MULTI-MODAL OCULAR RECOGNITION IN PRESENCE OF OCCLUSION IN MOBILE DEVICES

A Dissertation
in
Electrical and Computer Enginerring
and
Telecommunications and Computer Networking

Presented to the Faculty of the University
of Missouri–Kansas City in partial fulfillment of
the requirements for the degree

DOCTOR OF PHILOSOPHY

by
AHMAD SAEED MOHAMMAD

MS. EE., University of Missouri-Kansas City, USA, 2015

Kansas City, Missouri
2018

MULTI-MODAL OCULAR RECOGNITION IN PRESENCE OF OCCLUSION IN MOBILE DEVICES

Ahmad Saeed Mohammad, Candidate for the Doctor of Philosophy Degree

University of Missouri–Kansas City, 2018

ABSTRACT

The existence eyeglasses in human faces cause real challenges for ocular, facial, and soft-based (such as eyebrows) biometric recognition due to glasses reflection, shadow, and frame occlusion. In this regard, two operations (eyeglasses detection and eyeglasses segmentation) have been proposed to mitigate the effect of occlusion using eyeglasses.

Eyeglasses detection is an important initial step towards eyeglass segmentation. Three schemes of eyeglasses detection have been proposed which are non-learning-based, learning-based, and deep learning-based schemes. The non-learning scheme of eyeglasses detection which consists of cascaded filters achieved an overall accuracy of 99.0% for VISOB and 97.9% for FERET datasets. The learning-based scheme of eyeglass detection consisting of extracting Local Binary Pattern (LBP), Histogram of Gradients (HOG) and fusing them together, then applying classifiers (such as Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and Linear Discriminant Analysis (LDA)), and fusing the

output of these classifiers. The latter obtained a best overall accuracy of about 99.3% on FERET and 100% on VISOB dataset. Besides, the deep learning-based scheme of eyeglasses detection showed a comparative study for eyeglasses frame detection using different Convolutional Neural Network (CNN) structures that are applied to Frame Bridge region and extended ocular region. The best CNN model obtained an overall accuracy of 99.96% for ROI consisting of Frame Bridge.

Moreover, two schemes of eyeglasses segmentation have been introduced. The first segmentation scheme was cascaded convolutional Neural Network (CNN). This scheme consists of cascaded CNN's for eyeglasses detection, weight generation, and glasses segmentation, followed by mathematical and binarization operations. The scheme showed a 100% eyeglasses detection and 91% segmentation accuracy by our proposed approach. Also, the second segmentation scheme was the convolutional de-convolutional network. This CNN model has been implemented with main convolutional layers, de-convolutional layers, and one custom (lamda) layer. This scheme achieved better segmentation results of 97% segmentation accuracy over the cascaded approach.

Furthermore, two soft biometric re-identification schemes have been introduced with eyeglasses mitigation. The first scheme was eyebrows-based user authentication consists of local, global, deep feature extraction with learning-based matching. The best result of 0.63% EER using score level fusion of handcraft descriptors (HOG, and GIST) with the deep VGG16 descriptor for eyebrow-based user authentication. The second

scheme was eyeglass-based user authentication which consisting of eyeglasses segmentation, morphological cleanup, features extraction, and learning-based matching. The best result of 3.44% EER using score level fusion of handcraft descriptors (HOG, and GIST) with the deep VGG16 descriptor for eyeglasses-based user authentication.

Also, an EER enhancement of 2.51% for indoor vs. outdoor (In: Out) light settings was achieved for eyebrow-based authentication after eyeglasses segmentation and removal using Convolutional-Deconvolutional approach followed by in-painting.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the School of Graduate Studies, have examined a dissertation titled "Multi-Modal Ocular Recognition in presence of occlusion in Mobile Devices," presented by Ahmad Saeed Mohammad, candidate for the Doctor of Philosophy degree, and certify that in their opinion it is worthy of acceptance.

Supervisory Committee

Reza Derakhshani, Ph.D., Associate Professor and Committee Chair
Department of Computer Science & Electrical Engineering

Ghulam Chaudhry, Ph.D., Professor and CSEE Head
Department of Computer Science & Electrical Engineering

Cory Beard, Ph.D., Associate Professor
Department of Computer Science & Electrical Engineering

Zhu Li, Ph.D., Associate Professor
Department of Computer Science & Electrical Engineering

Baek-Young Choi, Ph.D., Associate Professor
Department of Computer Science & Electrical Engineering

Ahmed Hussan, Ph.D., Assistant Professor
Department of Computer Science & Electrical Engineering

CONTENTS

ILLUSTRATIONS

TABLES

ACKNOWLEDGEMENTS

This work would not have been conceivable without the financial support from the ministry of higher education and scientific research of Iraq, University of Missouri-Kansas City (UMKC), School of Computing and Engineering(SCE), ZOLOZ company (previously, EyeVerify, Co.).

I would like to thank my academic advisor Dr. Reza Derakhshani for his professional guidance and taught me a great deal about scientific research. Also, I would especially like to thank my committee, Dr. Ghulam Chaudhry, Dr. Cory Beard, Dr. Zhu Li, Dr. Baek-Young Choi, and Dr. Ahmed Hussan. Many thanks to Dr. Ajita Rattani to guide me in the process of research formation and writing. In addition, I am thankful to all of the people whom I have had the delight to work through classes and research and especially the lab member Narsi Reddy, Ala-Addin Nabulsi, Jesse Lowe, Mark Nguyen.

The most important to me and nobody has been more important to me than them, is my family. My heartfelt thanks to my parent, especially my father Dr. Saeed Mohammad, for their support and encouragement in whatever I pursue. Also, a special thanks to my uncle Dr. Raid Khalil Ahmad, who encourage me to continue my Ph.D. study. In addition, I would like to thank my supportive wife, Nibras, and my two wonderful children, Maryam and Yosif, for their patients.

CHAPTER 1

INTRODUCTION

There is a wide scale integration of biometric authentication in mobile devices. Mobile devices allow users to access their financial and personal information using biometric authentication system. Any authentication system could identify users according to the following parameters:

1. Thing you possess: The user has a thing which facilitates the access through the authentication system. For example, a house key to access your home.

2. Thing you know: The user should remember certain characters of phrase which is required for access when the authentication system ask about it. For example password.

3. Thing you have: The user posses different physical and behavioral modality which are unique and could be used to identify people known as Biometric. For example Face, eye-print, iris, gait, signature.

The biometric allows the secure access by establishing the user identity based on ”how you are” rather than ”what you possess” or ”what you know” [6].

1.1 Types of Biometric Traits

There are different types of biometric traits which are used in an authentication system. Some of these traits are based on a physiological characteristic such as the face, iris, eye-print. While the other biometric traits are based on a behavioral characteristic such as gait, signature, and voice. Moreover, there are some other biometric traits which could be used to support the above traits are called soft biometric. The soft biometric also could be used to continuous and passive authentication biometric system. Figure (1) shows a samples of these biometric traits. In addition, there are no biometric traits without challenging or drop-back, for example, iris trait could not be acquisition using an ordinary camera, and it required a close distance IR sensor to capture it. Moreover, face trait may be effected by aging [7].

1.1.1 Physiological Traits

This type of biometric traits based on a physiological characteristic which already exists in the human body such as the face, fingers, and palm. The usage of these traits dramatically increases with the mobile devices such as smartphones and the requirement of biometric secure access for business and personal information. There are many physiological traits which are still effectively involved in security applications such as the face, iris, ear, eye-print, palmprint.

Face trait is considered the first hard biometric which has a non-intrusive trait, and it contains the other biometric traits such as eyes and ears. Also, this trait could be captured from distance and extracted from video frames. Therefore this biometric trait is

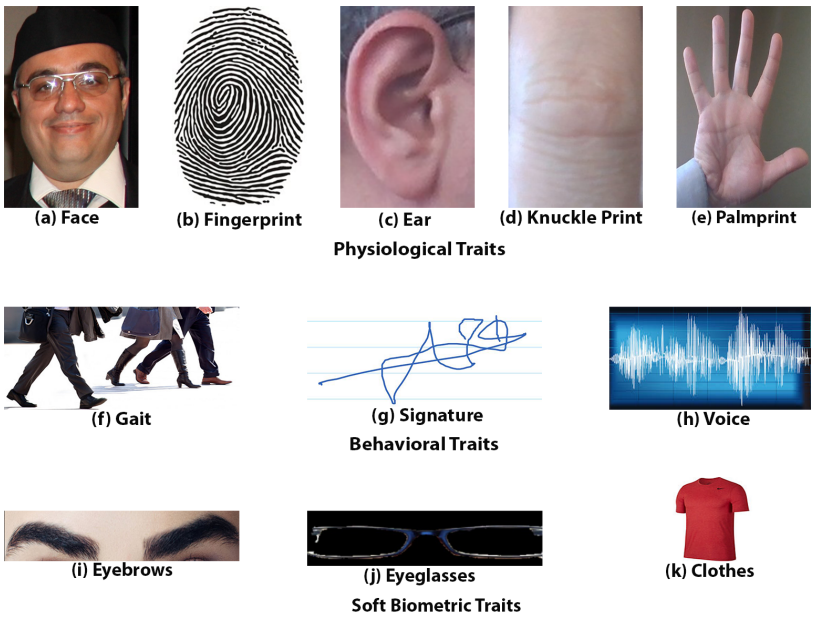


Figure 1: An Examples on Biometric Traits

used in surveillance and criminal identification applications [8].

The fingerprint is considered as a long time, unique, and stable hard biometric trait which made up of ridges, discriminative texture pattern (such as loop and whorl), and minutia which represents the ridge ending and ridge bifurcation. However, capturing a good quality of fingerprint still a difficult operation due to the sensors quality and the job nature of some people which have an indistinct texture of fingerprint such as worker [9].

The ear has a universal shape biometric trait and could be captured from distance and side profile, so that makes the traits useful in surveillance application in which people are running and have a side profile [10].

In addition, the iris is one of the best discriminative traits which has a donut shape region with furrows and ridges. However, this trait should be captured in Near Intra-Red

(NIR) high-resolution sensors [11]. Recent researches showed the possibility to implement an iris-based biometric system by acquiring the iris modality in light color setting or by creating a particular illumination environment [12].

Moreover, there is a non-facial part trait which forms the inner part of palm is called palmprint. This trait has a rich texture consists of ridges, minutia, lines, and delta points which enable this trait to be a unique trait [13].

In addition, there is another non-facial trait which forms the outer part of the finger is considered as a knuckle print. This trait has a baseline texture and could be captured by low-cost sensors, however, knuckle print suffered from illumination variation, rotation, and translation [14]. Table (1) shows the comparison of different hard biometric traits with respect to their advantages and challenges [15].

1.1.2 Behavioral Traits

The behavioral traits are another biometric traits that have been used in an authentication system. This type of traits depends on the behavioral characteristics of the human beings. These biometric behaviors are related to some measurements or responsibilities of human that help to identify subjects or group of subjects according to their behaviors. Some of these behavioral traits will discuss in this section such as Gait, and voice [16].

Gait represents one of the good behavioral traits in the biometric system since it could be acquired from a distant distance for still images or video streams. Also, this trait required a daylight light condition to identify people correctly. In addition, gait trait depends on the shape and gesture of subject walking style and it could be used in

surveillance applications as well could be involved with other biometric traits [17].

Signature is considered as an instance of individual verification. The signature trait depends on two main features which are static features, and dynamic features. The static features such as orientation, coordinates, writing angle, the shape of the constructed curve, and the number of breakpoints. While the dynamic features such as pen pressure, pen speed, and writing time. The signature trait could be used in different applications that required personal verification such bank checking and financial payment [18], however, this type of trait could be spoofed and forged easily so that some banks moved their verification process to mobile eye-print and periocular region modalities.

Voice is one of the naturally common behavioral of the human. Voice trait depends on several natural features to identify people such as voice texture, nasal tone, cadence, and speech influence. however, this trait could be imitated and faked easily due to its weak feature mentioned above, and a voice similarity between people [19]. Table (2) shows the comparison of different behavioral biometric traits with respect to their advantages and challenges.

1.1.3 Soft Biometric Traits

Soft biometric is considered in most cases as a descriptive trait [20] that identify people by creating a description label. This description label could be skin tone, human height, gender, ethnicity, body weight, fat measurement, the gist of an eyebrow, and the shape of eyeglasses. These labels could be combined with each other such as gender trait and ethnicity or could be fused with another hard physiological trait such as body

weight and fingerprint [21]. One of the main advantages of soft biometrics that it could be used to identify people based on the human description, Also less computation and memory usage by comparing with other hard physiological traits such as face [22]. Also, the soft biometric traits are not used only for personal identification; also it could be used Surveillance Environment [23].

Human weight was used as a soft biometric by fusing this trait with hard biometric such as a fingerprint to identify people by photoplethysmographic signals as a novel approach [24] [21]. Besides, soft biometric also called ancillary user information such as gender, human hight, eye color had been used for user authentication by fusing this information with fingerprint or face [25]. Moreover, ethnicity detection [26] could be used as a good supporter of primary biometric by fusing this personal information with the ocular region to identify people. Figure 2 shows a nice example of ethnicity support with Iraqi (middle-east ethnicity) person (Figure 2 (a)) that is so similar to the real suarez (a famous score player) as shown in figure 2 (b) [1].

The bifurcation of gender for people to obtain the descriptive label of sexual categories of male and female. These sexual labels could support the authentication system based on hard physiological traits if this system fail in verification process. For example, two workers people was verified using authentication system based on fingerprint, and both of them in different sexual categories. In this case the sexual label could help to verify these people [27].

Eyebrow represents a biometric facial part that it exist naturally above the ocular region. Also, eyebrow is one of recent biometric trait that both genders male and female

Figure 2: An Examples on ethnicity trait that support ocular recognition. [1]

poss it. Some of the novel studies considers this trait as a stand-alone biometric trait for biometric authentication system [28], [29], [30]. In addition, eyebrow could be considered as a soft biometric that could be used when the primary biometric is missing due to occlusion or eye half-closed for ocular recognition system according to some studies [31]. Another advantage of eyebrow trait that it give a trade-off between computational complexity and identification accuracy in comparison with face and other facial biometrics. The reason of mentioned advantage is related to a geometric fact that the region of eyebrow represents one-sixth of the full face region, and there is some searchers show this trade-off comparison with their reported results [28], [12].

Facial and ocular recognition systems have a real problem cased by eyeglasses. This is because eyeglasses reflection, shadow, and frame occlusion. Moreover, eyeglasses detection, segmentation, and removal will mitigation this occlusion problem that affect

the primary biometric recognition with facial the ocular regions. However, the eyeglasses detection and segmentation could be used to get a novel soft biometric trait, eyeglasses. The gist scene of eyeglasses, and the texture of eyeglasses material could be used as a stand-alone trait to identify people that wear eyeglasses for a short time passive authentication. In addition, the eyeglasses trait could be fused with other primary hard biometric trait such as face or ear for long time authentication.

1.2 Modes of Biometric Operation

The biometric system could be divided into three main modes which are enrollment, verification, and identification. Each of these modes has its requirement and specific operations. Any biometric system should start with enrollment operation to obtain the user information to the system. This personal information could be any modality or biometric traits such as the face, iris, or eyebrow. Next, the features will be extracted from these modality or trait and store in the system database as a template to be used in the next biometric operation. The verification or identification is the next operation that will operate after enrollment. In these two operations, the descriptors of the query modality or biometric trait are matched with the stored templates from the enrollment operation [32], [33].

1.2.1 Enrollment

This operation involves the enrollment or registration of a person to the system database. Enrollment consists of different steps 3 which are modality acquisition, quality

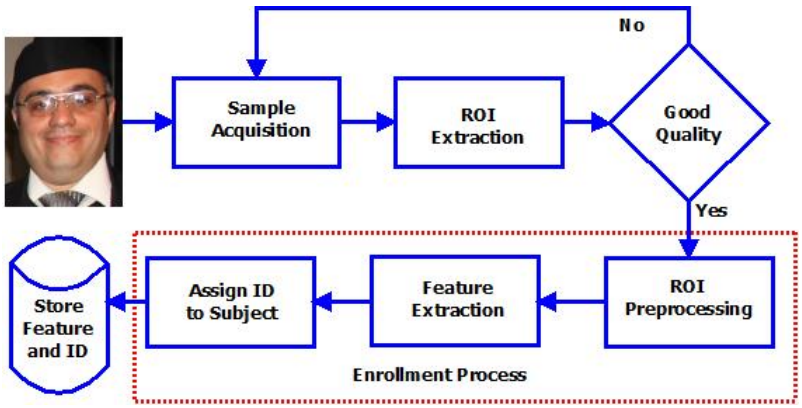


Figure 3: The block diagram of enrollment mode.

checking, preprocessing, feature extraction, and template storing. The modality or biometric trait is acquired as a region of interest (ROI) using different computer vision and cascaded classifiers such as viola-johns [34] or facial deep learning landmarks [35]. Next, the quality of acquired ROI will check with respect to a specific threshold, if the quality of ROI does not satisfy the requirement, then the modality or trait will be acquired again. Then, the features will be extracted and sorted as a template in the system database [36].

1.2.2 Verification

Verification is defined as *one-to-one* (1:1) system verification which the system match the identity of individual claim with its corresponding template. The result of the matching operation is the similarity score which represents the amount of similarity or distance between the claim and the corresponding template. Moreover, the matching is applied to the extracted feature of the biometric trait for the individual claim with the feature of the corresponding template. Then, the individual is considered as a genuine

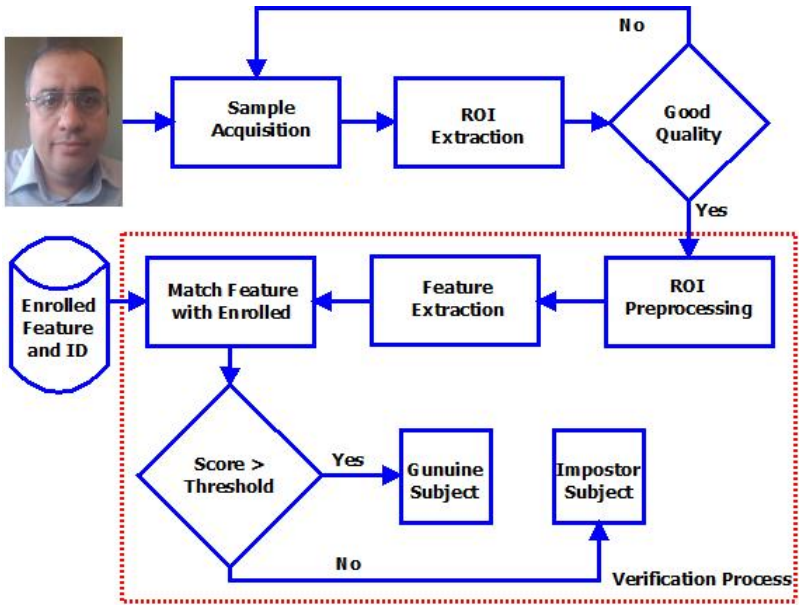


Figure 4: The block diagram of verification mode.

user if the matching score is above a certain predefined threshold and the claim is verified for this subject. Otherwise, this subject is considered as *impostor* user, and this subject is denied for access [37], [32]. Figure 4 shows the block diagram of this biometric mode.

1.2.3 Identification

This biometric mode is considered as one-to-many $(1 : N)$ system identification which the system match the identity of individual claim with all (N) stored templates. Also, the system attempt to determine whether the identity of the individual exists in the system database. In that case, if the matching score is above the predefined threshold, the claimed user is considered as *identify* user or *genuine* user. Otherwise, this user is considered as *impostor* user [37], [32]. Figure (5) shows the block diagram of this mode.

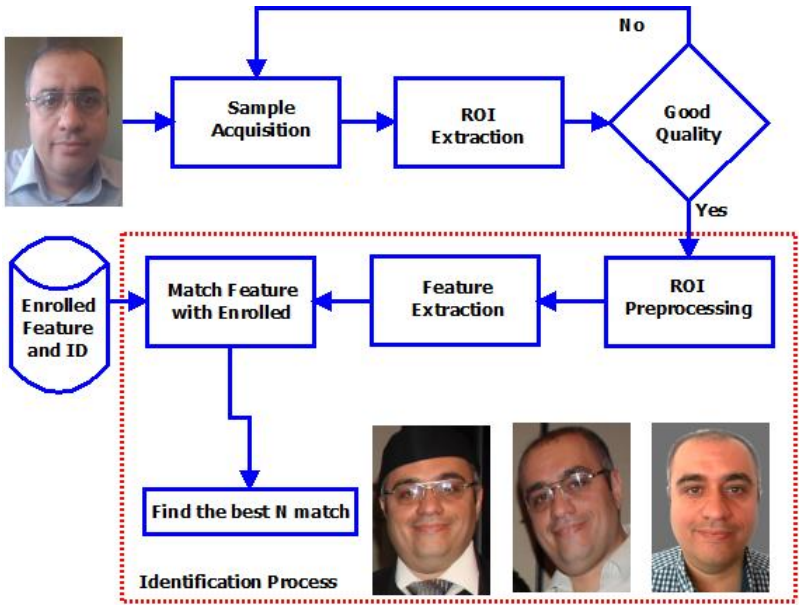


Figure 5: The block diagram of identification mode.

