probe dataset. The number of images in each set vary greatly and are commendable as compared to any other existing smartphone ocular databases. The data is collected in three different illumination conditions which include daylight illumination, dim light illumination and a mixed illumination scenario in office environment. In each of the different illumination and under different smartphones, two sets of data corresponding to enrolment and probe are collected in different session resulting in two disjoint set of data.

Table 5.3: Distribution of VISOB Periocular database

Details	Smartphone		
	iPhone 5S	Samsung	Oppo
Resolution	12 Mp	16 Mp	13 Mp
Number of subjects	569	584	584
Unique periocular instances	1138	1168	1168
Capture scenario	Daylight Illumination	Daylight Illumination	Daylight Illumination
Total images	10373	6495	7896
Capture scenario	Dim-Light Illumination	Dim-Light Illumination	Dim-Light Illumination
Total images	7314	8476	14979
Capture scenario	Office-Light Illumination	Office-Light Illumination	Office-Light Illumination
Total images	9598	9466	20450

5.5 Experiments and Results

In this section, we report the set of experiments and results obtained all the three different periocular databases collected using smartphones as discussed in the prior section 5.4. In the set of all the experiments reported in this chapter, each of the periocular image is considered as unique identity (subject) as each of the periocular image exhibits a significant amount of unique information stemming from an individual. In order to evaluate the performance of periocular recognition, we present the results with different state-of-art schemes used in earlier works which includes Binarized Statistical Image Features (BSIF) in the original version and it's block based variant (17×17 with 12 scales) [64, 71], SIFT and SURF [101]. We have employed RANdom SAmple Consensus (RANSAC) [46, 19] for SIFT and SURF features while BSIF features are compared using *Bhattacharya* distance.

DeSTiFF features from periocular region are collaboratively represented for the robust classification [163]. The collaborative representation based classification is computationally inexpensive and can easily be used when the number of samples in each class are imbalanced.

All the results of periocular verification in this chapter are presented in terms of Genuine Match Rate (GMR) at False Match Rate (FMR) of 10^{-2} and as an indicative performance metric, Equal Error Rate (EER) is also provided.

5.5.1 Experiments on ViSPer Database

This section presents the protocols and results obtained on ViSPer dataset. As discussed in Section 5.4, ViSPer dataset consists of images captured using three phones - iPhone 5S, Nokia Lumia 1020 and Samsung S5.

5.5.1.1 Protocols for ViSPer Dataset

Each unique periocular image is captured in ten different sessions from each smartphone. Of the 10 different captures, 5 different samples corresponding to first 5 captures are treated as the enrolment set and the remaining 5 samples are considered as the probe set. Thus, in the experimental protocols, we divide the images in two disjoint sets of reference and probe. Each subset consists of 5 captured samples. The detailed number of images in each set is provided in the Table 5.4.

Table 5.4: Database division for experiments on ViSPer dataset

Smartphone	Per Subject		Total	
	Enrolment	Probe	Enrolment (samples × subjects)	Probe (samples × subjects)
Nokia	5	5	$5 \times 152 = 760$	$5 \times 152 = 760$
iPhone	5	5	$5 \times 152 = 760$	$5\times 152 = 760$
Samsung	5	5	$5 \times 100 = 500$	$5 \times 100 = 500$

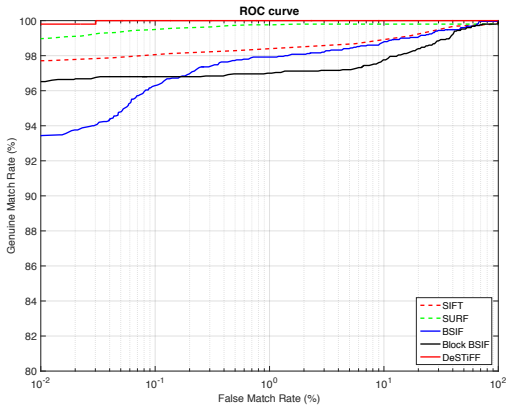
5.5.1.2 Results on ViSPer Dataset

Table 5.5 presents the results obtained using various state-of-art methods. It can be observed from the Table 5.5 that the *DeSTiFF* features for periocular verification out-performs other state-of-art algorithms for data stemming from different smartphones. Nonetheless, the state-of-art methods provide reasonably good performance that are comparable with presented features employing *DeSTiFF*.

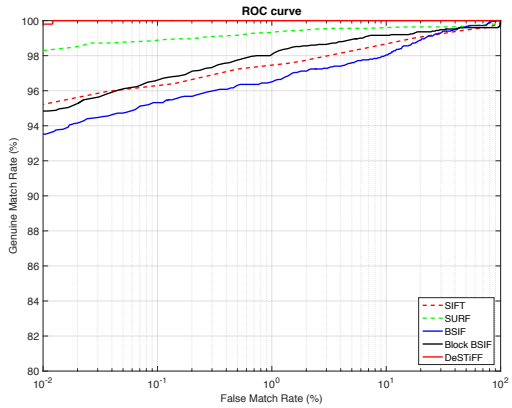
Table 5.5: Verification performance of various algorithms on ViSPer Database

Algorithm	Nokia		iPhone		Samsung	
	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%) EER
SIFT	95.96	2.06	97.72	1.59	97.19	1.83
SURF	98.32	0.70	99.00	0.33	98.12	1.31
BSIF	93.52	2.75	93.44	1.92	93.24	1.43
Block-BSIF	94.84	1.51	96.52	2.84	96.64	1.95
DeSTiFF	**99.76**	**0.01**	**99.80**	**0.02**	**99.78**	**0.01**

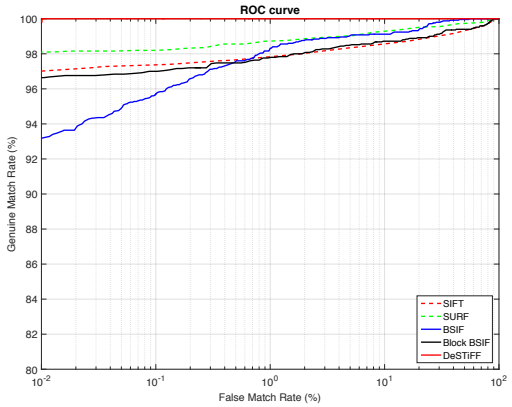
Table 5.5 presents the results that indicate 99.8% GMR at $FMR = 0.01\%$ with an EER of 0.01%, 0.02% and 0.01% for images captured from Nokia, iPhone and Samsung respectively. Further, F1igure 5.6 presents the performance of various algorithms on three subsets of ViSPer dataset. As discussed earlier, it can be noted that the $DeSTiFF$ features out-perform the verification accuracy of other algorithms in all three sets corresponding to images captured from iPhone, Nokia and Samsung.



(a) iPhone 5S



(b) Nokia Lumia 1020



(c) Samsung Galaxy S5

Figure 5.6: ROC curves for various algorithms on ViSPer database

5.5.2 Experiments on MICHE-I Periocular Database

This section presents the protocols and the results obtained on MICHE-I database. MICHE-I dataset consists of the images captures using iPhone and Samsung from both rear and frontal

camera. It has to be noted that, the images in MICHE-I database are partial face images of the subjects and thus, we have pre-processed the images to locate the eye-region which results in periocular region consisting of eye and the eyebrow.

5.5.2.1 Protocols on MICHE-I Periocular Database

In this chapter, we have used the protocols described in our works on this database [76] where the data is separated based on the placement of camera and the illumination condition. Thus, iPhone data is divided as 'iPhone-Indoor-Rear', 'iPhone-Indoor-Frontal', 'iPhone-Outdoor-Rear' and 'iPhone-Outdoor-Frontal'. Likewise, data captured from Samsung is subdivided into four sets. It has to be noted a subset of 50 users are employed for the experiments as number of samples across rest of the subjects are not equal. Further, we divide the dataset to contain three images in the enrolment set and rest of the one image for probe set. The final division of the dataset for the experiments are provided in Table 5.6.

Table 5.6: Database division for experiments on MICHE-I dataset

Illumination	Smartphone	Camera	Per Unique Periocular		Total	
			Enrolment	Probe	Enrolment	Probe
Outdoor	iPhone	Front	3	1	50*3 = 150	1 × 50 = 50
		Rear	3	1	50 × 3 = 150	1 × 50 = 50
	Samsung	Front	3	1	50 × 3 = 150	1 × 50 = 50
		Rear	3	1	50 × 3 = 150	1 × 50 = 50
Indoor	iPhone	Front	3	1	50 × 3 = 150	1 × 50 = 50
		Rear	3	1	50 × 3 = 150	1 × 50 = 50
	Samsung	Front	3	1	50 × 3 = 150	1 × 50 = 50
		Rear	3	1	50 × 3 = 150	1 × 50 = 50

5.5.2.2 Results on MICHE-I Periocular Database

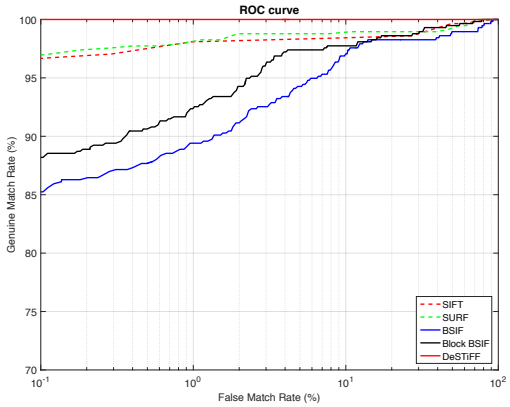
Table 5.7 presents the results obtained on all the dataset acquired in the indoor condition with mixed illumination and Table 5.8 presents the results obtained on the data captured in outdoor-illumination. It has to be noted that the performance is reported at $FMR = 0.1\%$ as the dataset is limited in size. As the data is relatively challenging as compared to the ViSPer dataset, a degraded performance in the state-of-art techniques can be observed. The trends of lower performance is seen in the data captured in both indoor and outdoor illumination for state-of-art techniques. Further, it can be noted from the Table 5.7 and Table 5.8 that the proposed feature extraction algorithm ($DeSTiFF$) has performed consistently well in all of the different cases. The robustness can be evidently observed from the lowest EER (close to 0%) and higher GMR (close to 100%) obtained on all different datasets.

Table 5.7: Verification performance of various algorithms on the data captured in indoor illumination in MICHE-I periocular dataset

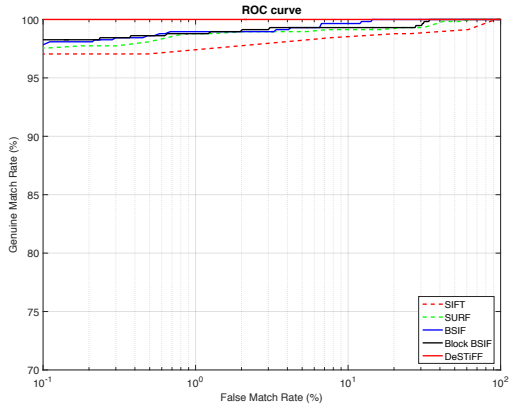
Algorithm	iPhone - Outdoor- Front		iPhone - Outdoor- Rear		Samsung - Outdoor- Front		Samsung - Outdoor- Rear	
	GMR (%) @ FMR=0.1%	EER (%)	GMR (%) @FMR=0.1%	EER (%)	GMR (%) @ FMR=0.1%	EER (%)	GMR (%) @ FMR=0.1%	EER (%)
SIFT	96.69	1.44	92.53	6.14	97.05	2.11	93.36	4.56
SURF	97.04	1.62	95.14	4.27	97.57	1.23	95.10	3.73
BSIF	85.24	5.48	93.58	4.79	97.92	1.05	72.57	7.60
Block-BSIF	88.19	3.41	94.79	4.18	98.26	1.21	84.55	4.38
DeSTiFF	**100.00**	**0.09**	**100.00**	**0.00**	**100.00**	**0.00**	**100.00**	**0.00**

Table 5.8: Verification performance of various algorithms on the data captured in outdoor illumination in MICHE-I periocular dataset

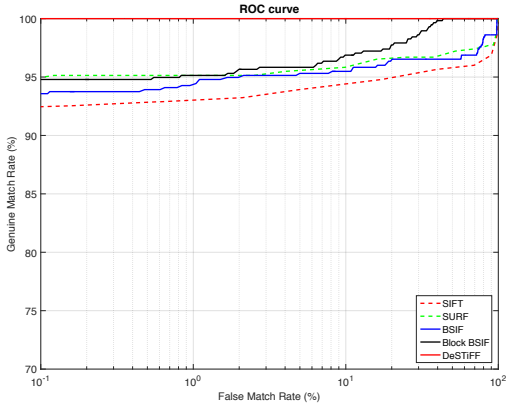
Algorithm	iPhone - Outdoor- Front		iPhone - Outdoor- Rear		Samsung - Outdoor- Front		Samsung - Outdoor- Rear	
	GMR (%) @ FMR=0.1%	EER (%)	GMR (%) @ FMR=0.1%	EER (%)	GMR (%) @ FMR=0.1%	EER (%)	GMR (%) @ FMR=0.1%	EER (%)
SIFT	97.74	1.62	91.08	4.26	96.53	1.85	96.53	2.22
SURF	97.92	1.01	91.43	2.89	95.31	2.46	97.22	2.38
BSIF	97.22	1.04	69.62	5.02	94.44	2.43	94.27	3.61
Block-BSIF	97.05	1.04	82.99	3.54	96.01	1.93	93.06	2.07
DeSTiFF	**99.48**	**0.09**	**100.00**	**0.02**	**100.00**	**0.00**	**99.48**	**0.11**



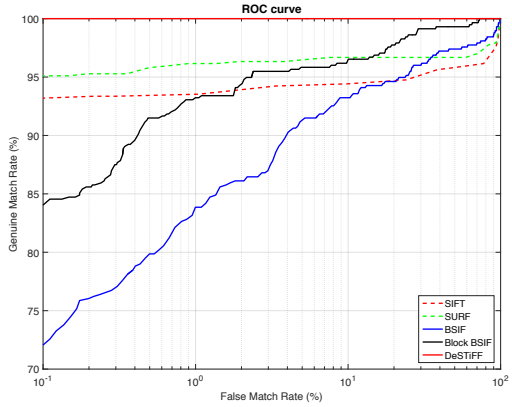
(a) iPhone 5S - Data from front camera.



(b) Samsung Galaxy S5 - Data from front camera.



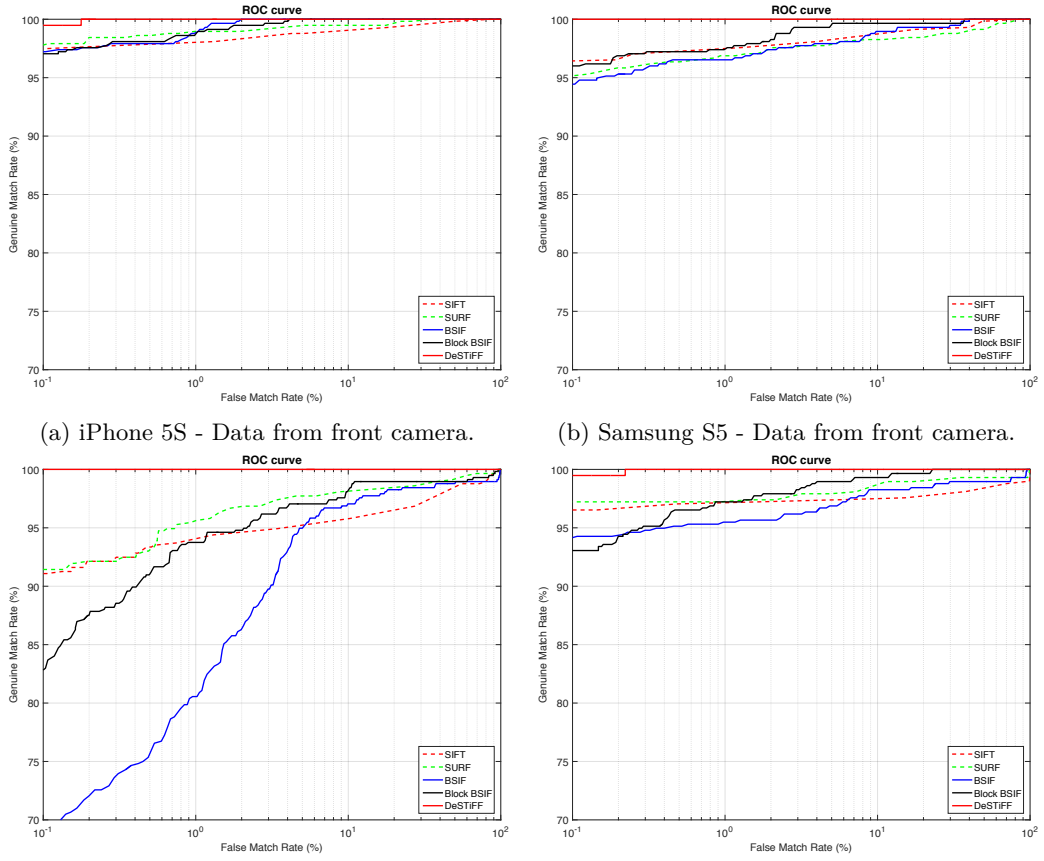
(c) iPhone 5S - Data from rear camera.



(d) Samsung Galaxy S5 - Data from rear camera.

Figure 5.7: ROC curves for various algorithms on the data captured in indoor illumination in MICHE-I Periocular Dataset

Figure 5.7 and Figure 5.8 present the ROC of various algorithms on the data captured in indoor and outdoor illumination respectively. The near accurate results obtained using $DeSTiFF$ presents a significant improvement over state-of-art methods and it can be clearly observed from the ROC presented. The EER obtained on all the subsets are close to 0% and the GMR at $FMR = 0.01\%$ equals 100% in majority of the subset data signifying the robustness of the newly presented approach.



(a) iPhone 5S - Data from front camera.

(b) Samsung S5 - Data from front camera.

(c) iPhone 5S - Data from rear camera.

(d) Samsung S5 - Data from rear camera.

Figure 5.8: ROC curves for various algorithms on the data captured in outdoor illumination in MICHE-I Periocular Dataset

5.5.3 Experiments on VISOB Periocular Database

VISOB dataset presents the images captured using three different phones under three different lighting conditions, which correspond to regular daylight illumination, dim light illumination and mixed illumination (working/office space illumination). Further, the data is captured such that the enrolment and probe correspond to disjoint set of images for each periocular characteristic captured from a subject. The number of samples for each individual vary across the subsets of the dataset. As discussed in the Section 5.4, VISOB dataset presents the largest set of periocular data captured from the smartphones. In this section, we present the experimental results obtained using different techniques on the VISOB dataset. For the sake of simplicity in presenting the results, we present the verification performance corresponding to a particular illumination condition across all different phones in one section. Further, as we treat one periocular characteristics as an unique instance for a subject, we present the results of left and right periocular characteristics independently. The set of results obtained using the *Deep Sparse Filtering* on VISOB dataset has been listed in the Appendix C for a reference comparison of the performance based on *DeSTiFF*.

5.5.3.1 Protocols for VISOB Dataset

The VISOB dataset consists of disjoint set of enrolment and probe samples which was collected in two different sessions. Thus, in our current work, we use the data stemming from first session (corresponding to enrolment) for creation of enrolment database and similarly,

the data collected from second session was used as probe samples. The complete partition of the dataset is provided in the Table 5.9. As the number of samples captured for each individual varies, we have presented the total number of images in each set of reference and probe.

Table 5.9: Database division for experiments on VISOB dataset

Smartphone	Illumination	Left		Right		Total	
		Enrolment	Probe	Enrolment	Probe	Enrolment	Probe
iPhone	Day-Light	2622	2536	2648	2567	5270	5103
	Dim-Light	1865	1763	1897	1789	3762	3552
	Office-Light	2522	2561	2523	2292	5045	4553
Samsung	Day-Light	1582	1587	1648	1678	3230	3265
	Dim-Light	2074	2007	2220	2175	4294	4182
	Office-Light	2225	2336	2418	2456	4673	4793
Oppo	Day-Light	1963	1985	1963	1985	3926	3970
	Dim-Light	3749	3742	3748	3740	7497	7482
	Office-Light	5284	4962	5269	4935	10553	9897

5.5.3.2 Preprocessing of VISOB Dataset

As the data is highly challenging and presents number of degradation factors such as low illumination, in this thesis, we pre-process the images to enhance the quality of images. We have employed block-based adaptive histogram equalization (CLAHE)[104] with a clip limit of 4 in a window of size 8 pixels [18] to improve the visibility of texture for the set of experiments corresponding to $SIFT$ and $SURF$ to detect the key-points in a robust manner.

5.5.3.3 Experiments on data from day light illumination

Table 5.10 presents the results of the verification for images captured using iPhone, Samsung and Oppo in the daylight illumination. It can be observed that the state-of-art techniques employing SIFT, SURF and BSIF provide lower verification accuracy as compared to $DeSTiFF$. The drop in the performance is drastic as compared to the performance obtained on MICHE-I and ViSPer database. This drop can be attributed to large size and unconstrained nature of images in the database. Nonetheless, the $DeSTiFF$ features provide a reliable performance across all the data captured from different smartphones. Figure 5.9 presents the ROC depicting the verification performance for data captured using all three different smartphones. Higher verification performance of $DeSTiFF$ features can be clearly observed by the significant difference over state-of-art techniques.

From Table 5.10, the following key observations need to be noted:

- GMR of 97.66% and 97.85% is obtained on the data captured (left and right periocular images respectively) in daylight illumination using iPhone from proposed $DeSTiFF$ features while best performance from state-of-art results in average of 87% from SIFT features.
- GMR of 96.50% and 97.85% is obtained on the data captured from Samsung with proposed $DeSTiFF$ features while best performance from state-of-art results in $GMR = 75.55\%$ from SIFT features for left periocular images and $GMR = 73.63\%$ from BSIF features for right periocular images .

- Oppo provides lowest performance among all the data captured in daylight illumination with GMR of 95.94% and 95.00% for left and right periocular images. The general decrease in state-af-art performance can be seen for Oppo as compared to same state-of-art technique applied on data captured from other smartphones while $DeSTiFF$ features still provide significantly higher performance.
- Lowest EER is achieved from $DeSTiFF$ features for all images from different smartphones.

5.5.3.4 Experiments on data from dim light illumination

Another set of data corresponds to dim light illumination, where the images are acquired in poor illumination condition. Table 5.11 presents the results corresponding to the captured images in dim light illumination. It can be noticed by comparing the Table 5.10 and Table 5.11 that the verification performance further decreases for state-of-art techniques for the dim light images as compared to daylight illuminated images. The performance obtained using $DeSTiFF$ features have performed well under such poor illumination conditions. Figure 5.9 presents the ROC depicting the verification performance for data captured using all three different smartphones. Higher verification performance of $DeSTiFF$ features can be observed which is significantly higher than state-of-art techniques as depicted in Figure 5.10.

From Table 5.11, the following key observations need to be noted:

- GMR of 97.15% and 96.90% is obtained on the left and right periocular data respectively in dim light illumination using iPhone from proposed $DeSTiFF$ features while best performance from state-of-art results in GMR of 75.13% and 72.85% from SIFT features.
- GMR of 96.50% and 97.85% is obtained on the data captured from Samsung with proposed $DeSTiFF$ features while best performance from state-of-art results in $GMR = 64.40\%$ from bock based BSIF features for left periocular images and $GMR = 66.45\%$ from BSIF features for right periocular images .
- Oppo provides a performance of $GMR = AA\%$ and $GMR = BB\%$ for left and right periocular images. The data captured from the Oppo results in lower GMR as against data from iPhone and Samsung.
- EER is observed to be lowest from $DeSTiFF$ features across the images from different smartphones.

5.5.3.5 Experiments on data from office light illumination

Apart from regular illumination conditions, this section describes the results obtained when the periocular images are captured under office illumination condition. The illumination under such a scenario can be from pure artificial illumination or from a mixture of artificial illumination and natural illumination from daylight (commonly entering from windows/transparent walls with glass). Table 5.12 presents the results obtained from different state-of-art techniques and $DeSTiFF$ features. Under the general comparison with the performance obtained from the data captured in day light, the verification accuracy has decreased in this illumination. The significance of the $DeSTiFF$ features is exemplified by the robust performance even under difficult illumination proving the applicability of $DeSTiFF$ features for many applications. The robust performance can further be observed from ROC presented in Figure 5.11 which depicts the clear margin of separation from state-of-art techniques.

Table 5.11 presents the key observations for data captured in office light illumination:

- GMR of 95.85% and 95.78% is obtained on the left and right periocular data respectively in office light illumination using iPhone from proposed $DeSTiFF$ features while best performance from state-of-art results in GMR of 75.13% and 72.85% from SIFT features.
- GMR of 96.50% and 97.85% is obtained on the data captured from Samsung with proposed $DeSTiFF$ features while best performance from state-of-art results in $GMR = 64.40\%$ from bock based BSIF features for left periocular images and $GMR = 66.45\%$ from BSIF features for right periocular images .
- Oppo provides lowest performance among all the data captured in daylight illumination with GMR of 89.24% and 90.45% for left and right periocular images. The data captured from the Oppo results in lower GMR as against data from iPhone and Samsung under the office illumination conditions.

Table 5.10: Verification performance of various algorithms on the data captured in daylight illumination in VISOB Periocular Dataset

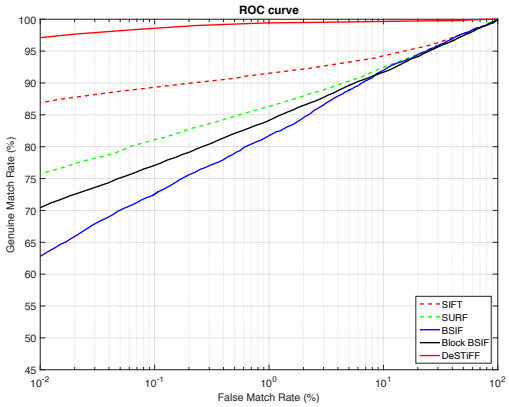
Algorithm	iPhone - Left		iPhone - Right		Samsung -Left		Samsung - Right		Oppo -Left		Oppo - Right	
	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)
SIFT	87.48	7.30	86.91	5.66	75.55	7.56	67.07	11.30	48.99	15.82	48.13	14.81
SURF	76.13	8.11	77.26	7.65	73.01	10.30	69.86	12.94	29.79	17.42	30.76	17.51
BSIF	62.01	9.48	60.59	9.56	62.88	8.62	65.24	7.98	51.50	22.02	49.74	21.37
Block-BSIF	70.50	8.77	70.81	8.76	70.46	8.43	73.63	7.09	58.90	12.13	58.00	13.54
DeSTiFF	**97.66**	**0.76**	**97.85**	**0.72**	**96.50**	**0.91**	**97.85**	**0.64**	**95.94**	**1.09**	**95.00**	**1.40**

Table 5.11: Verification performance of various algorithms on the data captured in dim light illumination in VISOB Periocular Dataset

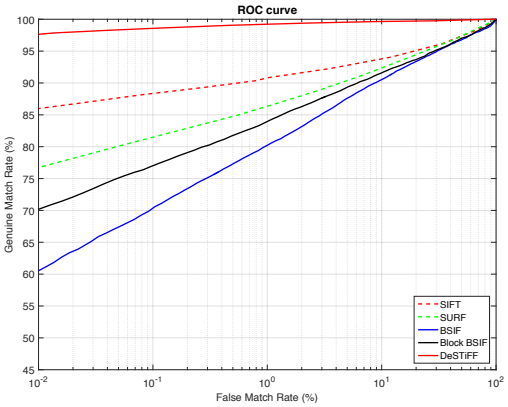
Algorithm	iPhone - Left		iPhone - Right		Samsung -Left		Samsung - Right		Oppo -Left		Oppo - Right	
	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)
SIFT	75.13	8.30	72.85	9.11	54.21	12.54	37.66	11.79	47.13	12.13	50.62	10.51
SURF	49.20	18.61	51.47	18.50	40.65	21.22	38.95	21.28	37.75	18.18	38.44	16.39
BSIF	64.36	8.87	64.78	8.82	60.84	9.24	65.57	6.74	57.64	16.86	56.18	17.06
Block-BSIF	64.16	7.07	65.60	7.37	64.40	6.89	66.45	6.71	67.65	7.72	64.74	8.40
DeSTiFF	**97.15**	**0.48**	**96.90**	**0.80**	**98.17**	**0.65**	**98.69**	**0.75**	**97.50**	**1.14**	**98.13**	**1.06**

Table 5.12: Verification performance of various algorithms on the data captured in mixed (office) illumination in VISOB Periocular Dataset

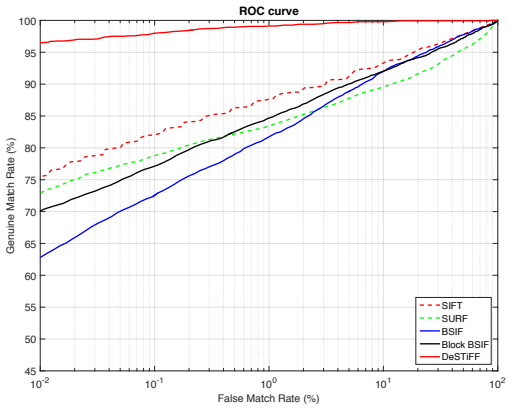
Algorithm	iPhone - Left		iPhone - Right		Samsung -Left		Samsung - Right		Oppo -Left		Oppo - Right	
	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)	GMR (%) @ FMR=0.01%	EER (%)
SIFT	67.60	10.16	68.78	10.53	25.77	19.32	26.65	20.34	45.61	16.22	45.55	18.37
SURF	50.69	16.55	55.76	16.11	31.43	21.82	32.13	21.30	34.14	20.75	33.10	23.34
BSIF	46.56	11.59	48.41	11.25	41.46	12.94	46.46	12.27	37.90	22.10	35.67	27.82
Block-BSIF	57.67	10.45	59.78	10.20	47.85	13.40	52.47	11.05	41.95	24.33	40.48	28.12
DeSTiFF	**95.85**	**1.41**	**95.78**	**1.45**	**93.53**	**2.42**	**95.17**	**1.95**	**89.24**	**3.58**	**90.45**	**4.51**



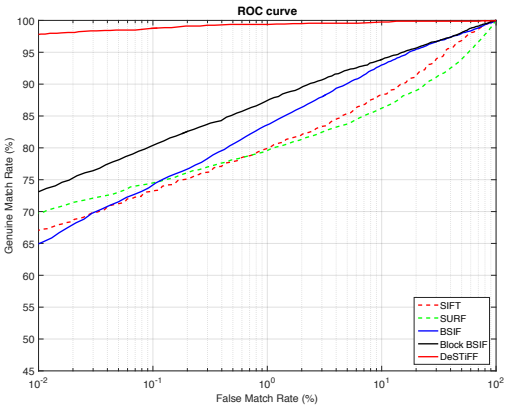
(a) iPhone - Left periocular data.



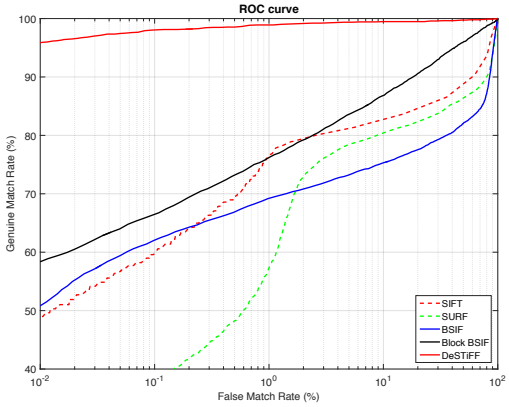
(b) iPhone - Right periocular data.



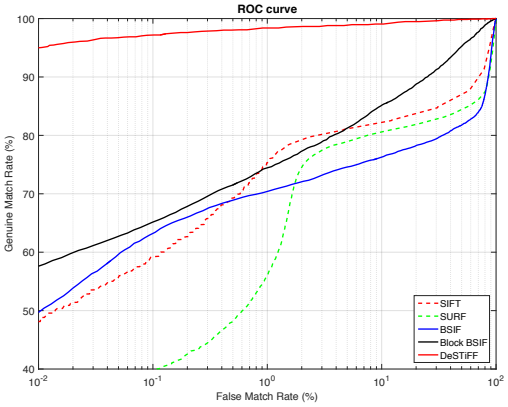
(c) Samsung - Left periocular data.



(d) Samsung - Right periocular data.

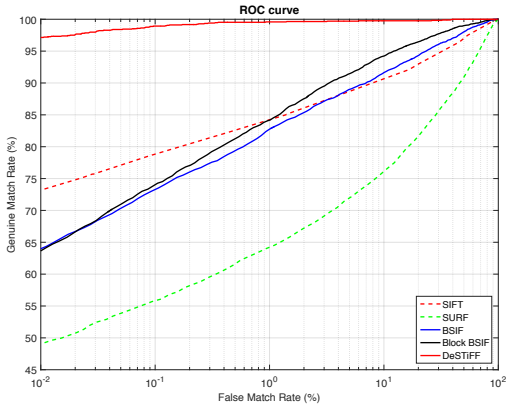


(e) Oppo - Left periocular data.

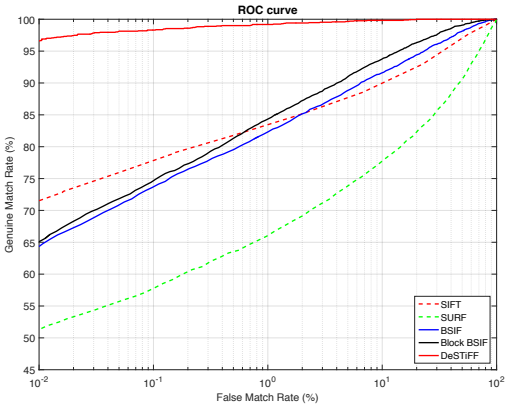


(f) Oppo - Right periocular data.

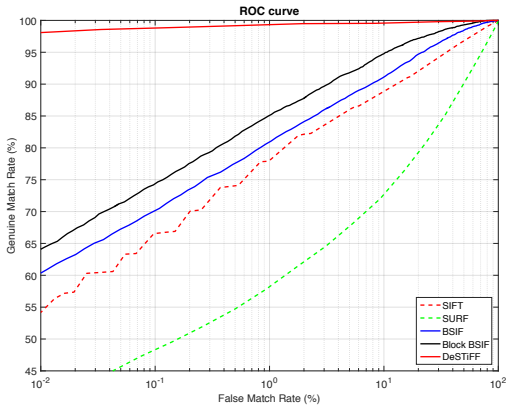
Figure 5.9: ROC curves for various algorithms on the data captured in daylight illumination in VISOB Periocular Dataset



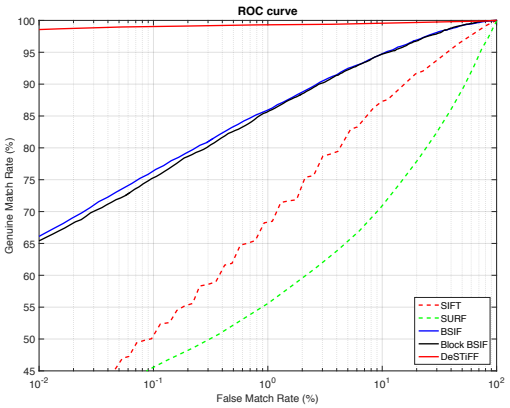
(a) iPhone - Left periocular data.



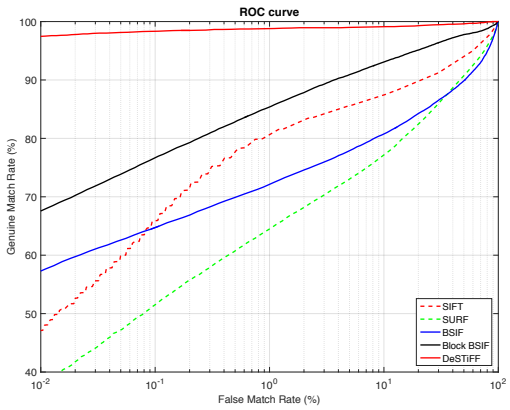
(b) iPhone - Right periocular data.



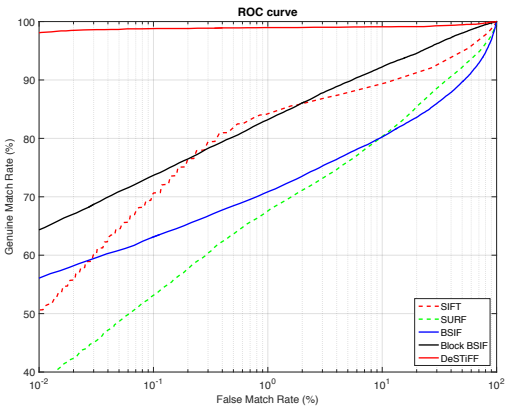
(c) Samsung - Left periocular data.



(d) Samsung - Right periocular data.

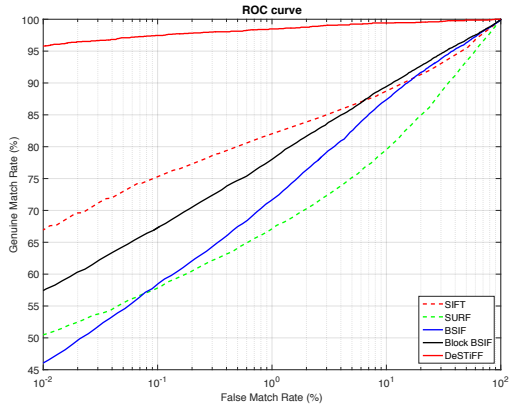


(e) Oppo - Left periocular data.

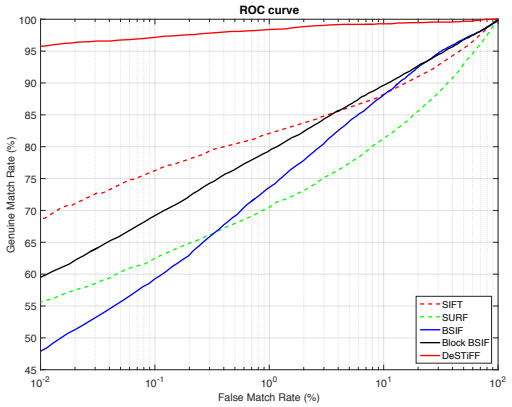


(f) Oppo - Right periocular data.

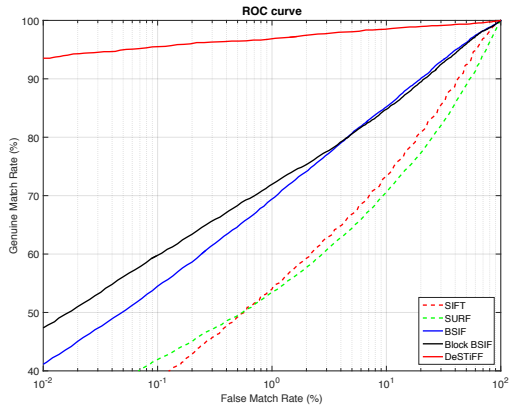
Figure 5.10: ROC curves for various algorithms on the data captured in dim light illumination in VISOB Periocular Dataset



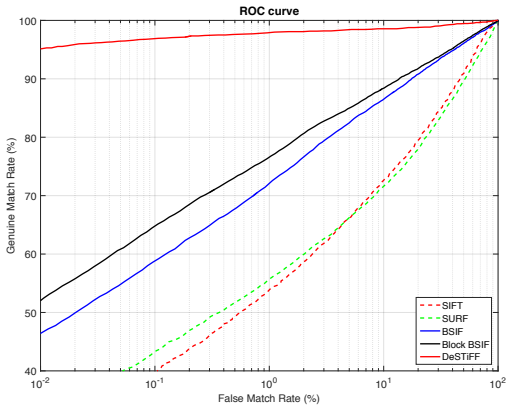
(a) iPhone - Left periocular data.



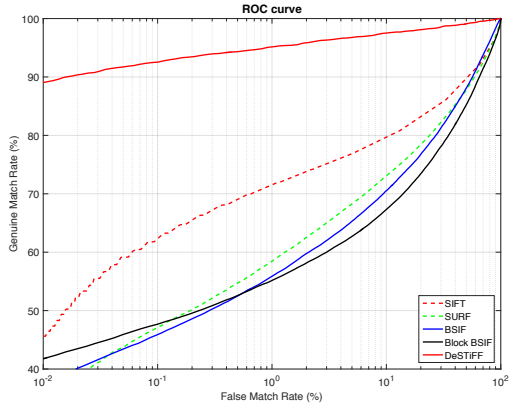
(b) iPhone - Right periocular data.



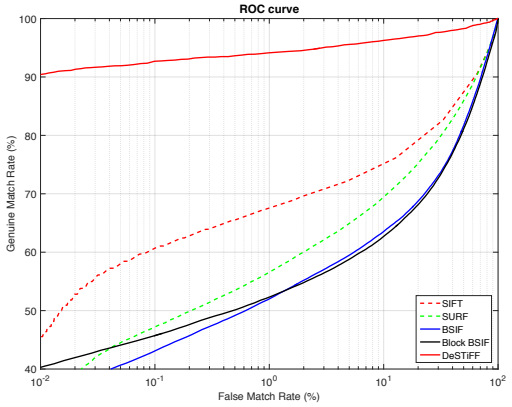
(c) Samsung - Left periocular data.



(d) Samsung - Right periocular data.



(e) Oppo - Left periocular data.



(f) Oppo - Left periocular data.

Figure 5.11: ROC curves for various algorithms on the data captured in office illumination in VISOB Periocular Dataset

5.5.4 Discussion

This section presents an analysis for the experimental results discussed in the previous section. As it can be deduced from the results presented in Tables (Table 5.5, Table 5.7, Table 5.8, Table 5.10, Table 5.11, Table 5.12), the verification accuracy decreases generally when the data is not ideal as in the case of adverse illuminations or poor illumination for state-of-art schemes. *DeSTiFF* features have performed consistently well on all different

databases, even when the data is heavily degraded as in the case of VISOB dataset.

In the case of MICHE-I, the images are captured in unconstrained manner and the same applies for the VISOB database. However, the key difference in VISOB and other datasets employed in this work, is that VISOB data is captured in highly challenging illumination. The challenging illumination condition impacts the verification accuracy and the trend can be validated by the observed low performance for VISOB dataset, when state-of-art algorithms are employed. Although, pre-processing techniques including block-based CLAHE are employed for VISOB database to enhance the periocular image, the preliminary observation suggests that the challenging illumination degrades the verification performance. A possible future work in this direction should investigate the impact of illumination, especially for the unconstrained periocular verification.

The significance of $DeSTiFF$ features stem from two aspects that include multi-representation of periocular image using set of *deep sparse filters* and the localization of time and frequency features. Coupled with such reliable features and collaborative representation, $DeSTiFF$ features present a promising direction for the path forward in periocular recognition for smartphone captured data.

5.6 Conclusion

The problems of iris recognition in visible spectrum has lead to explore alternative characteristics around the eye region. Periocular region, which includes the region around eye along-with eyelids, folds formed by eyelids and shape of eye has been well explored to supplement the lower performance of iris recognition in visible spectrum. In this chapter, we have explored periocular characteristics as a stand-alone characteristics for verification in visible spectrum.

Further, we have systematically demonstrated the applicability of periocular characteristics for the smartphone based verification scenario using the periocular data captured in various conditions. The experiments carried out on two public databases and our new periocular database (ViSPer) have indicated the applicability of state-of-art techniques in periocular recognition to smartphone data. However, it was demonstrated that the performance of state-of-art techniques degrades due to low quality and complex nature of data emerging from poor illumination and unconstrained nature of data capture. A new feature extraction based on *deep sparse filtering* and localization of time-frequency features are presented in this chapter in conjunction with collaborative representation. The newly presented features $DeSTiFF$ have consistently proven to perform well under challenging conditions on all publicly available databases. The feature representation has not only demonstrated the robustness on smaller dataset, but also on large scale data as in VISOB database.

A preliminary observation from the set of experiments on VISOB database indicates the necessity of eyebrows for good accuracy. However, a detailed study needs to confirm the preliminary observations signifying the importance of eyebrows for periocular recognition. Another important observation from this chapter is the need to study the influence of illumination. The images captured in poor illumination have resulted in lower verification performance as compared to images captured in well illuminated images. A future work in this direction needs to investigate on pre-processing algorithms such that the current state-of-art algorithms can be well exploited to suit the state-of-art periocular verification systems.

Chapter 6

Presentation Attack Detection for Ocular Biometrics on Smartphones

Biometric systems are vulnerable towards the presentation attacks at capture level. In this chapter, three new mechanisms of presentation attack detection are presented. The techniques are validated using the set of experiments on the publicly available ocular databases designed to address presentation attacks.

6.1 Introduction

International Standardisation at ISO/IEC defines Presentation Attack as "Presentation of an artefact or human characteristic to the biometric capture subsystem in a fashion that could interfere with the intended policy of the biometric system" [52]. Generalizing from the ISO definition of presentation attacks, the biometric systems operating via smartphones as data capture unit can be considered vulnerable to the presentation attacks at capture level for various biometric characteristics [103, 131, 68, 67]. The biometric systems are presented with the artefacts to gain access to secure system projecting it as a normal or bona fide presentation. Without *apriori* knowledge, the biometric system compares it against the enrolled samples and authenticates the user based on the obtained comparison score. If the artefact of the genuine enrollee in the database is presented to the sensor, the probability of getting access into the system is very high as the biometric sample enrolled closely resembles the artefact sample.

It is essential to have robust attack detection mechanisms in a biometric system. The threat of presentation attack should not be compromised for the sake of providing usability of biometric system with ease on smartphones. The vulnerability of such biometric systems working with the data captured in the smartphones are well illustrated in the earlier works [103, 131, 68, 67]. The key aspect for such attack stems from the fact that biometric data is captured in an unsupervised manner when any individual uses smartphone for secure access authentications such as banking applications. As the face data can be easily obtained by any impostor from different sources such as social media [26], it can be used to attack the smartphones operating as biometric capture device. Inherently, the face region presents the ocular characteristics which can again be used to attack ocular biometric systems. Alternatively, the videos and images from the social media can be replayed on the electronic display units to attack the biometric capture devices such as smartphones. In extreme cases, when the biometric systems employ centralized storage to maintain the biometric data, the impostors can gain access to biometric samples when the storage unit is compromised. Thus, the original enrolment data can itself be used to attack the capture devices of biometric data [144, 115, 110, 5, 6, 87, 164].

The presentation attacks span across many different modes which range from low cost printed artefacts to near-original quality replay attacks through high quality electronic display screens. On generalizing different ways to carry presentation attacks for smartphone based biometric systems, we can classify them as print-attack and electronic-screen-attacks [26, 40, 41, 42]. Another form of presentation attacks can emerge from subjects wearing textured contact lens where the texture corresponds to actual iris pattern under normal

presentation of genuine subjects[114, 159]. The high level classification of commonly encountered presentation attacks in the hierarchy is listed as below:

- Print attacks
  - Low-quality printed attacks
  - High-quality printed attacks
- Electronic screen attacks
  - Still image attacks
  - Video replay attacks

Any biometric system can be attacked with any or all different attack modes mentioned above. The ocular biometric systems, operating in visible spectrum via smartphones/regular-cameras have been challenged using simple low cost printed artefacts [68, 131, 67]. MobILive 2014 (Mobile Iris Liveness Detection Competition) [131] was specifically designed to address presentation attacks using printed images. Smartphone based biometric systems can also operate by capturing the video of biometric characteristics. Many biometric systems which employ video based authentication have been robustly built to detect the liveness of the subject for print attacks by analyzing the motion [5, 6, 87, 164]. To attack the video based biometric systems, the artefact video can be replayed using any electronic screen enabled devices such as tablets, smartphones, display monitors [12, 68, 67, 70].

The problem of presentation attack has been well addressed in Near-Infra-Red and regular (RGB) iris imaging devices operating in visible spectrum for the print attacks [99, 42, 30, 45, 41, 81, 49, 160, 27, 68, 67]. The print attacks are successfully detected by quantifying the quality artefacts of image and ocular features of the presented image [99, 42, 138].

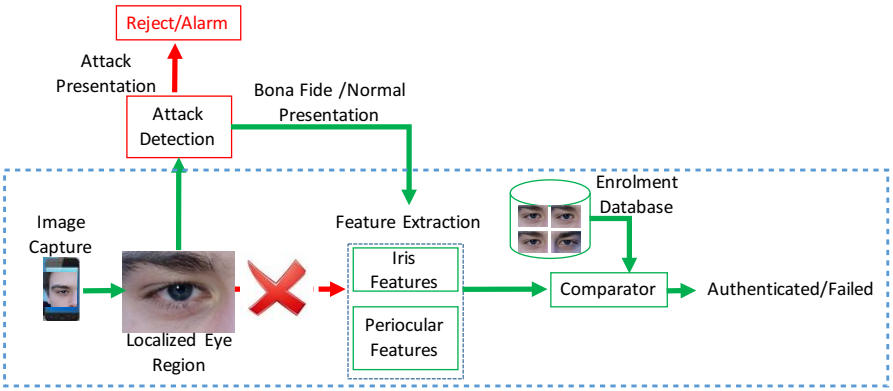


Figure 6.1: General architecture of a ocular/iris biometric system with incorporated PAD module.
*Note - The block enclosed in the dashed blue line indicates a conventional verification scheme.

Figure 6.1 illustrates the general architecture of ocular/iris biometric system with integrated presentation attack detection module. Typical ocular(*aka periocular*) and/or iris biometric system involves all of the components enclosed in dashed-blue line where an image is captured, ocular region is localized following the feature extraction from iris and/or periocular region and compared against the samples present in the enrolment database to verify or reject the identity. The systems can be attacked by presenting artefact samples which can compromise the security level of the system. Thus, it is essential to introduce a

Presentation Attack Detection (PAD) module, which verifies the liveness of presented ocular and/or iris characteristics.

In this chapter, presentation attacks and presentation attack detection modes for ocular (periocular) biometrics on smartphones are discussed from two perspectives for the data captured from smartphones operating in visible spectrum. In the first place, we demonstrate the vulnerability of the ocular systems towards presentation attacks. With the motivation of addressing the presentation attacks for ocular biometric systems operating in visible spectrum, we explore various techniques based on motion and textural features for image and video presentation attack detection in ocular biometric system.

In the rest of the chapter, we discuss the available databases in Section 6.2 for ocular biometrics on smartphones designed specifically for presentation attack detection. In the Section 6.3, we present the vulnerability of the biometric systems towards the possible attacks of known types. The obtained results with various presentation attack detection techniques are given in the Section 6.4. Finally, in Section 6.5, we discuss major summary of presentation attack detection techniques for ocular biometrics on smartphones along with the notable remarks.

6.1.1 Contributions

The key contributions from this chapter can be summarized as:

1. A detailed analysis of video based presentation attack detection for iris/ocular recognition on smartphones in the visible spectrum with experiments on relatively large video iris database acquired using two different smartphones in the visible spectrum. The database consists of 152 unique iris patterns obtained from 76 subjects which was collected during the course of this thesis and is distributed freely for research work.

2. Presents two algorithms based on micro-texture features and one motion based approach for detecting video presentation attacks on smartphone systems.

3. The code for the proposed approaches of presentation attack detection based on texture features are distributed freely for non-profitable research purpose to promote reproducible research.

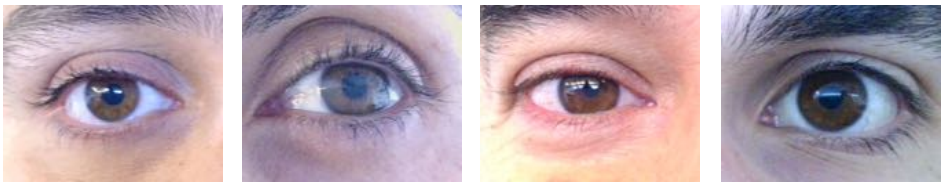
6.2 Presentation Attack Databases

This section provides a summary of the available ocular databases captured using smartphones specifically designed for presentation attacks captured using smartphones. The two publicly available ocular databases are presented, which provide two different attack artefacts - image based attacks (either electronic or printed attacks) and video based attacks to surpass the biometric system. Whereas MobILive 2014 Database concentrates on the print attack, Presentation Attack Video Iris Database (PAVID) collected during the course of this thesis focuses on video presentation attacks in video based authentication scenarios, specifically using an electronic screen.

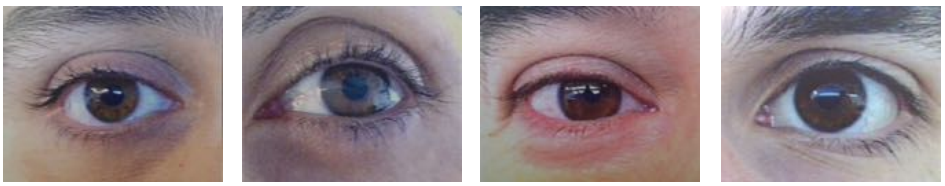
6.2.1 MobILive 2014 Database

The MobILive 2014 dataset was released in conjunction with the 1^{st} Mobile Iris Liveness Detection Competition (MobILive) organized in the context of IJCB2014 [131, 130]. MobILive 2014 dataset consists of images from 100 volunteers collected using *Asus EeePad Transformer* tablet. Four ocular samples were used to capture each eye corresponding to each subject resulting in 800 live ocular images in the dataset.

The artefact dataset of MobILive 2014 Database, also referred as MobBIOfake database [131] consists of artefact images collected from 100 subjects. Each of the live ocular image is printed using a color printer and reacquired using the *Asus EeePad Transformer* tablet.



(a) Live samples



(b) Artefact/Spoof samples

Figure 6.2: Sample images from MobILive 2014 dataset; (a) Sample images from normal/live(bona fide presentation (b) Artefact or attack images corresponding to normal images in (a)

Table 6.1: Composition of MobILive 2014 Dataset

	Bona Fide Presentations	Attack Presentations
Number of subjects	100	100
Unique eye instances	200	200
Samples per eye	4	4
Total images	800	800

Thus, the dataset comprises of 1600 samples of which 800 are live samples and rest of 800 are artefact samples. The artefact samples were obtained from printed images of the original ones captured with the same handheld device and in similar conditions. Table 6.1 presents the composition of the MobiLIve2014 dataset. Figure 6.2 presents the sample images from the MobILive 2014 Database. The Figure 6.2(a) presents the images corresponding to normal images while the Figure 6.2 (b) presents the artefact samples corresponding to the images in Figure 6.2(a).

6.2.2 Presentation Attack Video Iris Database (PAVID)

Another database with smartphone ocular images designed to address presentation attack detection is the Presentation Attack Video Iris Database (PAVID) collected during this thesis [68, 67]. This database is constructed to explore video based presentation attacks on the ocular biometric systems in smartphones operating in visible spectrum [68, 67]. The PAVID database consists of 2 parts which correspond to bona fide and attack presentation videos of the iris/ocular characteristic. The live or bona fide presentation videos are the recordings of the ocular characteristics and the artefact videos are the recordings of video replays. Unlike most of the other databases in ocular biometrics, PAVID database is specifically tailored to address video based attacks.

Video based ocular biometric systems can be considered superior to image based system as there are per transaction attempt number of frames available. With higher number of frames available, the decision module can employ decisions on the entire video, individual frames or aggregated frames using independent classifier or ensemble methods. Thus, in this chapter, we discuss on two independent approaches of detecting presentation attacks for the PAVID database:

1. Motion based aggregate decision module for a number of 25 frames from the video.
2. Frame based independent decision along with the majority voting for a number of 25 frames from the video.