1.3 Performance Analysis of Biometric System

The performance of any biometric system affected by different variation parameters such resolution of acquisition sensor, occlusion of wearable object such eyeglasses, variation of light condition such as indoor/outdoor settings, pose profile of individuals such high/low poses, user cooperation in different biometric modes (section 1.2). In that case, all mentioned variation parameters made the performance of any biometric system not to be a perfect system, that is never to be 100% accuracy. Therefore, the measurements for performance evaluation is required in different biometric operation or modes. These measurements represent the amount of system accuracy and sometimes represents the error amount due to a specific variation or computation.

There are different types of rates by which the biometric system is evaluated which

are [38]:

1. False Rejection Rate (FRR)

   This rate also called as *Type I* error. *Type I* error or false reject happens if the *genuine* user is verify or identify as an *impostor* user due the similarity score is below the predefined threshold. The rate is computed using equation (1.2).

$$FAR = \frac{false \quad reject \quad attempts \quad made \quad by \quad genuine \quad users}{Total \quad number \quad of \quad attempts}. \tag{1.1}$$

2. False Acceptance Rate (FAR)

   This rate also called as *Type II* error. *Type II* error or false acceptance occurs when the *impostor* user is recognized as a *genuine* user due to the similarity score is above the predefined threshold. Therefore, FAR or False Match Rate (FMR) represent the probability of incorrect access of unauthorized individual made by a biometric system. The rate is computed using equation (1.2).

$$FAR = \frac{false \quad sucessful \quad attempts \quad made \quad by \quad genuine \quad users}{Total \quad number \quad of \quad attempts}. \tag{1.2}$$

3. Equal Error Rate (EER)

   EER is the obtained value where the False Reject Rate (FRR) is equal to the False Acceptance Rate (FAR). This value could determine the accuracy level of the biometric system is reciprocal manner. In that case, a low value of EER represents a high value of accuracy level for a particular biometric system. Also, EER value is calculated using the Genuine Acceptance Rate (GAR) as shown in equation (1.3) with False Acceptance Rate (FAR) in the same manner.

$$GAR = FRR - 1. \tag{1.3}$$

4. Failure to Enroll Rate (FTE)

   Sometimes the biometric system locks the generation ability to create or register multiple copies of templates for a particular modality or biometric trait of one or more individual. There are different reasons for this failure such as the absence of biometric features for a particular individual, and the sensor could not be acquired an excellent image quality during the enrollment, the protocol of enrollment, and the long-term enrollment.

5. Failure to Acquire Rate (FTA)

   This rate represents the inability of the biometric system to acquire, detect, or locate the required region of interest (ROI). This is because of the low quality of images, and the adaptive threshold of the quality is checking for extracted ROI.

1.4 Challenges associated to Biometrics in Mobile Environment

Many challenges associated to biometric systems in the mobile environment such as the occlusion of wearable objects (i.e. eyeglasses, sunglasses, hat, and earrings), images acquired in different light conditions (i.e. indoor versus outdoor light settings), different profile or poses during "selfie" images.

1.4.1 Eyeglasses Occlusion

Eyeglasses occlusion represents a large factor for face and ocular recognition. The reasons for this effect are the frame occlusion, light reflection on eyeglasses, and the shadow due to frame and glasses. Furthermore, inaccurate eye pair detection may happen due to the presence of eyeglasses which cover the ocular and periocular region of the face. A significant amount of literature shows that the face recognition depends on eye pair detection for face registration and alignment. Moreover, the performance of ocular and periocular recognition systems may degrade due to occlusion of these important biometric region by eyeglasses [39], [40], [12].

1.4.2 Lighting variations

Most of the acquisition process in biometric system demands high-quality images, and keeping the original characteristic of biometric feature for a modality or trait, to satisfy this requirement, a sufficient source light condition is preferred in visible and infrared. Further, high performance for the biometric system could be obtained using controlled light condition database. In a mobile environment, it is difficult to obtain a good control over light condition. To mitigate this problem, some research studies suggest to applied some computer vision filter such as CLAHE to soften this problem [41], [42].

1.4.3 Poses in "selfie" images

There are different profiles of poses for "selfie" images that acquired by a mobile device such as horizontal poses (i.e., right and left poses), and vertical poses (i.e., high,

medium, and low poses). The horizontal poses have more impact on some of the biometric traits such as eyebrow. However, the biometric could select one eyebrow (such as right), and leave the other one. Also, most of the "selfie" images have vertical poses rather than horizontal poses. In case of vertical poses, the low pose affects the acquisition process to obtain some of the biometric traits such as blood vessels in scalar portion [43].

1.5 Limitation of Biometric in Mobile Environment

There are different types of limits that have a valuable impact on any biometric system. The sources of these limitations are related to the acquisition process, the matching/classification process, or the biometric trait itself. Some of these limitation is shown bellow [33], [19]:

1. Inter-class similarity: This limitation happens because of the overlapping of extracted feature which corresponding to multiple different enrolled subjects. FAR could a good indicator for this limitation.

2. Intra-class variations: The main reason for this limitation is the variation in illumination and pose during the acquisition process. Lifespan and aging (i.e., face aging) are another reason for the intra-class variation.

3. Noisy acquired Data: The noisy data occurs mainly in the acquisition process due to an unsuitable configuration for sensors, or due to temporal variation caused by long-term matching. Noisy data may affect the matching process and cause incorrect identification in a biometric system.

4. Object occlusion: Occlusion for some biometric traits affect on trait acquisition and matching processes. Some of these objects naturally exist such hair, and another object unnaturally exists eyeglasses.

5. Spoofing attaches: Spoofing in the mobile environment occurs when someone creates a fake physical trait of another subject such a facial image of another subject and try to spoof the biometric system during the recognition process.

6. Non-universality: Some individual does not have a particular biometric trait. For example, most the worker have a weak or poor fingerprint.

7. Facial Plastic Surgery: This new limitation represents a significant problem in the recognition process for the face-based biometric system. This limitation happens because the impostor user depends on plastic surgery with facial characteristics such as facial shape, eyes size, eyebrows shape, nose, and chin.

8. Hormone Replacement: Another new limitation achieve by a medical operation on the hormone inside the human body to change the gender. This limit affects the biometric system which uses face and or gender biometric trait.

1.6 Multimodal Biometric and Level of Fusion

1.6.1 Multimodal Versus Unimodal Systems

Mainly, there is two type of biometric system namely, unimodal and multimodal. Unimodal use one biometric trait while the multimodal use of more than one modality or fused biometric traits. Different limitations may affect the unimodal-based system such

noisy data, a limited degree of freedom, intra-class variation, and spoof attack. The previously mentioned problems that are related to the unimodal-based biometric system could be enhanced by involving a different source of biometric information such as different biometric traits.

Multimodal biometric system mitigates the disadvantage of the unimodal-based system by involving multiple sources of biometric information. In that case, the impostor user can not spoof multiple biometric traits of the genuine user at the same time. Also, if one of biometric information is noisy, then the other information (i.e., other biometric traits) can compensate for the noisy one. Moreover, a different type of biometric traits can increase the degree of freedom and raise the performance of the biometric system.

1.6.2 Fusion Levels

Different biometric traits can be fused at acquisition level, feature level, matching level, rank level and decision level as listed below:

1. Acquisition level: The data from a different sensor can be normalized and fused to obtain new data for the same enrolled individual in the biometric system. An example of this level, a fingerprint can acquire using different sensors such finger scanner sensor and solid state sensor.

2. Feature level: The features of a different trait for each user extracted separately. Then, these features are concatenated, normalized, and stored as a template in the database of the biometric system. For example, Local Binary Pattern (LBP) feature of an eyebrow and Binarized Statistical Image Feature (BSIF) of iris could be fused

together.

3. Match score level: The match score of different modality can be fused to obtain one final score. For instance, the match score of a right eyebrow could be fused with the match score of a left eyebrow to form the final score of both eyebrows.

4. Rank level: This fusion is compatible with different identification classifier. A rank assigned for each enrolled identity and the fusion involve many rands with each identity. In that case, a new rank determined from these many ranks, and this will support the final decision of the classifier.

5. Decision level: Fuse the decision of different recognized trait by voting. For example, a biometric system identifies people using face, iris, eyebrow. Then, for each modality has its own decision (thresholded score) namely, [1, 0, 1] for face, iris, and eyebrow respectively in visible light. In that case, fusion could be applied by majority voting to obtain one on the final decision.

1.7 Motivation and contribution of this work

The occlusion of a wearable device such eyeglasses has a real impact on the performance of ocular and periocular biometric system. Also, eyeglasses detection represent the first step of eyeglasses segmentation. In this work, eyeglasses detection using a different approach (non-learning based, learning based) was introduced. Besides, eyeglasses segmentation have been implemented using different deep learning models with cascaded and non-cascaded models. Moreover, two authentication system have was introduced namely

eyebrows-based user authentication, and eyeglasses-based user authentication. Both of these biometric system used for continues and short-team authentication. Also, an enhancement has been applied to these authentication systems using eyeglasses in-painting and removal.

1.7.1 Eyeglasses Detection and Segmentation using Deep Learning

Different approaches have been introduced for eyeglasses detection and segmentation is the summary of the contribution for this part (as it will be explained in details in the next chapters):

1. **Eyeglasses detection**:
    - **Non-Learning based Approach**: This approach was implemented by applying cascaded computer vision filters such as CLAHE filter, Sobel filter, and binary thresholding filters.
    - **Leaning based Approach**: This approach was implemented using classification classifiers such as Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA).
    - **Deep Leaning Approach**: This approach was implemented using (18) squeezed model of Convolutional Neural Networks (CNN) applied into two different regions namely, frame bridge region, and extended ocular region.
2. **Eyeglasses Segmentation**:

- **Cascaded Convolutional Network**: This approach involved two Convolutional Neural Networks (CNN). The first one for eyeglasses detection, and weight generation to be used in the second network. The second network uses the generated weight with convolutional layers and upsampling layers.
- **Convolutional and Deconvolutional Network**: This approach was implemented using deep learning network consist of convolutional, deconvolutional, and a user-defined custom layer.

1.7.2 User Authentication using Eyebrows and Eyeglasses

Two user authentication systems have been implemented using soft-biometric traits namely, eyebrows and eyeglasses. Different local and global descriptors have used to obtain the binary, gradient direction, and the shape of these modalities. Also, non-learning based and learning based matches have been evaluated in this work.

1.7.3 Soft-Biometric Enhancement with Eyeglasses Removal

Eyeglasses in-painting and removal was applied after predicting the eyeglasses masks using deep learning approach. The experimental results was evaluated before and after eyeglasses mitigation.

1.8 Outline of Dissertation Topics

Chapter (1) gives an introduction to the biometric system, Modes fo biometric operation, perform analysis for a biometric system, challenges associative to biometric mobile-based, biometric limitation, multimodal biometric with fusion levels, motivation

and contribution.

Chapter (2) provide background on machine learning classifier, and feature descriptors on computer vision.

Chapter (3) shows schemes for eyeglasses detection, and eyeglasses segmentation using a non-leaning-based, and learning-based method.

Chapter (4) introduce a user authentication using two soft biometric traits which are eyebrows and eyeglasses. This authentication implemented for short-term and continuous biometric system.

Chapter (5) gives a conclusion and future work for eyeglasses mitigation, soft-based authentication, and the usage of deep learning in biometric systems.

Table 1: The comparison of different hard biometric traits with respect to their advantages and challenges

Modality	**Advantage**	**Challenges**
Face	Most common. Non-intrusive Could be capture using cheap sensors Widely acceptable traits	Face pose and illumination Facial expressions Aging effect
Fingerprint	Unique and discrimiative Lesser cooperation in acquistion Could be capture using cheap sensors	Getting good quality of figger-prints
Ear	Non-intrusive Could be capture using cheap sensors Universal shape Highly accepatable traits	Illumniation variation Pose and translation Hair and eye rings occlusion
Iris	Highly discriminative Unique texture Highly discriminative	Illumination variation Eyelid and eyelash occlusion Illumination variation Specular reflection Off-angle iris recognition
Palmprint	Could be capture using less cost sensors Dicriminative and unique feature	Illumniation variation, rotation, and translation Problem of occlusion
Knuckleprint	Naturally protected Could be capture less cost sen-sors	Illumniation variation, rotation, and translation

Table 2: The comparison of different behavioral biometric traits with respect to their advantages and challenges

Modality	Advantage	Challenges
Gait	Non-Instrusively. Could be acquired from distance. Good user acceptance.	Background, and clothes variation. Nature of walking surface.
Signature	Good individuals acceptance	Aging, and emotion variation. Nature of writing surface
Voice	Well Individual acceptance	Aging, and emotion variation. Medical issues that are related to voice quality

CHAPTER 2

BACKGROUND IN MACHINE LEARNING AND COMPUTER VISION

2.1 Introduction to Machine Learning

Machine learning is an approach that supply systems the ability to learn and improve its knowledge from the subset of data, validate this knowledge, and generate the final decision with unseen data. Besides, machine learning can be divided into supervised, semi-supervised, unsupervised algorithm as listed below [44]:

- Supervised learning: The algorithm learns from example input (train subset) alone with desired output (also it called labels or target), after training the algorithm, the classifier can generate the final output by providing unseen examples (test subset). Also, the evaluation can be made by comparing the final output and the label of unseen examples. Predicting of heart ischemia is an application of this learning.

- Semi-supervised learning: The algorithm learns from both labeled and unlabeled training data. This learning algorithm is used when the label acquisition required a relevant source to obtain it. Protein sequence classification is an example of this learning since the inferring process for protein function required intensive task by a human.

- Unsupervised learning: The algorithm learns from unlabeled training data, and it

can not validate the output for unlabeled test data. However, it can provide the inferences from the whole dataset to describe hidden structures. Clustering approach is an example of unsupervised learning such audio recording clustering for microphone surveillance system.

2.1.1 Machine Learning Applications

In recent days, Machine learning has grown rapidly and become one of primary source of information technology in our life. Dramatically, the amount of data increase rapidly, so it is crucial to involve a smart analysis of these data for participating in technology development in smart devices and mobile phones. Machine learning could be found in many aspects of our life. One of the known machine learning application is web page ranking. This application can be defined as an operation that sends a query to a search engine implemented by a particular machine learning method and returning the relevant pages to that query. Another widely used application for machine learning is security application. This application can be used as an access control for financial and personal information. Since most people nowadays use a mobile device such as smartphones or tablet to access their information, the machine learning involves inside these devices to deploy this job. Another application is named entry recognition which performs an identification for entries such place, title, names inside documents. For instance, modern e-mail clients such as Apple's mail posses the ability to identify the address in mail filling than in the address book of smartphone [45].

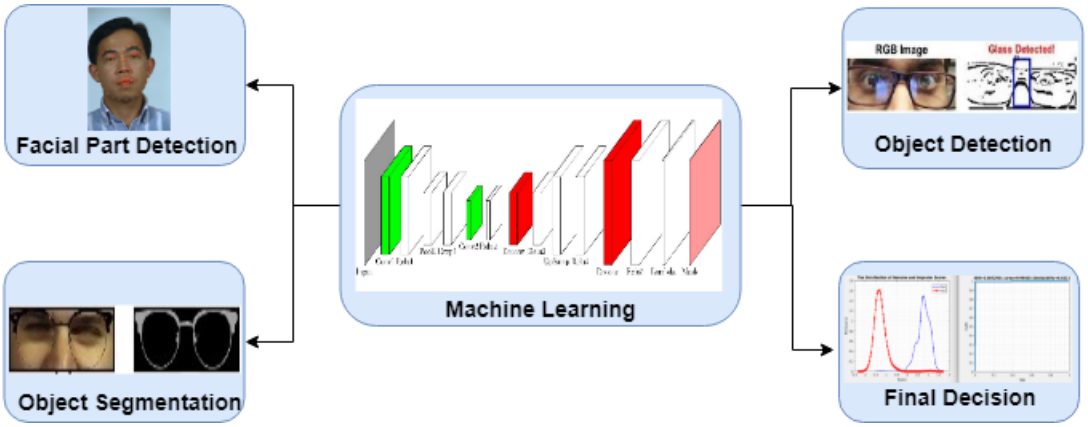


Figure 6: The role of machine learning in biometric.

2.1.2 Role of Machine Learning in Biometrics

Machine learning plays an important role in the different aspect of biometric system 6. There are many applications of machine learning in a biometric system such as facial part detection, wearable object detection and segmentation, and making the final decision of biometric system [46].

Face and facial part detection are one of the important applications for ROI prepare. There are different approaches to ROI detection using machine learning such as viola-johns [34], and Dlib landmarks and face detection [35].

Another application of machine learning is the detection of a wearable object such as sunglasses and eyeglasses. This could be implemented using different classifiers such as Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), Linear Discriminant Analysis (LDA), and Quadratic Discriminant Analysis (QDA) [47]. This will be explained in details in Chapter 3.

Moreover, any biometric system output the decision at the final stage for verification and identification. At this stage, the biometric system link or match the feature of the acquired sample with the feature of a query sample to obtain the final decision for a specific identity. In that case, the matching operation could be implemented using a non-learning based method by calculating the distance between the acquired and query feature. However, machine learning could be considered a good match between the acquired feature (training set), and the query feature (test set) [48].

2.1.3 Deep learning and Biometrics

Deep learning is a machine learning with computational model consists of multiple processing layers which poss the ability to process raw data such as voices or images in a different level of representation and abstraction [49].

Deep learning has a dramatic development and involved in different applications in our life due to several reasons. The first reason is the hardware evolution in computational devices such as Graphical Processing Unit (GPU) which posses multiple cores for computation. For example, a GPU GTX 1090 Ti has streaming pascal Multiprocessors for a total of 3840 CUDA cores. Also, the ability of Convolutional Neural Networks (CNNs) to process raw data such as image matrix due to convolutional layers inside CNN. While the traditional machine learning such Multi-Layer Perceptron (MLP) need an extracted features to process it. Also, the deep learning has different data representation and abstraction for each stage (layer). For instance, the heat map image could be obtained from the activation layer after convolutional layer while the output categories could be acquired

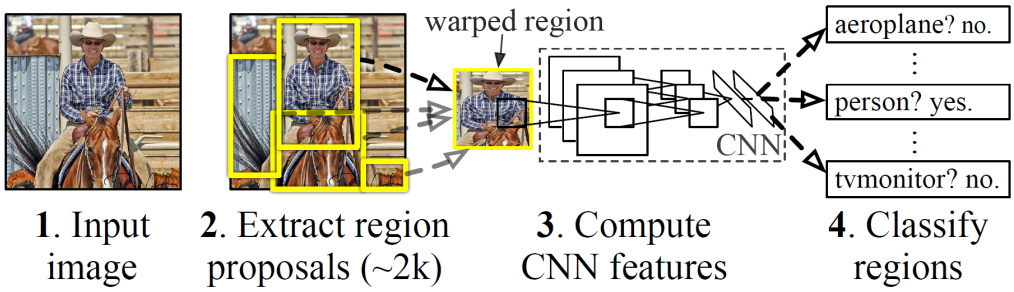


Figure 7: The steps of object detection using DCNN. [2]

from full connected layers [50].

There are different biometric applications used deep learning as a machine learning which are object detection, object segmentation, feature extraction, and feature matching. Object detection is one of the biometric application which uses deep learning machine. In this case, deep learning model consists of two main parts which are the convolutional layers part and the full connected layers part (MLP part). The convolutional part is responsible for the feature extraction from the raw data while the full connected part will classify these features into categories [2]. Figure(7) shows an example for object detection steps using deep learning machine [2].

The second biometric application which uses the deep learning machine is object segmentation. A machine learning could classify any image in pixel level (pixel by pixel classification) which is called semantic image segmentation. The semantic segmentation could be applied in different computer vision applications such as understanding the whole imaging scene, inferring the relations between objects inside the image, removing the impact of object occlusion such as eyeglasses from the important biometric region such as ocular region, and extracting biometric trait such as eye region from facial images.

In general, DCNN is not enough to localize and segment objects inside image accurately, research combine DCNN with other machine learning task such as fully connected (CRF) [51] to obtain more accuracy in semantic image segmentation. In the same manner, other research follows the core DCNN segmentation model by pre-trained CNN model such as Visual Geometry Group (VGG16) [52].