In both cases, we employ the PAVID database with two independent and disjoint subsets of real and artefact dataset as discussed in the upcoming section.

6.2.2.1 PAVID - Real Iris Video Database

PAVID is a relatively large-scale iris video database acquired using two smartphones operating in visible spectrum. The significance of the PAVID database lies in three folds:

- Provides video iris database collected using smartphones in visible spectrum.
- Provides a database with relatively higher number of subjects.
  76 *subjects* & 152 *unique ocular instances*.
- Provides presentation attack database to assess the robustness of the PAD techniques.

Table 6.2: Distribution of bona fide iris video database in PAVID

	Smartphone	
	Nokia Lumia 1020	iPhone 5S
Number of subjects	76	76
Unique eye instances	152	152
Enrolment video	152	152
Probe Video	152	152

PAVID database [67, 68] consists of videos captured from 152 unique eye instances from 76 subjects using two new smartphones - Nokia Lumia 1020 and iPhone 5S. The total distribution of iris videos is presented in the Table 6.2. Each unique eye instance is captured in two different sessions that correspond to enrolment and verification. In each of the sessions, a video of duration $1-3$ seconds is acquired for each subject. The captured eye videos are processed such that at least 25 frames are obtained between the blinks. The video frames further are used to localize the eye region using the Haar cascade based eye detector [150]. This is essential to remove the background captured due to the larger field of view of the camera. The located boundary of eye region from the first frame is propagated across other frames in the video to localize the eye region. Finally, these videos for each subject are used as the enrolment videos. In a similar manner, the ocular/iris video is obtained for a verification session. Figure 6.3 presents the sample frames of the normal access videos captured using the iPhone 5S and Nokia Lumia 1020 along with corresponding artefacts for each phone.

6.2.2.2 PAVID - Artefact Iris Video Database

The database also consists of an artefact subset which has videos of iris captured using two smartphones. However, the videos are obtained from the electronic screen of devices which are capable of playing high quality/near-original quality videos. To simulate realistic attacks in a verification scenario,the artefact database is constructed under the assumption that the video from the enrolment database is available to the impostor. In this scenario, the impostor can use the enrolment video to generate the printed iris artefacts or replay the iris video. However, when the biometric system employs video based authentication, one has to design the attacks using video based approaches. Going by such an argument, the artefact database is created by replaying the iris video on the high-quality display enabled *iPad* and presenting it to smartphones (i.e., the biometric sensor in this case). The replay

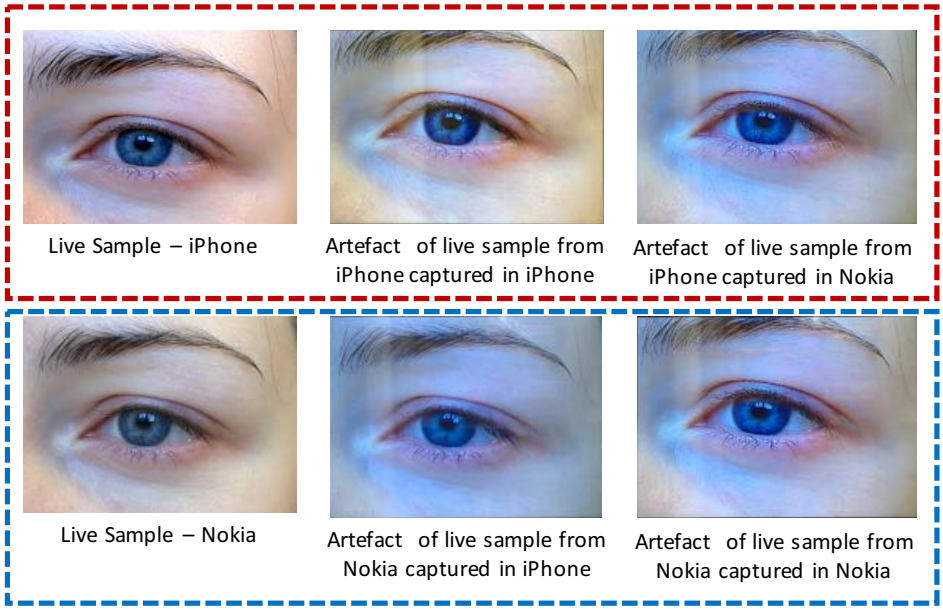


Figure 6.3: Sample frames from ocular videos from PAVID dataset; The top row indicates the live (bona fide) sample captured from iPhone along with the artefact samples (replayed using iPad) captured from iPhone and Nokia correspondingly. The bottom row indicates the live sample captured from Nokia along with the artefact samples (replayed using iPad) captured from iPhone and Nokia correspondingly.

attack database consists of 4 different attack subsets such that the enrolment videos obtained from iPhone replayed to iPhone and also to Nokia. Similarly, the enrolment videos obtained from Nokia are replayed to Nokia and iPhone as well. Under each replay attack subset, a total of 152 iris videos are present, which make a total of 608 artefact iris videos in total in the PAVID database. Table 6.3 provides an overview of the different subsets in the PAVID database.

Table 6.3: Composition of PAVID artefact database

Source obtained from	Capture smartphone / Sensor Attacked	Number of videos
iPhone	iPhone	152
	Nokia	152
Nokia	iPhone	152
	Nokia	152

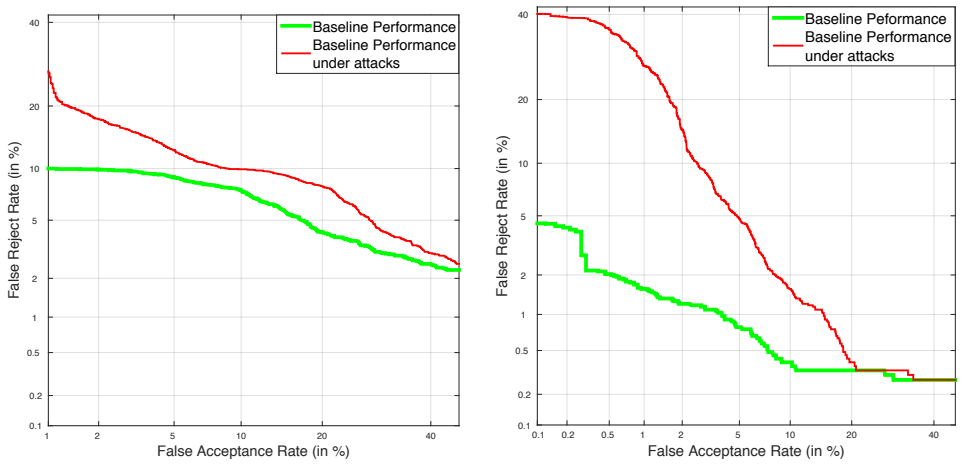
6.3 Vulnerability of Ocular Biometric Systems Towards Presentation Attacks

In order to make the ocular/iris biometric systems robust against presentation attacks, the first step is to assess the vulnerability of the system towards these attacks. The performance of a biometric system is measured using the False Accept Rate (FAR) versus the False Reject Rate (FRR) [54]. Compliant to a general system, the vulnerability towards presentation attacks can be evaluated by presenting artefact samples and measuring the FAR-FRR using

a specific baseline system. Vulnerability of the MobILive 2014 dataset is directly implied in the competition - 1^{st} Mobile Iris Liveness Detection Competition (MobILive) in the context of IJCB2014 [131, 130].

6.3.1 Vulnerability Analysis with PAVID

This section discusses the vulnerability biometric system operating with ocular characteristics using artefacts from PAVID. To present the simplistic operations, a sample frame from the probe video is compared against the reference frame in the enrolment database. The baseline performance with normal iris videos and the baseline performance of a system, when iris videos are used for attacking the sensor of the system is analysed in this section. Such an analysis is based on a trivial idea of measuring the genuine and impostor scores under normal presentation. Furthermore, it comprises of measuring the genuine and impostor scores when the artefacts are used to attack the sensor. The genuine score is obtained by comparing the reference frame from enrolment video with all the frames from probe video. These scores are used to obtain the Detection Error Trade-off (DET) curves of the system. To simplify the number of comparisons, we consider one frame from iris video of enrolment as the reference image and 25 frames from probe video as the probe samples. A similar approach is used to compute the genuine and impostor scores when the replay attack video is used. In this work, the baseline system is evaluated using Local Binary Pattern [96] feature extractor and Sparse Reconstruction Classifier [117, 68, 76] on the periocular characteristics.

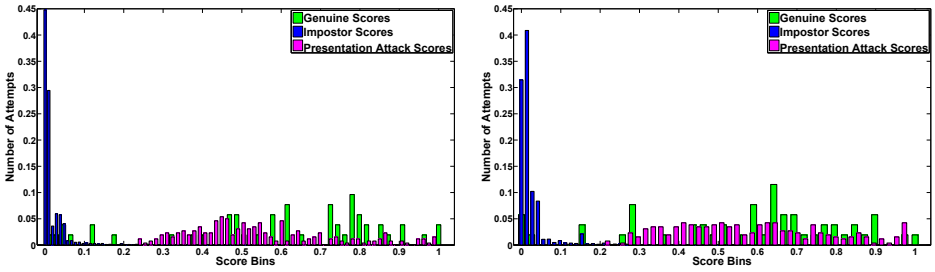


(a) DET Curve indicating the system performance for iPhone 5S

(b) DET Curve indicating the system performance for Nokia Lumia 1020

Figure 6.4: Baseline scores and performance of the systems employing iPhone 5S and Nokia Lumia 1020 as capture sensors

As indicated in the Figure 6.4, the baseline provides an Equal Error Rate (EER) close to 8% for iPhone 5S while the EER is around 2% for Nokia Lumia 1020 which are represented in green curves in Figure 6.4(a) and Figure 6.4(b). Further, the artefact samples are presented to the baseline system and the FAR-FRR is measured, which is indicated in the red curve in the Figure 6.4. As it can be noted from the Figure 6.4, the artefacts are accepted by the system to a greater extent, which results in EER of around 10% for iPhone 5S while it results at 5% EER for Nokia Lumia 1020. The EER obtained from the artefact presentation competes closely with the normal bona fide presentations and thereby indicates the vulnerability of the biometric system.



(a) (b)

Figure 6.5: (a) Score distributions for genuine, impostor and presentation attack iris videos for iPhone 5S (b) Score distributions for genuine, impostor and presentation attack iris videos for Nokia 1020

Alternatively, the vulnerability of the system can be demonstrated using the distribution of the genuine and impostor scores when both the live/bona fide and artefact biometric samples are presented to the biometric system. Figure 6.5 presents the genuine and impostor scores for iPhone 5S and Nokia Lumia 1020. It can be observed from the Figure 6.5 that there is a significant overlap of the genuine scores for bona fide presentation and artefact presentation. The overlap indicates high vulnerability of the ocular system towards video attacks and re-iterates the need for robust techniques to detect the presentation attacks.

6.4 Presentation Attack Detection Techniques

Presentation attack detection techniques can be broadly classified on the basis of image or video attacks. The image based biometric systems use single images for classifying the presentation as attack or bona fide presentation. In case of video based biometric system, the techniques can rely upon single frame or multiple frames within the video. In the upcoming section, the techniques based on multi-frame (video) strategy are discussed and subsequently single frame (image) based techniques are discussed.

6.4.1 Algorithm 1: Video Based Scheme for Presentation Attack Detection

Earlier studies on videos based presentation attacks for face biometrics have presented PAD modules, where the video is decomposed into several frames and the decision is made using individual frames [5, 6, 87, 164]. A recent work has devised a method to differentiate the motion characteristics under normal presentation of subject and motion characteristics under presentation attack using Eulerian Video Magnification [12].

In this section, we discuss an approach of presentation attack detection leveraging the frames from entire video using a similar approach. As the video based ocular biometric system can be attacked using electronic screens, it is intuitive to employ the frequency based approaches to detect the additional frequency of the electronic display and different motion characteristics. Specifically, the ocular video is decomposed using Fourier Transform to obtain the phase and magnitude component. The phase component of the ocular video is magnified to emphasize the variations in phase using modified EVM. The magnification amplifies the phase component in video which is further analysed to make the decision. The specific details involved in deciding the ocular video as presentation attack video or bona fide presentation is discussed in the upcoming sections.

6.4.1.1 Modified Eulerian Video Magnification for PAD

Eulerian Video Magnification (EVM) can magnify the temporal variations in videos by decomposing it spatially and applying temporal filter [157]. In the case of presentation attacks from an electronic screen, the replayed video presents different information that can be attributed to the display frequency of electronic screen. However, such information is very subtle and is hard to detect. If such information is magnified using linear magnification approaches in the spatial domain, the resulting videos contain larger amount of noises[151] . An alternative approach is to employ phase information to determine the presentation attacks. Phase based approaches provide robust performance, which are not sensitive to noises as in the case of amplitude based approaches [98]. The intuition in using the phase information is to detect the additional frequency emitted when the video is played from an electronic screen.

Inspired by the success of EVM in various cases to magnify the motion [157, 151], in this chapter, we have adopted a modified version of EVM to enhance the small variations in phase component of video frame. The modified version of EVM uses the phase information as input. The video frame is decomposed using Fourier transform to obtain the phase information which is further fed to the EVM. The decomposed phase information is spatially filtered using Laplacian pyramids and temporally filtered using the Butterworth lowpass filter to magnify the variations in phase of each frame in the video. The enhanced phase variation in the video is used to estimate the liveness of subject as normal presentation or attack presentation. The algorithm devised to detect the presentation attack is presented in detail in the section below.

6.4.1.2 Estimation of Liveness Score for PAD

Given the phase magnified video consisting of N number of frames obtained using modified EVM, we perform a series of operations to detect the presentation attack. Since, performing computations on each of the frame is relatively expensive in terms of memory and speed, each frame is downscaled to a smaller size. We have downscaled the frames to a size of 100×100 pixels based on the experimental trials on the development database. The magnified phase variation of each frame is normalized to have the values in the range of 0 *to* 1. Let F be the magnified phase variation of frame obtained from the EVM, then j^{th} normalized frame $NorF(j)$ is given by:

$$NorF(j) = \frac{F(j) - min(F(j))}{max(F(j)) - min(F(j))} \quad \text{where } j = 1 : N \tag{6.1}$$

The normalized frame is further divided into non-overlapping blocks of specific size, $bx \times by$ as shown in Figure 6.6. We have employed a block size of 20×20 in this chapter based on the experimental trials on the development database. This results in k number of blocks and thus, each frame results in $k = 25$ *blocks* in our work. The normalized phase information of the block is further referred as normalized block phase variation and is represented as $NorFB(j)_k$ corresponding to j^{th} frame.

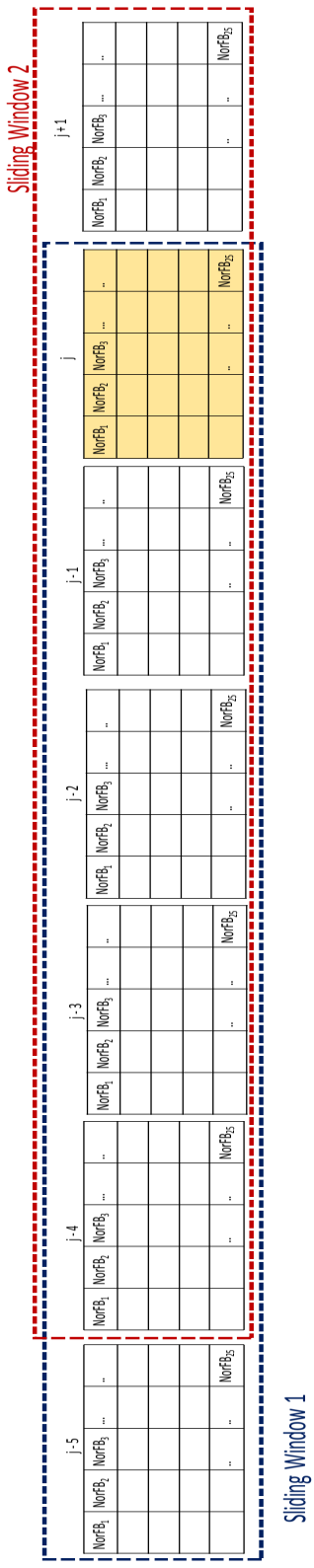


Figure 6.6: Schematic of the block based phase variation information

To effectively identify the presentation attack, a sliding window approach with 6 frames is employed, out of which 5 previous frames are used for making a decision on the present frame. The sliding window is propagated until 30 frames by incrementing one frame at a time. The sliding window is used to detect the rate of the change in the phase with respect to time. The size of the window was chosen based on the experimental trials conducted on development database. As illustrated in Figure 6.6, for any given current j^{th} normalized frame ($NorF(j)$), differential phase variation for a block k is computed using 5 previous frames $NorF(j-1)$ *to* $NorF(j-5)$. For a particular block k, the differential phase information with respect to 5 previous frames is given by:

$$\begin{aligned} DPI(j-5)_k &= NorFB(j)_k - NorFB(j-5)_k \\ DPI(j-4)_k &= NorFB(j)_k - NorFB(j-4)_k \\ DPI(j-3)_k &= NorFB(j)_k - NorFB(j-3)_k \\ DPI(j-2)_k &= NorFB(j)_k - NorFB(j-2)_k \\ DPI(j-1)_k &= NorFB(j)_k - NorFB(j-1)_k \\ &\text{for } k = 1, 2, \cdots 25 \end{aligned} \tag{6.2}$$

The final differential phase variation for a particular block in a frame j is obtained by determining the maximum of all the computed differences given by Equation 6.2.

$$\begin{aligned} DPI(j)_k = max\{DPI(j-5)_k, \cdots DPI(j-1)_k\} \\ \text{for } k = 1, 2, \cdots 25 \end{aligned} \tag{6.3}$$

The cumulative phase information, CPI is obtained for the entire frame j by summing all differential phase information across all the blocks b in the frame j.

$$CPI(j) = \sum_{x=1}^{k} DPI(j)_x \tag{6.4}$$

The cumulative phase information given by Equation 6.4 is computed in a similar manner for all the frames by employing the sliding window with 6 frames as described earlier.

In order to have the obtained score mapped to fixed interval values, the cumulative phase information is further normalized to a value between 0.5 *to* 1. The normalized cumulative phase information is used to determine the presentation attack. If the obtained value is above the threshold, the video is classified as a presentation attack. For a set of obtained CPI corresponding to a particular frame j, we apply single sided logistic or sigmoid function to obtain normalized CPI represented as $NCPI$:

$$NCPI(j) = \frac{1}{1 + e^{-CPI(j)}} \tag{6.5}$$

The normalized cumulative phase information ($NCPI$) is compared against a threshold value indicated by $Th = 0.7$ to obtain the liveness decision LD for a particular frame j. The threshold value of 0.7 is obtained from the empirical trials on development database.

$$LD(j) = \begin{cases} 1, & \text{if } NCPI(j) \leq Th \\ 0, & \text{otherwise} \end{cases} \tag{6.6}$$

The frames with $LD = 1$ are classified as bona fide presentation or live subjects and frames with other values are classified as the presentation attacks videos. The obtained liveness decision at various frames can be used to decide on the presentation category as bona fide presentation or attack presentation.

6.4.1.3 Experiments and Results on Video based PAD

The video based PAD scheme is evaluated on the PAVID dataset (31 subject subset) with iris/ocular videos captured in visible spectrum using smartphones. A subset of the database is employed to evaluate the video based PAD. The subset of the database consists of 62 unique live iris video and 62 unique attack iris videos from two different smartphones. Under the assumption that presentation attack videos can be obtained from various sources, the attack videos are recaptured from live iris videos obtained from iPhone 5S and Nokia Lumia 1020. The bona fide/live videos are replayed on the high quality display device to attack the biometric system employing either iPhone 5S or Nokia Lumia 1020. The attacks thus presents a scenario where live video is obtained from a particular smartphone, say iPhone 5S, and the video from the same smartphone is used to attack the system by presenting it. An alternative case is where the live iris video and spoof iris video correspond to videos originating from different smartphones. The second case gains importance in the light of the fact that people tend to change the smartphones quite often due to availability of better features, reduced cost and limited shelf life. In order to assess the robustness of proposed system, we consider both situations and propose two different protocols to evaluate the proposed PAD scheme.

Protocol 1

In the protocol 1, the biometric system is attacked by the videos originating from same smartphone. In accordance to this protocol, the biometric system employing iPhone 5S is challenged by the attack videos by replaying the videos from iPhone 5S. In terms of similar arguments, the system employing Nokia Lumia 1020 are attacked using the replay videos of Nokia Lumia 1020. This protocol takes care of the attacks based on the same sensors. Since the live video and presentation attack video originating from the same sensor are highly identical in terms of quality, this protocol intends to gauge the robustness of the technique in identifying highly identical artefacts.

Protocol 2

Under the assumption that the impostor can use an iris video obtained using different smartphone to attack the smartphone based visible spectrum ocular/iris biometric system, we propose to evaluate a situation where the attack videos and reference videos originate from different smartphones. Thus, in a system employing iPhone 5S as the biometric sensor, the attack video corresponding to Nokia Lumia 1020 is replayed and vice-versa. This protocol evaluates the reliability of the proposed technique to address the cross sensor presentation attacks in smartphone based visible spectrum iris recognition.

6.4.1.4 Results on video based PAD

All the results related to the video based presentation attack detection have been presented in terms of Attack Presentation Classification Error Rate (APCER) and Bonafide Presentation Classification Error Rate (BPCER) [53] as discussed in the Chapter 2. The results are also presented in terms of Average-Classification-Error-Rate (ACER) which is described as the average of APCER and BPCER. ACER is defined by the Equation 6.7 as:

$$ACER = \frac{APCER + BPCER}{2} \tag{6.7}$$

Based on the empirical trials on development database, a fixed threshold of $Th = 0.7$ is used to obtain the liveness decision. The technique is able to achieve an ACER of 0%. The obtained threshold on development database is well suited for the decision after frame number 11. The same threshold when applied on the testing database, we obtain the

$ACER = 0\%$ indicating general applicability of proposed scheme for presentation attack detection on video based smartphone iris/ocular recognition in visible spectrum.

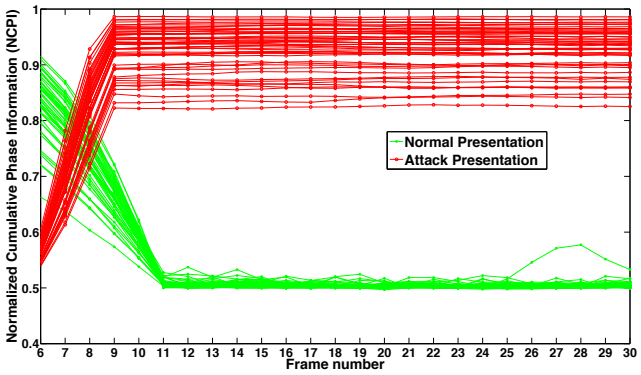


Figure 6.7: Liveness score for iPhone data.

The Table 6.4 presents various ACER obtained on testing database when different frames starting from 6 to 11 are considered with a threshold of $Th = 0.7$. The results are indicated from frame number 6 as the frames 1 *to* 5 are used to make the first decision at frame 6. From the obtained results, the best possible and reliable frame for making a decision is frame number 11 which provides ACER of 0% for all cases. The determined liveness score is observed to be constant after 11^{th} frame from the experiments. It can be observed from Figure 6.7 that the liveness score obtained using the proposed scheme is robust after frame number 11 with $Th = 0.7$ for reference iris videos captured from iPhone 5S and presentation attack videos are obtained by replaying videos from iPhone 5S on iPad.

Table 6.4: Presentation classification error rates with a Normalized Cumulative Phase Information (NCPI) threshold of 0.7.
*Note: Frame number 1 to 5 are used to make the decision on frame number 6 in the proposed approach.

Reference Video	Attack Video	Classification Error Rate (%)																	
		Frame 6			Frame 7			Frame 8			Frame 9			Frame 10			Frame 11		
		APCER	BPCER	ACER	APCER	BPCER	ACER	APCER	BPCER	ACER	APCER	BPCER	ACER	APCER	BPCER	ACER	APCER	BPCER	ACER
Nokia	Nokia	100.00	100.00	100.00	100.00	98.07	99.04	100.00	75.00	87.50	50.00	19.23	34.62	3.85	0.00	1.92	0.00	0.00	0.00
	iPhone	100.00	100.00	100.00	100.00	96.23	98.12	100.00	73.60	86.80	51.92	18.48	35.20	0.00	0.00	0.00	0.00	0.00	0.00
iPhone	iPhone	100.00	98.07	99.04	100.00	92.31	96.15	100.00	78.85	89.42	55.76	13.46	34.61	0.00	0.00	0.00	0.00	0.00	0.00
	Nokia	100.00	97.40	98.70	100.00	89.70	94.85	100.00	73.62	86.81	51.92	17.30	34.61	1.92	0.00	0.96	0.00	0.00	0.00

6.4.2 Frame/Image based Presentation Attack Detection

Another class of presentation attacks are based on the images or single frames of video. The attacks are mainly designed to target the biometric systems by presenting single image for the entire duration of probe data acquisition. Thus, it is essential to devise robust attack detection techniques to handle artefacts presented in the form of image. In this chapter, we present two PAD mechanisms, a.k.a, spoof attack detection algorithms to handle such image based attacks via electronic screens.

6.4.2.1 Algorithm 2: Laplacian Pyramid Decomposed Frequency Response Features

In this section, we present an approach to address the presentation attacks for ocular biometrics on smartphones that leverages on the features from space and frequency domain in Laplacian space [67]. The Laplacian Pyramid Frequency Response (LPFR) feature [67] used for PAD algorithm is depicted in the Figure 6.8. Given the image/video of the subject under verification scenario, we first decompose each image/frames into Laplacian pyramids of multiple scales. Each of the resulting images at the specific scale is used to obtain a Short Term Fourier Transform (STFT) response at four different orientations. The response corresponding to 4 different orientation is encoded as a single response image, and the features are obtained by taking the histogram as described in this section. The features are used to classify the presentation category as bona fide or attack presentation using a SVM classifier. The motivation and intuition behind the proposed technique are thoroughly discussed in this section.

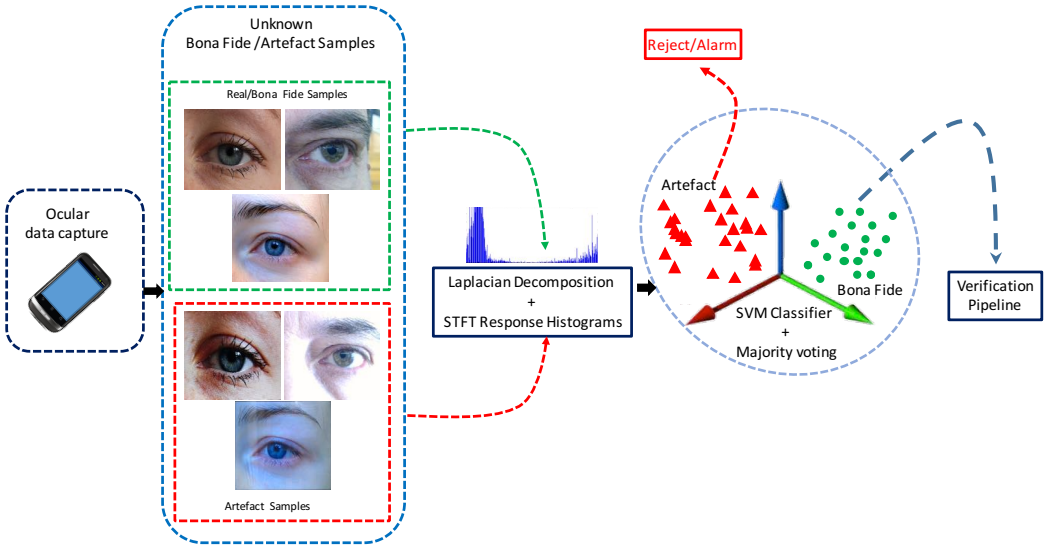


Figure 6.8: Presentation attack detection based on Laplacian Pyramid Frequency Response

Laplacian Pyramids

Laplacian pyramid decomposition of the image was initially developed with the idea of encoding the image using local operators at many scales with identical basis functions [21]. The significance of the Laplacian pyramid decomposition comes from the fact that the elements of an image are localized both in space and frequency domain. Further, the Laplacian pyramid can be used effectively to represent images as a series of band-pass filtered images that are sampled successively at sparser representations [21]. Although, the

frequency content of the image is well localized using Laplacian Pyramids, the orientation information of each frequency content is not obtained.