Another exciting application of deep machine learning for a biometric system is feature extraction. It is essential to find the best-extracted features for biometric trait obtained from sensors such as facial images, voice, and human behaviors. Also, the conventional features of computer vision showed un-relenting progress to develop the accuracy of biometric systems [53]. Besides, the deep models such as Google networks, and VGG networks introduce new significant progress with the development of computer hardware such as GPU. This progress of deep models outer-perform the conventional features descriptors. Pre-Trained DCNN models such as VGG16, VGG19, AlexNet, ZFNet, LeNet-5, ResNet and GoogleNet used to extract micro-structure features from the image scene. These pre-trained models vary in their architectures which are consist mainly of stages of the convolutional, and max-pooling layer that allows it to filter the information at a different level. For instance, VGG16 (figure(8)) is a pre-trained CNN model consists of (16) layers involving mainly five stages of convolutional and max-pooling layers, three stages of fully connect (MLP) layers, with one softmax layer for classification. In that case, the convolutional and max-pooling layers are used for feature extraction by obtaining the information from the last max-pooling layer with a pre-trained model.

In addition to feature extraction, feature matching could be obtained using deep

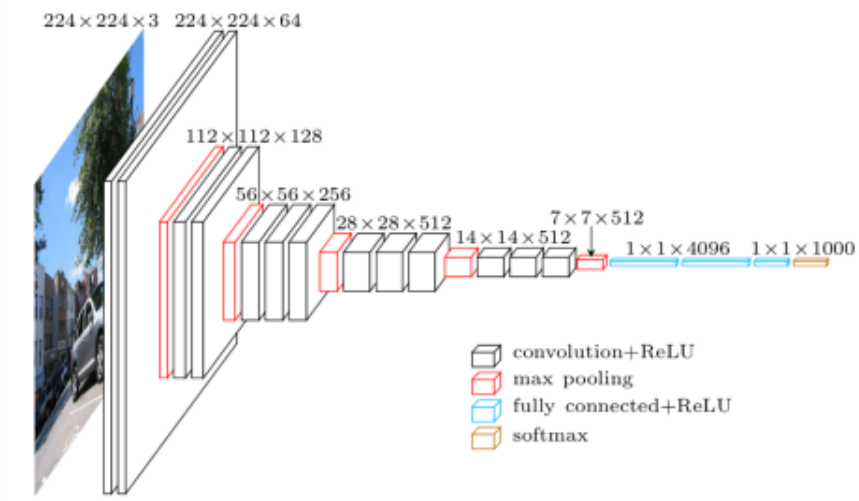


Figure 8: The architecture of VGG16 Model. [3]

models. In that case, the deep learning models could act as a feature comparison between the deep feature template which is acquired and stored during the enrollment process and the input query during verification or identification process. In a traditional biometric system, the feature matching process obtained using distance function such as Euclidean distance while the deep learning models could match features using fully connected layers which are Multi-Layer Perceptrons (MLP) and usually called learning distance metrics. An excellent example for this application is MatchNet. This network consists of three main parts which are feature network, metric network, and MatchNet training [4].

Figure (9) shows the architecture of MatchNet network [4] which consists of five convolutional layers, three max-pooling layers, three fully connect layers, and one softmax layer. In that case, the convolutional layer and max-pooling layer were used for extracting the deep features while the fully connected layer and softmax layer was used

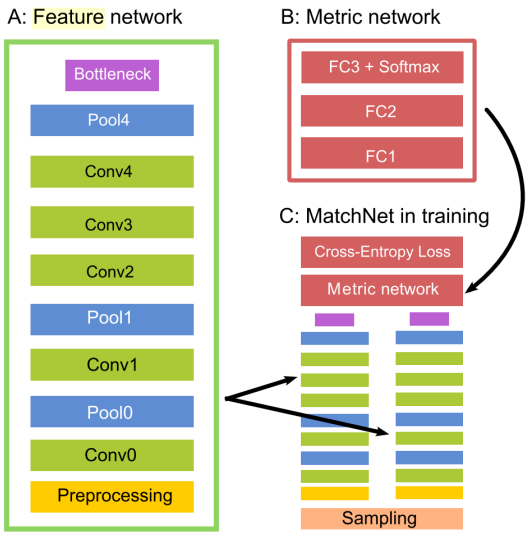


Figure 9: The architecture of MatchNet Model. [4]

for feature matching.

2.2 Feature Descriptors in Computer Vision

Any biometric trait such as a face, iris, or eyebrow could be acquired as raw data, and this data could be represented as a specific number of features such as binary, categorical, or continues. A feature is the equivalent phase of the biometric trait or modality which could be stored as template or input query in a biometric dataset. However, there are some challenges related to find the best data representation or feature which is related to different parameters such environment nature (i.e., mobile environment), sensor type (i.e., NIR sensor for iris), biometric trait type (i.e., physical or behavioral traits), and other parameters [54, 55].

2.2.1 Important Considerations for Feature Extraction

To achieve a good performance of features matching for a biometric system, the following preprocessing operations and considerations should be covered [54]:

- **Standardization**: Suppose there are two characteristic biometric traits such as the hight and width of human face measured in millimeter, and centimeter respectively. In that case, these traits should be standardized in order to use it with the biometric system, so both of these traits should a unique unit measure to have a resealable matching or fusion. Also, a statistical standardization (also called Z-score standardization) could be applied on the extracted feature that depends on the mean μ_i and standard deviation σ_i of feature f_i, so the standardized feature $f_i^{'}$ could be obtained using this equation $f_i^{'} = (f_i - \mu_i)/\sigma_i$.

- **Normalization**: In order to obtain an efficient features from image x with n pixels with different intensities, it is better to divided the image by the total number of count (image distribution) into to have a normalized and uninformed the image before the feature extraction $x' = x/||x||$. This operation will allow the feature to be dependable from image size.

- **Scaling**: A binary scaling could be applied on the extracted feature to scale the feature between two specific range (i.e., zero and one). This scaling depends on the minimum min_i and the maximum max_i of the values of the features. The scaled feature $f_i^{'}$ could be obtained using this equation $f_i^{'} = (f_i^{'} - min_i)/(max_i - min_i)$.

- **Signal Enhancement**: Signals with different dimensions such voice (1D signal),

gray image (2D signal), and colored image (3D signals) have a common noise metric called signal-to-noise ratio $SNR = Power_{Signal}/Power_{Noise}$ which could be improved by reducing the amount of noise via applying some preprocessing such filters, background remove, de-noising [56], or morphological image analysis [57].

- **Extraction of Structured Features**: Structured features could be extracted using specific convolutional kernels via deep models. These features could achieve a significant improvement to the biometric system and possess good representation for the raw data.

- **Reduction**: Some of the local features or deep features have a very high dimension (feature size) which contain a redundant representation for the raw data; also it caused a time-consuming for features matching. Therefore, data reduction techniques could be used to avoid the above problems. For instance, Principal Component Analysis (PCA) and Multidimensional Scaling (MDS) are a common technique for data reduction.

- **Discretization**: The operation that involves converting the continuous data into finite discrete data. In that case, the features will be understandable and simplified.

There are different categories of feature descriptors such local, global, and deep features as explained in next sections.

2.2.2 Local Image Descriptors

Local image descriptors are obtained from image locally by dividing the whole image into parts (sub-images or cells) in repeated form, then these features could be encoded in a specific representation (such as histogram) to be invariant to image transformation such as translations, rotation, scaling, and affine [58].

Three parameters should be applied on feature extraction for local features to obtain a good quality features which are repeatable, precise, and distinctive. The feature extraction should be repeated locally to every part of the image (image cells), and these features should be computed accurately for each cell. Also, these features should be unique on different images in which it could be distinctive during the matching process. To extract a local feature, distinctive key points should be located inside the image, then the regions around these key points should be defined. Consequently, the region content should be normalized, then the descriptor could be computed inside these regions [59].

There are some common local image descriptors which are widely used by biometric researches as shown below:

- **Histogram of Oriented Gradient (HOG)**: This image descriptor was first time introduced and applied on detection problem, namely, pedestrian detection in still images by two researchers Bill Triggs and Navneet Dalal [60]. After that, other research shows more work which leads to human, animals, and vehicle detection not only for still images but for videos and films [60]. Figure (10) shows a visualization of HOG. HOG operation involved different steps. First, the image is divided into small square parts called (cells), then the gradient histogram is computed and

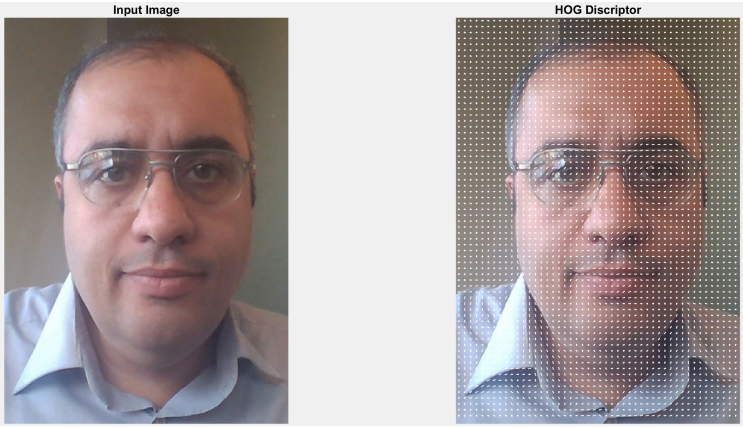


Figure 10: The HOG visualization

normalized for each cell, and finally, the descriptor is return [61].

- **Local Binary Pattern (LBP)**: This descriptor is considered as the best image descriptor to recognize texture pattern in images. It was originally introduced by Ojala [62] which was applied on images with multiresolution grayscale. In that research, LBP shows a robust work against rotation and multi-resolution scale to classify images with different textures. Figure (11) shows the steps of LBP operation which starts to divide an input image into (N) blocks, and the descriptor operates labels the 3x3 neighborhood inside each block. Then, LBP operator create a binary threshold mask (M_{th}) by comparing the centric value (g_{center}) of the 3×3 cell (M_{cell}) with each value in this cell (g_i), if ($g_{center} <= g_i$) a 1 label will be assigned. After that, the binary threshold mask (M_{th}) is multiplied by the binary weighted mask (M_w) to produce the resulted pattern (Mp). The resulted mask value is summed to form one value in the histogram of a certain block. Moreover,

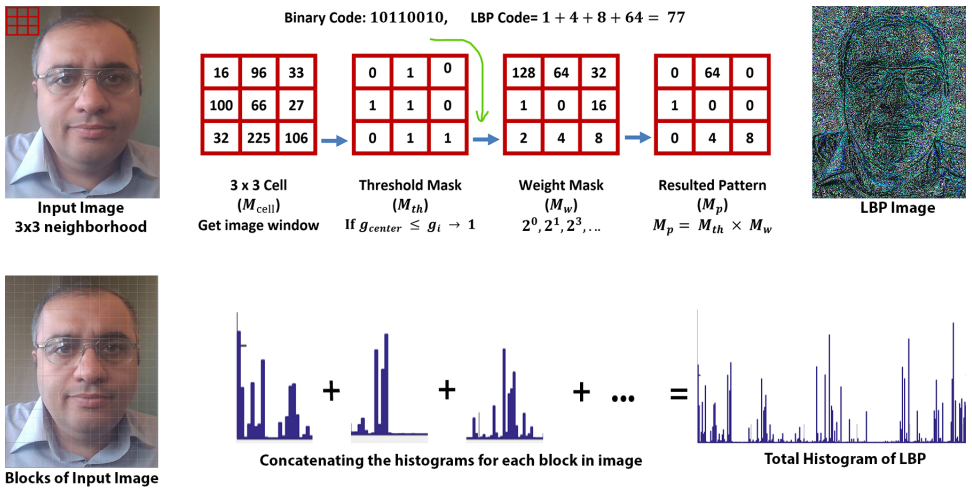


Figure 11: The Steps involved in LBP feature extraction.

the histogram of each N block is concatenated to form the final histogram of LBP.

2.2.3 Global Image Descriptors

These descriptors poss the capability of generalizing a representation for an entire image as a single vector which may describe the color, texture, and shape of an image. In contrast, local image descriptors which are operated on different image division (image blocks or cell), and are used to calculate the feature at different points of interest to recognize objects inside the image. Global image descriptors show a more robust and suitable performance on similarity image applications such as Google Similar Image (GSI) application, in this case, the images could be pre-filtered using text similarity [63].

There are some common global image descriptors which are widely used by biometric researches as shown below:

- **GIST**: Oliva and Torralba [64] introduced this global image descriptor. In their

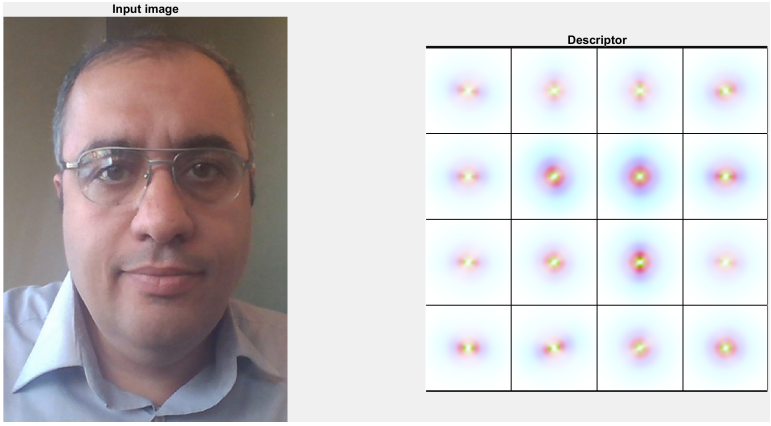


Figure 12: The visualization of GIST image descriptor.

work, they proposed a recognition model to recognize the scenes for real-world applications. This model computes the image representation (they called it *Spatial Envelope*) at very low dimension by passing the segmentation, region, and individual object inside the scene. In that case, the spatial envelope model could show certain information such as identity or shape of the object, and probable semantic category. GIST descriptors use a 32 Gabor filters with 4 scales, and 8 orientation. Then, these features maps are divided into 16 region, and for each region, the average of these features is calculated. After this, a concatenation of 16 averaged feature with 32 features map is applied to produce a GIST feature vector of size $16 \times 32 = 512$ [65]. Moreover, GIST shows a robust performance in different applications such as classification of traffic scenes [66], web-scale image search [67], object recognition [68]. Figure (12) shows a visualization for GIST image descriptor.

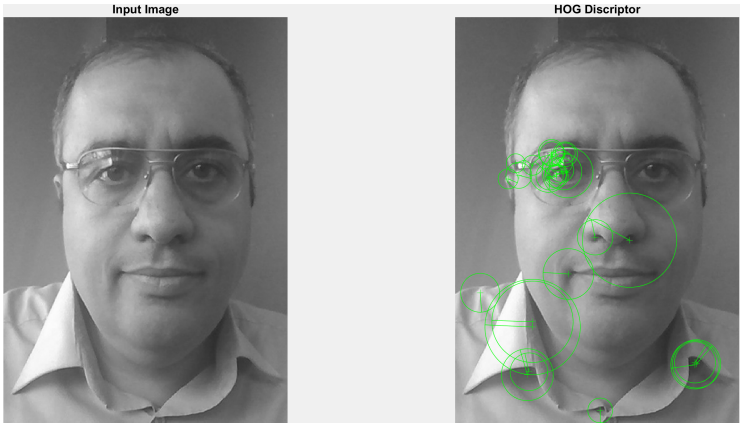


Figure 13: The visualization of SURF image descriptor.

- **Speeded-up Robust Features (SURF)**: This image descriptor was first introduced on 2006 by Herbert Bay [69]. SURF is considered as an updated work of previous image descriptor named Scale Invariant Feature Transform (SIFT) in which some enhancements applied into SIFT to introduce another faster and robust descriptor over SIFT. First of all, SURF localize the interesting point inside the image using the Hessian detector which is a matrix calculation of convolution of Gaussian second order derivative with image I at a point x. On another side, SIFT uses Laplacian of Gaussian (LoG) with Difference of Gaussians (DoG) instead. Then, the SURF descriptor is calculated at the interesting point using Haar wavelet quickly using integral images [70]. Also, this descriptor could be used in different applications such as forgery detection [71]. Figure (13) shows a visualization of this descriptor.

Figure 14: The visualization of deep image descriptor.

2.2.4 Deep Learning Image Descriptors

Recently, the deep image representation shows a robust and outer-performance over the handcraft descriptors such as local and global image descriptors which were mentioned in the previous sections (section 2.2.2, and section 2.2.3). Deep CNN models have been used for image retrieval to extract a micro-structure feature using the convolutional layers. A pre-trained CNN-based model used to obtain deep features using cascaded convolutional and max-pooling layers. In this case, these CNN-based models usually acquired local features from the input image using networks which were previously trained on big dataset such as ImageNet [72]. Moreover, the deep features could be used in different applications such as visual tracking [73], and image classification [74]. Figure (14) shows the visualization of deep image descriptor.

There are some common pre-trained CNN models for deep image descriptors

which are widely used by biometric researches as shown below [75]:

- **VGGNet**: The Visual Geometry Group Networks (VGGNet) was first introduced by developed by Simonyan and Zisserman [75] in 2014. Two models were introduced VGG16 and VGG19 which have 16 and 19 layers respectively. The filter size which used in the convolutional layer was 3×3, and these two models have been trained in 4 GPU for 2-3 weeks. Moreover, the VGGNet was implemented with 138 million parameters trained on ImageNet dataset. Also, the VGG16 model is the common and preferred model for feature extraction since it shows a robust performance in different light variation [76]. The standard size of input image for this model is $224 \times 244 \times 3$, and feature size is $7 \times 7 \times 512 = 25,088$ (see figure (8)).

- ResNet: The Residual Neural Network (ResNet) model was first implemented by Kaiming He who introduced a novel model with skip connection (gated recurrent units) which could give features with huge batch normalization [77]. Also, this model was implemented with 50 layer; however, this model had a lower complexity comparing with VGGNet [72].

CHAPTER 3

EYEGLASSES DETECTION AND SEGMENTATION

3.1 Introduction to Eyeglasses Detection and Segmentation

Mobile devices are playing a significant role in daily life, not only for communications but also for entertainment, work, or social networks. Along with the rapid increase in smart-phone applications, the number of sensitive data that these devices are accessing is also increasing (e.g., bank accounts, personal e-mails, photographs etc.). Thus there's a need to protect the access to such sensitive data and services. Biometrics is natural, and secure alternative mechanism over passwords for such protection [78], [79].

In this regard, smartphone devices show a dramatic advancement in the technology. Also, these devices pose advanced sensors such as cameras and microphones which allow usage of biometrics as a means of access control [40]. In this context, face and ocular biometrics have received significant attention. This is because this biometrics can be acquired using the regular RGB camera of the smartphones.

Face biometrics is one of the most popular and widely adopted biometric traits due to being natural and non-intrusive. General face recognition pipeline consists of image acquisition, face detection, image pre-processing, feature extraction and matching. Face detection consists of detecting the facial region of interest (ROI) from the image. Then the detected face is preprocessed to account for intra-class variations such as those caused by lighting and pose, followed by a feature extraction and matching [80].

Figure 15: (A) reflective glasses, (B) shadow, and (C) dark images causing challenges in accurate eyeglasses detection. The upper and lower facial regions have been masked to preserve volunteers' privacy (VISOB database).

The challenges associated with face recognition can be attributed to factors such as pose, facial expression, illumination variations and the presence of intervening structural components such as eyeglasses among other things [80], [39]. These challenges have attracted researchers from various backgrounds such as psychology, pattern recognition, computer vision, and computer graphics.

Eyeglasses are considered as a confounding factor of face recognition systems. This is because eyeglasses due to frame occlusion and reflections may cover a significant part of the face over the ocular region. Furthermore, the presence of eyeglasses may cause inaccurate eye pair detection, a detrimental outcome given that most of the face recognition systems depend on precise eye pair detection for accurate face detection and registration, Occlusion of the ocular and periocular regions, due to the presence of eyeglasses may result in degraded biometric performance given the importance of these regions [39], [40], [12].

To mitigate the effect of eyeglasses, researchers have been working on eyeglasses detection [81], [5], [82], [83], [84], localization, and removal [85]. Accurate eyeglass detection is an important first step towards its localization and removal. The challenges

associated with accurate eyeglasses detection are due to reflections, shadows, and dark images due to low lighting or skin tone (Figure 15).

3.2 Previous Work on Eyeglasses Detection and Segmentation

Fernández et al., [81] proposed a detection scheme for eyeglasses using a learning-based approach. In this paper, a local feature extraction method which is the local binary pattern (LBP) and Supports Vector Machines (SVM) which achieve a detection accuracy of 98.65% on the LFW database.

Wu et al., [5] proposed AdaBoost approach for detecting eyeglasses with Support Vector Machine (SVM); also two feature extraction has been in this work which are Haar and Gabor features. Their experimental have been evaluated on the FERET database that showed an overall correct detection accuracy in the range of 95.5% to 98.0%.

Jing et al., [82] combined edge information which includes the strength and orientation with geometric features (like convexity, symmetry, smoothness, and continuity with a deformable contour) for eyeglasses scheme. The proposed scheme has been applied on a subset of frontal face images, reporting on eyeglasses detection rate of 99.5%.

Wu et al. [83] obtained the 3D plane of eyeglasses frame using the Hough transform with three dimensions that have been applied to stereo face images. The highest accuracy of eyeglasses frame detection was 80%.

Jang et al., [84] used the probability to detect the existence of eyeglass frame by calculating the likelihood of eyeglasses using edge information via computing Fisher's criterion value on the different region inside the inter-ocular area. Fisher's value gives a

cue to detect the eyeglasses existence. This work has been evaluated on Olivetti Research Ltd (ORL) database which contains a 10 images for overall a 40 subjects. The resulted fisher's values ranged from 4.9 to 9.6.

Mohammad et al., [86] proposed two schemes for eyeglasses detection, namely, non-learning based approach (consists of multiple filters applied on the frame bridge region), and learning based approach with two levels of fusion (feature level and decision level). The reported accuracy of 99% for VISOB database, and 97.9% for FERET database with non-learning based, also an overall detection accuracy of 100% for VISOB database, and 99.3% for FERET database with learning based.

Table (3) summarizes the previous work on eyeglasses detection. As can be seen, most of the existing work used either non-learning based or a learning-based method using traditional classifiers such as Multi-layer Perceptron (MLP) or Support Vector Machine (SVM). Most of the proposed approaches were evaluated on in-house datasets. The reported results vary between 80% to 100

On another hand, some researchers have been worked on eyeglasses removal. Table (4) shows the summary of the previous work on eyeglasses segmentation. The existence eyeglasses segmentation approaches used a classical method of computer vision and data reduction such as Principle Component Analysis (PCA) to achieve their goal. Also, the primary motivation of eyeglasses segmentation work is to improve the performance of a facial recognition system.

Chenyu Wu [85] proposed an image editing approach based on intelligent (Markov-Chain Monte Carlo) and face synthesis system to detect, locate, and remove eyeglass automatically from the frontal face image. The reported result shows a 90.3% of mean error for eyeglasses localization for eyeglasses be removed applied to their in-house database, and FERET database.

Du et al. [88] proposed a novel eyeglasses removal scheme by detecting the eyeglasses region first on frontal facial images, then the region synthesized by recursive error compensation of PCA reconstruction. Their scheme produces an image that does not contain eyeglasses trace or the shadow, and eyeglass reflection. authors reported a Root Mean Square (RMS) error of 7.5% to 5% between captured eyeglasses facial images and reconstructed facial images.

Wong et al. [89] proposed two reconstructed algorithms based on a fusion of thermal infrared (IR) images and visible images. The first eyeglasses removal algorithm based on visible information to remove eyeglass from a thermal facial image, and the second one based on the refined visible information. The result of the reconstruction algorithm shows an accuracy of correct reconstructed of 88% on Equinox and DHUFO dataset.

Cheng et al. [90] proposed an eyeglasses removal using morphological operations. They found the mask of eyeglasses by applying a dilation operation followed by image binarization after spitting the eyeglasses into two regions (right, and left). Then they filled the two selected blobs to remove the eyeballs, the resulted eyeglasses mask the left white area. The removal of eyeglasses may fail when the driver's head rotate over 30 degrees.

Jia et al. [91] suggested an algorithm for eyeglasses detection and removal based

on phase congruency and progressive inpainting. They constructed an elliptical model on the upper part of the face to shield eyes using prior information of facial feature; they detected the eyeglass frame texture via phase congruency. The final step of this algorithm is progressive inpainting to remove the eyeglasses frame from a face. This algorithm shows an improvement of 5% on facial recognition rate after removing the eyeglass frame in CAS-PEAL database.

Heo et al. [92] implemented an approach via two levels of fusions, data fusion, that fuse visual and thermal infrared (IR) images), and decision fusion, that combines the matching scores of individual face recognition modules. They detected the eyeglasses region in the thermal images by a fitting method as an ellipse shape, and then they replaced with a template of eye pattern to enhanced the data fusion. The reported results show an improvement of face recognition after glasses removal on the thermal face and data fusion. The experimental analysis is performed on the NIST and Equinox dataset.

Wong et al. [93] implemented a removal method for eyeglasses based on Active Appearance Model (AAM). They extracted the eyeglasses region using AAM search for the texture and color information. Then synthesized the facial image via PCA reconstruction. Their results show a 5% improvement on face recognition after eyeglasses removal.

Park et al. [94] created a removal method based on a recursive PCA reconstruction that applied on a color frontal facial image to generate a glasses and obtain reconstruction image for compensation. The reported results evaluated on KFDB database which show a 10% improvement over previous PCA reconstruction work.

Park et al. [95] proposed a glass removal based on recursive error compensation

using PCA reconstruction by first extracting glasses region via color and shape information; then they generated a natural looking of a facial image without glasses using recursive error compensation of PCA reconstructed. This paper shows an enhancement of 5% for facial recognition after removing the eyeglasses.

Table (4) summarizes previous work on eyeglasses segmentation. As can be seen, existing eyeglasses segmentation approaches used classical computer vision methods (such as Markov-Chain) along with data reduction schemes such as Principal Component Analysis (PCA) for eyeglasses segmentation. The performance of the proposed eyeglasses segmentation was evaluated using performance enhancement on existing facial recognition systems. The reported results of facial recognition improvement range between 5% to 10%.

3.3 The Proposed Approaches on Eyeglasses Detection and Segementation

Three main categories of eyeglasses detection have been proposed which are non-Learning-based, learning-based, deep learning-based. Next sections explain in details these approaches from a low level to a high level.

3.3.1 Non-Learning-Based Eyeglasses Detection

The first proposed scheme uses Viola-Jones to detect ocular ROI using facial geometric information, followed by glass detection (Figure (16)). Figure (17) shows the Viola-Jones [96] based ROI detection using geometric information. In this scheme, the maximum width (W_{max}) has been calculated using the maximum of (w_1) and (w_2) (see Figure 17) as detected by Viola-Jones. The minimum width (W_{min}) is estimated in the

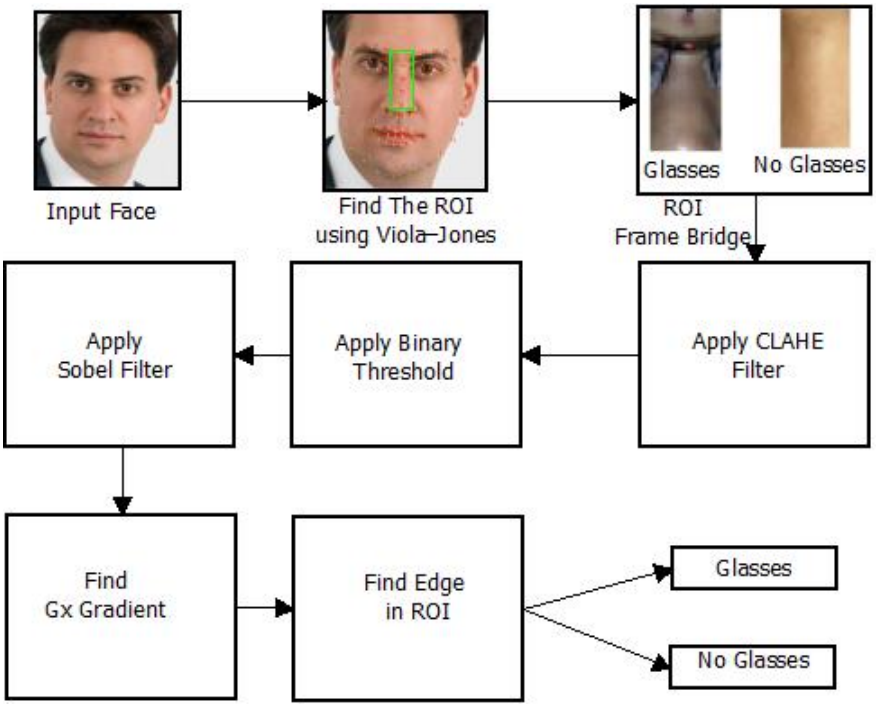


Figure 16: The block diagram for our non-learning based scheme.

same fashion. Further, the minimum width (W_{min}) was divided by the maximum width (W_{max}) to estimate the ratio (W_r) as explained in equation (3.1).

$$W_r = \frac{W_{min}}{W_{max}} \tag{3.1}$$

Other facial parts, such as the nose and the eyebrows were used as alternative landmarks to detect the ocular ROI when the earlier ROI finder fails due to reflection on glasses, as shown in Figure (18).

Under-exposed images are mitigated by applying CLAHE [3] to grayscale image. After that, Sobel filter is applied [97], and the gradients (Gx) and (Gy) are estimated. Further, the threshold for glass detection is selected in an adaptive manner based on the

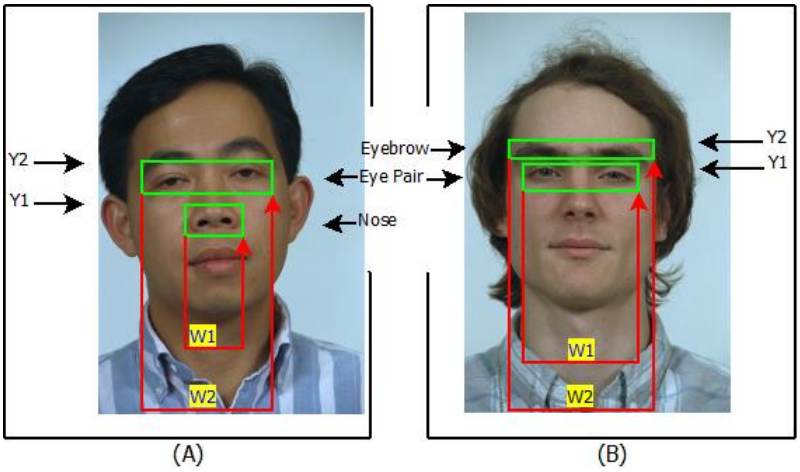


Figure 17: Viola-Jones based nose detector applied to eye pair region estimation; (A) original face, (B) detected nose region, and (C) detected eye pair region using facial geometric information.

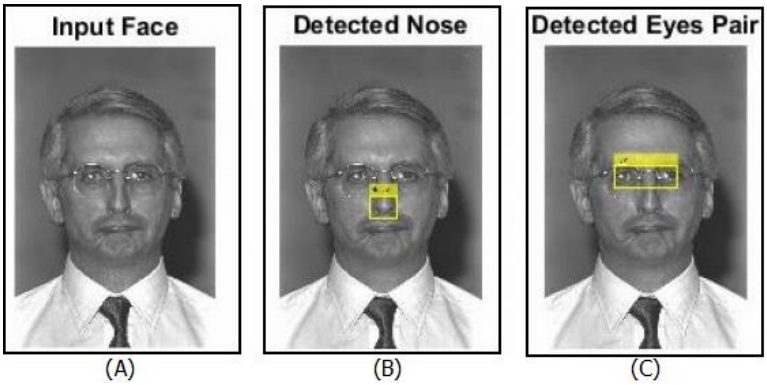


Figure 18: Viola-Jones based nose detector applied for eye pair region estimation; (A) original face, (B) detected nose region, and (C) detected eye pair region using facial geometric information.

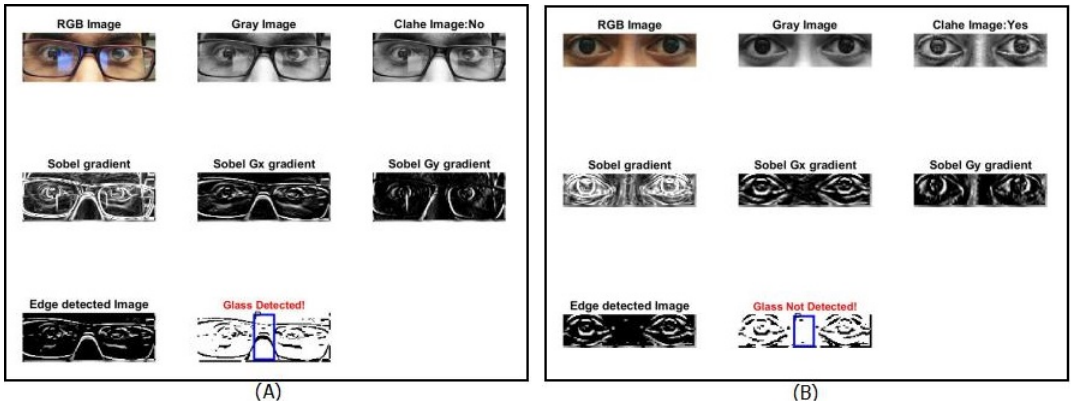


Figure 19: The proposed non-learning-based eyeglasses detection approach on sample images (A) with and (B) without eyeglasses (VISOB database).

illumination and skin tone, as shown in Figure (19).

3.3.2 Learning-Based Eyeglasses Detection

The learning-based scheme consists of three main steps (a) ROI detection, (b) Local Binary Pattern (LBP) and Histogram of Gradient (HOG) based feature extraction from the ROI, and (c) applying supervised learning using Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and Linear Discriminant Analysis (LDA) [98], [99], [100], [101], then fusing the output of these classifiers, as shown in Figure (20).

For ROI detection, a facial landmark localization was used using Dlib-ml library [35] instead of Viola-Jones and generated 68 landmarks for localizing eyeglasses region and the that of eyeglasses bridge using the location of the nose, ear, and eyebrows (see Figure (21)). The Dlib's 68 landmarks seem to be more robust than Viola-Jones in the presence of reflection or over and underexposure.

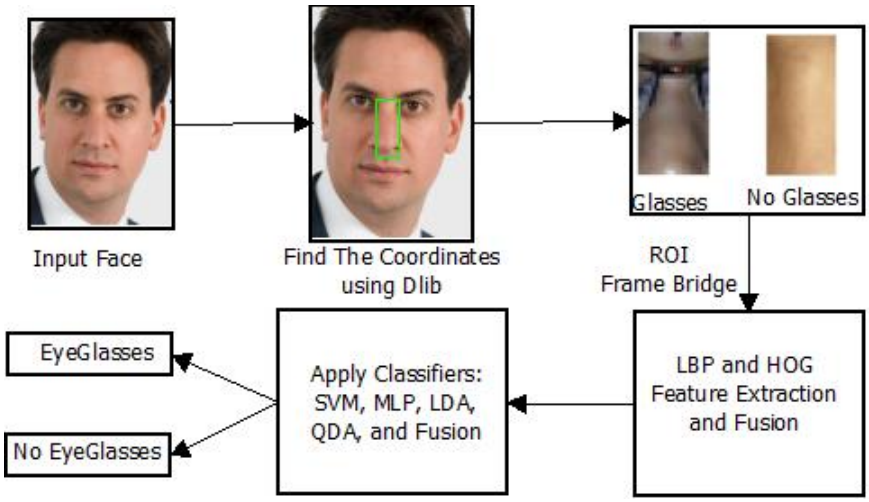


Figure 20: The block diagram of learning based eyeglasses detection scheme.

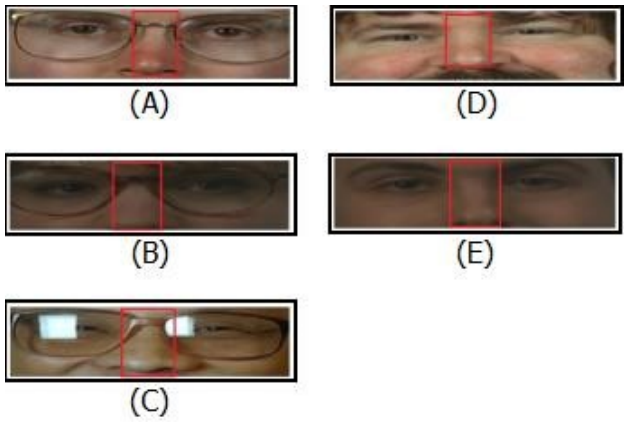


Figure 21: ROI detection using landmarks from Dlib (A) normal condition, (B) dark condition, (C) reflection, (D) no eyeglasses in office light, and (E) no eyeglasses in low light condition.

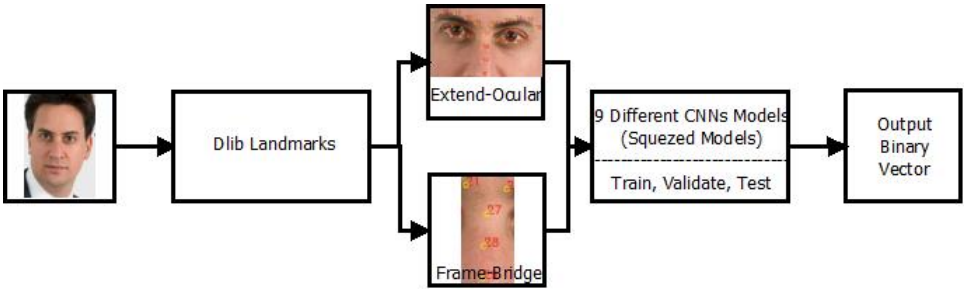


Figure 22: The block diagram of the glass detection approach using different CNN models.

3.3.3 Eyeglasses Detection using Squeezed CNN

In this work, two regions of interest (ROIs) was cropped using Dlib library [35]. These ROIs are the frame bridge (the majority part of the nose), and the extended ocular region (see Figure 22).

Next, these images have been normalized, randomized, shuffled, and divided into three datasets namely, training, validation, and test set. Training and validation sets are used to train and validate different CNN models. The test set is used to evaluate their accuracy. Table 5 shows nine different Convolutional Neural Network (CNN) models created using different kernel size, kernel numbers and number of convolutional layers. The last seven CNN model represents the squeezed model over the first two CNN models.

The first CNN model (Mod-01) has been designed with six 2D convolutional layers, the number of channels (filters) are (26, 16, 16, 16, 16, 16), and the size of each channel (filter) is (5×5). The benefit of these convolutional layers is to extract the microstructure feature from the ROI. Also, this model contains two fully connected layers for detection. While the second CNN model (Mod-02) has been implemented with four

2D convolutional layers, the number of filters are (26, 16, 16, 16), and the size of each filter is (5×5).

The rest seven squeezed CNN models have been designed with two fully connected layers, filter (channel size) of 3×3. However, the primary structure different between these squeezed CNN model is the number of channels for each 2D convolutional layer.

Model (Mod-03-S) has six 2D convolutional layer with filter numbers of (26, 16, 16, 16, 16, 16), while Model (Mod-04-S) has four 2D convolutional layer with filters number of (26, 16, 16, 16), Model (Mod-05-S) has only two 2D convolutional layer with filters number of (26, 16). The last four squeezed models (Mod-06-S, Mod-07-S, Mod-08-S, and Mod-09-S) have only one 2D convolutional layer with filters number of (16, 6, 3, 1) channels respectively.

These models have been squeezed to reduce the size of the model and the weight (network parameters) as shown in Figure 23, and Figure 24 for frame bridge based ROI, and the extended ocular region as ROI, respectively.

Figure 25 shows the number of operations which represent the training time cost for different implemented models that have been applied to two different ROIs namely, frame bridge and extended ocular region. The highest number of operations are in Mod-02 in both ROIs which are 85.56 and 509.02 million operations respectively, while the lowest number of operations are in the mod-09-s model for frame bridge ROI (8.66 million operations), and mod-07-s for extended ocular ROI (18.79 million operations). Also, the accurate and reasonable cost is mod-06 in both ROIs. However, a GeForce GTX 1070

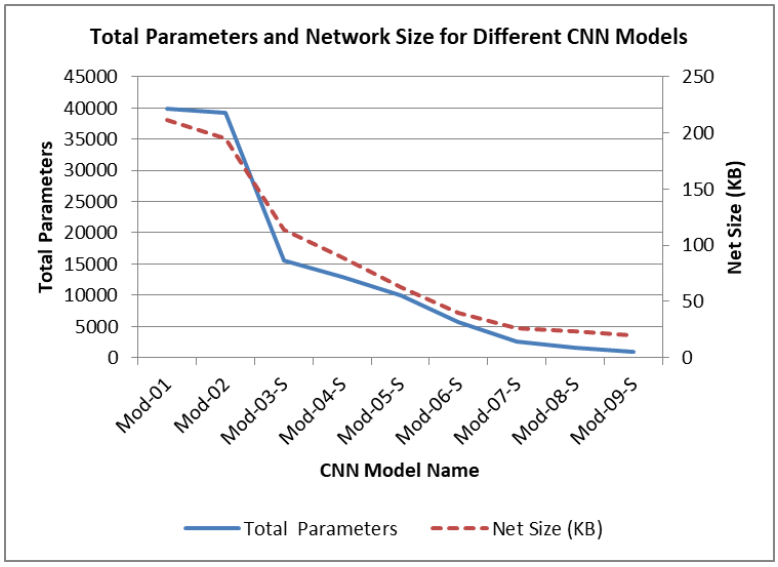


Figure 23: Total parameters and network size (KB) for different CNN models using Frame Bridge as ROI.