Given an image I, a set of Gaussian pyramid images can be represented using $\{G_l\}$ for l levels and the levels correspond to lower resolution with each level in the range 1 to l. With the progression of each layer, the high frequency edge details are removed and low frequency components are retained. The lowest component of Gaussian pyramid G_0 corresponds to the original image I and the G_k is the downsampled version of G_{k-1} which is twice the width and height of G_k. Similarly, Laplacian pyramid represents the images at different levels l with different scale spaces. Each scale space corresponds to multiple frequency bands. An image in a particular level k is the difference between the Gaussian pyramid image of level k and level $k+1$ which can be defined as:

$$L_k = G_k - upsample(G_{k+1}) \tag{6.8}$$

Upsampling of image involves doubling the size of the image in each dimension using a smoothing kernel. The image in the top-most part of the pyramid corresponds to a minute version of the original image I.

Algorithm for PAD

In the case of any natural image or video frames, there exists a substantial amount of edge information which contributes to the frequency information of that image or frame [133]. However, when the same images are printed using low/high-resolution printers or when the same images are displayed on an electronic screen, the images present frequency information which is different from the original frequency distribution of the image. This additional frequency information is inherently present in the artefacts generated by printing the live samples or replaying the live samples in the context of biometric samples. Intuitively, localizing this frequency makes the separation of normal presentation versus attack presentation. In order to localize this frequency, we employ Laplacian pyramids at 5 different scales with binomial filter kernel of size 9. It has to be noted that various kernels with different size were evaluated and the size of the kernel was fixed based on the empirical trials conducted on the development database. We have also explored the Steerable Pyramid and Gaussian pyramid decomposition which resulted in similar but lower accuracy as compared to Laplacian pyramid decomposition.

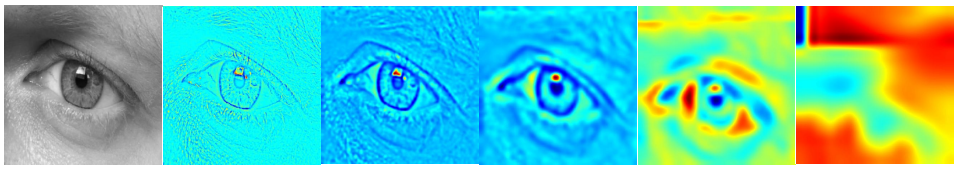
Laplacian pyramid decomposition separates the lower and higher frequency in well-defined components. We have employed a scale of $n = 5$ in this work. The difference of the obtained low pass and high pass filter at each scale is used to localize the frequency information further by analyzing STFT response corresponding to four different orientations $\Phi = \{0°, 45°, 90°, 135°\}$.

If an image at a particular scale s of the Laplacian pyramid is represented by I_s, we obtain the STFT response of the image. The STFT of the image at scale s, which is represented by F_s is the image resulting to response of frequency components in four different orientations such that $\Phi = \{0°, 45°, 90°, 135°\}$. The filter response obtained from each orientation are separated for real and complex values subsequently. Each of the responses denoted by b is finally encoded to form the final response map as given by FR_s where i corresponds to different orientation angles given by $\Phi = \{0°, 45°, 90°, 135°\}$.

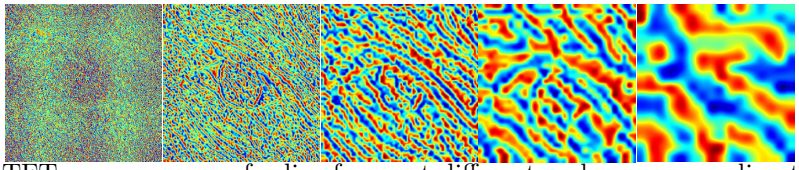
$$FR_s = \mathrm{Re}(\sum_{i=1}^{4}(b_i) * (2^{(i-1)})) + \mathrm{Im}(\sum_{i=1}^{4}(b_i) * (2^{(i-1)})) \tag{6.9}$$

The feature vector FV, of the image at a particular scale s is formed by obtaining the histogram of the response map at scale FR_s.

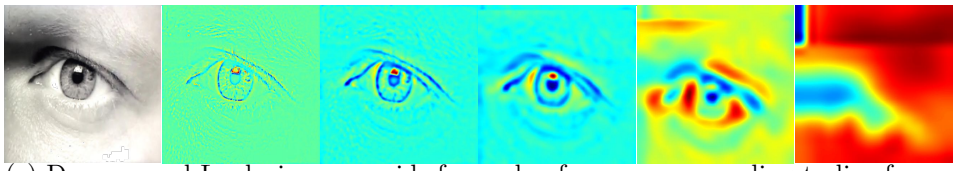
$$FV_s = \sum_{i=0}^{255}\{FR_s\}_i \tag{6.10}$$



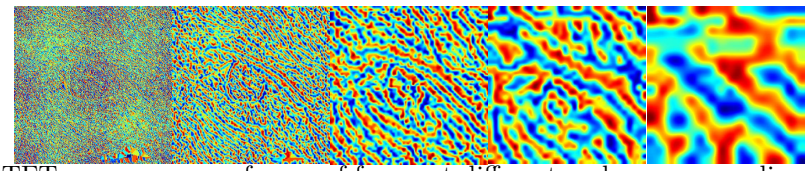
(a) Decomposed Laplacian pyramids for live frame at different scales (Scale 1 to 5 from left to right)



(b) STFT response maps for live frame at different scales corresponding to (a)



(c) Decomposed Laplacian pyramids for replay frame corresponding to live frame shown in (a) at different scales (Scale 1 to 5 from left to right)



(d) STFT response maps for spoof frame at different scales corresponding to (c)

Figure 6.9: Decomposition of images into Laplacian pyramids of scale 5 and corresponding STFT response maps in each scale. [*Note: Images from all scales are resized to uniform size for the purpose of illustration only].

The final feature vector for the frame or image is formed by concatenating the feature vectors of images from scale 1 to n and orientation $\Phi = \{0°, 45°, 90°, 135°\}$. The final feature vector FV_f can be represented as :

$$\begin{aligned} FV_f = \{&FV_{s=1,\Phi=0°}, FV_{s=1,\Phi=45°}, \\ &FV_{s=1,\Phi=90°}, FV_{s=1,\Phi=135°}, \\ &..., FV_{s=n,\Phi=0°}, FV_{s=n,\Phi=45°}, \\ &FV_{s=n,\Phi=90°}, FV_{s=n,\Phi=135°}\} \end{aligned} \tag{6.11}$$

The final feature vector given by Equation 6.11 is used to represent the image for classification purposes. Figure 6.9 presents the Laplacian pyramid decomposition at five different scales and its corresponding STFT response maps. Figure 6.9(a) presents a multi-scale Laplacian pyramid images for a sample frame from live video and Figure 6.9(b) presents the corresponding STFT response map of live frame at different scales. Figure 6.9(c) and Figure 6.9(d) illustrate the Laplacian pyramid and STFT response map of replay attack video frame. It can be observed from the figure that subtle changes in the frequency information along various orientations of the frame of replay attack video can be highly enhanced at different scales in STFT response maps.

6.4.2.2 Algorithm 3: Laplacian Pyramid Color Adaptive Quantized Hybrid Patterns

A key feature observed in the visible spectrum imaging devices is the ability to capture the color information inherently. The color information can significantly differ from real (bona fide) presentation to attack presentation when captured using the cameras which can record color intensity. Thus, we have explored color based texture descriptors in Laplacian space for classifying the presentation attack samples from bonafide samples which is termed as Laplacian Decomposed Color Adaptive Hybrid Pattern (LCAHP).

Predominantly used class of descriptors- Local binary patterns (LBP) [4] have gained great success in texture classification and the texture encoding schemes have been explored for presentation attack detection [6, 86, 87]. One deficiency in LBP schemes is that it is not robust against the noise as the scheme itself relies on encoding spatial structure based only on local information which is sensitive to noise. LBP also employs fixed quantization thresholds which add to sensitivity of noise in capturing robust texture information.

Recently proposed class of descriptors, Adaptive Hybrid Pattern (AHP) addresses the sensitivity of noise on the texture encoding [166]. They are designed robustly to make use of both local primitive features and global spatial structure. The descriptors further uses adaptive quantization to obtain unique features employing the concepts of angular quantization. The key difference as compared to LBP descriptor is the vector quantization which is adaptive to local patches. Such a strategy of adaptive quantization helps in obtaining the information from microfeature level of images. Additionally, adaptive quantization is based on equal probability quantization to achieve the maximum partition entropy [166].

Based on the success of the adaptive hybrid patterns in classifying the texture [166], we explore them to obtain the distinctive features of normal ocular images as against the artefact ocular images presented using either printed images or images presented on electronic display screen. As the ocular authentication systems operating in the visible spectrum capture color information, extracting the descriptors across color channels can improve the texture features representation. Intuitively, using the color information boosts the performance of the artefact detection as each channel present different information under normal presentation and attack presentation. The distinctiveness of the color images is sub-optimally used in biometrics and it was recommended to use image in CIE-Lab color space to gain maximum information in three different channels [62]. We transform the image to CIE-Lab color space [62] followed by the extraction of AHP descriptors.

Additionally, in the case of images under normal presentation and attack presentation, the underlying information remains same for the artefact image and the bona fide image in low frequency domain. High frequency content in those images change due to the presence of edges, half-tones patterns in the printed artefacts while information such as Moire pattern, aliased screen pattern can be observed in the electronic screen attack. This information contributes to differentiate the artefact images from bona fide images. Thus, we separate the high frequency content by applying Laplacian decomposition before extracting the AHP. AHP can be obtained with different radius and neighbourhood pixel configuration. To extract highly discriminative information, we obtain the AHP features with different radius to encode the texture features and concatenate them. However, the computation time can increase significantly with the increase in the radius and neighbourhood.

Figure 6.10 presents the attack detection scheme based on Color Adaptive Hybrid Patterns. It can be seen from the Figure 6.10 that the image is first decomposed into the CIE-Lab color space followed high frequency selection in Laplacian scale-space images. Further, adaptive hybrid patterns are extracted. The features are used to learn classifiers for each channel in CIE-Lab and one classifier for all the in concatenated form. A majority voting is performed to decide if the presentation is a bona fide or attack presentation.

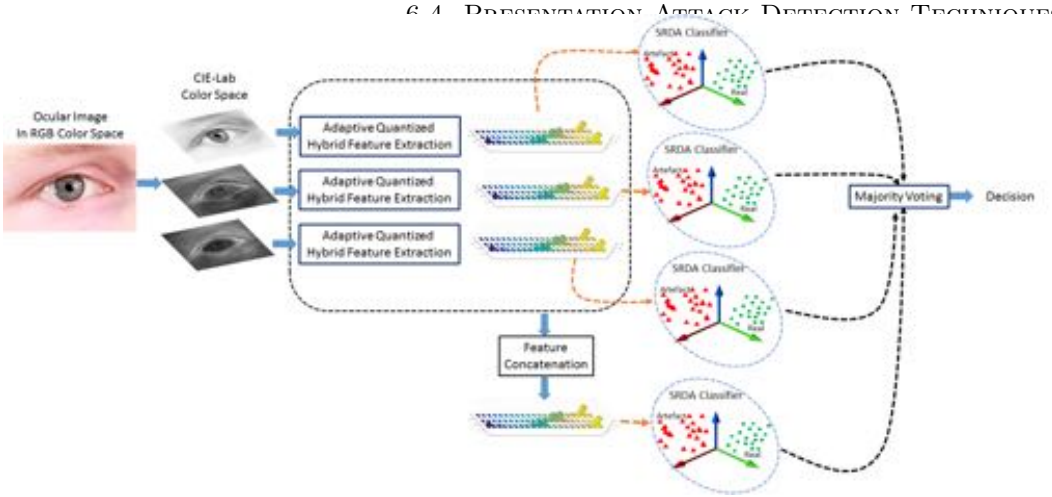


Figure 6.10: Presentation attack detection scheme based on Color Adaptive Hybrid Patterns.

6.4.3 Classification of features using Support Vector Machines

Since the presentation attack detection boils down to a two class problem, one can learn a classifier in different ways. In order to classify the features of bona fide presentation versus the features of the attack presentation (aka, spoof attack), we explore Support Vector Machine (SVM) with a polynomial kernel [24]. The detailed configurations of the SVM classifier employed in this work is presented in Table 6.5.

Table 6.5: Configurations of the SVM employed in this work

Name	Parameter
Package	LIBSVM [24]
svm type	nu-svc
kernel	Polynomial Kernel
degree	3

6.4.4 Classification of features using Spectral Regression Discriminant Analysis

Spectral Regression Discriminant Analysis (SRDA) is another well explored classifier to classify the normal presentations against attack presentations based on the efficiency and the classification accuracy [22, 7]. Motivated by accuracy of classification using SRDA, in this chapter, we employ SRDA classifier to differentiate the normal ocular presentations against the artefact presentations.

6.4.5 Experiments and Results

This section provides the details on the performance of various algorithms on two different presentation attack databases captured using smartphones. We also present comparative performance of various state-of-art algorithms which are based on Binarized Statistical Image Features (BSIF) [109], Local Binary Patterns (LBP) [86] and image quality features [41].

6.4.5.1 Results on MobiLive 2014 Database

The MobILive 2014 database provides a disjoint/independent training set and testing set which consist of 50 subjects in each test. Different teams have submitted their algorithm on this database in the MobiLive 2014 IJCB competition [131] which have used different

features. Further, BSIF-SVM [109] indicated the performance of $ACER = 0\%$. The results obtained on the same dataset using the LPFR-SVM, LCAHP-SVM and LCAHP-SRDA algorithms [67] performed with the same $ACER$ of 0%. The complete set of results are tabulated in the Table 6.6.

Table 6.6: Performance of the proposed scheme on Mobilive 2014 dataset

Techniques proposed by teams	FAR (or APCER)	FRR (or BPCER)	Mean Error Rate (or ACER)
HH	29.25	7.00	18.13
IrisKent	0.25	3.75	2.00
Liv-IC-INICAMP	0.50	2.00	1.25
Federico II	1.25	0.00	0.63
GUC	0.75	0.00	0.38
IIT Indore	0.50	0.00	0.25
M-BSIF-SVM [111]	**0.00**	**0.00**	**0.00**
Proposed LPFR-SVM	**0.00**	**0.00**	**0.00**
Proposed LCAHP-SVM	**0.00**	**0.00**	**0.00**
Proposed LCAHP-SRDA	**0.00**	**0.00**	**0.00**

Table 6.7: Division of PAVID database for experiments using ocular videos acquired from each smartphone

	Smartphone	
	Nokia Lumia 1020	iPhone 5S
Real Iris Videos		
Development	20	20
Training	50	50
Testing	82	82
Artefact Iris Videos for each attack		
Development	40 (20 x 2)	40 (20 x 2)
Training	40 (20 x 2)	40 (20 x 2)
Testing	224 (112 x 2)	224 (112 x 2)

6.4.5.2 Results on PAVID database

In order to effectively evaluate the different set of algorithms for presentation attack detection PAVID, the whole database of 152 unique eye patterns (i.e. instances) obtained using a particular smartphone is divided in three sets: Training set, Development set and Testing set. The training set comprises of 50 unique eye patterns that were used only for training the SVM classifier. The development dataset comprises of 20 unique eye patterns that are used to tune any parameters associated with the presentation attack detection algorithms. The development set if further used to determine the filter kernel for the Laplacian pyramid, the size of the window, scales for the pyramid and parameters of classification methods. The testing dataset comprises of 82 unique eye patterns that are

solely used to evaluate the presentation attack detection algorithm proposed in this work. The detailed division is provided in Table 6.7.

The Table 6.8 presents results obtained on the PAVID database using various state-of-the-art methods such as (IQM-SVM) [41], LBP-SVM [86], and BSIF-SVM [111]. It can be observed in the Table 6.8 that the LCAHP-SRDA [70] has emerged as the best technique for PAD out of all the techniques available in state-of-the-art schemes. Lowest ACER is obtained consistently across all different attacks from the proposed technique. The best ACER of 0% is obtained when the system employing iPhone as the primary sensor is attacked using enrolment videos captured using iPhone. Similarly, an ACER of 0% is obtained when enrolment video captured using Nokia phone is used to attack the iris recognition system employing iPhone as capture sensor. The obtained results indicate the applicability of the LPFR-SVM and LCAHP-SRDA approach for detecting presentation attacks in real life verification scenarios when adapted to video based iris recognition systems, specially working on the smartphones. The techniques are also tested on other artefact databases using regular RGB cameras and NIR cameras in our earlier work[70, 67].

Table 6.8: Classification error rates obtained using various schemes for PAVID database

Real Video	Artefact Video	IQM-SVM [41]			LBP -SVM [86]			BSIF-SVM [111]			Proposed - LPFR-SVM			Proposed - LCAHP-SRDA		
		BPCER	APCER	ACER	BPCER	APCER	ACER	BPCER	APCER	ACER	BPCER	APCER	ACER	BPCER	APCER	ACER
iPhone	iPhone	57.31	11.6	34.45	4.87	0.89	2.88	6.09	9.82	7.955	1.21	1.78	1.49	**0.00**	**0.00**	**0.00**
	Nokia	76.92	10.71	43.81	3.84	3.54	3.69	2.56	8.92	5.74	1.28	0.00	0.64	**0.00**	**0.00**	**0.00**
Nokia	iPhone	76.92	4.5	40.71	3.84	4.51	4.175	2.56	10.81	6.68	1.28	4.46	2.87	**0.00**	**0.00**	**0.00**
	Nokia	57.31	3.57	30.44	4.87	2.67	3.77	6.09	0.89	3.49	1.21	2.68	1.95	**0.00**	**0.00**	**0.00**

6.5 Discussions and Conclusion

Presentation attacks pose high level of threats to existing biometric systems. In this chapter, we have depicted the vulnerability of ocular biometric systems operating in visible spectrum towards presentation attacks. The advancement in types of attacks range from simple print attacks to near-real quality electronic screen attacks with both image based and video based presentations.

In this chapter, we have discussed both kind of attacks from electronic display - image attacks and video replay attacks. We have employed two databases for the experimental work in this thesis of which, one is publicly available and other is a new database (PAVID) constructed in the due course of this thesis which is now freely distributed for non-profitable research work. The new database corresponds to large scale video based artefacts consisting of 152 unique patterns acquired using 2 different smartphones - iPhone 5S and Nokia Lumia 1020. Further, we have systematically demonstrated the vulnerability of video based ocular biometric system on smartphones to video replay artefacts using electronic replays. The video based attacks are addressed using phase magnified videos with a tailored decision module. The newly proposed technique is evaluated on the subset of PAVID dataset and has demonstrated the applicability for real-life use case.

This chapter also presented two techniques for frame/image based presentation attacks. For frame based artefacts, we have employed MobiLive 2014 dataset [131] and PAVID dataset. We have evaluated the state-of-art techniques in detecting video replay attacks using features obtained from both quality and texture features. From the set of experiments, the features employing spatial and frequency features from Laplacian images have proven robust in detecting artefact presentations. The features are well classified with the use of popular SVM classifier along with majority voting. The proposed texture based methods - LPFR-SVM and LCAHP-SRDA methods have provided the lowest ACER which is close to 0% for video replay attacks on ocular recognition systems when the data was captured using smartphones in visible spectrum.

The success of the texture based methods in the current chapter can largely be attributed to the strong feature descriptors which can localize key characteristics such as additional texture contributed by printers in the case of printed artefact and other specific patterns in case of electronic screens such as moire pattern. Further, decomposing the image using Laplacian Pyramid provides the scale-space information where the micro-texture information can be observed easily. The discriminant information from each of the scales in Laplacian pyramid is optimally used with two texture descriptors presented in Algorithm 2 and Algorithm 3. The first set of texture based descriptors (Algorithm 2) localizes the time and frequency features across 4 different orientations which can detect the patterns emerging due to electronic display.

The color intensity information differs in the case of normal presentation and the artefact presentation which can be used to identify presentation attack. However, obtaining the color information alone may not suffice to detect attacks reliably. Thus, in the second class of textural descriptors presented in Algorithm 3, the texture information in different color channels from different scales of Laplacian pyramid is obtained in an adaptive manner. The region adaptive texture information along with the global information is obtained which are highly discriminative in nature for bona fide and artefact image. Further, adaptive quantization of the pattern contributes significantly to identify artefacts from bona fide images.

Although the results indicate the promising nature of the proposed methods, there is a necessity for continued research to detect artefacts robustly to make smartphone based ocular biometrics highly secure, especially when the artefact data is unknown. The known set of artefacts can be handled well by devising algorithms whereas the unknown artefacts still pose a serious threat. Thus, the need for generalizable solutions for presentation attacks including unknown artefacts in smartphone remains a problem which needs significant research.

The generalizability also needs to be investigated with respect to database-independent techniques that can be applied in real-life scenarios using large scale data obtained in visible spectrum whereas most of the current algorithms are tailored specifically for different databases.

Chapter 7

Multi-biometric Authentication System for Smartphones Using Face, Periocular and Iris

Multi-biometric systems are used to achieve better authentication accuracy for secure access controls. This chapter presents such a multi-biometric authentication system on the smartphones. The unimodal biometric performance using face, periocular and iris characteristics are first demonstrated which is followed by a robust multi-biometric system with high biometric performance. The systems are realized fully on the computationally limited smartphones (compared to desktop platforms) which are evaluated for real-life applicability using set of experiments.

7.1 Introduction

Earlier works have successfully demonstrated that the quality of images captured using smartphone embedded cameras are of sufficiently high quality to be accepted as the biometric sample for various use cases [34, 141, 147, 13, 73, 113, 128, 129]. Motivated by the earlier works including the previous chapters of this thesis, we explore a multi-biometric system that employs face, periocular and iris images all captured with embedded smartphone cameras for authentication purpose.

As the face image is captured from a close distance, one can always obtain periocular characteristics and iris information with significant details. Using periocular information in addition to face under difficult circumstances (Pose, Illumination and Expression) can maintain and even improve the recognition accuracy of a biometric system[165, 102]. The robustness and performance of the ocular characteristics including the iris is well illustrated in the previous chapters (Chapter 3 and Chapter 5). Further within the face image one can obtain the visible spectrum iris representation with sufficiently high resolution [34, 75]. As iris is known to provide very reliable recognition performance, we make use of the iris pattern, whenever this can be segmented. Considering the verification performance of such systems individually employing face, periocular and iris, in this chapter we explore them for applicability on smartphone based authentication system. Specifically, in this chapter, we systematically explore the unimodal performance of face, iris and periocular region independently. Further, this chapter also presents a holistic multi-biometric system, which uses face, periocular region and iris region in a combined manner.

Unlike most of the earlier works in this direction where the biometric data was collected using the smartphone and evaluated offline on the desktop computing platforms [34, 141, 147, 13, 73, 113, 128, 129], in this chapter, the entire framework from capture to computation of comparison score is carried out on the smartphones. The main motivation behind the implementation of the multi-biometric system on smartphone is to gauge the real-life applicability. Further, the key idea in devising the multi-biometric authentication system is to employ the state-of-art feature extraction scheme. Thus, in this chapter, we employ the state-of-art feature extraction schemes like Scale Invariant Feature Transform (SIFT), Speeded Up Robust Features (SURF) and Binarized Statistical Image Features (BSIF) for extracting the features in based on following factors:

1. Earlier works have employed these techniques to obtain reliable performance in

biometric systems [34, 141, 147, 13, 73, 102] which can be considered as a basis for using it on smartphone biometric system.

2. The implementation of these methods are available openly for the usage and thus any systems can leverage the merits of these feature extraction techniques by tuning to specific applications.

3. As these feature extraction methods are not demanding in terms of computation, they can be easily adapted to mobile/smartphones platforms with limited computational complexity.

4. Determining comparison score between two set of features using these descriptors can be obtained using low computation distance metrics as against expensive dictionary approaches.

In the similar manner, we employ 2D Gabor features with Hamming Distance comparison following the works of iris recognition [29] for smartphone based iris recognition.

Although highly robust algorithms for iris and periocular recognition are presented in Chapter 3 and Chapter 5, it has to be noted that these methods can be demanding for computationally limited devices like smartphones. Specifically, the sequence of operations in classification involves in representing the signals in sparse manner or learning the feature set in a collaborative manner which are not optimally designed for systems with low memory and computational power. Thus, in this chapter, we have made use of publicly available algorithms that are not demanding and can easily be ported to run on smartphone platforms such that the entire pipeline of image capture to comparison score computation is available on smartphone.

In the remainder of this chapter, Section 7.2 presents the architecture of multi-biometric system for the authentication system. Section 7.3 presents the database employed to evaluate multi-biometric authentication system. Section 7.4 details the experiments carried out and the corresponding results obtained. Section 7.6 provides the remarks and summary from the current chapter.

7.1.1 Contributions

The contributions of this chapter can be summarized as:

- Explores multi-modal biometric system for authentication on smartphones using face, periocular and iris characteristics.
- Multi-biometric system presented here is tested using 78 subjects on two different devices to gauge the applicability in real-life authentication applications.
- The presented system is fully realized and tested on smartphones. The entire pipeline starting from image capture including segmentation, feature extraction and comparison are implemented on the smartphones. This chapter thus validates the applicability of presented system on all of the current smartphones capable of providing images of sufficient biometric quality.

7.2 Multi-biometric Authentication System on Smartphones

This section presents architecture and principles of multi-biometric authentication system on smartphones. First, we list explicit advantages of the using multi-biometric system on smartphones and later present the schematic of the proposed multi-biometric system.

7.2.1 Advantages of Multi-biometric Authentication System

Multi-Biometric systems have proven to perform better than systems based on single biometric characteristics especially in a unconstrained acquisition scenario [126]. The availability of all the intended biometric characteristics can be optimally acquired when the acquisition is cooperative. In case of an unconstrained acquisition conditions, relying on a single biometric characteristic may not yield best performance for authentication. Further, when the biometric systems are intended to capture face region, they inherently capture periocular region with acceptable quality [126, 58] and in an ideal capture distance, iris characteristics can be obtained with good details. Further, various works have demonstrated the use of periocular region under the challenges of uncooperative/non-standard face or iris capture [101, 102, 63, 11]. Thus intuitively, the performance of the system can be improved by employing all the characteristics from face that include face, periocular (left and right) and iris (left and right) [102, 11, 88], specifically in the scenarios listed below:

- When the subject captures the face image with non-standard pose resulting in partial face capture, the multi-biometric system can rely on periocular region or iris (if available).
- When the subject captures the face image in non-uniform illumination on one side of the face, the textural details on the complete face may not be fully available and thus, multi-biometric system can employ ocular region which is not affected due to illumination and shadows [102].
- When the subject captures the face image with the expression such as with smile, the periocular region is affected less as compared to face region [101].
- When the subject captures the image by holding it very close to camera, the appearance of the face image can be impacted by geometrical distortions in face and thereby, the multi-biometric systems can employ iris and ocular region as the image is captured in closer range [88].

7.2.2 Schematic of Multi-biometric Authentication System

The multi-biometric (*a.k.a multi-modal*) authentication system for smartphones is illustrated in the Figure 7.1. The system combines face, periocular and iris recognition subsystem as the core components. When a particular subject wishes to enrol into the system, the image is captured and provided to the face detection subsystem. This subsystem works synchronously with the capture device or camera by providing continuous feedback following the architecture in our earlier work[73]. If the face is not detected in the captured frame, the face detection subsystem sends continuous feedback to recapture. Once the captured face sample is determined to be of sufficient quality by the Haar cascade based face detector, the exact region of face is localized[18].