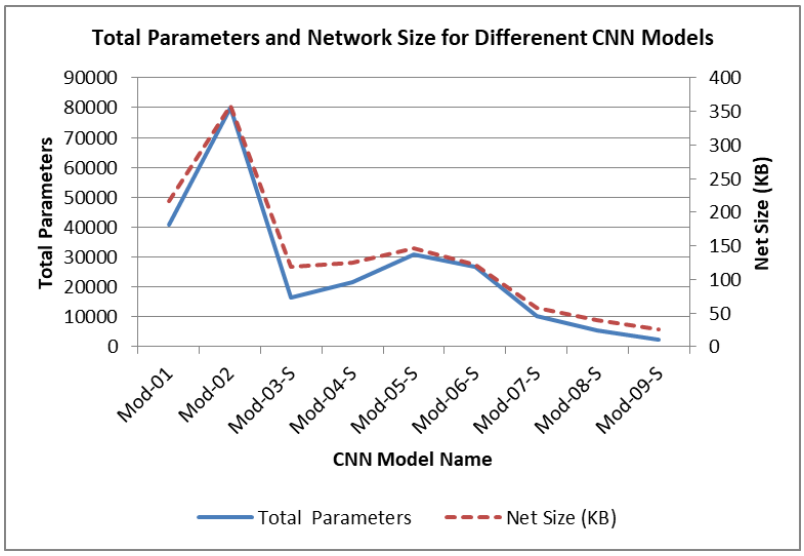


Figure 24: Total parameters and network size (KB) for different CNN models using extended ocular region as ROI.

8GB Turbo GPU has been used to train these model with a speed of 6.5 TFlops. The cost (number of operations) have been calculated using the equations listed below [10]. Equation 3.2 calculate the floating point operations per second ($FLOPS_1$) for Multi-layer Perceptron (MLP) or Fully connected (FC) layer, max-pooling, and convolutional layers. Equation 3.4 calculate the $FLOPS_2$ of batch layer. Also, equation 3.4 calculate the $FLOPS_3$ of rectifier linear unit (Relu) layer.

$$FLOPS_1 = InputSize \times LayerParameters \tag{3.2}$$

$$FLOPS_2 = InputSize \times (LayerParameters + 6) + 4 \tag{3.3}$$

$$FLOPS_3 = InputSize \times Filter_Number \times 2 \tag{3.4}$$

3.3.4 Cascaded CNN for Eyeglasses Detection and Segmentation

The proposed eyeglasses detection and segmentation uses cascaded convolutional neural networks (CNNs). Also, the primary motivation for this proposed approach is to prepare a trained CNN for eyeglasses segmentation. Moreover, this approach consists of four main stages, namely region of interest (ROI) detection, cascaded CNNs for eyeglasses detection, weight generation, and glasses segmentation, followed by mathematical and binarization operations (Figure **??**).

The ocular regions of interest (ROI) are cropped using Dlib library [35] that detects 68 coordinates from an input face image.

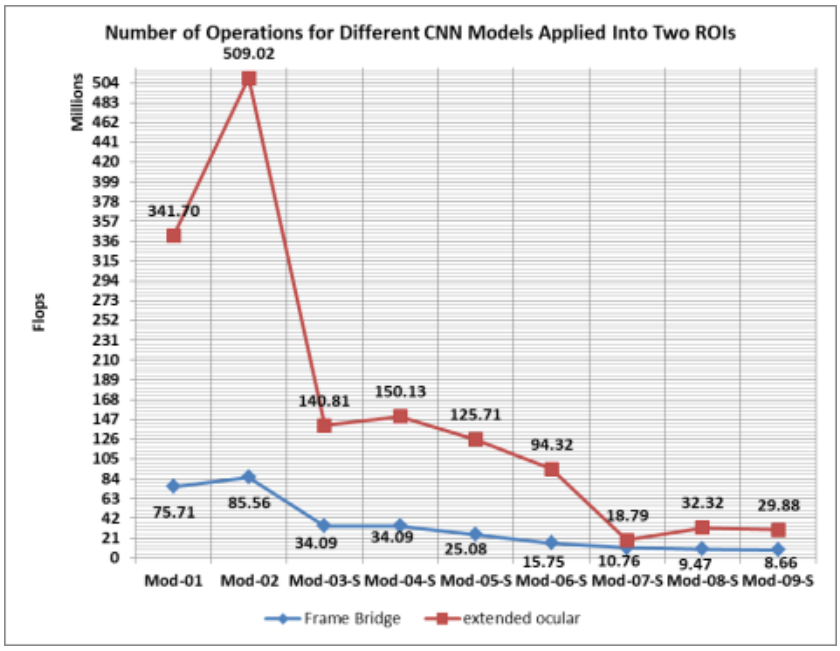


Figure 25: The number of operation for different CNN models using frame bridge and extended ocular region as ROIs.

3.3.4.1 CNN Architecture

The second stage of the proposed scheme was implemented leveraging the first CNN to generate the trained weights for the next CNN, and this network (CNN-01) is also used as eyeglasses detector. The CNN-01 was designed to deal with large-scale images to learn the weights of the microstructure features of those images. Table 6 details the architecture of CNN-01. The CNN-01 layers are then defined as follows:

1. The input ocular image of size $120 \times 480 \times 3$ is fed to the first convolutional layer which is constructed with 26 filters of size 5×5 and 2×2 stride to extract a microstructure feature from the input image. This convolutional layer is followed by the rectified linear unit (ReLU), 2×2 stride max pooling, and drop-out layer.

2. The output of first drop-out layer of size $30 \times 120 \times 26$ is fed to the second convolutional layer which is implemented with 16 filters of size 3×3 and 2×2 stride for the second level of feature extraction. This convolutional layer is followed by a rectified linear unit (ReLU), 2×2 stride max pooling, and drop-out layer.

3. The output of second drop-out layer of size $7 \times 30 \times 16$ is flattened to 3360 and is fed to two fully connected layers which contain 16, and 2 neurons with ReLU as activation layers.

4. The final layer is a soft-max activation layer which is used to obtain the probability of eyeglasses existence.

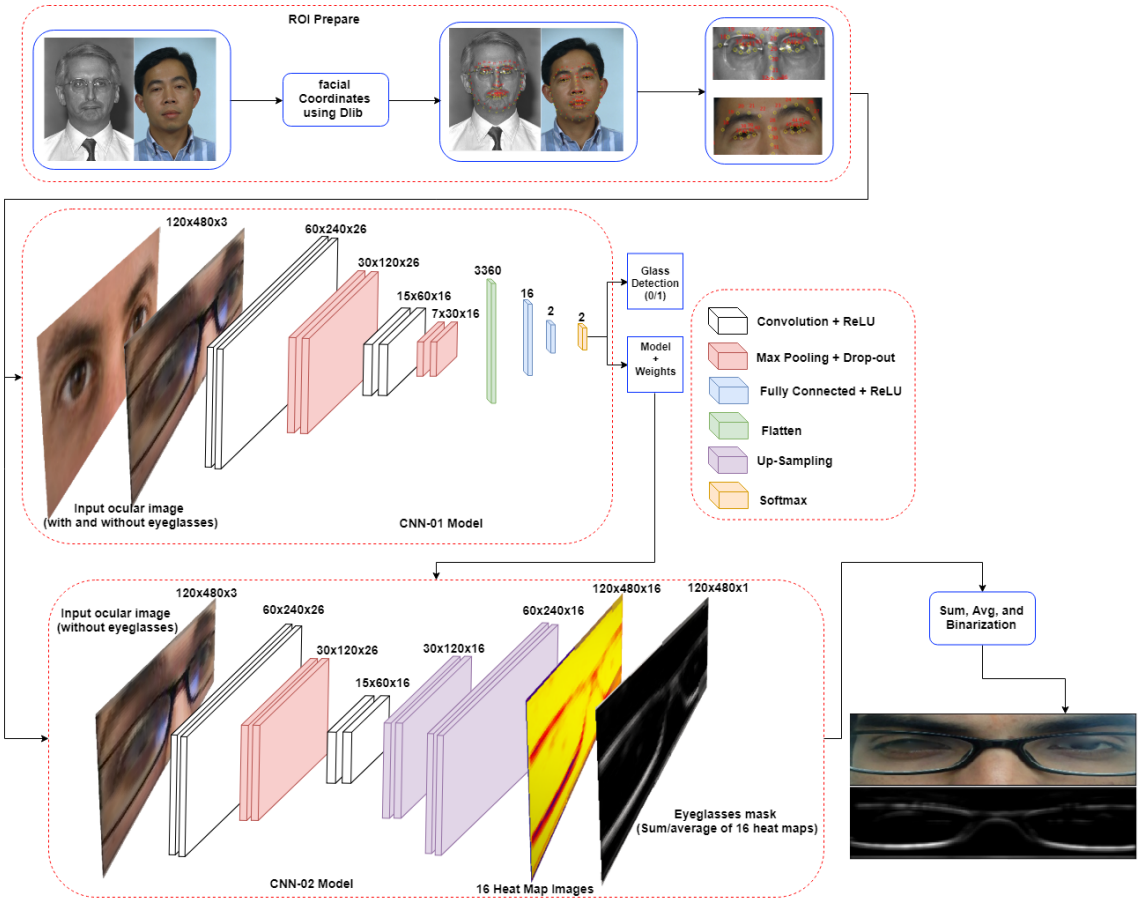


Figure 26: The block diagram of detailed architecture of the cascaded CNNs for eyeglasses detection and segmentation.

Figure (26) shows CNN architecture for eyeglasses detection (CNN-01) and the second CNN in our proposed network is the (CNN-02) network. This network was designed to generate sixteen heat maps from the up-sampling layer (last layer) for eyeglasses segmentation. Table 7 summarizes the CNN-02 architecture. The CNN-02 is consisting of the following:

1. The input ocular image of size $120 \times 480 \times 3$ is fed to the first convolutional layer which is constructed with 26 filters of size 5×5 and 2×2 stride to extract a micro-structure feature from the input image. This convolutional layer is followed by a rectified linear unit (ReLU), 2×2 stride max pooling, and drop-out layer.

2. The output of first drop-out layer of size $30 \times 120 \times 26$ is fed to the second convolutional layer which is implemented with 16 filters of size 3×3 and 2×2 stride for second level feature extraction. This convolutional layer is followed by a rectified linear unit (ReLU), 2×2 stride max pooling, and drop-out layer.

3. The output of second drop-out layer of size $7 \times 30 \times 16$ is fed to up-sampling layer to construct a sixteen heap map images of size $120 \times 480 \times 16$. This network was used with the trained weights from the previous network, CNN-01.

3.3.4.2 CNN Training and Testing

The proposed (CNN-01) model was trained on input ocular images with and without eyeglasses existence with a batch size of 128 using Adam optimizer [102] with an initial learning rate of 0.006. The weights were initialized using Gaussian distribution

with variance scaling to the size of the weight matrix. Also, The data have been labeled as 0, 1 for non-glasses, and glasses respectively for CNN-01 network training. Then, the saved (CNN-01) model and its weights were supplied into (CNN-02) model for frame prediction using input ocular image with eyeglasses existence.

3.3.4.3 Post-processing and Mask Generation

The summation and the averaging operations were applied to the output of CNN-02 to combine the sixteen heat map images into one. Then, a zero-one normalization and binarization operations were applied to the sum and average images to generate the segmentation mask for glasses. The normalization is essential to scale the values of resulting images from CNN-02 to standard binary scale.

3.3.5 Convolutional-Deconvolutional CNN for Eyeglasses Segmentation

Figure 27 shows the overall operations for the second scheme of eyeglasses frame segmentation. This scheme consists of six main stages, namely ROI extraction, frame in-painting, CNN model application, mask clean-up, segmentation and frame removal.

The ocular regions of interest (ROI) are cropped using Dlib landmark [35] as explained in the previous session. A convolutional deconvolutional neural network has been implemented as mentioned in the next section.

3.3.5.1 CNN Architecture

Table 8 and Figure 28 show the details architecture of Convolutional Deconvolutional Neural Network. This network is defined as following:

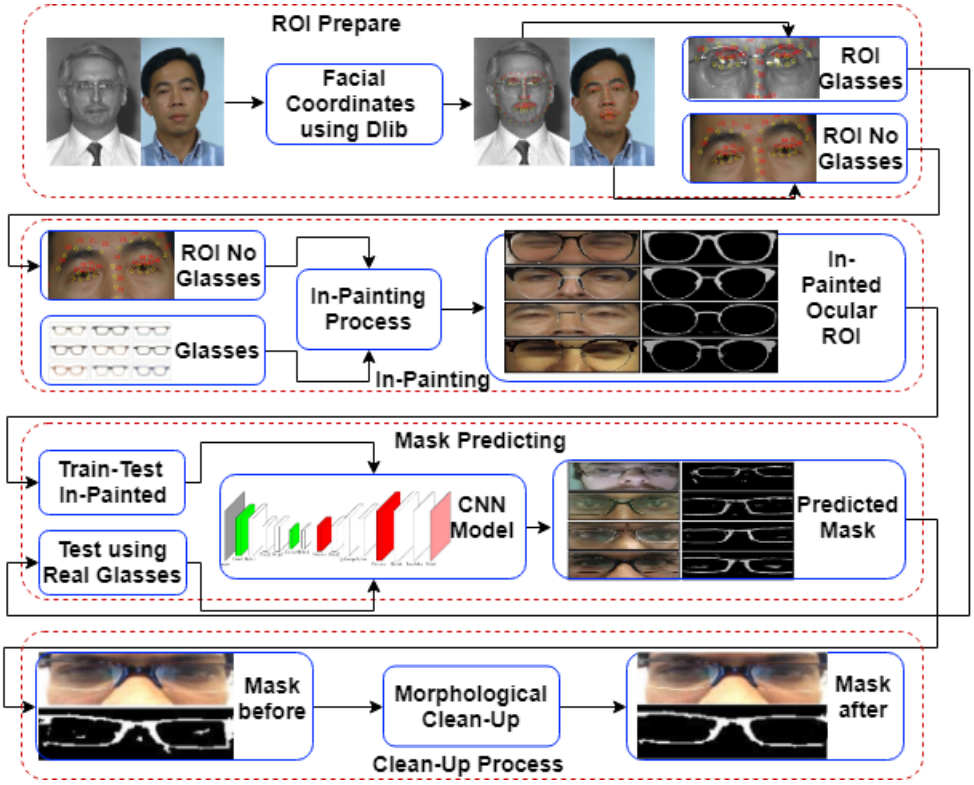


Figure 27: The block diagram of the proposed scheme of glasses segmentation using convoluational-deconvoluational neural network.

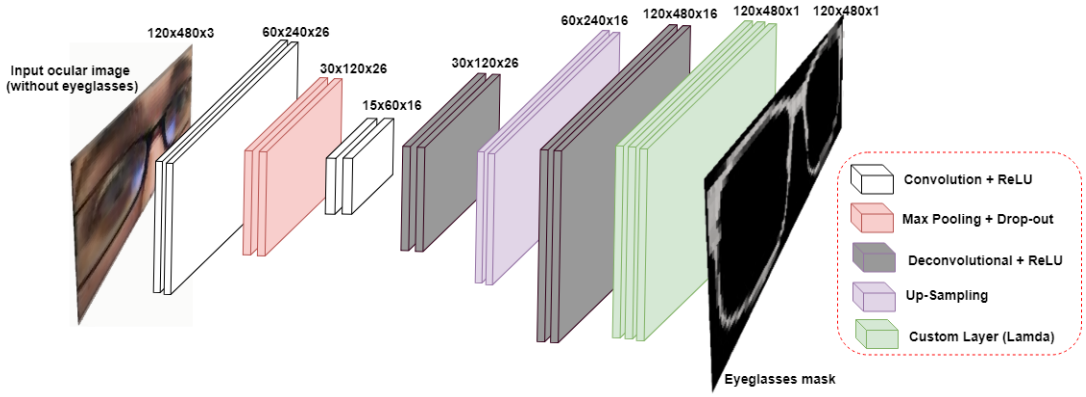


Figure 28: Detailed architecture of the convolutional deconvolutional neural network for eyeglasses segmentation.

1. The input ocular image of size $120 \times 480 \times 3$ is fed to the first convolutional layer which is constructed with 26 filters of size 5×5 and 2×2 stride to extract micro-structure features from the input image. This convolutional layer is followed by a rectified linear unit (ReLU), 2×2 stride max pooling, and drop-out layer.

2. The output of first drop-out layer of size $30 \times 120 \times 26$ is fed to the second convolutional layer which is implemented with 16 filters of size 3×3 and 2×2 stride for the second level of feature extraction. This convolutional layer is followed by a rectified linear unit (ReLU).

3. The output of ReLU of size $15 \times 60 \times 16$ is fed to the De-convolutional layer of 2×2 strides followed by ReLU layer.

4. A rectified input of size $30 \times 120 \times 16$ is fed to up-sampling layer of 2×2 strides followed by ReLU layer.

5. The resulted input of size $60 \times 240 \times 16$ is fed to the De-convolutional layer of 1×1 strides followed by ReLU layer.

6. The final layer is Lamda (customized) layer which is customized to satisfy four main operations which are the summation of the sixteen heat maps, apply Gaussian low pass filter, applying Otsu's binarization, finding the loss metric using intersection over the union between the predicted binary mask and the reference target binary mask.

3.3.5.2 CNN Training and Testing

The model has been designed to predict the binary mask for the in-painted test set, and the real glasses images after training and validating the proposed model with more than 3 million images. The predicted binary mask is post-processed using open area and connected components based morphological operations.

3.3.6 Database Information

3.3.7 Visible Ocular Biometric Database (VISOB)

This database consists of eye images of 550 healthy adult subjects obtained using three smart-phones: Oppo N1, Samsung Note 4, and iPhone 5s. Participants were asked to take a "selfie" via front-facing cameras of the phones in two sessions that were about 10 to 15 minutes apart. The distance between the camera and volunteers face was 8 to 12 inches apart [103].

For each session, the images were taken in three indoor lighting conditions: bright, normal and dim. For volunteers wearing eyeglasses, each capture was repeated with and without eyeglasses.

3.3.8 Face Recognition Technology Database (FERET)

The FERET database [104] was collected between August 1993 and July 1996 in 15 different sessions. The database contains 1564 sets of facial images, the total number of images are 14,126 images from 1199 subjects and 365 duplicate sets of images. The second set of images are duplicate set in which a person already in the database and was

usually taken on a different day. This database contains both grayscale and color images. The number of frontal face images from the FERET database is 1978 from 989 subjects, where 233 wore eyeglasses.

3.3.9 In-pained Eyeglasses Frames for Eyeglasses Segmentation

The cropped ocular region has been in-painted with the best purchases (200) frames in which (100) male glasses and (100) female glasses has been collected from different websites of eyeglasses stores. The glasses have been registered inside the extended ocular region using the Dlib landmarks of eyebrows, eyes, and nose.

Figure 29 shows in-painted databases (VISOB, FERET01, and FERET02) which have been divided, trained, validated, and evaluated using the proposed CNN model. In next stage, the three databases have been concatenated which consists more than 4.1 million images, divided as 80% (3301280) training set, 15% (619040) validation set, and 5% (4269) testing set for the CNN model. The reason for the small test set since we have tested the model with a large number of real glasses (not in-painted glasses).