The localized facial region is further submitted to the face recognition and periocular recognition subsystems. Along with the processing in these two subsystems, the iris recognition subsystem is activated, if the iris is represented with sufficient quality. For a data subject to whom the visible spectrum representation of the iris pattern is insufficient, as in the case of dark irises, the iris recognition subsystem shall not attempt to enrol an iris reference.

OSIRIS v4.1 is employed for the segmentation due to the robustness demonstrated in segmenting the images from both visible spectrum iris samples [135, 76, 117]. In this work, a lighter version of OSIRIS v4.1 is implemented, specifically to run on smartphones (Android and iOS platforms) devices under limited computation. As the segmentation task of the iris on smartphones is a challenging problem, this work contributes significantly by providing the open source iris segmentation scheme for smartphone environments operating at minimal

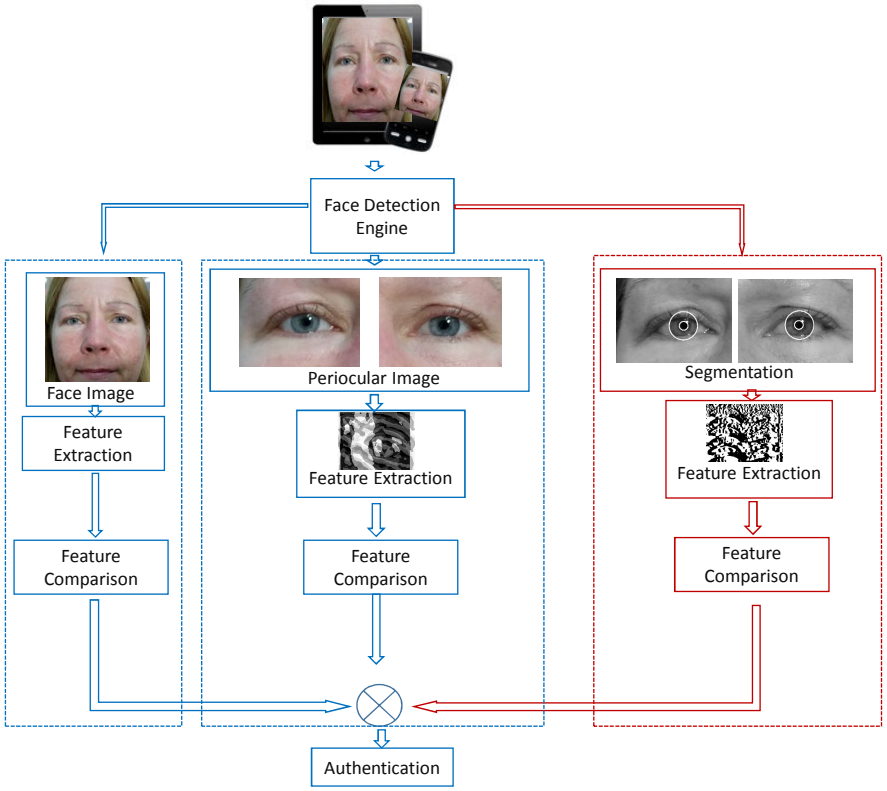


Figure 7.1: Schematic representation of multi-biometric authentication system; The blocks in blue color indicate imperative contributions to the authentication process (i.e. to the decision subsystem) and the blocks in red color indicate an optional contribution in case the iris pattern has been segmented successfully.

response time. The segmented iris texture is further processed using Daugman's rubber sheet expansion technique [29]. The iris pattern is normalized to a dimension of 512×64 pixels. Once the subject is enrolled, all templates corresponding to face, periocular and iris are stored in the smartphone embedded database.

When the subject wants to authenticate, the image is acquired and the face is detected as illustrated previously. The features are extracted from the face and periocular region. Depending on the visibility and quality of the iris texture pattern, iris features are extracted as well. SIFT, SURF and BSIF features are extracted for face and periocular region while 2D Gabor features are extracted from iris. Probe feature vectors are compared against the reference templates stored on the smartphone. We have employed Fast Approximate Nearest Neighbor Search with Hierachical K-means Tree to determine the similarity of feature vectors in case of SIFT and SURF [93]. Bhattacharya distance [23] is employed to measure the similarity score between two histograms of BSIF features. The scores obtained from all modalities are fused at score level and submitted to the decision subsystem to finally authenticate the data subject.

7.3 Multi-Biometric Smartphone Database

In order to validate the performance of the multi-biometric system running on smartphone platforms, in this chapter, we have created a database with images captured from two different devices, the smartphone - Samsung Galaxy S5 and the tablet - Samsung Galaxy Note 10.1. The specification of hardware is listed in the Table 7.1. It has to be specifically

noted that the multi-modal authentication system was evaluated using the data captured from 78 subjects[75]. The database was divided into a development and testing dataset. The development database consisting of 32 subjects was used to tune the parameters for feature extraction algorithms and weights for different fusion schemes. As there is no training involved, there arises no requirement to reserve a partition of the database for training purposes. The partition of the entire database is provided in the Table 7.2.

Table 7.1: Specifications of different hardware used in this chapter.

Device	Operating System	Screen Size	Back Camera
Samsung Galaxy S5	Android v4.4.2	1080 x 1920 pixels 5.1 inches	16 MP, 5312 x 2988 pixels
Samsung Galaxy Note 10.1	Android v4.4.2	800 x 1280 pixels, 10.1 inches	5 MP, 2592 x 1944 pixels

Table 7.2: Division of database into development and testing; *Note: 15 indicates 15 different sessions of which 5 correspond to reference image and 10 correspond to probe images

Camera	Total Subjects	Development set		Testing set	
		Subjects	Subjects × Images	Subjects	Subjects × Images
Smartphone - Samsung S5					
Back	78	32	32×15	46	46×15
Back Assisted	78	32	32×15	46	46×15
Tablet - Samsung Note 10.1					
Back	78	32	32×15	46	46×15
Back Assisted	78	32	32×15	46	46×15

For both the development and testing set, each subject is enrolled into the system by capturing 5 reference samples on each of the two different portable devices as mentioned in the Table 7.1. The enrolled subject on each of the device is authenticated by using 10 probe samples captured in different attempts ranging from single day to several days. For both reference and probe samples the image was acquired using the back camera of the device, as it provides higher image resolution and it is expected to have the same resolution for the front camera in near future due to continuous improvement in smartphone cameras.

Thus, in order to obtain an optimal baseline performance of the system, images in this database were also captured by a trained expert in similar settings for each attempt of enrolment and probe which is hereafter referred as assisted acquisition. Therefore, two set of reference images and probe images were acquired for each device. For each subject, there are in total 15 images (5 reference images and 10 probe images) captured from a single device in a single setting of capture (i.e., self acquisition or assisted acquisition). The experimental protocol on this database is explained in the upcoming section.

Table 7.3: Total images from each device in the database

Device	Back Camera	Rear Camera (Assisted)
Samsung Galaxy S5	2340	2340
Samsung Galaxy Note 10.1	2340	2340

7.4 Experiments and Results

The multi-biometric system is evaluated as a stand-alone biometric system in this chapter. The system is first evaluated seeking the biometric performance for three unimodal characteristics and later for multi-biometric approach. In order to obtain unimodal performance systematically, three sets of experimental protocols are designed:

1. In the first set, experiments are designed to evaluate the system using face based authentication.
2. Second set of experiments evaluates the performance of the system when the periocular region alone is used for recognition.
3. Another set of experiments are conducted to measure the reliability of iris based recognition on a smartphone scenario in unconstrained illumination.

Each of these experiments is described in the sections below. Further, the last set of experiments are designed to combine all the different biometric characteristics with score level fusion, which are detailed in Section 7.4.5.

7.4.1 Evaluation protocol

All experiments in this work are based on the database described in the Section 7.3. Each of the images from the reference set is compared against the probe image to obtain the genuine and imposter score. The enrolment and the probe set are disjoint in both development set and testing set where 5 images are present in enrolment set for a subject and 10 images in probe set. Thus, for each subject in testing set of 46 users, 50 genuine scores are computed ($5\ enrolment \times 10\ probe$). Similarly, $2,250$ impostor scores are obtained ($5\ enrolment \times 46 subjects \times 10\ probe$). The detailed distribution of the genuine and impostor scores from the number of comparisons are listed in the Table 7.4.

Table 7.4: Details of the number of samples and distribution of genuine-impostor composition from the testing dataset

Camera	Total Subjects	Reference Image	Probe Image	Total Images	Geniune Comparisons	Imposter Comparisons
Smartphone Samsung S5						
Back	46	5	10	690	2300	103500
Back Assisted	46	5	10	690	2300	103500
Smartphone Tablet						
Back	46	5	10	690	2300	103500
Back Assisted	46	5	10	690	2300	103500

7.4.2 Experiments on Smartphone Based Face Recognition

The presented system captures the image with the rear camera of smartphone based on the optimal focus computed using the preview frame in the camera's view. Once the image is captured, the user is presented a choice to either keep or discard the image in order to have sufficient visible quality. Capturing facial images is very challenging with respect to pose and illumination changes. In this set of experiments, we explore the face recognition performance accuracy under the assumption that neither pose and illumination are explicitly controlled nor that we deliberately introduce weak poses or ill-suited lighting. Table 7.5 presents the performance for face recognition obtained using various feature extraction methods.

Since the capture is not regulated by specific conditions and assumed to be cooperative[1], in this chapter, we have also analysed the data obtained in the similar settings by a trained expert. The data collected by trained expert is considered as hereafter referred "Assisted" in the rest of the chapter. The main motivation to study the performance when captured by an trained expert is the compare the reliability of the proposed system when used in unconstrained (semi-cooperative) settings. Further, such a study also reveals the agonistic nature of the algorithms when the data can be degraded arising out of non-standard interaction with the device (smartphone) and number of other factors including the illumination, motion blur while capture process is underway.

Table 7.5: Biometric performance in terms of Genuine Match Rate and Equal Error Rate for unimodal approaches.

Camera	Feature Extraction	Face		Right Periocular		Left Periocular		Both Periocular	
		FMR @ 0.01%	EER	FMR @ 0.01%	EER	FMR @ 0.01%	EER	FMR @ 0.01%	EER
Smartphone - Samsung S5									
Back	SIFT	76.36	5.18	57.34	**6.63**	45.35	7.07	65.89	4.40
	SURF	45.03	10.21	70.56	6.55	59.28	6.20	76.00	4.52
	BSIF	**87.55**	**4.65**	**75.86**	7.01	**76.00**	**5.80**	**83.24**	**4.14**
Back Assisted	SIFT	88.43	1.88	65.43	5.00	70.83	5.03	82.04	3.04
	SURF	52.91	5.13	**84.04**	**4.00**	**80.22**	**4.86**	**90.52**	**3.13**
	BSIF	**94.39**	**1.61**	79.00	5.56	72.09	5.73	83.96	3.57
Tablet - Samsung Note 10.1									
Back	SIFT	92.83	2.62	31.43	9.40	61.91	8.40	72.96	6.85
	SURF	81.83	3.34	77.26	6.01	74.30	7.54	85.35	5.04
	BSIF	**94.61**	**2.43**	**77.39**	**5.91**	**77.30**	**6.75**	**87.04**	**4.71**
Back Assisted	SIFT	95.57	**1.81**	30.83	8.79	48.52	6.87	72.17	5.18
	SURF	79.91	1.96	71.30	5.31	51.09	5.47	82.00	3.87
	BSIF	**96.65**	2.03	**82.91**	**5.00**	**85.78**	**4.78**	**89.96**	**3.74**

The data obtained from the back camera in self aquisition (i.e. non-assisted capture mode) provides an EER of 4.65% corresponding to GMR of 87.55% at FMR = 0.01% with BSIF features. The data obtained in the assisted mode has an EER of 1.61% corresponding to a GMR of 94.39% at FMR = 0.01%. Figure 7.2 (a) and (b) present the Receiver Operating Characteristic (ROC) curves for face based recognition for self acquisition and assisted acquisition from Samsung S5 while Figure 7.2 (c) and (d) present it for the Samsung Note 10.1 tablet.

It can be observed that the face recognition is again validated for promising recognition performance for biometric authentication on smartphones in line with the earlier works[141, 13, 147]. It can also be noted that when the subject uses the back camera to capture the face image in a self acquisition mode, the challenges due to pose alignment cause a slightly lower performance but comparable to the images captured under the same settings by a trained expert. In the similar manner, Figure 7.2 presents the obtained performance on database collected using the Samsung Galaxy Note 10.1. Following the previous conventions of the reporting the results, it can be observed that the Figure 7.2 presents the performance for both self acquired images and images captured by expert (assisted mode).

[1]Banking applications are typically designed such that the users are cooperative to provide the biometric data to the authentication system.

(a) BA - Face based recognition in smartphone

(b) BS - Face based recognition in smartphone

(c) BA - Face based recognition in tablet

(d) BS - Face based recognition in tablet

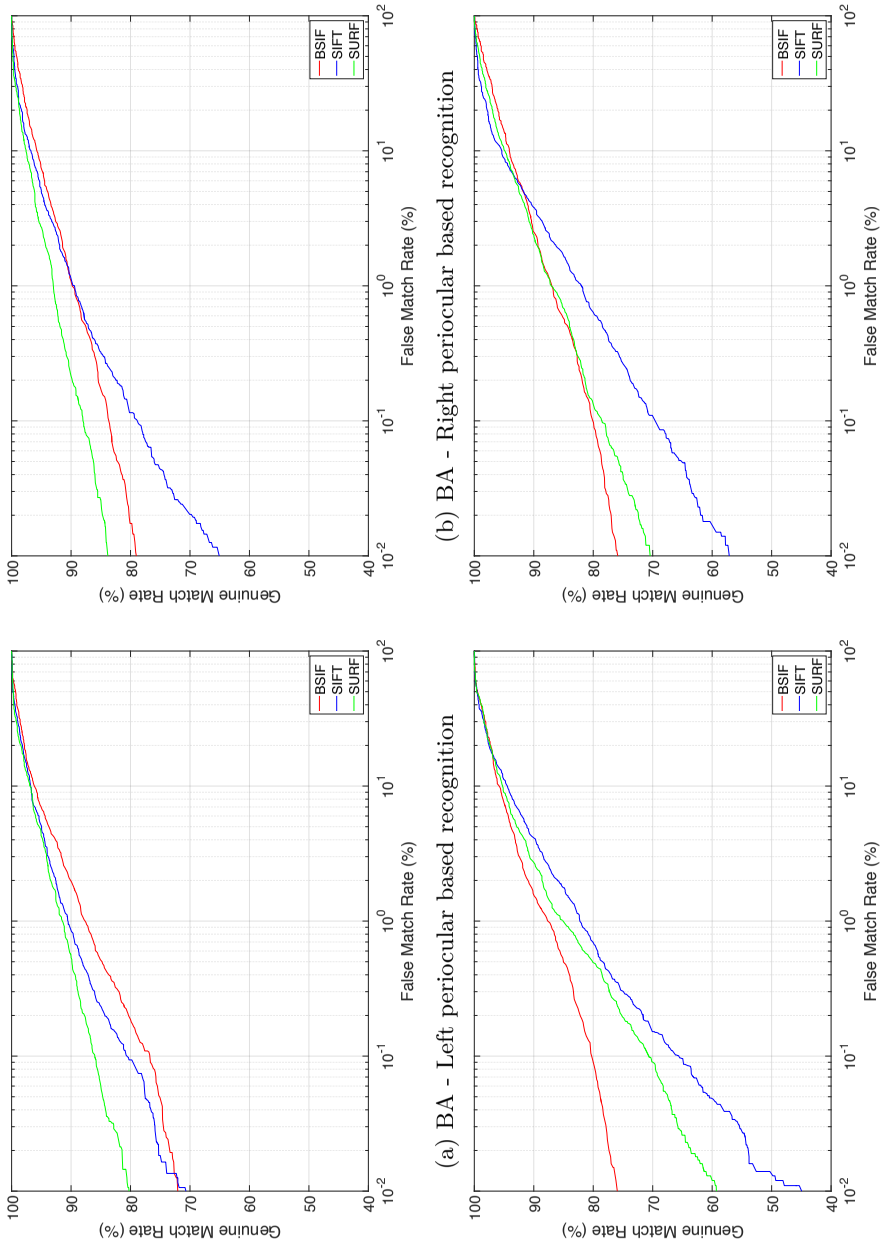
Figure 7.2: ROC curves for various unimodal recognition employing face characteristics; (a) & (c) correspond to assisted acquisition using the back camera; (d)-(f) correspond to self acquisition using the back camera;
*BA - Assisted acquisition from back camera, *BS - Self acquisition from back camera

7.4.3 Experiments on Smartphone Based Periocular Recognition

As discussed in earlier sections, the problem of non-uniform illumination, pose changes and various expression is known to degrade the performance of face recognition systems [165]. The problem becomes more prominent, when the capture device is not fixed as in a border crossing scenario or authentication for an e-commerce application. When the face is illuminated in a non-uniform manner, one of the two periocular regions which is unaffected can still be used for recognition as discussed in earlier works [102]. The intrinsic advantage in using periocular information is that two periocular regions can be used for one subject complementing or substituting the information from the overall face image.Thus, we also explore periocular based recognition subsystem to address the challenges arising out of non-uniform illumination in this section.

Table 7.5 presents the recognition performance in terms of EER and GMR for periocular features with various feature extraction techniques. Images of periocular region from right side of the face provides an EER of 6.55% with back camera (SURF) and 4% with assisted acquisition (SURF). Table 7.5 also presents the recognition performance when the left periocular image is used. An EER of 5.80% is obtained with back camera (BSIF) and EER of 4.86% is obtained with back camera in assisted mode (SURF). Similar results can be seen for the periocular recognition with tablet device. Combining both the periocular region as one unique characteristic (by concatenating the periocular region next to each other) further boosts the authentication performance as indicated in the Table 7.5. An average gain of around 1.5% in EER can be seen when both periocular information is fused to treat it as an single modality.

Figure 7.3 and 7.4 present the ROC curves for periocular based recognition for different acquisition modes with Samsung S5 and Samsung tablet. The periocular based recognition can perform close (comparable) to the accuracy of face based recognition as it can be observed from the experiments.



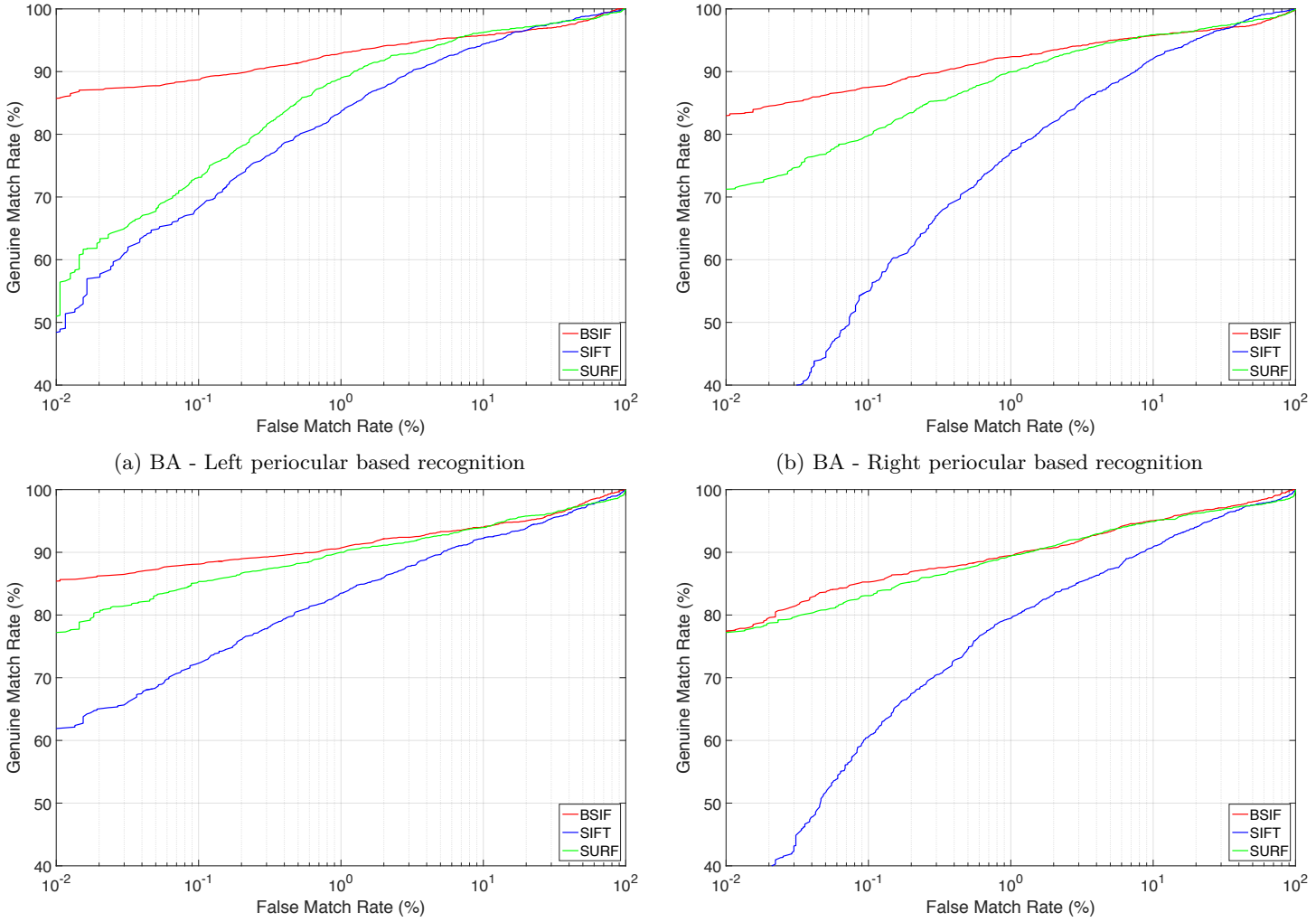
(a) BA - Left periocular based recognition

(b) BA - Right periocular based recognition

(c) BS - Left periocular based recognition

(d) BS - Right periocular based recognition

Figure 7.3: ROC curves for various unimodal recognition employing periocular characteristics on Samsung S5; (a)-(c) correspond to assisted acquisition using the back camera; (d)-(f) correspond to self acquisition using the rear(back) camera; *BA - Assisted acquisition from rear camera, *BS - Self acquisition from rear camera



(a) BA - Left periocular based recognition

(b) BA - Right periocular based recognition

(c) BS - Left periocular based recognition

(d) BS - Right periocular based recognition

Figure 7.4: ROC curves for various unimodal recognition employing face and periocular characteristics on Samsung Note 10.1 tablet; (a)-(c) correspond to assisted acquisition using the back camera; (d)-(f) correspond to self acquisition using the back camera; *BA - Assisted acquisition from back camera, *BS - Self acquisition from back camera

7.4.4 Experiments on Smartphone Based Iris Recognition

Motivated by the recent works on iris recognition in visible spectrum and the possibility of using the iris information on a smartphone [108, 34, 8, 76], in this chapter, we also explore the iris recognition from the images obtained on smartphone.

The key aspect in this analysis is that the entire iris recognition pipeline starting from image capture to segmentation until verification is tested on the framework developed for smartphones. Further, since each person has two unique iris patterns, we have explored the performance of the smartphone based iris recognition using each individual eye. We have employed 2D Gabor features with Hamming distance for similarity score computation [29] to make use of state-of-art technique.

Unlike the case of periocular region where both the images can be used for recognition purpose, the availability of iris depends highly on the accuracy of the segmentation algorithm and visibility of the texture. Due to unconstrained nature of iris imaging in the visible spectrum, a number of challenges regarding the sample quality must be expected. Out-of-focus imaging and motion blur are more prominent under smartphone based iris imaging than it is the case for conventional near infrared iris imaging. Thus, the general algorithm performance metrics FMR and FNMR are insufficient to report the effective performance under the presence of iris images that can not be segmented. All the non-segmented iris images are treated as Failure-to-Acquire (FTA). Data subjects that due to dark iris patterns cannot be enrolled with the visible spectrum samples must be treated as Failure-to-Enroll (FTE) as discussed in the Chapter 2. We thus present the results using Generalized False Accept Rate (GFAR) and the Generalized False Reject Rate (GFRR) according to International Standard ISO/IEC 19795-1 [54] along with indicative metric of Generalized Equal Error Rate (GEER).

Table 7.6: Performance of iris recognition

Camera	Smartphone - Samsung S5			Tablet - Samsung Note 10.1		
	Iris	1 - GFRR (%) @ GFAR=0.01%	GEER (%)	Iris	1 - GFRR (%) @ GFAR=0.01%	GEER (%)
Back Assisted	Left	38.95	22.67	Left	35.93	24.20
	Right	48.80	23.72	Right	32.67	26.44
Back	Left	42.56	23.85	Left	39.70	22.74
	Right	39.90	22.89	Right	43.16	21.88

Table 7.6 lists the algorithmic performance of iris based recognition for right and left iris of both smartphone and tablet device. A GEER of 22.67% is obtained for left iris and 23.72% is obtained for right iris related to smartphone data. Almost comparable GEER is obtained for the back camera in self acquisition mode. Similar results are achieved for the tablet as indicated in the Table 7.6. Figure 7.5 provides the detailed graphical illustration of the iris based recognition performance which depicts the best accuracy closer to 50% indicating the challenge in wider deployment of iris recognition.

7.4.5 Experiments on Multi-modal Recognition

Face region comprises of periocular information which can be used independently or jointly with the face information. Non-standard and unconstrained face acquisitions always suffer from non-standard pose, illumination and angle. Even under such non-standard conditions, the use of the periocular region has proven to perform substantially or equivalently well as compared to face based verification [101]. In the similar terms, an earlier work has confirmed the performance of periocular information for authentication on smartphones [73]. Further, as the iris and face data contribute complementary information, we also explore multi-modal fusion using the face, periocular and iris data. Thus, under the non-uniform illumination on

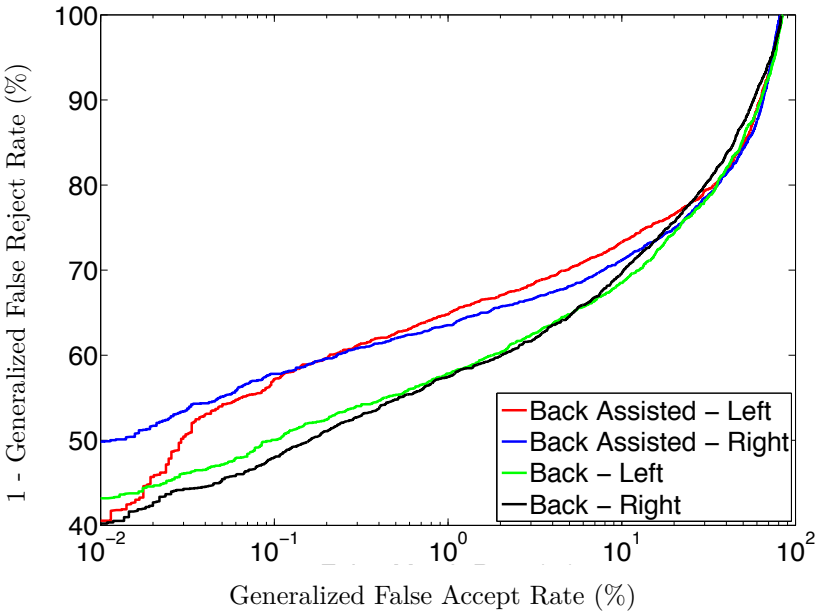


Figure 7.5: ROC plots of iris recognition

face, at least one of the features, either face, periocular or iris provides good features and/or comparison scores for recognition.

Based on the knowledge of fusion in biometrics, systems have devised two strategies for fusion of biometric characteristics [125]:

1. Fusion of independent biometric characteristics at feature level and combine them to obtain an aggregate score which is popularly referred as Feature level fusion.

2. Fusion of comparison scores from independent biometric features and obtain an aggregate score called as Score level fusion.

Although, two strategies are used in biometrics, comparison score level fusion has demonstrated the applicability in earlier works [125, 140, 58, 127]. It was also demonstrated in our work earlier that the feature level fusion provides minimal improvement as compared to score level fusion[74] which was in-line with the findings of other work [125, 140, 58, 127]. Thus, in this work, we employ comparison score level fusion in the multi-biometric system.

Figure 7.6 presents a simplistic view of comparison score level fusion scheme for multi-biometric system employing face, periocular and iris. As the comparison score from iris is subject to visibility of texture pattern, it may not be contribute always and thus, it is indicated in the red block in the paradigm of Figure 7.1. In the multi-biometric authentication framework presented in this chapter, five different characteristics which include face, left periocular, right periocular, left iris and right iris are employed. Iris provides complementary information as compared to the face and ocular region as the features descriptors differ from face and ocular region. Thus, in order to leverage the scores from different feature extraction of face and ocular region, we fuse the scores from three features. If the comparison scores from BSIF features are represented by C_b, SIFT features are represented by C_s and SURF features are represented by C_u, then the fused score for each modality is computed using weighted fusion as given by Equation 7.1.

$$F = w_b * C_{fb} + w_s * C_{fs} + w_u * C_{fs} \tag{7.1}$$

where w_b, w_s and w_u refer to weights of BSIF, SIFT and SURF respectively. Based on the trials conducted on development set, the w_b, w_s and w_u were set to 0.7, 0.15 and

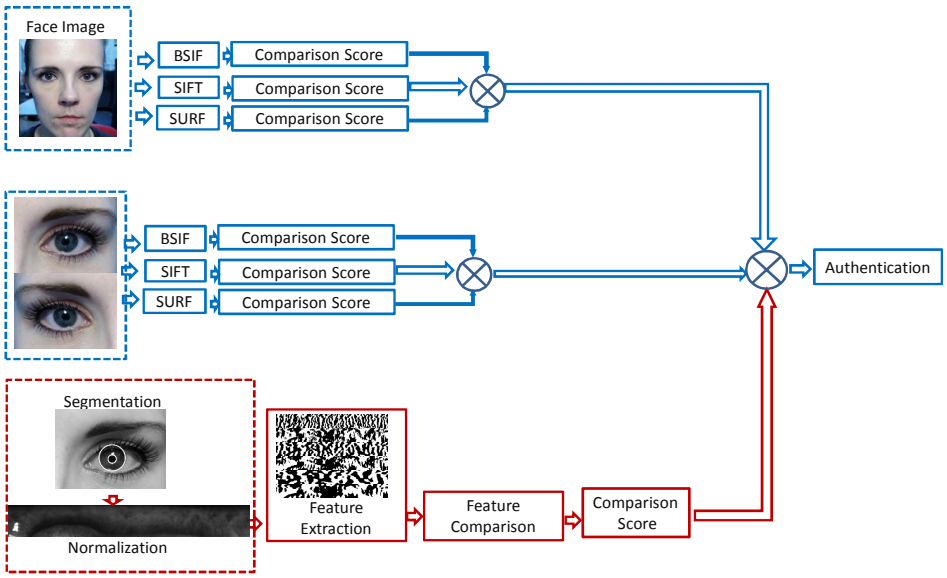


Figure 7.6: Schematic representation of multi-biometric authentication system with score level fusion; The blocks in blue color indicate imperative contributions to the authentication process (i.e. to the decision subsystem) and the blocks in red color indicate an optional contribution in case the iris pattern has been segmented successfully.

0.15 respectively. The determined weights were further used for the scores from periocular characteristics as given by Equation 7.2 for left and right periocular region given by P_l and P_r respectively.

$$P_l = w_b * C_{lb} + w_s * C_{ls} + w_u * C_{lu} P_r = w_b * C_{rb} + w_s * C_{rs} + w_u * C_{ru} \tag{7.2}$$

Further, the scores are normalized to the range of $0-1$. The multi-biometric score fusion is explored using Min, Max, $Product$ and $Dynamic\ weighting$ rule. If the fused score of face is provided as F, the fused score of periocular region is provided as P_l for left periocular and P_r for right periocular region and the comparison score for iris is represented as I_l for left iris and I_r for right iris respectively, then the final comparison score F_c can be obtained using one of the schemes given in the following subsections.

Min-score Fusion Rule

In this fusion scheme, the score corresponding to the minimum of all the obtained scores is used. The final score is obtained in accordance to Min-rule given by F_c as:

$$F_c = \arg\min\{F, I_l, I_r, P_r, P_l\}; \tag{7.3}$$

Max-score Fusion Rule

Under the Max-score fusion rule, the score corresponding to the maximum in the set of modality specific scores is used. The obtained final score in accordance to the Max-score rule is given as below:

$$F_c = \arg\max\{F, I_l, I_r, P_r, P_l\}; \tag{7.4}$$

Product-based Fusion Rule

Further, the product rule has been popularly explored in biometrics. In this scheme we compute the product score by multiplying the scores obtained for each modality. The

obtained final score under the product rule is given as:

$$F_c = F * I_l * I_r * P_r * P_l; \tag{7.5}$$

Dynamically Weighted-Score Fusion Rule

Since the scores of each modality contribute to the performance in various degrees, we explore a dynamic weighting scheme to make the recognition system robust. As discussed earlier, the performance of the system can be improved by incorporating multiple modalities [126]. At the same time, due to the various issue regarding capturing iris textures in the visible spectrum, it is likely that no iris data is available for a portion of subjects. Thus, in this work we propose a dynamic weighting scheme, where each modality is assigned a weight such that sum of all weights equals 1. Under circumstances where a particular modality does not contribute to the comparison score, the weight of that particular score is set to 0 and the weights are redistributed equally among all other modalities contributing to the recognition. Thus, the dynamic weighted fusion scheme is given as by Algorithm 2 and be summarized as Equation 7.6.

$$F_c = w_1 * F + w_2 * I_l + w_3 * I_r + w_4 * P_r + w_5 * P_l; \tag{7.6}$$

where $w_1 + w_2 + w_3 + w_4 + w_5 = 1$. For instance, if the score from right iris is missing, the weight w_3 is set to 0 and the new assignment of the weight is computed such that $w_1 + w_2 + w_4 + w_5 = 1$.

Algorithm 2 Dynamic Weight Distribution for Multi-biometric Characteristics

Initialize:
$w_i \leftarrow \frac{\sum_{i=1}^{5} w_i}{i}$, $i = 1, \ldots, 5$
where $\sum_{i=1}^{5} w_i = 1$
if All comparison scores among (F, I_l, I_r, P_r, P_l) are available **then**
$w_i = \frac{\sum_{i=1}^{5} w_i}{i}$, $i = 1, \ldots, 5$
else if One missing comparison score among (F, I_l, I_r, P_r, P_l) **then**
$w_i \leftarrow \frac{\sum_{i=1}^{4} w_i}{i}$, $i = 1, \ldots, 4$
else if Two missing comparison score among (F, I_l, I_r, P_r, P_l) **then**
$w_i \leftarrow \frac{\sum_{i=1}^{3} w_i}{i}$, $i = 1, \ldots, 3$
else if Three missing comparison score among (F, I_l, I_r, P_r, P_l) **then**
$w_i \leftarrow \frac{\sum_{i=1}^{2} w_i}{i}$, $i = 1, \ldots, 2$
else
$w_i = 1$

Table 7.7: Verification accuracy of multi-biometric score level fusion obtained by employing face, periocular and iris characteristics on the complete database. The performance can be compared against the uni-modal performance of face region alone given in the last row of this table.

Fusion Scheme	Camera	Samsung S5		Samsung Note	
		GMR (%) @ FMR = 0.01%	EER	GMR (%) @ FMR = 0.01%	EER
Min Rule	Back Assisted	99.17	0.43	88.57	3.43
	Back	97.12	0.93	88.13	4.34
Max Rule	Back Assisted	50.78	10.71	11.65	25.93
	Back	52.94	12.10	17.74	22.59
Product	Back Assisted	84.13	15.34	50.65	47.96
	Back	84.81	14.37	44.61	48.08
Weighted Fusion	Back Assisted	**99.13**	**0.43**	**99.20**	**0.48**
	Back	**97.98**	**0.68**	**97.91**	**0.92**
Face	Back Assisted	94.39	1.61	96.65	2.03
	Back	87.55	4.65	94.61	2.43

Table 7.7 presents the results for various multi-modal fusion schemes. It can be observed that the proposed system based on multi-modal biometric characteristics is robust in terms of recognition accuracy. The dynamic weighting fusion scheme provides best performance with an EER of 0.43% under assisted acquisition and an EER of 0.68% for data obtained in the self acquistion mode. Further, Figure 7.7 presents the ROC plots for multimodal biometric performance using dynamic weighted fusion scheme for data captured from Samsung S5 smartphone.

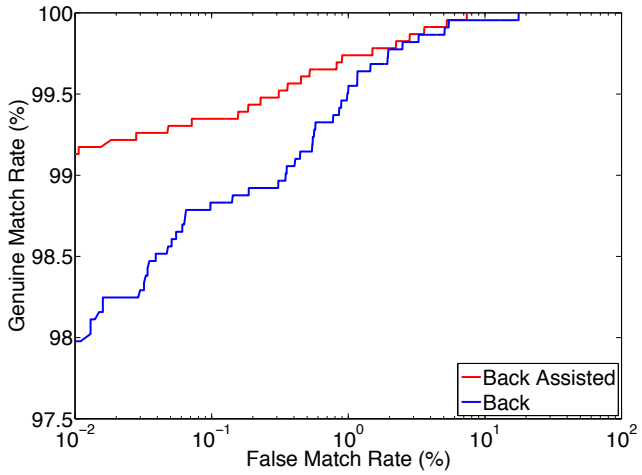


Figure 7.7: ROC plots for recognition based on dynamically weighted multi-modal fusion for Samsung S5 smartphone

7.5 Discussion

In this chapter, a multi-biometric authentication system specifically designed to work on smartphones is presented. The performance of uni-modal approach and multi-modal approach was demonstrated systematically with a set of experiments. The recorded performance indicates the feasibility of using such a system for authentication purposes. Further, the benchmarking with the data captured by a trained expert in similar settings of self-capture indicate that there is no drastic change in performance of the system validating the usability.

The set of experiments explored for fusion of comparison scores indicate the improvement of biometric performance. In the direction of this work, future works should explore the data captured in extremely different lighting conditions such as enrolment in regular illumination and verification attempt in harsh illumination. Another important direction is to capture the data in non-cooperative manner where partial face is available for verification and validate the proposed multi-biometric system.

7.5.1 Remarks on Execution Time

The speed of the prototype for proposed system implemented on Android platform (OS - KitKat 4.4) from capture to verification is based on the factors including the capture time of face image. Significant time (2-3 seconds) is spent for the capture of the face image. Once the face image is captured, the feature extraction and feature comparison happens in milliseconds. Further, the fusion of comparison scores is executed in fraction of milliseconds. The averaged time for each of the different operation is provided in the Table 7.8.

Table 7.8: Execution times for different operation in multi-modal biometric system

Operation		Time	
		Samsung Galaxy S5	Sasmung Note 10.1
Face Capture		$<$ 2 Seconds	$<$ 2 Seconds
Feature Extraction	SIFT	0.35 Seconds	0.38 Seconds
	SURF	0.32 Seconds	0.38 Seconds
	BSIF	0.26 Seconds	0.3 Seconds
Compariosn	SIFT	0.11 Seconds	0.13 Seconds
	SURF	0.16 Seconds	0.19 Seconds
	BSIF	0.08 Seconds	0.09 Seconds
Fusion		0.02 Seconds	0.02 Seconds

7.6 Conclusion

In this chapter, we have explored multi-biometric authentication framework for authentication on smartphones, specifically, fully realized and implemented on smartphones. Most of the earlier works have employed data from smartphones to carry out the recognition on desktop platforms, this chapter has illustrated the entire system on the smartphones. Further, this chapter has demonstrated through a series of experiments, the superior performance viable for many smartphone (tablet) devices which are capable of capturing the images of sufficient quality. In the first set of experimental protocols, uni-modal verification is investigated with face, left and right periocular region along with left and right iris region. The practical viability of using iris alone in visible spectrum was reiterated in this chapter which indicated lower performance due to low texture visibility and segmentation errors.

The best unimodal verification rate among all the experiments was obtained for face based authentication with GMR of 94.39% for Samsung S5 smartphone and GMR of 96.65% for Samsung Note tablet. Availability of three modalities such as face, periocular (both left and right) and iris (both left and right) has motivated us to perform score level fusion. The best verification rate of 94.91% GMR is obtained for smartphone and GMR of 96.58% is obtained for tablet. We have also explored comparison score level fusion using the scores obtained from all the features. The dynamically weighted score level fusion has provided an optimal performance with a GMR of 97.98% and 97.91% for smartphone and tablet respectively. The obtained performance is slightly lower ($\lessapprox 2\%$) than performance obtained

with the multi-biometric system when the images are acquired by trained expert indicating the usability under semi-cooperative environments.

The obtained performance serves as a good indicator of the proposed multi-biometric system for secure applications alternative to multi-factor authentication system. Possible future works should investigate the non-cooperative data capture within the system and fine-tune the algorithms to handle such unconstrained data. Additionally, future works can investigate better fusion approaches to handle missing scores due to missing characteristics such as non-segmented iris as a result of non-cooperative data capture. Handling such missing scores to improve the verification accuracy results in robust biometric systems.

Chapter 8

Conclusion

Advancements in the camera with superior quality optics on smartphone has provided a platform to capture biometric data in a contactless manner for authentication applications. In this thesis, we have evaluated and demonstrated the applicability of iris and periocular biometric data captured from smartphone in visible spectrum for authentication purposes. Further, we have presented a multi-biometric system for robust authentication on smartphones using iris, periocular and face characteristics. Based on the work presented together with the experimental validation, the following specific conclusions can be drawn which are listed below:

1. The challenge in adopting visible iris recognition on smartphone can be attributed to factors that include out-of-focus images, motion blur, partial availability of iris due to closure of eye and heavy pigmentation density (collagen fibrils and melanin) in iris. In addition, capturing the images by holding the smartphone at different distance from face and the way of interaction with the device by different users results in iris images which exhibit varying iris-pupil radius. As most of the segmentation schemes for iris recognition work in the known range of iris-pupil radius, the segmentation remains a challenge. In this thesis, we have addressed the challenge by estimating the radius using saliency based approach to localize coarse iris boundary that has improved segmentation accuracy as compared to standard OSIRIS v4.1.

   - The proposed technique has resulted in 70% accurate segmentation on average for MICHE-I database and further improved the EER by 5.8% for the data captured from different smartphones in different capture conditions.
   - On the VSSIRIS database, the method has resulted in 80% accurate segmentation while lowering the EER by 6.12% for images captured with different smartphones.

   This thesis has contributed to improve the open source OSIRIS v4.1 segmentation scheme by overcoming the need of manually providing the radius of initial estimate of iris boundary.

2. The iris features extracted need to be discriminative and robust enough to achieve good verification performance due to number of challenges listed above which results in partial iris availability in many cases. This thesis has contributed a new feature extraction scheme based on *deep sparse filtering* to extract robust features for reliable performance. The *deep sparse filtering* has experimentally demonstrated the improved performance on two different public databases as compared to performance obtained using other state-of-art techniques. Higher verification accuracy with $EER = 7.75\%$ was achieved with the new feature extraction scheme on VSSIRIS database on average for all the images captured using different phones. Further, a gain of 2% was obtained on EER for different datasets over other state-of-art methods on MICHE-I database.

   - This thesis has contributed to open-source feature extraction scheme to promote reproducibility and re-usability in biometric research community. The code for this can be availed for academic use together with the publication[76] or at: `www.nislab.no/biometrics_lab/code/deepsparse_iris`.

- The newly created iris database in visible spectrum for academic research from www.nislab.no/biometrics_lab/vssiris_db.

3. Visibility of texture is directly dependent on the pigmentation density in iris and higher pigmentation leads to low texture visibility in images when captured in visible spectrum, specifically on smartphones. Lower texture visibility in the captured image directly impacts the verification performance by resulting in lower values. In this thesis, we have presented an imaging set-up using a simple white LED to mimic the LED on the smartphones. The images captured using the LED based illumination have shown clear texture pattern and further resulted in good verification performance which is comparable to the performance obtained from iris images captured in NIR for corresponding subjects with high pigmentation density. The images obtained using the LED illuminated acquisition has demonstrated a generalized EER of 6.74% with a GMR of 91.86% at $FMR = 10^{-2}$ for the images captured with Nokia Lumia 1020 smartphone which is comparable to the same set of images captured with NIR device.

4. Although iris recognition in visible spectrum is promising with a good proportion of population with light colored and mildly pigmented iris, the challenges remain open when iris is used for larger population with high pigmentation density without additional illumination. Thus, in this thesis, we have demonstrated the use of periocular characteristics as alternative mode to authenticate using the data captured from smartphones in visible spectrum. The unconstrained periocular data with smartphones can be used optimally if the robust and discriminant features are extracted. Thus, in this thesis, we have proposed to employ time-frequency features of *deep sparse filter* response referred as $DeSTiFF$ for periocular images. The significance of new feature extraction technique ($DeSTiFF$) is validated through a series of experiments on two publicly available databases along with newly constructed periocular database - $ViSPer$ in the course of this thesis. The $ViSPer$ database is made available for the non-profitable research purposes. High verification of $GMR = 99.8\%$ was obtained on ViSPer database and close to 100% GMR was obtained at $FMR = 0.1\%$ on MICHE-I database signifying the applicability of new feature descriptors. The obtained verification accuracy of $GMR = 98\%$ at $FMR = 10^{-2}$ on large scale VISOB database further exemplifies the applicability of periocular recognition for smartphone captured data. A key observation is to note the degradation in verification performance when the data is captured in harsh illumination.

   - The preliminary version of this feature extraction was submitted to challenge session on periocular recognition in visible spectrum held in conjunction with ICIP-2016. The results obtained on the large scale database captured in non-standard and harsh illumination has demonstrated superior performance compared to state-of-art schemes [69].
   - $VisPer$ database is distributed freely to promote reproducible research and can be availed from: www.nislab.no/biometrics_lab/visper_db.
   - The implementation of $DeSTiFF$ feature extraction scheme to obtain robust and discriminant features for periocular images can be obtained from: www.nislab.no/biometrics_lab/code/destiff_periocular.

5. Further, driven by the fact that face can be captured to obtain both iris and periocular data with significant details, we have presented a multi-biometric authentication system using face, iris and periocular recognition. Based on the independent systems operating with iris and periocular, we have employed the state-of-art techniques to realize a multi-biometric authentication system using face, iris and periocular. The presented biometric system is designed to work on smartphones (including tablets) which are inherently limited by low computational power as compared to regular

desktop computing platforms. Performance derived from multi-biometric systems on smartphone platforms has shown acceptable biometric performance resulting in 99% GMR at $FMR = 10^{-2}$ by fusing the comparison scores of different biometric characteristics.

6. The biometric systems can be deemed reliable under the condition that they are not prone to attacks at various levels of operation. The systems can be attacked easily at the capture level, specifically for non-supervised data capture devices as in the case of smartphones. This thesis contributes robust algorithms to address the presentation attacks on the smartphone platforms for authentication systems employing ocular characteristics. The set of experiments on two public databases has resulted in 0% classification error of bona fide and attack presentation indicating the robustness of presented techniques to prevent the artefact data from being accepted into biometric system. The use of micro-texture features in different scale spaces has indicated promising feature in artefact classification.

    - This thesis contributes to promote reproducible research by disseminating the presentation attack database (PAVID) collected during this thesis and be availed from: www.nislab.no/biometrics_lab/pavid_db.
    - This thesis also distributes the implementation of PAD algorithms that can be obtained from: www.nislab.no/biometrics_lab/code/pad_lpfr_lachp.

8.0.1 General Conclusions

Smartphones are being used as an authentication unit for commercial applications such as banking and e-commerce applications using biometric data [148, 48]. In this thesis, iris, periocular and face characteristics captured using smartphone embedded cameras are explored for biometric authentication. Based on the experimental work carried out in this thesis, the following general conclusions are derived as listed in this section.

- Even though iris recognition is challenging in visible spectrum, especially using the data captured from smartphones, it can still be realized by devising robust feature extraction schemes. The state-of-art techniques for feature extraction specifically tailored for NIR spectrum do not provide good performance in visible spectrum due to low visibility of texture pattern and higher pigmentation density. The low performance can be attributed to high number of false matches in unconstrained iris data in visible spectrum. Robust methods such as *deep sparse filtering* proposed in this work has shown the effectiveness to reduce such false matches for visible spectrum iris recognition even under low visibility of texture. Texture extraction filters learnt using more sophisticated approaches of machine learning such as *deep sparse filtering* are expected to obtain discriminant features to yield improved or efficient results.

- The challenge of iris recognition remains open for heavily pigmented iris where innovative illumination in visible spectrum needs to be employed. Alternatively, periocular information captured in visible spectrum can be employed for authentication applications. Texture based features like $BSIF$ and key-point based descriptors like $SIFT$ and $SURF$ with simple comparison schemes perform well when the periocular data is captured in semi-cooperative manner while robust feature extraction schemes based on *deep sparse filtering* are required to achieve good verification performance, especially when the data is captured in harsh illumination conditions.

- Rather than employing smartphone to capture biometric data alone, biometric authentication system can itself be realized on computationally limited devices such as smartphones which can be used in everyday authentication applications. Multi-biometric systems have demonstrated good biometric performance when face,

periocular and iris characteristics are used as illustrated in this thesis. Simple features like $BSIF$, $SIFT$ and $SURF$ with score level fusion is promising for the path forward in smartphone based authentication in visible spectrum for real-life applications. The key factor to be noted is that such multi-biometric systems can provide alternate ways of engaging biometric authentication with no necessity for upgrades to have devices with integrated sensors.

- Despite the fact that systems are vulnerable to presentation attacks as the data is captured in non-supervised manner on smartphones, reliable counter-measures to detect the attacks assure security of the systems. Enhancing phase component with video magnification approaches like EVM can be used to detect electronic screen attacks. Texture features obtained using $STFT$ and AHP techniques across *Laplacian* scales can achieve good results in attack detection. Based on the experimental results in this thesis, it can be generalized that micro-texture features are highly applicable for PAD mechanisms to detect print attack and electronic screen attacks using image or video.

8.1 Future Works

Iris Recognition

- The set of filters learnt using *deep sparse filtering* in this thesis is based on the set of natural images.
  - The future works can investigate on learning the filters using domain specific data i.e., iris image data captured using smartphones in visible spectrum to bring out the merits as compared to filters learnt using natural images.
  - Unlike the current thesis, future works can investigate on choosing best filters from the pool of 256 filters to reduce number of features. Also, better grouping strategies may be devised instead of sequential grouping of filter responses.

  Further, as iris data obtained in visible spectrum is impacted by number of factors as compared to data from NIR spectrum, the robust feature classification are carried out using sparse representation in this thesis. The key challenge is the computational expense to adapt the sparse representation on smartphones where computational resources are limited. Future works in this direction can investigate on optimizing the sparse representation classifier for smartphone platforms which are typically limited in computational capacity [132].

- With regard to the texture visibility of heavily pigmented iris, the LED based illumination discussed in this thesis has provided promising results with set of experiments on a database of 31 subjects. Future works should consider to collect large-scale database to benchmark the performance against NIR spectrum to validate the biometric performance of iris characteristics in visible spectrum.

- The new generation of smartphones are now enabled with dual LED flash that emit twice as much light as a single LED of the same type which results in illuminating the subject 1.4 times further away than normal distance. The dual LED have two sources of light with different color temperatures where one of the LED serves as ambient light for the second LED. These features of dual LED can be explored for iris recognition in visible spectrum, specially for unconstrained iris capture where ambient light balance can be used to decide the intensity of illumination from other LED to obtain optimal iris texture.

- Another key work in this direction should study the quality factors for visible spectrum iris images captured using smartphones in similar lines of earlier work [105]. The

visibility of the texture plays vital role for iris recognition in visible spectrum, especially when the iris are dark colored (highly pigmented). The performance should be studied with respect to different quality of visible texture within the iris boundary. The common problems of partially closed eye-lids due to day-light illumination and the dilation of pupil due to uncontrolled illumination can result in low amount of usable iris texture that can actually be used for recognition purpose. Future works can thus establish acceptable texture, acceptable ratio of iris-pupil area and texture quality which can thereby establish set of quality metrics for using iris captured from smartphone embedded cameras in visible spectrum.

Periocular and Face Recognition

- The periocular recognition algorithms presented in this thesis have demonstrated robustness to the noisy and degraded images. While a class of features based on simple filters and distance metric have proven reliable in semi-cooperative data capture, robust feature extractors such as *deep sparse filtered* features represented in collaborative way are needed for heavily degraded data captured in harsh illumination. Discriminant feature descriptors and robust classification schemes are necessary to deal with heavily degraded and large scale data. Although, a strong feature descriptor is provided in this work, the future works in this direction should focus on adapting robust classification schemes on smartphone platforms [132]. They should implement and optimize the collaborative classification in a light weight fashion to make the periocular biometric systems robust in handling non-standard data.

- The challenges in using the face recognition arise from the way a particular subject interacts with the device to capture the data in the smartphone environment which results in partial face availability [88]. Thus, it is essential to study the biometric performance when the face image captured in the smartphone presents partial face and face with non-uniform illumination on a part of the face[88]. Future works should evaluate the robust algorithms proposed for periocular region using *deep sparse filtering* in this thesis for unconstrained face recognition on smartphone platforms.

- Further, the traditional face recognition systems have used well adopted quality metrics in automatic face recognition. When the face data is captured in a semi-cooperative manner, acceptable quality of image can be expected. However, when the face image is captured in unconstrained manner as in the case of smartphones, it is essential to decide the quality of image to be used in biometrics [88]. Thus, the future work in this direction should investigate on the quality factors for face recognition in unconstrained data capture on smartphones to discard non-standard data or devise robust algorithms to handle them.

Presentation Attack Detection

- Presentation attack detection in the current state-of-art techniques are tuned specifically to different datasets with disjoint training and testing set. The success of the techniques are largely based on learning good decision boundary by using good features with classifiers such as SVM or $SRDA$. However, these classifiers are expected to fail when an artefact other than the ones used for training the classifier itself are presented. Thus, the future works should investigate on achieving cross-database presentation attack detection by testing on different database than the one used for training. Additionally, the classifiers can be extended to discriminate n different class of artefacts to account for different attack potentials.

- Another future work in this direction can be towards removing the dependency on classifiers like SVM or $SRDA$ and instead look for inherent image characteristics to determine the attack artefacts without the explicit use of classifiers. This approach would make the PAD mechanisms universal instead of relying on learnt classifier model.

As a closing remark, the promising and upcoming field of authentication on the smartphones for secure applications via biometrics provides a number of opportunities.