3.4 Results of Eyeglasses Detection and Segmentation

The mentioned VISOB and FERET databases were used for performance evaluation of the proposed glass detection schemes. For VISOB, we randomly selected 20491 samples with and without glasses and 2274 samples with and without glasses, for training and validation, respectively, of the proposed scheme. The performance evaluation was done on unseen 1456 samples with and without glasses. For FERET, we randomly selected 1467 samples with and without glasses and 620 samples with and without glasses,

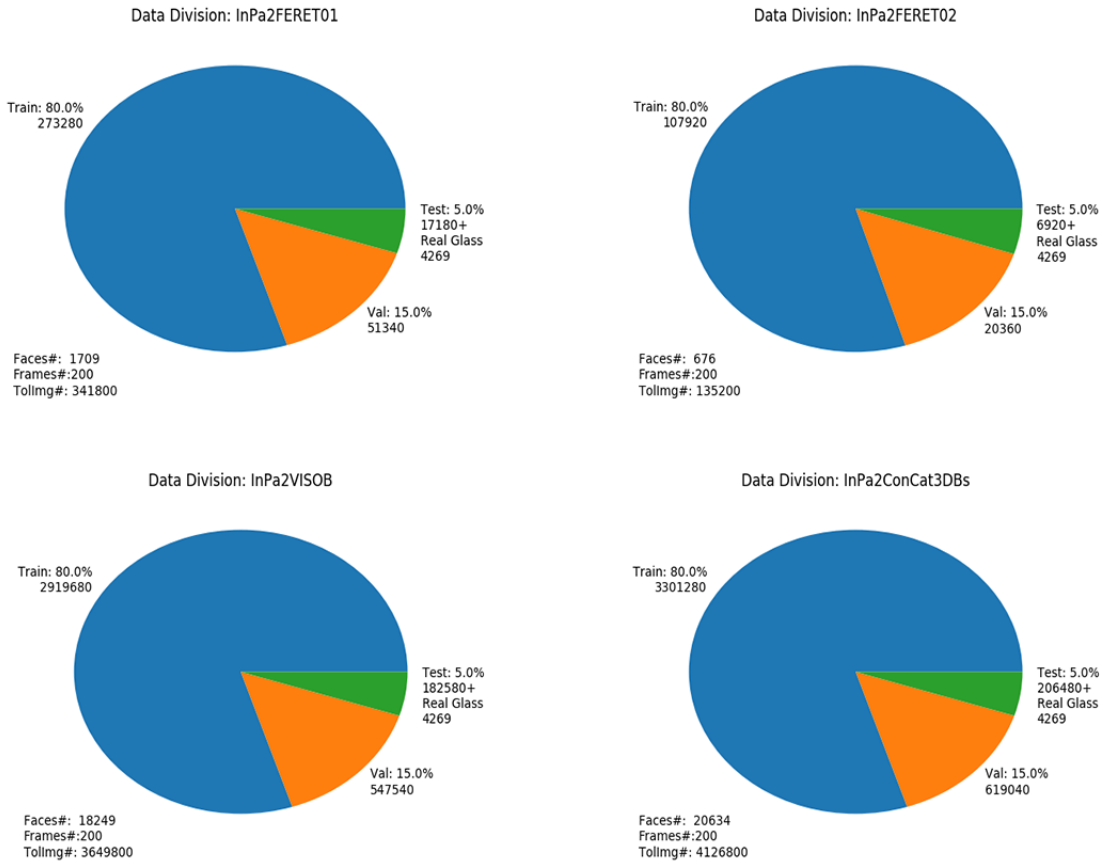
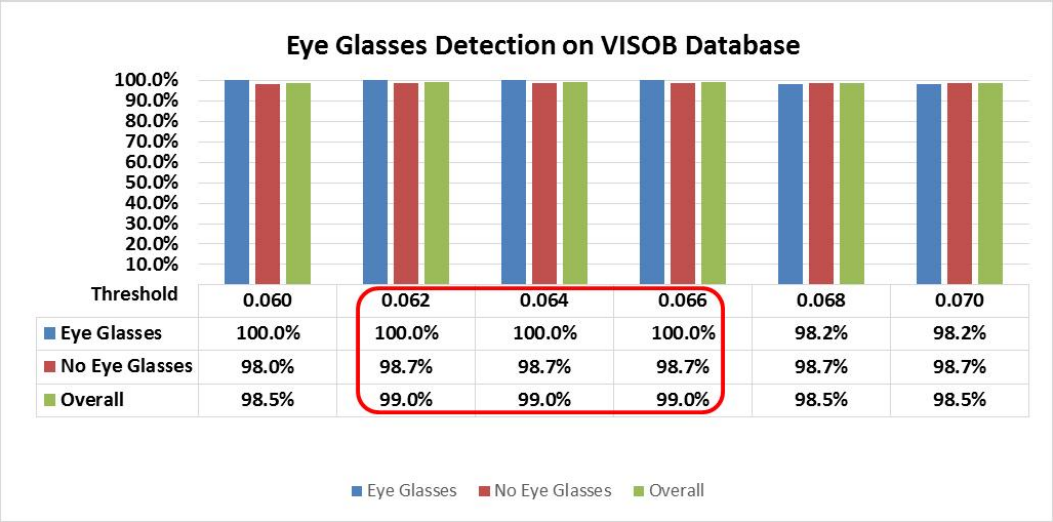


Figure 29: The In-Painting (FERET1, FERET2, VISOB, and Concatenated) databases.



Threshold	0.060	0.062	0.064	0.066	0.068	0.070
Eye Glasses	100.0%	100.0%	100.0%	100.0%	98.2%	98.2%
No Eye Glasses	98.0%	98.7%	98.7%	98.7%	98.7%	98.7%
Overall	98.5%	99.0%	99.0%	99.0%	98.5%	98.5%

Figure 30: The bar graph showing the performance of the proposed eyeglasses detection approach on VISOB database. The maximum overall obtained accuracy is 99.0%.

for training and validation, respectively, of the proposed scheme. The performance evaluation was done on unseen 616 samples with and without glasses.

3.4.1 Results of Non-Learning-Based Eyeglasses Detection

Figure 30 shows the performance of the proposed non-learning based eyeglasses detection approach on VISOB database. The eyeglasses detection accuracy is ascertained as a percentage of correctly classified ROI, correct non-glasses detection rate, correct glasses detection rate, and the overall accuracy; representing the True Positive Rate (TPR), True Negative Rate (TNR), and overall performance, respectively. It can be seen (Figure 14) that the best performances are obtained when the adaptive threshold used to detect eyeglasses varies in the range $[0.062, 0.066]$ with an overall accuracy of about 99.0%.

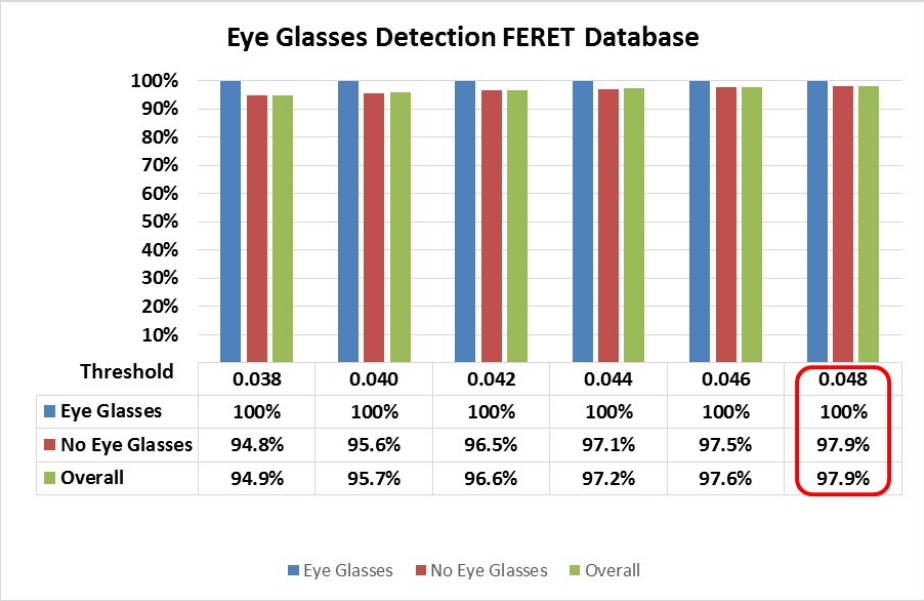


Figure 31: The bar graph showing the performance of the proposed eyeglasses detection approach on FERET database. The maximum overall obtained accuracy is 97.9%.

Figure 31 shows the performance of the proposed approach on the FERET database using our proposed method. The best performance of 97.9% accuracy for non-eyeglasses, 100% for eyeglasses, and 97.9% of the overall performance with adaptive threshold was obtained.

Table 9, shows a comparative evaluation of our proposed eyeglasses detection approach with an existing study [5] reporting their *best* results on FERET database. As can be seen, our method performs very well on challenging mobile VISOB database (Figure 14), and on FERET database (Figure 15) higher eyeglasses detection accuracy and no eyeglasses detection accuracy over [5].

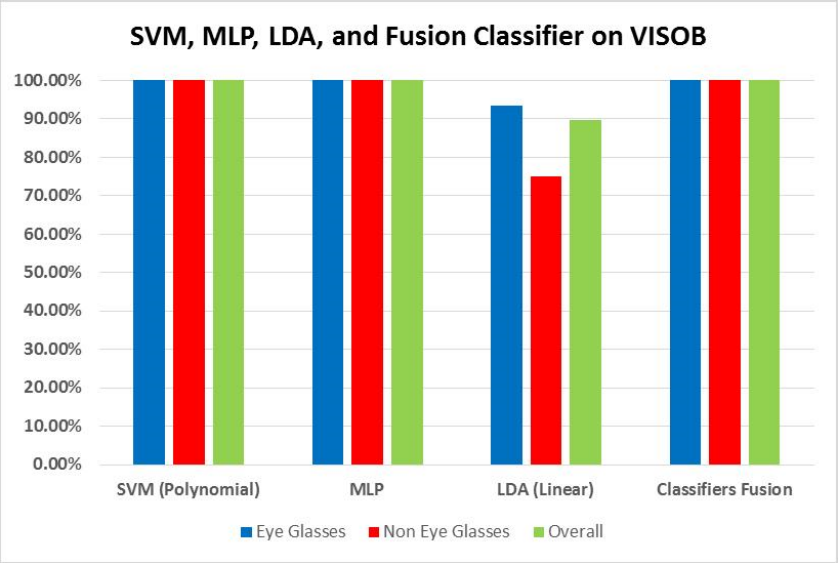


Figure 32: The bar graph showing the performance of the Learning based eyeglasses detection approach on VISOB database. The maximum overall accuracy is 100%.

3.4.2 Results of Learning-based Eyeglasses Detection

Figure 32 shows the performance of the proposed learning based eyeglasses detection approach on VISOB database. Again, the eyeglasses detection accuracy is ascertained as percentage of correctly classified ROI, correct non-glasses detection rate, correct glasses detection rate and the overall accuracy, representing True Positive Rate (TPR), True Negative Rate (TNR), and overall performance, respectively (Figure 32).

Figure 33 shows the performance of the proposed approach on the FERET database using an adaptive threshold for eyeglasses detection. The best performance obtained is 99.2% accuracy for non-eyeglasses detection, 100% for eyeglasses detection, and 99.3% for overall performance using SVM and fused classifiers.

Table 10 shows the performance of learning-based scheme for SVM, MLP, LDA

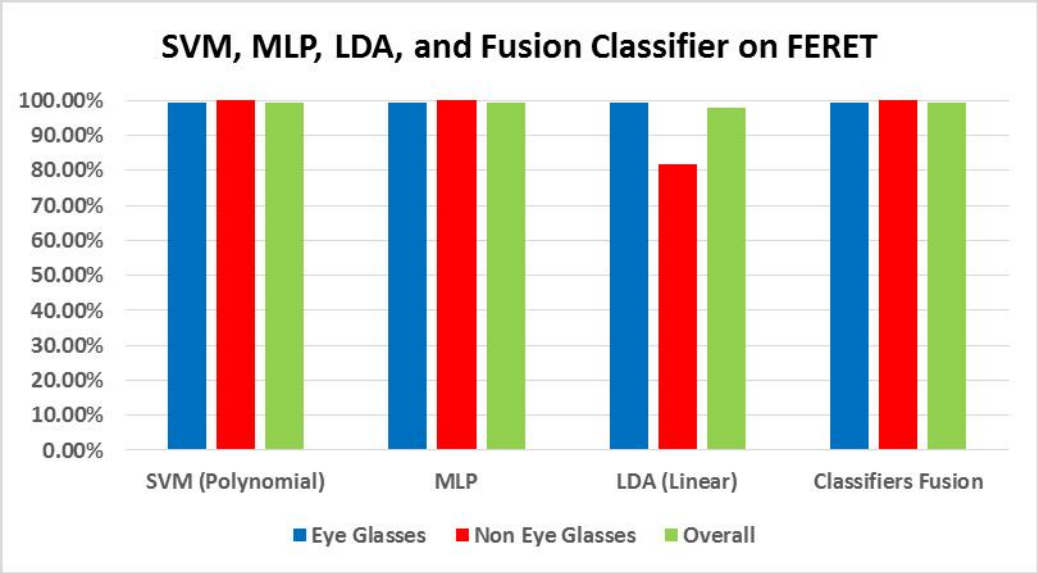


Figure 33: The bar graph showing the performance of the learning based eyeglasses detection approach on FERET database. The maximum overall accuracy is 99.3%.

and their fusion. SVM outperformed all other classifiers with an overall accuracy of 100%. Two level of fusion has been used in this scheme. At first level, feature fusion has been applied to concatenate LPB features with HOG features. At the second level, the output of SVM, MLP, and LDA have been fused at the decision level for majority voting. The proposed fusion classifier outperforms existing learning-based schemes based on individual features such as LBP [81].

3.4.3 Results of Squeezed CNNs for Eyeglasses Detection

Figure 34 shows the first evaluation results of nine different CNN models that have been applied on frame bridge region (the majority part of the nose region). The best results were obtained with the squeezed CNN model (Mod-06-S) with 99.97% of

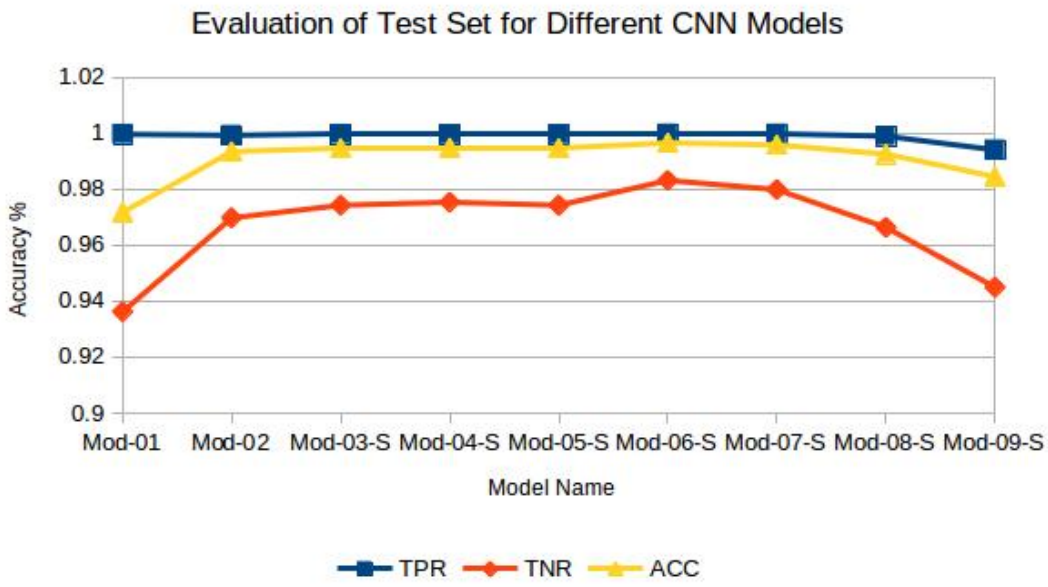


Figure 34: Evaluation of different CNN models using Frame Bridge as ROI.

True Positive Rate (TPR), 98.31% of True Negative Rate (TNR), and 99.96% of overall accuracy. This CNN model has a 5746 of total parameters (trainable, and non-trainable parameter), and 40.1 KB of network size (the file size of model and weights). The lowest result is obtained with the CNN model (Mod-01) with 99.96% of TPR, 93.62% of TNR, and 97.17% of overall accuracy. This CNN model has a 39878 of total parameters, and 211.2 KB of network size.

Figure 35 shows the evaluation results of nine different CNN models when applied to the extended ocular region (Figure 1). The best result has been obtained with the squeezed CNN model (Mod-06-S) with 99.95% of TPR, 99.33% of TNR, and 99.82% of overall accuracy. This CNN model has a 26578 of total parameters (trainable, and non-trainable parameter), and $122KB$ of network size (the file size of model and weights).

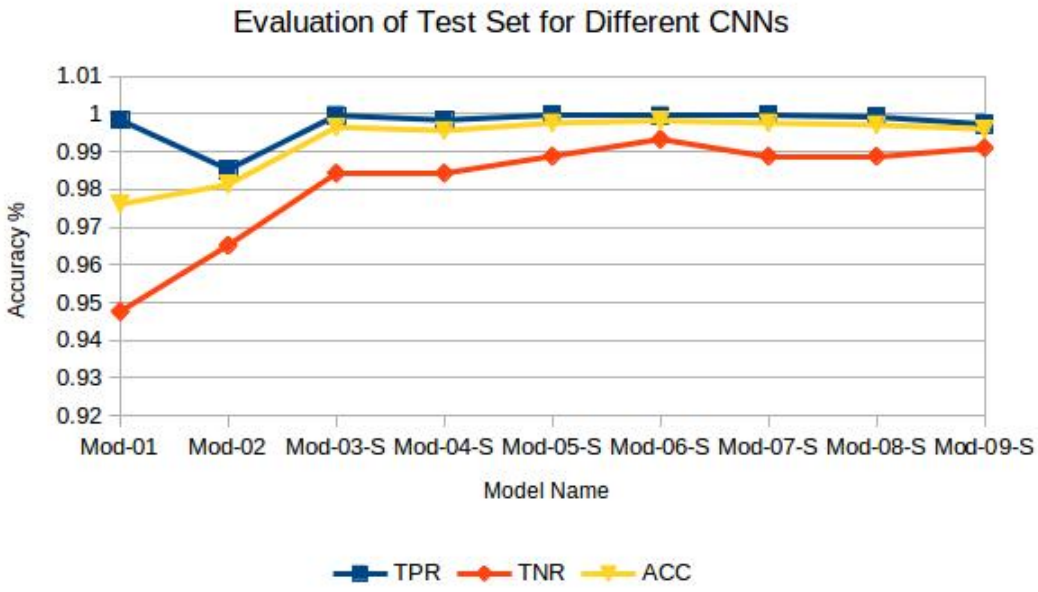


Figure 35: Evaluation of different CNN models using extended ocular region as ROI.

The lowest result has been obtained with the CNN model (Mod-01) with 99.96% of TPR, 93.62% of TNR, and 97.17% of overall accuracy. This CNN model has a 40838 of total parameters, and 216.3 KB of network size.

3.4.4 Results of Cascasded CNN for Eyeglasses Detection and Segmentation

The first cascaded network CNN-01 (section 3.3.4), was trained, validated, and tested using VISOB dataset with $16,000$ epochs and Adam optimizer. To evaluate CNN-01 (the weight generator, and eyeglasses detector), we randomly selected $23,826$ samples for training, and $1,191$ samples for validation, both with and without glasses. The performance evaluation of CNN-01 was done on unseen $1,192$ samples with 587 eyeglasses and 605 images without eyeglasses. We obtained an overall 100% glass detection accuracy

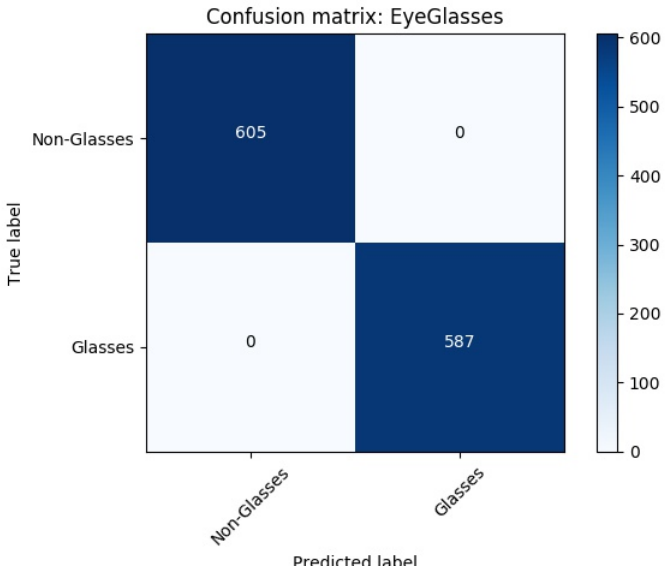


Figure 36: The performance of CNN-01 for eyeglasses detection.

(in Figure 36 using confusion matrix).

To evaluate CNN-02 (section (3.3.4)), an intersection over union ($Ratio_{Sum\ ||\ Avg}$ in Equation 3.5) was used as metric for eyeglasses segmentation. That perform pixel-wise operations between the target masks and the masks generated by the CNN.

$$Ratio_{Sum\ ||\ Avg} = \frac{\sum Ref_N \bigcap Out_N}{\sum Ref_N \bigcup Out_N} \tag{3.5}$$

In equation (3.5), Ref_N is the N target frame masks, Out_N is the N CNN output images after summing or averaging, $Ratio_{Sum}$ is the ratio of summing output masks, and ($Ratio_{Avg}$) is the ratio of averaging output masks. The first ratio ($Ratio_{Sum}$) in Equation 3.6 is intersection over union for the output mask after applying the summation operation on heatmap images obtained using Equation 3.5. While the second ration

($Ratio_{Avg}$) represents the intersection over union for the output mask after applying the averaging operation on heat map image.