Appendix

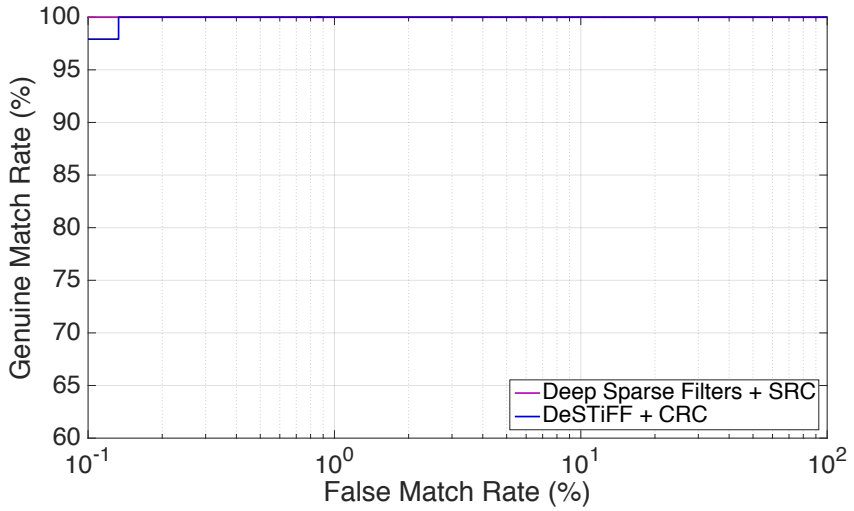
Appendix A

Deep Sparse Time Frequency Features for Iris Recognition

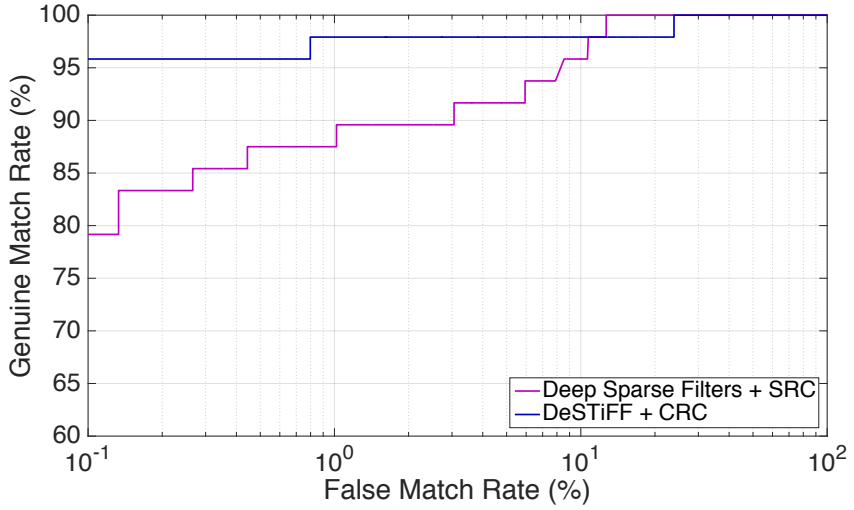
This appendix lists the performance of *Deep Sparse TimeFrequency Features* for iris recognition carried out on MICHE-I dataset and VSSIRIS database (detailed in Section 3).

A.1 Performance of DeSTiFF on MICHE-I Database

This section presents the results of *DeSTiFF* feature extraction on the MICHE-I iris database and compares it to performance of *Deep Sparse Filters*. In order to provide a fair comparison, the same set of protocols are followed as mentioned in the Section 3.5.3. The performance is listed as per the illumination and smartphone camera. Figure A.1 presents the performance of *Deep Sparse Filtering* along with *DeSTiFF* for the images captured using iPhone in the indoor illumination. Figure A.2 presents the performance of two methods for the images captured using iPhone in the outdoor illumination. Similarly, Figure A.3 and Figure A.4 present the comparative performance for images captured using Samsung phone under indoor and outdoor illumination respectively. It can be noted from the figures that, *DeSTiFF* performs better in terms of GMR at lower FMR indicating the superiority over the *Deep Sparse Filtering* method.

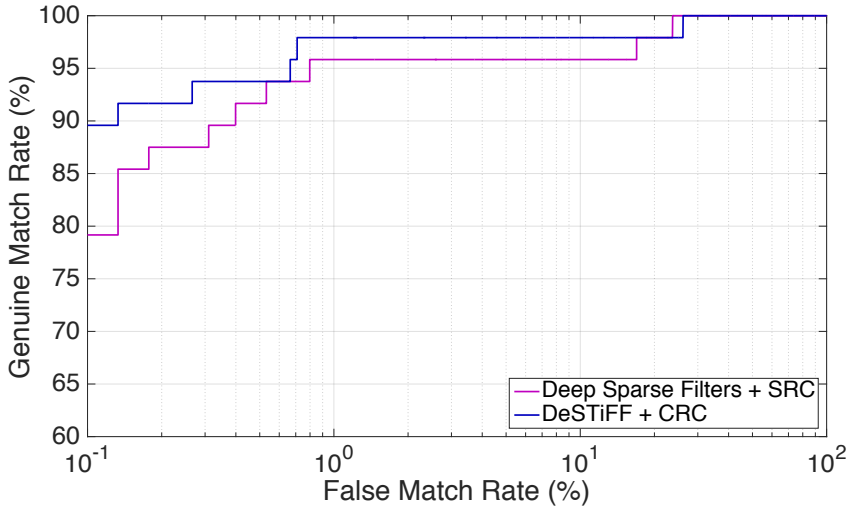


(a) iPhone-Indoor-Frontal camera

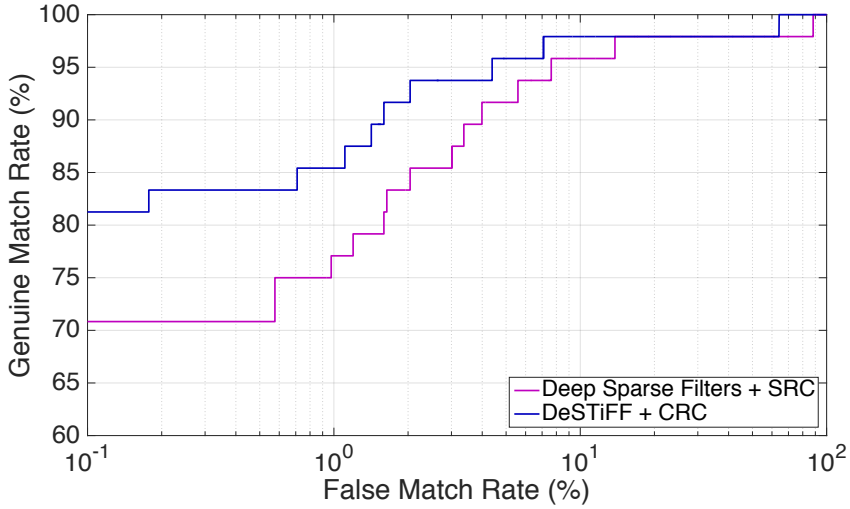


(b) iPhone-Indoor-Rear camera

Figure A.1: ROC curves obtained for various schemes applied on the iPhone images from MICHE-I database (Indoor illumination) [34].

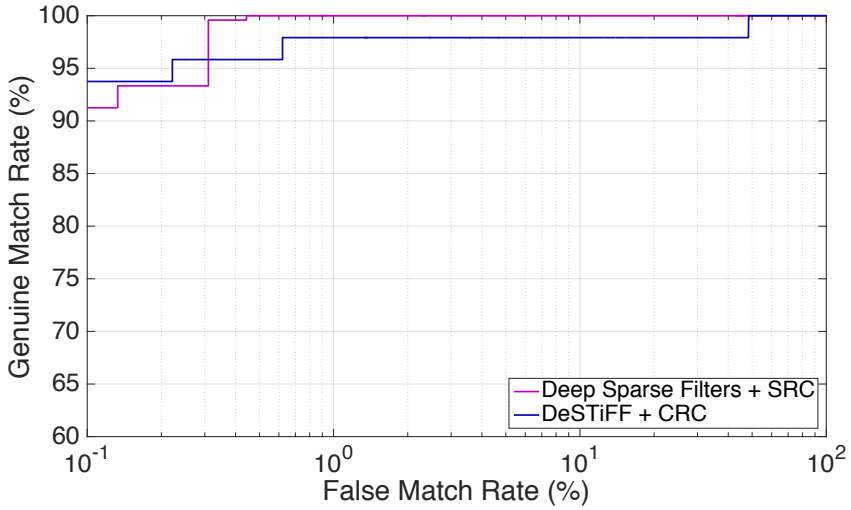


(a) iPhone-Outdoor-Frontal camera

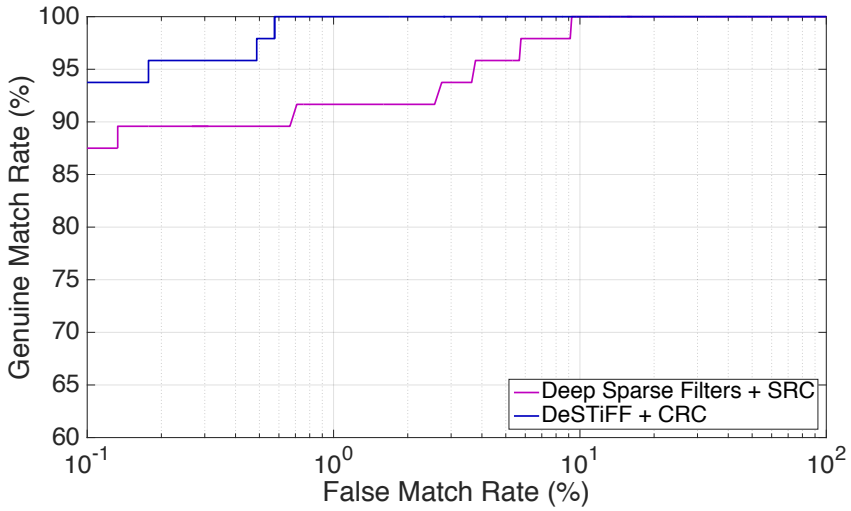


(b) iPhone-Outdoor-Rear camera

Figure A.2: ROC curves obtained for various schemes applied on the iPhone images from MICHE-I database (Outdoor illumination) [34].

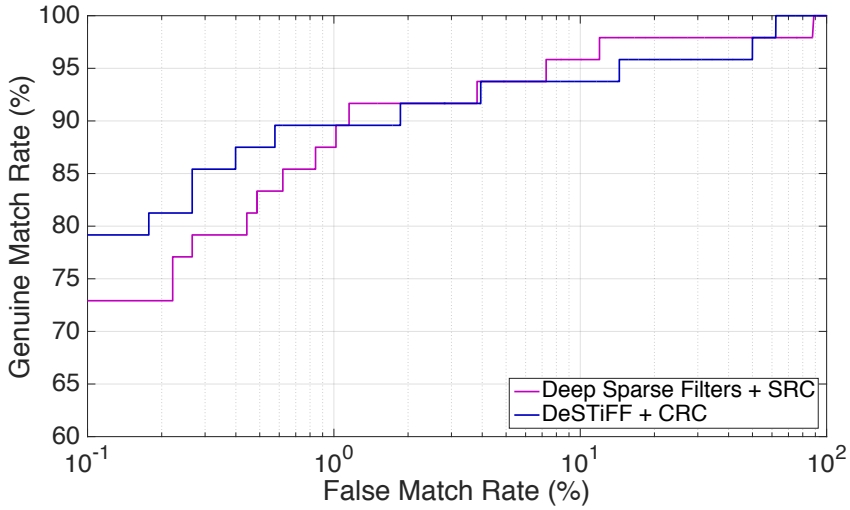


(a) Indoor-Frontal camera

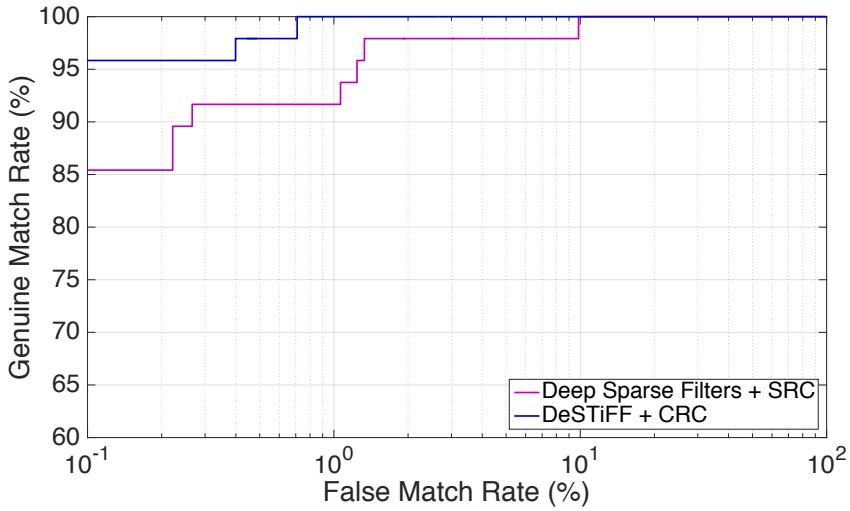


(b) Indoor-Rear camera

Figure A.3: ROC curves obtained for various schemes applied on the Samsung images from MICHE-I database (Indoor illumination)[34].



(a) Outdoor-Frontal camera

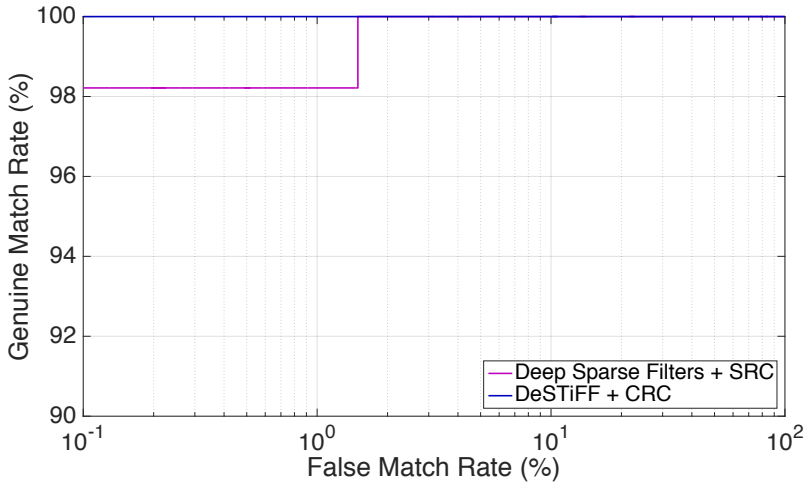


(b) Outdoor-Rear camera

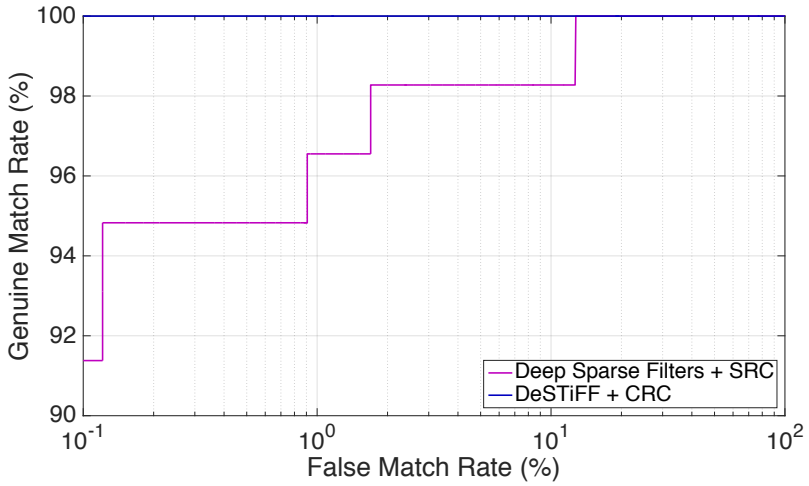
Figure A.4: ROC curves obtained for various schemes applied on the Samsung images from MICHE-I database (Outdoor illumination) [34].

A.2 Performance of DeSTiFF on VSSIRIS Database

This section presents the comparative performance of *DeSTiFF* and *Deep Sparse Filters* for iris recognition in visible spectrum using VSSIRIS database [76]. We follow the protocols mentioned in Section 3.5.4. It can be observed from Figure A.5, *DeSTiFF* performs robustly for the VSSIRIS database even at the lower FMR for images captured using both iPhone 5S and Nokia Lumia 1020 smartphones.



(a) iPhone 5S images from the VSSIRIS database



(b) Nokia Lumia 1020 images from the VSSIRIS database

Figure A.5: ROC curves obtained for various schemes applied on the VSSIRIS database

Appendix B

Deep Sparse Time Frequency Features for Iris Acquired Using LED

This appendix lists the performance of *Deep Sparse TimeFrequency Features* (*DeSTiFF*) for iris recognition using the iris database captured using white-LED as detailed in Chapter 4. We employ the images captured from 31 subjects from three different smartphones (iPhone 5S, Nokia Lumia 1020 and Samsung S4) using the proposed LED imaging set-up as described in Chapter 4. The experimental protocols are followed as detailed in Section 4.4.1. Figure B.1 presents the verification performance of *DeSTiFF* compared along with other state-of-art schemes. It can be noted that *DeSTiFF* performs equally well compared to state-of-art methods for images from iPhone and Nokia. Further, *DeSTiFF* outperforms the state-of-art methods for the images captured using Samsung smartphone.

(a) iPhone Iris Dataset

(b) Nokia Iris Dataset

(c) Samsung Iris Dataset

Figure B.1: ROC curves depicting the performance of various algorithms of iris recognition for the data captured using LED based approach.

Appendix C

Deep Sparse Filters for Periocular Recognition

This appendix lists the performance of *Deep Sparse Filters* for large scale periocular recognition carried out on VISOB database (detailed in Section 5.4.3). The protocols for this set of experiments are detailed in Section 5.5.3.1. Figure C.1 presents the performance of *Deep Sparse Filters* on all the images captured using iPhone. Similarly, Figure C.2 presents the performance of all images captured using Oppo smartphone while Figure C.3 presents performance for images captured from Samsung. The detailed comparison of the *Deep Sparse Filters* with other state-of-art techniques can be obtained from [69].

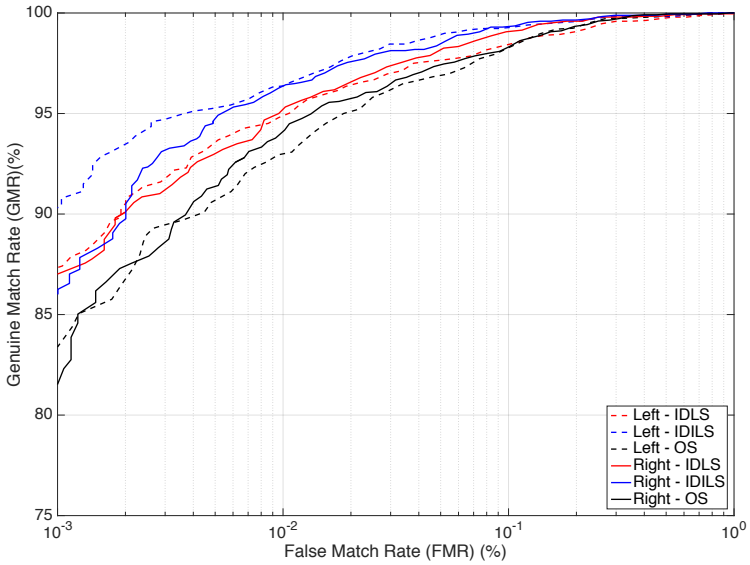


Figure C.1: ROC curves for depicting the performance of the *deep sparse filters* for periocular data captured using iPhone.

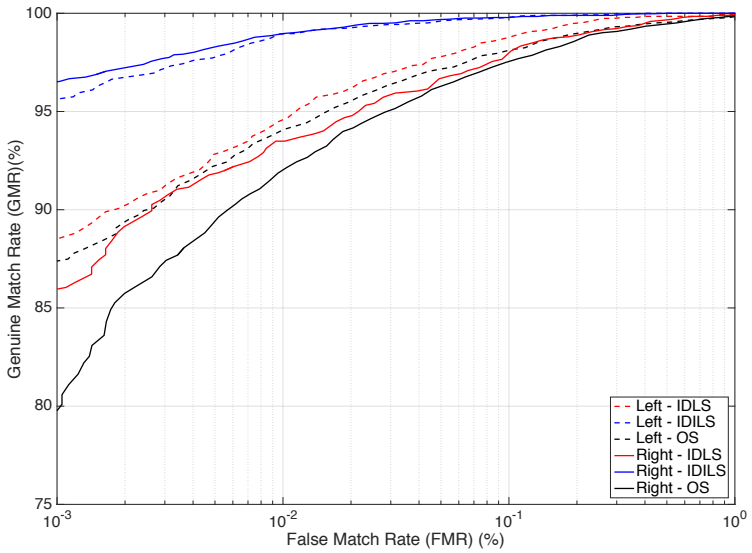


Figure C.2: ROC curves for depicting the performance of the *deep sparse filters* for periocular data captured using Oppo.

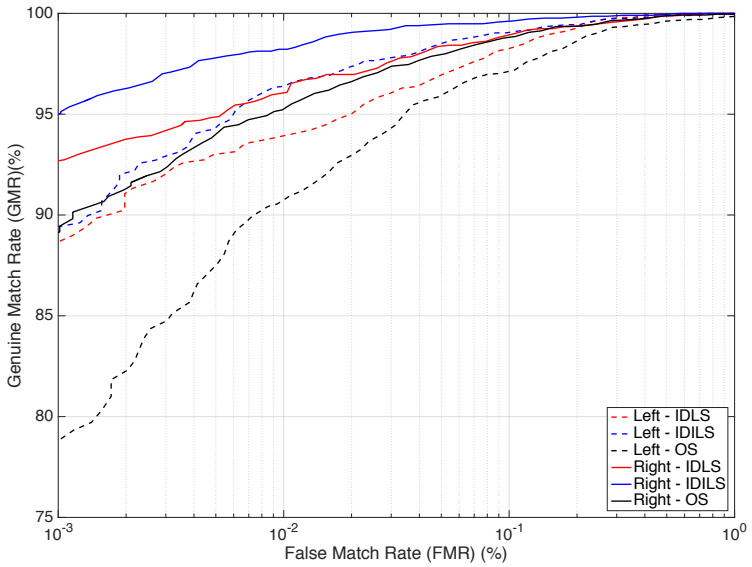


Figure C.3: ROC curves for depicting the performance of the *deep sparse filters* for periocular data captured using Samsung.

Appendix D

Deep Sparse Time Frequency Features for Multi-biometrics (Face and Periocular)

This appendix lists the performance of *Deep Sparse TimeFrequency Features* (*DeSTiFF*) for face recognition carried out on multi-biometric database collected using smartphone (detailed in Section 7.3 of Chapter 7). The protocols for this set of experiments are in-line with the protocols mentioned in Section 7.4.1. Figure D.1 presents the verification performance of face using different schemes for the images captured using Samsung Galaxy S5 smartphone and Samsung Note 10.1 tablet in self capture and assisted capture scenario. Figure D.2 presents the verification performance of periocular region using the images captured from smartphone. Similarly, Figure D.3 presents the performance of periocular region for the images captured from tablet. It has to be noted that *DeSTiFF* performs very good in lower FMR indicating the robustness of the extracted features.

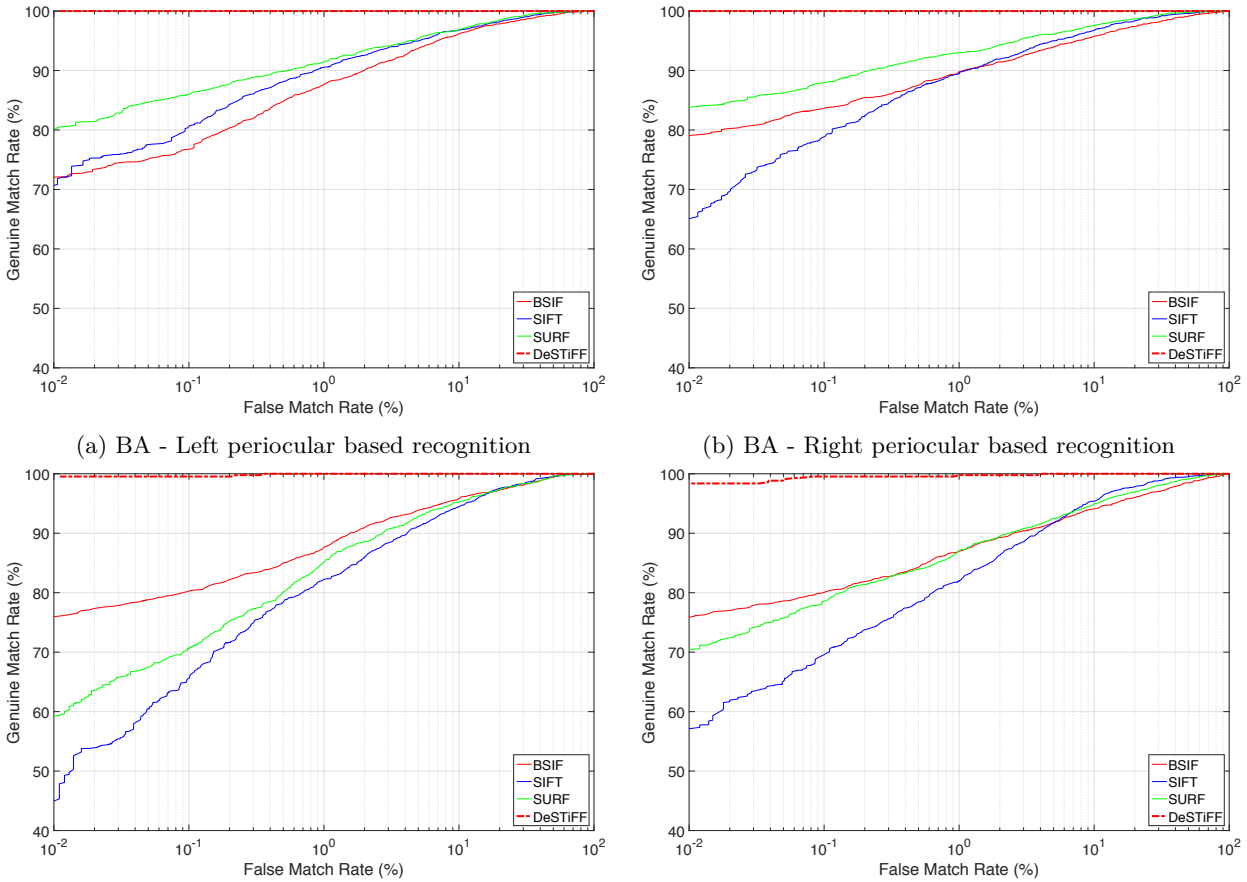
(a) BA - Face based recognition in smartphone

(b) BS - Face based recognition in smartphone

(c) BA - Face based recognition in tablet

(d) BS - Face based recognition in tablet

Figure D.1: ROC curves for various unimodal recognition employing face characteristics; (a) & (c) correspond to assisted acquisition using the back camera; (d)-(f) correspond to self acquisition using the back camera;
*BA - Assisted acquisition from back camera, *BS - Self acquisition from back camera



(a) BA - Left periocular based recognition

(b) BA - Right periocular based recognition

(c) BS - Left periocular based recognition

(d) BS - Right periocular based recognition

Figure D.2: ROC curves for various unimodal recognition employing periocular characteristics on Samsung S5; (a)-(c) correspond to assisted acquisition using the back camera; (d)-(f) correspond to self acquisition using the rear(back) camera; *BA - Assisted acquisition from rear camera, *BS - Self acquisition from rear camera

(a) BA - Left periocular based recognition

(b) BA - Right periocular based recognition

(c) BS - Left periocular based recognition

(d) BS - Right periocular based recognition

Figure D.3: ROC curves for various unimodal recognition employing face and periocular characteristics on Samsung Note 10.1 tablet; (a)-(c) correspond to assisted acquisition using the back camera; (d)-(f) correspond to self acquisition using the back camera; *BA - Assisted acquisition from back camera, *BS - Self acquisition from back camera

Bibliography

[1] Abate, A. F., Frucci, M., Galdi, C., and Riccio, D. Bird: watershed based iris detection for mobile devices. *Pattern Recognition Letters 57* (2015), 43–51.

[2] Adams, A., and Sasse, M. A. Users are not the enemy. *Communications of the ACM 42*, 12 (1999), 40–46.

[3] Adler, F. Physiology of the eye, mosby, st. *Louis, Mo* (1965).

[4] Ahonen, T., Hadid, A., and Pietikainen, M. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence 28*, 12 (2006), 2037–2041.

[5] Anjos, A., Chakka, M. M., and Marcel, S. Motion-based counter-measures to photo attacks in face recognition. *Institution of Engineering and Technology - Biometrics* (Apr. 2013).

[6] Anjos, A., and Marcel, S. Counter-measures to photo attacks in face recognition: a public database and a baseline. In *2011 International Joint Conference on Biometrics (IJCB)* (2011), IEEE, pp. 1–7.

[7] Arashloo, S. R., Kittler, J., and Christmas, W. Face spoofing detection based on multiple descriptor fusion using multiscale dynamic binarized statistical image features. *IEEE Transactions on Information Forensics and Security 10*, 11 (2015), 2396–2407.

[8] Barra, S., Casanova, A., Narducci, F., and Ricciardi, S. Ubiquitous iris recognition by means of mobile devices. *Pattern Recognition Letters* (2014).

[9] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. Speeded-up robust features (surf). *Computer vision and image understanding 110*, 3 (2008), 346–359.

[10] Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., et al. Greedy layer-wise training of deep networks. *Advances in neural information processing systems 19* (2007), 153.

[11] Bharadwaj, S., Bhatt, H. S., Vatsa, M., and Singh, R. Periocular biometrics: When iris recognition fails. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on* (2010), IEEE, pp. 1–6.

[12] Bharadwaj, S., Dhamecha, T. I., Vatsa, M., and Singh, R. Computationally efficient face spoofing detection with motion magnification. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2013), IEEE, pp. 105–110.

[13] Bicego, M., Lagorio, A., Grosso, E., and Tistarelli, M. On the use of sift features for face authentication. In *Conference on Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06.* (2006), IEEE, pp. 35–35.

[14] BIPLab, University of Salerno. Mobile Iris CHallenge Evaluation (MICHE I and II). http://biplab.unisa.it/MICHE/database/, 2013.

[15] Bowyer, K. W. The results of the nice. ii iris biometrics competition. *Pattern Recognition Letters 33*, 8 (2012), 965–969.

[16] Bowyer, K. W., Hollingsworth, K. P., and Flynn, P. J. A survey of iris biometrics research: 2008–2010. In *Handbook of iris recognition*. Springer, 2013, pp. 15–54.

[17] Boyce, C., Ross, A., Monaco, M., Hornak, L., and Li, X. Multispectral iris analysis: A preliminary study. In *2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)* (2006), IEEE, pp. 51–51.

[18] Bradski, D. G. R., and Kaehler, A. *Learning Opencv, 1st Edition*, first ed. O'Reilly Media, Inc., 2008.

[19] Brown, M., and Lowe, D. G. Recognising panoramas. In *ICCV* (2003), vol. 3, p. 1218.

[20] Burge, M. J., and Bowyer, K. *Handbook of iris recognition*. Springer Science & Business Media, 2013.

[21] Burt, P. J., and Adelson, E. H. The laplacian pyramid as a compact image code. *Communications, IEEE Transactions on 31*, 4 (1983), 532–540.

[22] Cai, D., He, X., and Han, J. Srda: An efficient algorithm for large-scale discriminant analysis. *IEEE transactions on knowledge and data engineering 20*, 1 (2008), 1–12.

[23] Cha, S.-H., and Srihari, S. N. On measuring the distance between histograms. *Pattern Recognition 35*, 6 (2002), 1355–1370.

[24] Chang, C.-C., and Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology 2* (2011), 27:1–27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[25] Cheng, M.-M., Mitra, N. J., Huang, X., Torr, P. H. S., and Hu, S.-M. Salient object detection and segmentation. Tech. rep., Tsinghua University, Oct 2011. Submission NO. TPAMI-2011-10-0753. Available from: http://mmcheng.net/salobj/.

[26] Chingovska, I. *Trustworthy Biometric Verification under Spoofing Attacks: Application to the Face Mode*. Ph.D. thesis, École Polytechnique Fédérale de Lausanne, 2015.

[27] Czajka, A. Database of iris printouts and its application: Development of liveness detection method for iris recognition. In *2013 18th International Conference on Methods and Models in Automation and Robotics (MMAR)* (Aug 2013), pp. 28–33.

[28] Daugman, J. How iris recognition works. In *Proceedings of International Conference on Image Processing 2002* (2002), vol. 1, IEEE, pp. I–33.

[29] Daugman, J. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology 14*, 1 (2004), 21–30.

[30] Daugman, J. Iris recognition and anti-spoofing countermeasures. In *7-th International Biometrics conference* (2004).

[31] Daugman, J. Results from 200 billion iris cross-comparisons. *University of Cambridge Technical Report UCAM-CL-TR-635* (2005).

[32] Daugman, J. Probing the uniqueness and randomness of iriscodes: Results from 200 billion iris pair comparisons. *Proceedings of the IEEE 94*, 11 (2006), 1927–1935.

[33] Daugman, J. G. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence 15*, 11 (1993), 1148–1161.

[34] De Marsico, M., Galdi, C., Nappi, M., and Riccio, D. Firme: face and iris recognition for mobile engagement. *Image and Vision Computing 32*, 12 (2014), 1161–1172.

[35] Dobeš, M., Martinek, J., Skoupil, D., Dobešová, Z., and Pospíšil, J. Human eye localization using the modified hough transform. *Optik-International Journal for Light and Electron Optics 117*, 10 (2006), 468–473.

[36] FAST PASS. Fastpass- a harmonized, modular reference system for all european automated border crossing points. https://www.fastpass-project.eu, 2012. Online; accessed 29 January 2014.

[37] Forney Jr, G. D. The viterbi algorithm. *Proceedings of the IEEE 61*, 3 (1973), 268–278.

[38] Frontex. Best Practice Technical Guidelines for Automated Border Control (ABC) Systems, 2015.

[39] Frucci, M., Nappi, M., Riccio, D., and di Baja, G. S. Wire: Watershed based iris recognition. *Pattern Recognition 52* (2016), 148–159.

[40] Galbally, J., Marcel, S., and Fierrez, J. Biometric antispoofing methods: A survey in face recognition. *IEEE Access 2* (2014), 1530–1552.

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

[42] Galbally, J., Ortiz-Lopez, J., Fierrez, J., and Ortega-Garcia, J. Iris liveness detection based on quality related features. In *2012 5th IAPR International Conference on Biometrics (ICB)* (2012), IEEE, pp. 271–276.

[43] Gottemukkula, V., Saripalle, S. K., Tankasala, S. P., Derakhshani, R., Pasula, R., and Ross, A. Fusing iris and conjunctival vasculature: Ocular biometrics in the visible spectrum. In *2012 IEEE Conference on Technologies for Homeland Security (HST)* (2012), IEEE, pp. 150–155.

[44] Goverment of India. Aadhaar-a unique identification number. *Research Cell: An International Journal of Engineering Science 4*, 2 (2011), 169–176.

[45] He, X., Lu, Y., and Shi, P. A fake iris detection method based on fft and quality assessment. In *Chinese Conference on Pattern Recognition, 2008. CCPR'08.* (2008), IEEE, pp. 1–4.

[46] Hess, R. An open-source siftlibrary. In *Proceedings of the 18th ACM international conference on Multimedia* (2010), ACM, pp. 1493–1496.

[47] Hinton, G. E., Osindero, S., and Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural computation 18*, 7 (2006), 1527–1554.

[48] Hoyos Labs. http://www.hoyoslabs.com/technology/#4F, 2016. Accessed: 2016-07-15.

[49] Hughes, K., and Bowyer, K. W. Detection of contact-lens-based iris biometric spoofs using stereo imaging. In *2013 46th Hawaii International Conference on System Sciences (HICSS)* (2013), IEEE, pp. 1763–1772.

[50] Hyvèarinen, A., Hurri, J., and Hoyer, P. O. *Natural Image Statistics*, vol. 39. Springer, 2009.

[51] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 2382-37:2012 Information Technology - Vocabulary - Part 37: Biometrics*. International Organization for Standardization, 2012.

[52] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC 30107-1. Biometric Presentation Attack Detection – Part 1: Framework*. Framework. International Organization for Standardization and International Electrotechnical Committee, August, 2016.

[53] ISO/IEC JTC1 SC37 Biometrics. *ISO/IEC DIS 30107-3. Biometrics-Presentation Attack Detection - Part 3*. Testing and Reporting, International Organization for Standardization and International Electrotechnical Committee, August, 2016.

[54] ISO/IEC TC JTC1 SC37 Biometrics. *ISO/IEC 19795-1:2006. Information Technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework*. International Organization for Standardization and International Electrotechnical Committee, Mar. 2006.

[55] Itti, L., Koch, C., Niebur, E., et al. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence 20*, 11 (1998), 1254–1259.

[56] Jain, A., Flynn, P., and Ross, A. A. *Handbook of biometrics*. Springer Science & Business Media, 2007.

[57] Jain, A. K., and Kumar, A. Biometrics of next generation: An overview. *Second Generation Biometrics 12*, 1 (2010), 2–3.

[58] Jain, A. K., and Ross, A. Multibiometric systems. *Communications of the ACM 47*, 1 (2004), 34–40.

[59] Jain, A. K., Ross, A., and Prabhakar, S. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology 14*, 1 (2004), 4–20.

[60] Jillela, R., Ross, A., Boddeti, N., Kumar, B. V., Hu, X., Plemmons, R., and Pauca, P. An evaluation of iris segmentation algorithms in challenging periocular images. *Handbook of Iris Recognition, Springer, Heidelberg (to appear, 2012)* (2002).

[61] Jillela, R. R., and Ross, A. Segmenting iris images in the visible spectrum with applications in mobile biometrics. *Pattern Recognition Letters 57* (2015), 4–16.

[62] Jones, III, C. F., and Abbott, A. L. Optimization of color conversion for face recognition. *EURASIP J. Appl. Signal Process. 2004* (Jan. 2004), 522–529. Available from: http://dx.doi.org/10.1155/S1110865704401073.

[63] Juefei-Xu, F., and Savvides, M. Unconstrained periocular biometric acquisition and recognition using cots ptz camera for uncooperative and non-cooperative subjects. In *Applications of Computer Vision (WACV), 2012 IEEE Workshop on* (2012), IEEE, pp. 201–208.

[64] Kannala, J., and Rahtu, E. Bsif: Binarized statistical image features. In *21st International Conference on Pattern Recognition (ICPR) 2012* (2012), IEEE, pp. 1363–1366.

[65] Kiran B. Raja. Robust Iris Recognition Using Light-field Camera. Master's thesis, Master Erasmus Mundus in Color in Informatics and Media Technology (CIMET), UJM-France and GUC-Norway, 2013.

[66] Kiran B Raja, Raghavendra, R., and Busch, C. 'iris imaging in visible spectrum using white led. In *2015 IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)* (2015), IEEE, pp. 1–8.

[67] Kiran B Raja, Raghavendra, R., and Busch, C. 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)* (2015), IEEE, pp. 1–8.

[68] Kiran B. Raja., Raghavendra, R., and Busch, C. Video presentation attack detection in visible spectrum iris recognition using magnified phase information. *IEEE Transactions on Information Forensics and Security PP*, 99 (2015), 1–1.

[69] Kiran B. Raja, Raghavendra, R., and Busch, C. Collaborative representation of deep sparse filtered feature for robust verification of smartphone periocular images. In *23rd IEEE International Conference on Image Processing (ICIP 2016).* (Oct. 2016), pp. 1 –5.

[70] Kiran B Raja, Raghavendra, R., and Busch, C. Color adaptive quantized patterns for presentation attack detection in ocular biometric systems. In *Proceedings of the 9th International Conference on Security of Information and Networks, SIN-2016, USA* (2016), ACM, pp. 1–8.

[71] Kiran B. Raja, Raghavendra, R., Cheikh, F. A., and Busch, C. Robust iris recognition using light field camera. In *IEEE conference on The Colour and Visual Computing Symposium 2013* (2013), IEEE.

[72] Kiran B. Raja, Raghavendra, R., Cheikh, F. A., Yang, B., and Busch, C. Robust iris recognition using light field camera. In *IEEE conference on The Colour and Visual Computing Symposium 2013* (2013), IEEE.

[73] Kiran B. Raja, Raghavendra, R., Stokkenes, M., and Busch, C. Smartphone authentication system using periocular biometrics. In *BIOSIG 2014 - Proceedings of the 13th International Conference of the Biometrics Special Interest Group, September 2014, Darmstadt, Germany* (2014), pp. 15–26.

[74] Kiran B. Raja, Raghavendra, R., Stokkenes, M., and Busch, C. Fusion of face and periocular information for improved authentication on smartphones. In *18th International Conference on Information Fusion (FUSION 2015), Washington DC, USA* (2015).

[75] Kiran B. Raja, Raghavendra, R., Stokkenes, M., and Busch, C. Multi-modal authentication system for smartphones using face, iris and periocular. In *IEEE International Conf. Biometrics (ICB), Phuket, Thailand* (2015).

[76] Kiran B. Raja, Raghavendra, R., Vemuri, V. K., and Busch, C. Smartphone based visible iris recognition using deep sparse filtering. *Pattern Recognition Letters 57*, 0 (2015), 33 – 42.

[77] Klare, B., and Jain, A. K. On a taxonomy of facial features. In *2010 Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)* (2010), IEEE, pp. 1–8.

[78] Ko, J.-G., Gil, Y.-H., Yoo, J.-H., and Chung, K.-I. A novel and efficient feature extraction method for iris recognition. *ETRI journal 29*, 3 (2007), 399–401.

[79] Kollias, N. The spectroscopy of human melanin pigmentation. In *Journal of Investigative Dermatology* (1994), vol. 102, BLACKWELL SCIENCE INC 238 MAIN ST, CAMBRIDGE, MA 02142, pp. 268–268.

[80] Krichen, E. Lef3a: Pupil segmentation using viterbi search algorithm. In *2012 5th IAPR International Conference on Biometrics (ICB)* (2012), IEEE, pp. 323–329.

[81] Lee, S. J., Park, K. R., and Kim, J. Robust fake iris detection based on variation of the reflectance ratio between the iris and the sclera. In *2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference* (2006), IEEE, pp. 1–6.

[82] Lowe, D. G. Object recognition from local scale-invariant features. In *The proceedings of the seventh IEEE international conference on Computer vision, 1999.* (1999), vol. 2, Ieee, pp. 1150–1157.

[83] Lowe, D. G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision 60*, 2 (2004), 91–110.

[84] Ma, L., Tan, T., Wang, Y., and Zhang, D. Personal identification based on iris texture analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence 25*, 12 (2003), 1519–1533.

[85] Ma, L., Wang, Y., and Tan, T. Iris recognition using circular symmetric filters. In *16th International Conference on Pattern Recognition (ICPR) 2002* (2002), vol. 2, pp. 414–417.

[86] Maatta, J., Hadid, A., and Pietikainen, M. Face spoofing detection from single images using micro-texture analysis. In *2011 international joint conference on Biometrics (IJCB)* (2011), IEEE, pp. 1–7.

[87] Maatta, J., Hadid, A., and Pietikainen, M. Face spoofing detection from single images using texture and local shape analysis. *IET Biometrics 1*, 1 (2012), 3–10.

[88] Mahbub, U., Patel, V. M., Chandra, D., Barbello, B., and Chellappa, R. Partial face detection for continuous authentication. *arXiv preprint arXiv:1603.09364* (2016).

[89] Marcel, S., Cool, C. M., Atanasoaei, C., Tarsetti, F., Pesán, J., Matejka, P., Cernocky, J., Helistekangas, M., and Turtinen, M. MOBIO: Mobile biometric face and speaker authentication. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition* (San Francisco, CA, USA, 2010).

[90] Martin, A., Doddington, G., Kamm, T., Ordowski, M., and Przybocki, M. The det curve in assessment of detection task performance. Tech. rep., DTIC Document, 1997.

[91] Masek, L., and Kovesi, P. Matlab source code for a biometric identification system based on iris patterns. *The School of Computer Science and Software Engineering, The University of Western Australia 2*, 4 (2003).

[92] MorphoTrust USA. http://www.morphotrust.com/IdentitySolutions/ForFederalAgencies/Officer360/ArrestandCustody360/Mobile-Eyes.aspx, 2015. Accessed: 2015-07-03.

[93] Muja, M., and Lowe, D. G. Fast approximate nearest neighbors with automatic algorithm configuration. In *VISAPP (1)* (2009), pp. 331–340.

[94] Ngiam, J., Koh, P. W., Chen, Z., Bhaskar, S. A., and Ng, A. Y. Sparse filtering. In *NIPS* (2011), vol. 11, pp. 1125–1133.

[95] O'Gorman, L. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE 91*, 12 (2003), 2021–2040.

[96] Ojala, T., Pietikainen, M., and Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence 24*, 7 (2002), 971 –987.

[97] Ojansivu, V., and Heikkilä, J. Blur insensitive texture classification using local phase quantization. In *International conference on image and signal processing* (2008), Springer, pp. 236–243.

[98] Oppenheim, A. V., and Lim, J. S. The importance of phase in signals. *Proceedings of the IEEE 69*, 5 (1981), 529–541.

[99] Ortiz-Lopez, J., Galbally, J., Fierrez, J., and Ortega-Garcia, J. Predicting iris vulnerability to direct attacks based on quality related features. In *2011 IEEE International Carnahan Conference on Security Technology (ICCST)* (2011), IEEE, pp. 1–6.

[100] Padole, C. N., and Proenca, H. Periocular recognition: Analysis of performance degradation factors. In *2012 5th IAPR International Conference on Biometrics (ICB)* (2012), IEEE, pp. 439–445.

[101] Park, U., Jillela, R., Ross, A., and Jain, A. K. Periocular biometrics in the visible spectrum. *IEEE Transactions on Information Forensics and Security 6*, 1 (2011), 96–106.

[102] Park, U., Ross, A., and Jain, A. K. Periocular biometrics in the visible spectrum: A feasibility study. In *3rd IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS'09)* (2009), pp. 1–6.

[103] Patel, K., Han, H., and Jain, A. K. Secure face unlock: Spoof detection on smartphones. *IEEE Transactions on Information Forensics and Security* (2016).

[104] Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J. B., and Zuiderveld, K. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing 39*, 3 (1987), 355–368.

[105] Proença, H. Quality assessment of degraded iris images acquired in the visible wavelength. *IEEE Transactions on Information Forensics and Security 6*, 1 (2011), 82–95.

[106] Proença, H. Iris recognition: what is beyond bit fragility? *IEEE Transactions on Information Forensics and Security 10*, 2 (2015), 321–332.

[107] Proença, H., and Alexandre, L. A. Ubiris: A noisy iris image database. In *Image Analysis and Processing - ICIAP 2005*. Springer, 2005, pp. 970–977.

[108] Proenca, H., Filipe, S., Santos, R., Oliveira, J., and Alexandre, L. A. The ubiris. v2: A database of visible wavelength iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence 32*, 8 (2010), 1529–1535.

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

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

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

[112] Raghavendra, R., and Busch, C. Learning deeply coupled autoencoders for smartphone based robust periocular verification. In *23rd IEEE International Conference on Image Processing (ICIP 2016)*. (Oct. 2016), pp. 1 –5.

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

[114] Raghavendra, R., Kiran B. Raja, and Busch, C. Ensemble of statistically independent filters for robust contact lens detection in iris images. In *Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing* (2014), ACM, p. 24.

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

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

[117] Raghavendra, R., Kiran B. Raja, Yang, B., and Busch, C. Combining iris and periocular recognition using light field camera. In *2nd IAPR Asian Conference on Pattern Recognition (ACPR2013)* (2013), IEEE.

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

[119] Ratha, N., Connell, J., and Bolle, R. An analysis of minutiae matching strength. In *Audio- and Video-Based Biometric Person Authentication*, J. Bigun and F. Smeraldi, Eds., vol. 2091 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2001, pp. 223–228. Available from: http://dx.doi.org/10.1007/3-540-45344-X_32.

[120] Rathgeb, C., and Uhl, A. Secure iris recognition based on local intensity variations. In *Image Analysis and Recognition*. 2010, pp. 266–275.

[121] Rathgeb, C., and Uhl, A. Context-based biometric key generation for iris. *IET computer vision 5*, 6 (2011), 389–397.

[122] Rathgeb, C., Uhl, A., and Wild, P. Iris recognition: From segmentation to template security. advances in information security, vol. 59, 2013.

[123] Rattani, A., Derakhshani, R., Saripalle, S. K., and Gottemukkula, V. Icip 2016 competition on mobile ocular biometric recognition. In *Image Processing (ICIP), 2016 IEEE International Conference on* (2016), IEEE, pp. 320–324.

[124] Ross, A. Iris recognition: The path forward. *Computer 43*, 2 (2010), 30–35.

[125] Ross, A., and Jain, A. Information fusion in biometrics. *Pattern recognition letters 24*, 13 (2003), 2115–2125.

[126] Ross, A., and Jain, A. K. Multimodal biometrics: An overview. In *2004 12th European Signal Processing Conference* (2004), IEEE, pp. 1221–1224.

[127] Ross, A. A., Jain, A. K., and Nandakumar, K. Information fusion in biometrics. *Handbook of Multibiometrics* (2006), 37–58.

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

[129] Santos, G., Grancho, E., Bernardo, M. V., and Fiadeiro, P. T. Fusing iris and periocular information for cross-sensor recognition. *Pattern Recognition Letters 57* (2015), 52–59.

[130] Sequeira, A. F., Murari, J., and Cardoso, J. S. Iris liveness detection methods in mobile applications. In *2014 International Conference on Computer Vision Theory and Applications (VISAPP)* (2014), vol. 3, IEEE, pp. 22–33.

[131] Sequeira, A. F., Oliveira, H. P., Monteiro, J. C., Monteiro, J. P., and Cardoso, J. S. Mobilive 2014-mobile iris liveness detection competition. In *2014 IEEE International Joint Conference on Biometrics (IJCB)* (2014), IEEE, pp. 1–6.

[132] Shen, Y., Hu, W., Yang, M., Wei, B., Lucey, S., and Chou, C. T. Face recognition on smartphones via optimised sparse representation classification. In *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks, IPSN-14* (2014), IEEE, pp. 237–248.

[133] Simoncelli, E. P., and Olshausen, B. A. Natural image statistics and neural representation. *Annual review of neuroscience 24*, 1 (2001), 1193–1216.

[134] Sturm, R. A., and Larsson, M. Genetics of human iris colour and patterns. *Pigment cell & melanoma research 22*, 5 (2009), 544–562.

[135] Sutra, G., Dorizzi, B., Garcia-Salicetti, S., and Othman, N. A biometric reference system for iris, osiris version 4.1.

[136] Tan, C.-W., and Kumar, A. Human identification from at-a-distance images by simultaneously exploiting iris and periocular features. In *2012 21st International Conference on Pattern Recognition (ICPR)* (2012), IEEE, pp. 553–556.

[137] Tan, C.-W., and Kumar, A. Unified framework for automated iris segmentation using distantly acquired face images. *IEEE Transactions on Image Processing 21*, 9 (2012), 4068–4079.

[138] Tan, C.-W., and Kumar, A. Integrating ocular and iris descriptors for fake iris image detection. In *2nd International Workshop on Biometrics and Forensics, Malta* (2014), IEEE.

[139] Tan, T., Zhang, X., Sun, Z., and Zhang, H. Noisy iris image matching by using multiple cues. *Pattern Recognition Letters 33*, 8 (2012), 970–977.

[140] Tan, X., and Triggs, B. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE transactions on image processing 19*, 6 (2010), 1635–1650.

[141] Tao, Q., and Veldhuis, R. Biometric authentication for a mobile personal device. In *2006. 3rd Annual International Conference on Mobile and Ubiquitous Systems-Workshops* (2006), IEEE, pp. 1–3.

[142] Thavalengal, S., Bigioi, P., and Corcoran, P. Evaluation of combined visible/nir camera for iris authentication on smartphones. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2015), pp. 42–49.

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

[144] Toth, A. B., and Galbally, J. *Anti-Spoofing: Iris*. Springer-Verlag Berlin Heidelberg, 2005, pp. 970–977.

[145] van Hateren, J. H., and van der Schaaf, A. Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences 265*, 1394 (1998), 359–366.

[146] Vapnik, V. Statistical learning theory. 1998, 1998.

[147] Venkataramani, K., Qidwai, S., and Vijayakumar, B. Face authentication from cell phone camera images with illumination and temporal variations. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 35*, 3 (2005), 411–418.

[148] VeriEye SDK. http://www.neurotechnology.com/verieye.html, 2015. Accessed: 2015-07-03.

[149] Viola, P., and Jones, M. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001. CVPR 2001.* (2001), vol. 1, IEEE, pp. I–511.

[150] Viola, P., and Jones, M. Robust real-time face detection. *International Journal of Computer Vision 57* (2004), 137–154.

[151] Wadhwa, N., Rubinstein, M., Durand, F., and Freeman, W. T. Phase-based video motion processing. *ACM Transactions on Graphics (TOG) 32*, 4 (2013), 80.

[152] Wei, H., Chen, L., and Ferryman, J. Biometrics in ABC: counter-spoofing research.

[153] Weickert, J., and Scharr, H. A scheme for coherence-enhancing diffusion filtering with optimized rotation invariance. *Journal of Visual Communication and Image Representation 13*, 1 (2002), 103–118.

[154] Wiskott, L., Fellous, J.-M., Kuiger, N., and Von Der Malsburg, C. Face recognition by elastic bunch graph matching. *IEEE Transactions on pattern analysis and machine intelligence 19*, 7 (1997), 775–779.

[155] Woodard, D. L., Pundlik, S., Miller, P., Jillela, R., and Ross, A. On the fusion of periocular and iris biometrics in non-ideal imagery. In *2010 20th International Conference on Pattern Recognition (ICPR)* (2010), IEEE, pp. 201–204.

[156] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., and Ma, Y. Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence 31*, 2 (2009), 210–227.

[157] Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., and Freeman, W. Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics (TOG) 31*, 4 (2012), 65.

[158] Xu, J., Cha, M., Heyman, J. L., Venugopalan, S., Abiantun, R., and Savvides, M. Robust local binary pattern feature sets for periocular biometric identification. In *2010 Fourth IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS)* (2010), IEEE, pp. 1–8.

[159] Yadav, D., Kohli, N., Doyle, J. S., Singh, R., Vatsa, M., and Bowyer, K. W. Unraveling the effect of textured contact lenses on iris recognition. *IEEE Transactions on Information Forensics and Security 9*, 5 (2014), 851–862.

[160] Yambay, D., Doyle, J. S., Bowyer, K. W., Czajka, A., and Schuckers, S. Livdet-iris 2013-iris liveness detection competition 2013. In *2014 IEEE International Joint Conference on Biometrics (IJCB)* (2014), IEEE, pp. 1–8.

[161] Yan, J. J., Blackwell, A. F., Anderson, R. J., and Grant, A. Password memorability and security: Empirical results. *IEEE Security & privacy 2*, 5 (2004), 25–31.

[162] Zhang, B., Shan, S., Chen, X., and Gao, W. Histogram of gabor phase patterns (hgpp): a novel object representation approach for face recognition. *IEEE Transactions on Image Processing 16*, 1 (2007), 57–68.

[163] Zhang, L., Yang, M., and Feng, X. Sparse representation or collaborative representation: Which helps face recognition? In *2011 IEEE International Conference on Computer Vision (ICCV)* (2011), IEEE, pp. 471–478.

[164] Zhang, Z., Yan, J., Liu, S., Lei, Z., Yi, D., and Li, S. Z. A face antispoofing database with diverse attacks. In *2012 5th IAPR International Conference on Biometrics (ICB)* (2012), IEEE, pp. 26–31.

[165] Zhao, W., Chellappa, R., Phillips, P. J., and Rosenfeld, A. Face recognition: A literature survey. *ACM Comput. Surv. 35*, 4 (Dec. 2003), 399–458. Available from: http://doi.acm.org/10.1145/954339.954342.

[166] Zhu, Z., You, X., Chen, C. P., Tao, D., Ou, W., Jiang, X., and Zou, J. An adaptive hybrid pattern for noise-robust texture analysis. *Pattern Recognition 48*, 8 (2015), 2592–2608.