The overall segmentation accuracy obtained using Equation 3.6 represents the average of two ratios.

$$OverallRatio = \frac{Ratio_{Sum} + Ratio_{Avg}}{2} \tag{3.6}$$

Using these metrics, 91% segmentation accuracy was obtained (see Equation 3.6).

Figure 37 shows sample eyeglasses segmentation masks obtained using cascaded approach, for ocular images under different lighting conditions. As mentioned previously, factors such as glasses with heavy reflections, thin frames, dim images, frame shadow, an absence of some part of the frame, and frameless glasses affect eyeglasses segmentation accuracy. Our proposed network obtained decent segmentation results for many of these challenging cases (see Figure 38).

3.4.5 Resutls of Convolutional Deconvoluational Neural Network

To evaluate the convolutional deconvoluational Neural Network, equation 3.5, and equation 3.6 have also been used. Using these matrices, 97% segmentation accuracy was obtained.

The test evaluation has been implemented using two different test sets. The first test set is the 5% of unseen images of conducted on the in-painted version of FERET01, FERET02, VISOB and the concatenation of them. The second test set is the unseen real eyeglasses images of VISOB dataset. Also, the in-painting process was implemented

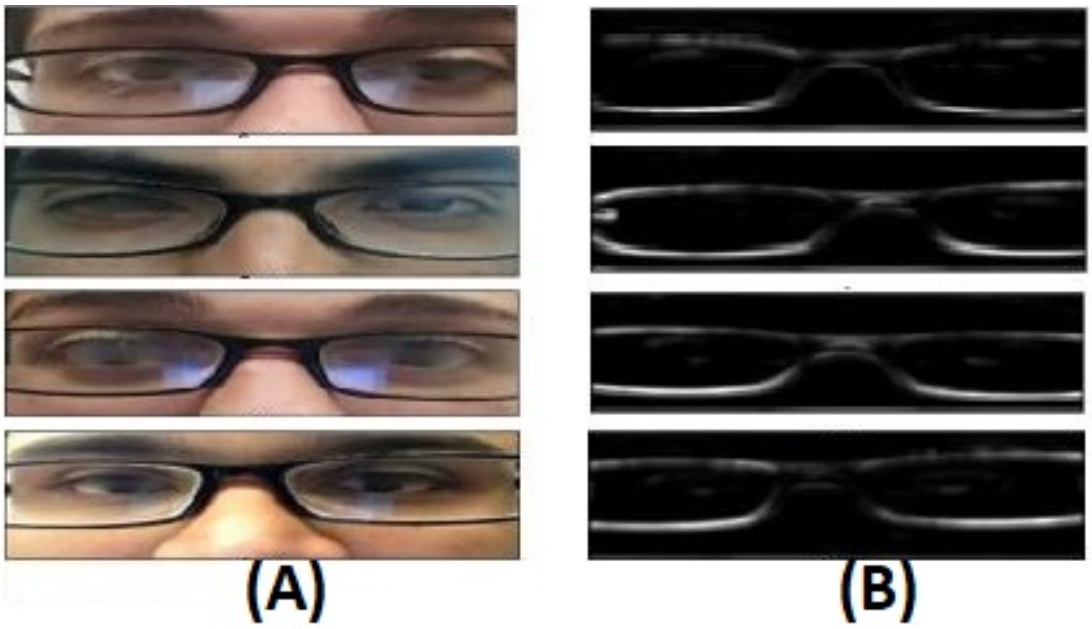


Figure 37: Example of eyeglasses segmentation obtained using cascaded approach, for ocular images under different lighting conditions.

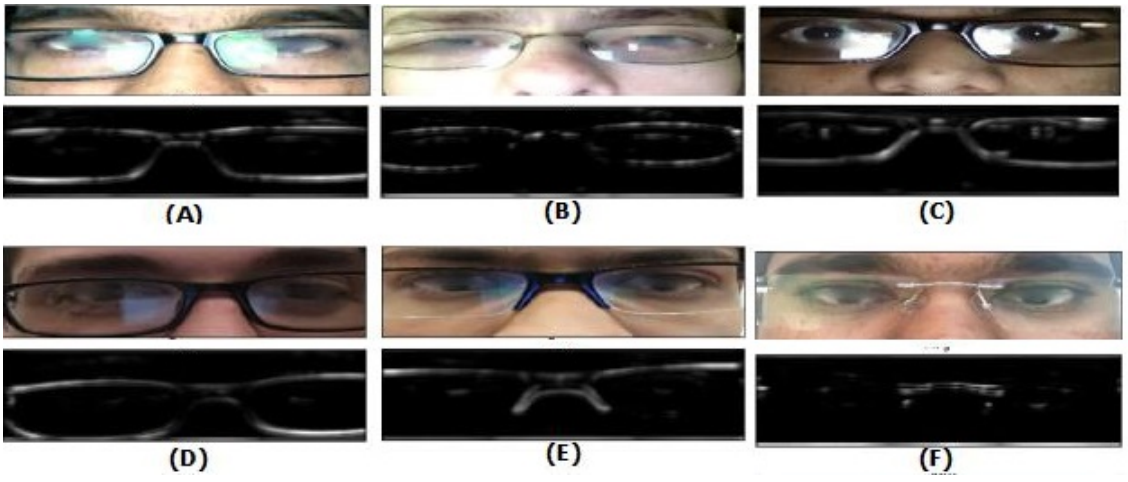


Figure 38: Sample results of eyeglasses segmentation using cascaded approach for challenging factors such (A) reflection, (B) reflection along with thin frames, (C) dim image with reflection, (D) frame shadow, (E) absence of the lower part of the frame, and (F) frameless glasses.

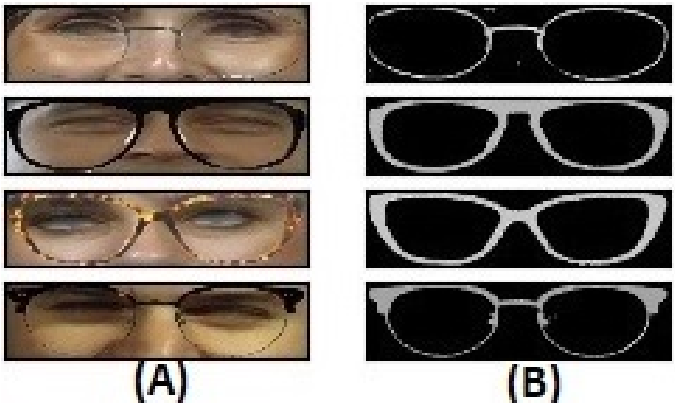


Figure 39: Predicted masks (B) of eyeglasses frames (A) for test set of in-painted FERET database using convolutional-deconvolutional approach.

using the geometric of facial parts (eyebrow, ear, and nose) supported by Dlib library [35].

Figure 39 shows sample results of predicted masks of the eyeglasses frame fin in-painted eyeglasses version of FERET database. As can be seen, CNN predicted a most challenging skinny frame shown in the first row. Figure 40 shows the predicted mask of eyeglasses frames on the in-painted version of VISOB dataset. As can be seen, different eyeglasses shapes and frame thickness have been predicted precisely. Figure 41,and 41 shows the frame mask prediction for real eyeglasses from subset of VISOB dataset.

3.4.5.1 Eyeglasses Removal using In-painting

Figure 43 shows a sample of eyeglasses segmentation using convolutional de-convolutional approached mentioned above to obtained the predicted mask of eyeglasses frame followed by in-painting operation.

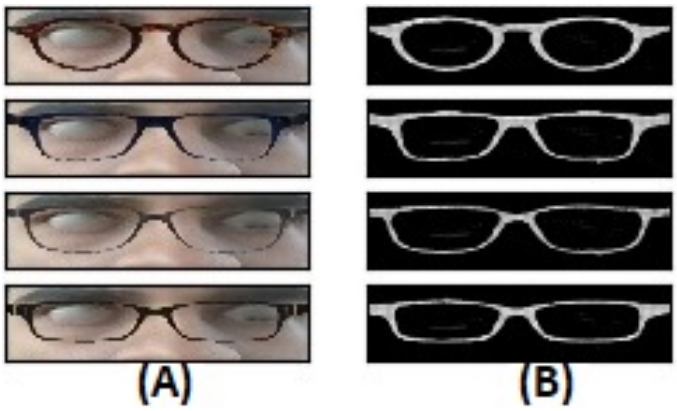


Figure 40: Predicted masks (B) of eyeglasses frames (A) on test set of in-painted VISOB database obtained using convolutional-deconvolutional approach. approach.

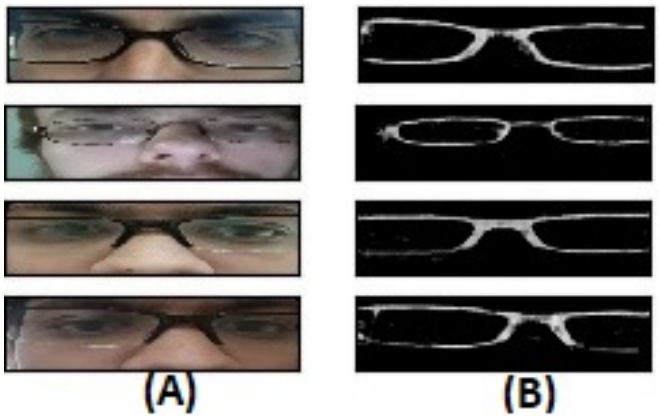


Figure 41: Predicted masks (B) of eyeglasses frames (A) on test subset of real eyeglasses obtained from VISOB dataset using convolutional-deconvolutional approach.

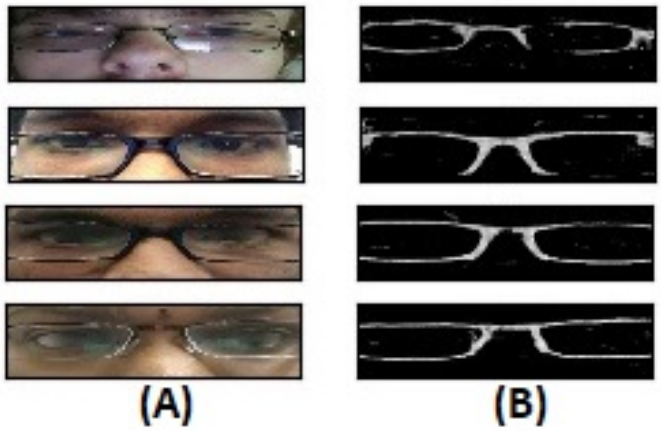


Figure 42: Predicted masks (B) of eyeglasses frames (A) on test subset of real eyeglasses obtained from VISOB dataset using convolutional-deconvolutional approach.

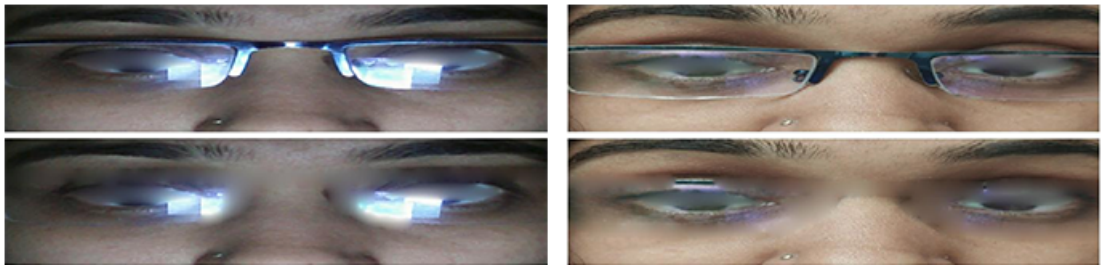


Figure 43: Eyeglasses frame removal using in-painting after frame prediction convolutional deconvoluational approach.

Table 3: Summary of previous work on eyeglasses detection

Ref.	Method	Database	Subjects	Images	Accuracy
[81]	LBP and SVM	LFW database	5749	13233	98.65%
[5]	AdaBoost, Haar, Gabor, and SVM	FERET	989	1978	95.5%-98%
[82]	Convexity, symmetry, and smoothness with deformable contour	In-house	419	1257	99.50%
[83]	Hough Transform	In-house	19	512	80%
[84]	likelihood probaility of edge information	ORL	40	400	Fisher value of 4.9-9.6
[86]	Cascaded Filters, LBP, HOG, SVM, MLP, LDA, and QDA	FERET + VISOB	989 + 550	1978 + 120000	97.9%-100%
[87]	Squeezed models of CNNs	FERET + VISOB	989 + 550	1978 + 120000	94.6%-99.9%

Table 4: Summary of previous work on eyeglasses segmentation

Ref.	Method	Database	Subjects	Images	Accuracy
[85]	Markov-Chain Monte Carlo	In-house + subset FERET	274 + 20	548 + 40	Mean Error = 90%
[88]	Recursive error compensation of PCA	In-house	130	260	RMS Error = 7.5%-5%
[89]	Reconstructed algorithms	Equinox and DHUFO	33	1320	Correct Reconstucted 88%
[90]	Morphological operations	In-house (video frame)	4	3500	Fail with rotation
[91]	Phase congruency and progressive inpainting	In-house + CAS-PEAL	30 + 40	60 + 80	Improvement of 5% on facial recognition rate
[92]	Two level of fusion	Equinox	90	3244	Face success rate=97%
[93]	Active Appearance Model (AAM)	In-house	20	600	5% Improvement on face recognition
[94]	Recursive PCA reconstruction	KFDB	33	264	10% Improvement of face recognition
[95]	Recursive error compensation	KFDB	33	264	10% Improvement of face recognition

Table 5: The structure of the nine proposed CNN models.

Model Name	No of Conv2D	No. of Channels	Filter size	No. of FC
Mod-01	6	26, 16, 16, 16, 16, 16	5×5	2
Mod-02	4	26, 16, 16, 16	5×5	2
Mod-03-S	6	26, 16, 16, 16, 16, 16	3×3	2
Mod-04-S	4	26, 16, 16, 16	3×3	2
Mod-05-S	2	26, 16	3×3	2
Mod-06-S	1	16	3×3	2
Mod-07-S	1	6	3×3	2
Mod-08-S	1	3	3×3	2
Mod-09-S	1	1	3×3	2

Table 6: Detailed architecture of the proposed CNN-01 network for weight generation and eyeglasses detection.

Layer	Output Shape	#Parameters
Input	$120 \times 480 \times 3$	-
Conv2D	$60 \times 240 \times 26$	$1,976$
Relu	$60 \times 240 \times 26$	-
Maxpooling	$30 \times 120 \times 26$	-
Dropout	$30 \times 120 \times 26$	-
Conv2D	$15 \times 60 \times 16$	$10,416$
Relu	$15 \times 60 \times 16$	-
Maxpooling	$7 \times 30 \times 16$	-
Dropout	$7 \times 30 \times 16$	-
Flatten	3360	-
Fully Connected	16	$53,776$
Fully Connected	2	34
Soft-Max	2	-
Total Parameters		$66,202$

Table 7: Detailed architecture of the proposed CNN-02 network for glasses segmentation

Layer	**Output Shape**	**#Parameters**
Input	$120 \times 480 \times 3$	-
Conv2D	$60 \times 240 \times 26$	1,976
Relu	$60 \times 240 \times 26$	-
Maxpooling	$30 \times 120 \times 26$	-
Dropout	$30 \times 120 \times 26$	-
Conv2D	$15 \times 60 \times 16$	10,416
Relu	$15 \times 60 \times 16$	-
UpSampling	$30 \times 120 \times 16$	-
UpSampling	$60 \times 240 \times 16$	-
Output	$120 \times 480 \times 16$	-
Total Parameters		$12,392$

Table 8: Detailed architecture of the proposed Convolutional DeConvolutional network for glasses segmentation

Layer	**Output Shape**	**#Parameters**
Input	$120 \times 480 \times 3$	-
Conv2D	$60 \times 240 \times 26$	$1,976$
Relu	$60 \times 240 \times 26$	-
Maxpooling	$30 \times 120 \times 26$	-
Dropout	$30 \times 120 \times 26$	-
Conv2D	$15 \times 60 \times 16$	$10,416$
Relu	$15 \times 60 \times 16$	-
DeConv2D	$30 \times 120 \times 16$	$6,416$
Relu	$30 \times 120 \times 16$	-
UpSampling	$60 \times 240 \times 16$	-
Relu	$60 \times 240 \times 16$	-
DeConv2D	$120 \times 480 \times 16$	$6,416$
Relu	$120 \times 480 \times 16$	-
Lamda	$120 \times 480 \times 1$	-
Output	$120 \times 480 \times 1$	-
Total Parameters		$25,224$

Table 9: Comparative assessment of the proposed approach with existing study in [5]

Method Name	eyeglasses [%]	No eyeglasses [%]	Overall [%]
Proposed Method on VISOB	100	98.7	99.0
Proposed Method on FERET	100	97.9	97.9
Bo el al., [5] on FERET	97.1	97.7	97.2

Table 10: Overall accuracy of the proposed learning based schemes for eyeglasses detection using three different classifiers i.e., SVM, MLP, LDA and their fusion.

Classifier	Database	EyeGlasses [%]	No EyeGlasses [%]	Overall [%]
SVM	VISOB	100	100	100
SVM	FERET	99.2	100	99.3
MLP	VISOB	100	100	100
MLP	FERET	99.2	100	99.3
LDA	VISOB	93.5	75.0	89.7
LDA	FERET	99.2	100	97.9
Fused Classifier	VISOB	100	100	100
Fused Classifier	FERET	99.2	100	99.3

CHAPTER 4

USER AUTHENTICATION USING SOFT-BIOMETRIC

4.1 Introduction

With the modern mobile technology revolution, the increasing number of individuals are using their smartphone for sensitive applications and transactions. However, mobile phones are misplaced, lost and stolen more than other computing devices. Therefore, efforts have been directed at the development of biometrically secure mobile access and transactions [105]. The use of biometric technology in mobile devices has been referred to as mobile biometrics, encompassing both the sensors that acquire biometric signals and software algorithms for their verification. According to Acuity Market Intelligence forecast[1], mobile biometric revenue is expected to surpass 33 billion dollars by 2020, not just for unlocking the device but to approve payments and as a part of multi-factor authentication services.

The eyebrow is one of the novel biometrics that naturally exists in the human face for all genders. Studies have shown the potential of eyebrows as stand-alone biometrics for recognizing individual [106–108]. Other studies have considered eyebrows as a soft-biometric trait to be used when a primary biometric trait is unavailable due to occlusion and eyes half-closed for ocular biometrics [109]. Eyebrow region offers the best

[1] http://www.acuity-mi.com/GBMR_Report.php

trade-off between computational complexity and accuracy in comparison to face or ocular biometrics. This is because the eyebrow region is one-sixth of the full face region and existing studies suggest a minimal drop in recognition accuracy of eyebrows over face biometrics [12, 106].

This paper aims to develop a method of user authentication using eyebrows for smartphone devices. *This is the first large-scale study evaluating the potential of eyebrows biometric for mobile user authentication.* The advantage of eyebrow based mobile user authentication include:

1. Eyebrows can be extracted using the front facing RGB camera available in the mobile device.

2. Eyebrow is computationally efficient and offers fast throughput. Therefore, it can also be employed as a primary or soft-biometric trait in combination with primary biometrics traits such as face and ocular region.

3. Eyebrows can also be employed for continuous user authentication [110] to ensure that the user primarily authenticated is still the user under control of the device.

To this aim, local Histogram of Oriented Gradients (HOG) [60] and global GIST [67] descriptors are evaluated as features extracted from eyebrow ROIs and used with support vector machine (SVM) for user authentication. Experimental results suggest minimum Equal Error Rate (EER) of 3.23% and Area under Curve (AUC) of 0.9916 obtained from score level fusion using (GIST) [67] descriptor obtained from left and right eyebrows.

Besides, comparative evaluation with the existing methods [111] is conducted by evaluating the proposed method on FERET dataset. The experimental results suggest minimum Equal Error Rate (EER) of 1.08% and Area under Curve (AUC) of 0.9990 obtained using score level fusion of left and right eyebrows with HOG [60].

Eyeglasses can also be used as a soft-biometric trait for short-term user authentication. Soft biometric could be divided into three categories namely, facial traits (such eyebrow, eye color , hair color, and skin color), body traits (such as height, weight, leg length, and arm length), and accessories (such as eyeglasses, cloth color, hat color, hat shape) [112].

4.2 Previous Work

Juefei-Xu et al., [106] used eyebrows trait as a stand-alone modality for biometric system. In this work, three different transformations involved which are Walsh-Hadamard (WHT), Discrete Cosine Transform (DCT), and Discrete Fourier Transform (DFT) with Local Binary Pattern(LBP). Also, L1 norm used as a distance measure. This work evaluated on the FRGC-02 dataset, and the reported results show a verification rate of 8.2% at a false acceptance rate of 0.1%.

Li et at., [111] design an approach for eyebrows localization and identification using a Fast Fourier Transform (FFT) with the fast matching template. Their result shows a verification performance of 98%. Also, Li et at., [113] implemented an eyebrows-based recognition system using a Hidden Markov Model (HMM) and localized the eyebrow

trait using a manual polygonal shape. The obtained result was 92.60% for a small in-house dataset of 27 subjects.

Yujian et al., [114] implemented an eyebrows identification that matches the acquired features with query features which extracted using Discrete Fourier Transform (DFT) for 32 subjects in-house dataset. The best identification accuracy was 93.75%.

Guo et al., [115] evaluated three different descriptors namely, Gradient Orientation (GO) histogram, LBP and SIFT applied on 30 subjects dataset for eyebrow recognition. The results vary between 77% and 80% for right and left eyebrow respectively.

Le et al., [111] implement an eyebrow segmentation and recognition system using an Active Shape Model (ASM) with a 2D profile along with F-measure. The best result was a 99.4% of performance accuracy evaluated on MBGC dataset.

Jun-Bin et al., [116] extracted the eyebrow features with biorthogonal-wavelet and matched the features using Radial Base Function (RBF) kernel inside Support Vector Machine (SVM). Their FAR of 29.58% and FRR of 8.22% applied on 100 subjects in-house dataset.

Youg et al., [108] investigate the performance of eyebrow recognition using a sparse weight of Sparsity Preserving Projection (SPP). Their accuracy was 92.50% and 80.63% applied on open eyes and closed eyes respectively for 32 subjects in-house dataset.

Bharadwaj et al., [109] implemented a novel eyebrow-based biometric system using GIST descriptor and LBP and matched the histograms of these descriptors applied to the UBIRIS-02 dataset. Their accuracy was 70.82% and 63.77% for GIST and LBP respectively.

Table (11) shows the summary of the previous work on user authentication using eyebrow biometrics. Most of the existing approaches used a traditional computer vision descriptors such as Local Binary Pattern (LBP) with Fourier transform applied on an in-house database. The reported results vary between 70% to 98

4.3 Proposed Work of User Authentication

Two approaches of user authentication were implemented using different soft-biometric traits. The first biometric trait used is eyebrows which represent a facial soft-biometric which naturally exist in the face. The second trait is the eyeglasses which represent an accessory based soft-biometric trait which could be used to identify users that wear glasses.

4.3.1 Non-Learning Eyebrows-based User Authentication

The proposed approach is divided into three steps as shown in Figure 44. As the first step, the region of interest (ROI), right and left eyebrow, has been cropped from VISOB dataset using Dlib library [35].

Next, ten local descriptors namely Rotated Local Binary Pattern (RLBP), Completed Local Binary Pattern (CLBP), Local Ternary Pattern (LTP), Local Oriented Statistical Information Booster (LOSIB), Locally Uniform Comparison Image Descriptor (LUCID), Local Phase Quantization (LPQ), Local Binary Pattern (LBP), Binarized Statistical Image Feature (BSIF), Pattern of Oriented Edge Magnitudes (POEM), and Histogram of Oriented Gradients (HOG), are extracted as features (descriptors) from these ROIs at four scales (2 to 5 times of the original size).

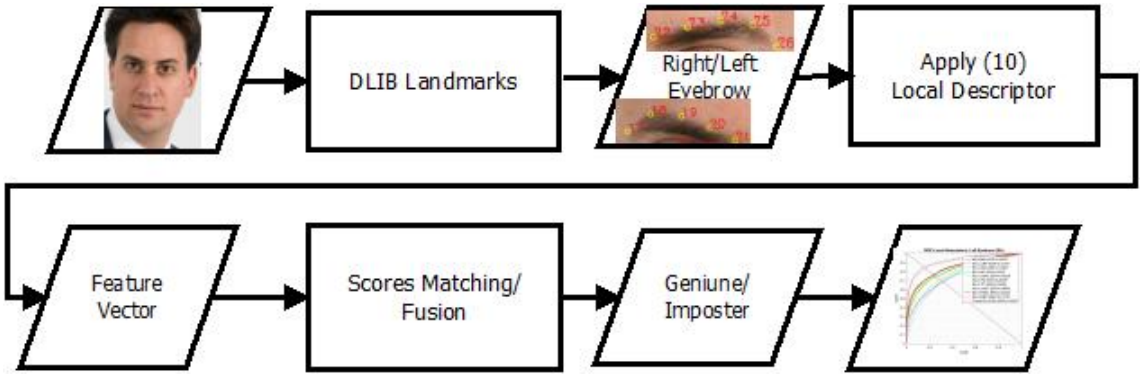


Figure 44: The block diagram of the proposed Non learning-based method for user authentication based on eyebrows.

Then, these features are matched, individually and in combination, using a pair of enrollment and test ROIs using Euclidean distance. Fusion of these descriptors is done at score level using a simple sum rule. The evaluation of the proposed approach is done using the Equal Error Rate (EER), the area under the curve (AUC) and ROC Curves. Besides, a global (GIST) descriptor has been evaluated using non-learning based matching for eyebrows as ROI.

4.3.2 Learning Eyebrows-based User Authentication

Figure 45 shows the overall steps of user authentication based on eyebrows. The eyebrow which represents the region of interest (ROI) was extracted from a "selfie" face image using a pre-trained facial landmark detector [35]. Then, descriptive features were extracted from the left and right eyebrows using deep, local and global descriptors along with Support Vector Machine (SVM) training and classification. The VGG16 feature vector was obtained from the last max pooling layer of size $(3 \times 10 \times 512)$ and was

flattened to (1×15360). After that, the flattened vector was reduced using Principal Component Analysis (PCA) to a size of (1×1600) as shown in figure (46).

Deep descriptor (VGG16), global descriptor (GIST), and global descriptor (HOG) have been used and described as follows:

- **VGG16:** is a convolutional neural network (CNN) which consists of 16 layers, and it developed by Visual Geometry Group (VGG) at Oxford University, also it called (OxfordNet) [117]. Figure (46) shows the VGG16 network for eyebrow feature extraction. This network has 14 convolution and ReLU layers, 4 max-pooling layers, 3 fully connected and ReLU layers, and 1 softmax layer for object classification. Also, the VGG16 was trained with ImageNet which is a Large-Scale Hierarchical Image dataset [**?**]. For feature extraction, the required feature vector was obtained from the last max pooling layer and was reduced using PCA.

- **GIST:** a GIST descriptor [67] was implemented using a convolution method for the image with 32×32 filter size of Gabor at 4 scales and 8 orientations. The convolution produced 32 feature maps of the same size of the input image. Then these features were split into 16 regions each (by a 4×4 grid), and feature values within each region are averaged. These 16 averaged values of all 32 feature maps were concatenated to form 16×32=512 GIST descriptor. Figure (47) shows the output image obtained by applying GIST on the extended ocular region (eyeglasses frame) ROI.

- **HOG:** This descriptor was computed by dividing the image window into small

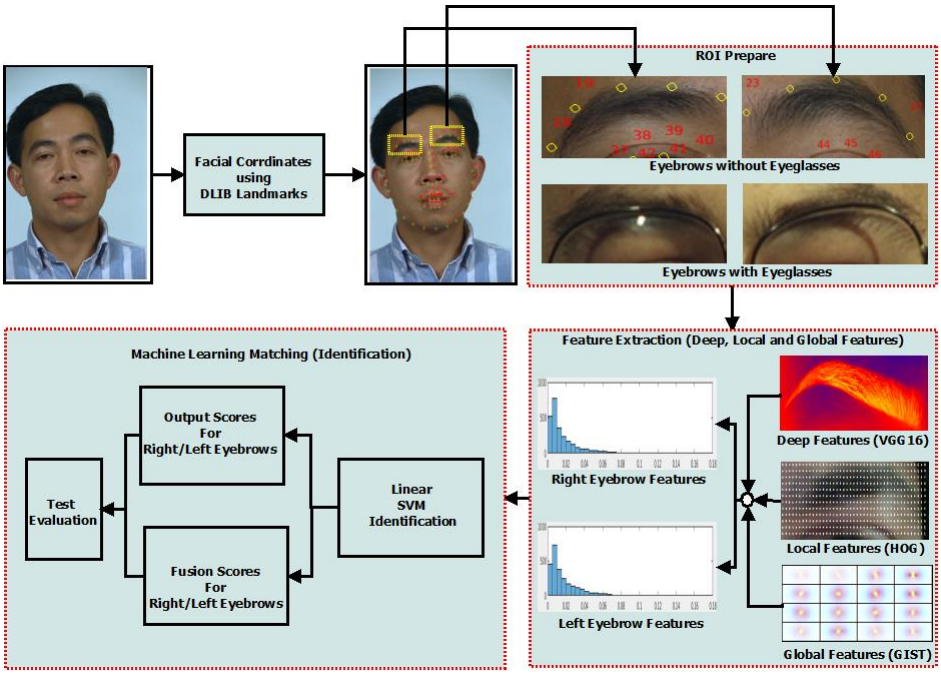


Figure 45: The block diagram of the proposed learning-based method for user authentication based on eyebrows.

cells, for each cell collecting a local 1-D histogram of gradient directions or edge orientations over the pixels of the cell [60]. The combined histogram entries form the representation. In this approach, the ROI was divided into 16×16 cells size, then the histogram of gradient orientation was computed within these cells. This operation formed a concatenated histogram of size 3600 from the ocular region (eyeglasses frame) of size 120×480. Figure (47) shows the output obtained on applying HOG on the eyebrow ROI.

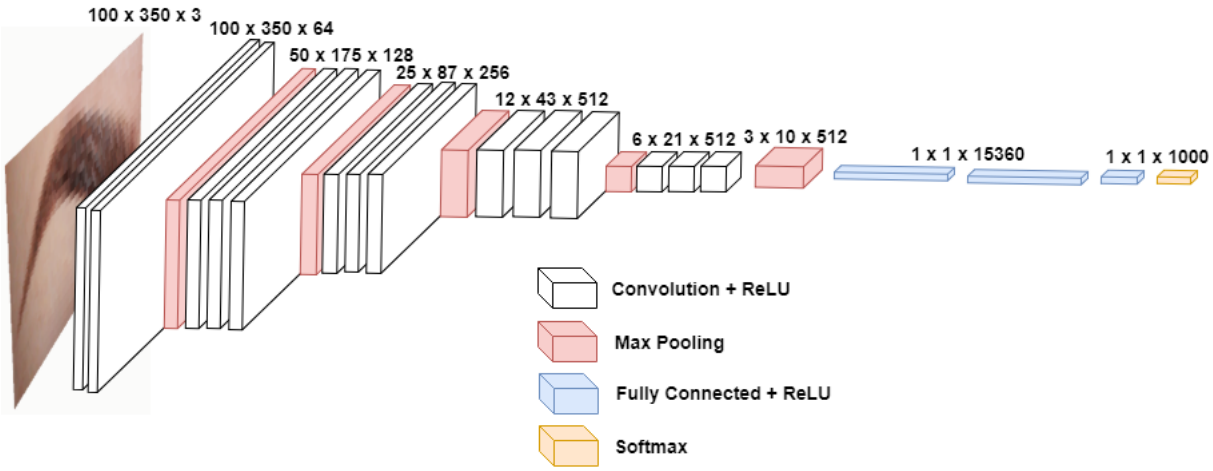


Figure 46: Detailed architecture of the VGG16 CNN network for eyebrow features extraction

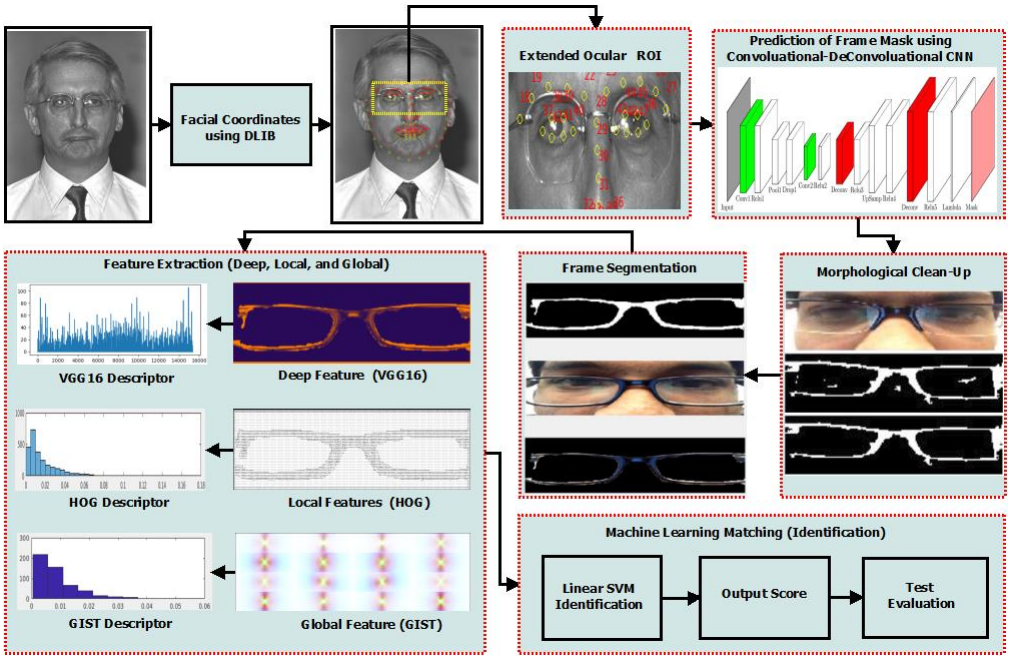


Figure 47: The block diagram of the proposed approach on eyeglasses based user authentication.

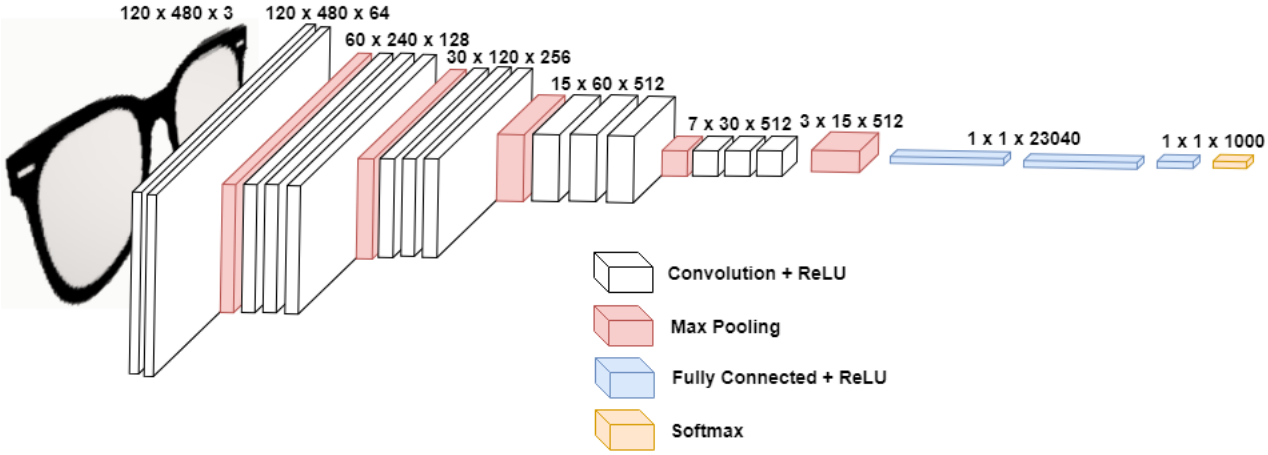


Figure 48: Detailed architecture of the VGG16 CNN for eyeglasses features extraction

4.3.3 Learning Eyeglasses-based User Authentication

Figure 47 shows the overall of the proposed approach for user authentication based on the eyeglasses frame. This approach was implemented with six main stages, namely, Region of Interest (ROI) extraction, eyeglasses frame prediction, mask clean-up, frame segmentation, features extraction, and learning-based classification.

The extended ocular region (ROI) was cropped using geometric information of eyebrows, and nose. These geometric coordinates were calculated using Dlib library [35]. Then the ROI was used in the proposed approach for eyeglasses frame prediction (mentioned on section 3.3.5). The predicted frames were cleaned using a morphological operation based on the opening area with binary eight neighbors connected components.

Consequently, the cleaned frame was multiplied by the RGB channels of the original ROI (extended ocular region) to form the colored image of the eyeglasses frame (figure 47). The local and global feature descriptors were extracted from the colored image

of the eyeglasses frame, namely, Histogram of Oriented Gradient (HOG), GIST descriptors, respectively. Then, the extracted features were used for the Support Vector Machine (SVM) training and classification to authenticate the subjects based on his/her eyeglasses frame shape. Further, deep features were extracted using VGG16 convolutional Neural Network (CNN) as shown in figure (48). The VGG16 feature vector was obtained from the last max pooling layer of size (3) and was flattened to (1×23040). After that, the flatted vector was reduced using Principal Component Analysis (PCA) to size (1). We evaluated deep descriptor (VGG16), global descriptor (GIST), and local descriptor (HOG) for eyeglasses based user authentication.

4.4 Results

4.4.1 Experiments Conducted

To evaluate the efficacy of eyeglasses detection, eyeglasses segmentation, and user authentication, the following experiments were conducted:

1. Evaluating of different local image features, and global GIST feature using the proposed non-learning-based approach on Eyebrow based user-authentication.

2. Evaluating the mitigation of eyeglasses impact using the proposed learning-based approach on Eyebrow based user-authentication.

3. Evaluating segmented learning eyeglasses-based user authentication.

4. Evaluating the score level fusion of learning eyebrows-based and learning eyeglasses-based User authentication.

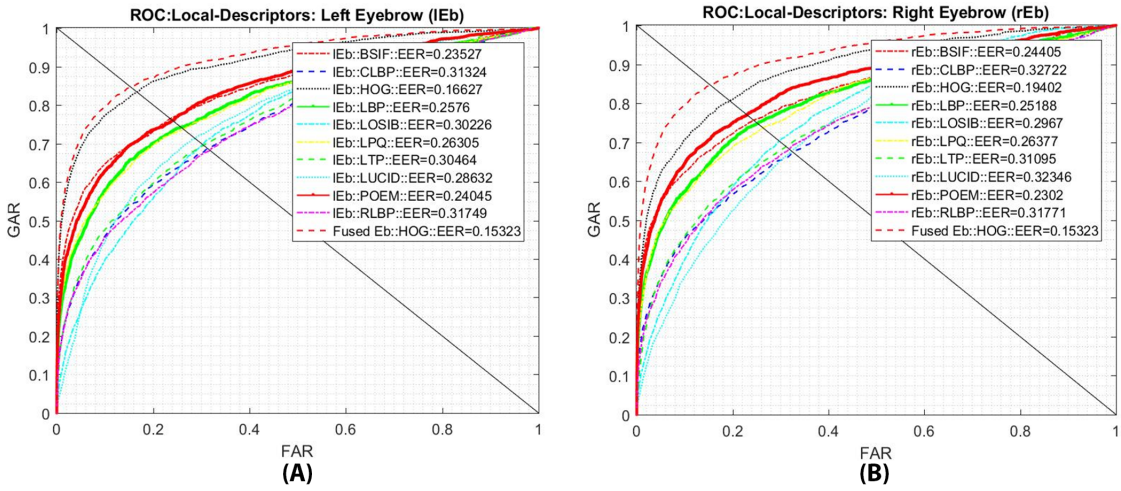


Figure 49: ROC of ten local descriptors for non-learning eyebrow-based user authentication applied on ROI consisting of (A) left eyebrow, and (B) right eyebrow and the fusion of the best descriptor (HOG) for ROI consisting of left and right eyebrow.

The experiments have been evaluated using different light settings such as indoor (In), outdoor (Out), and all combined conditions (All).

4.4.2 Non-Learning Eyebrow-based User Authentication

The proposed approach is evaluated (Experiment (1) in section 4.4.1) on Visible Ocular Biometric Database (VISOB) dataset [103]. Figure 49 (A) shows ROC Curves of ten local descriptors pertaining to left eyebrow. Also, figure 50 shows AUC and EER values of left eyebrow for these descriptors. As it can be seen AUC ranges from 0.75 to 0.90, and the EER ranges from 31.7% to 16.6%.

Similarly, Figure 49 (B) shows the ROC Curves of these ten local descriptors for left eyebrow. Also, figure 51 shows AUC and EER values for these descriptors. It can be seen that AUC ranges from 0.74 to 0.88, and the EER ranges from 32.7% to 19.4%. HOG

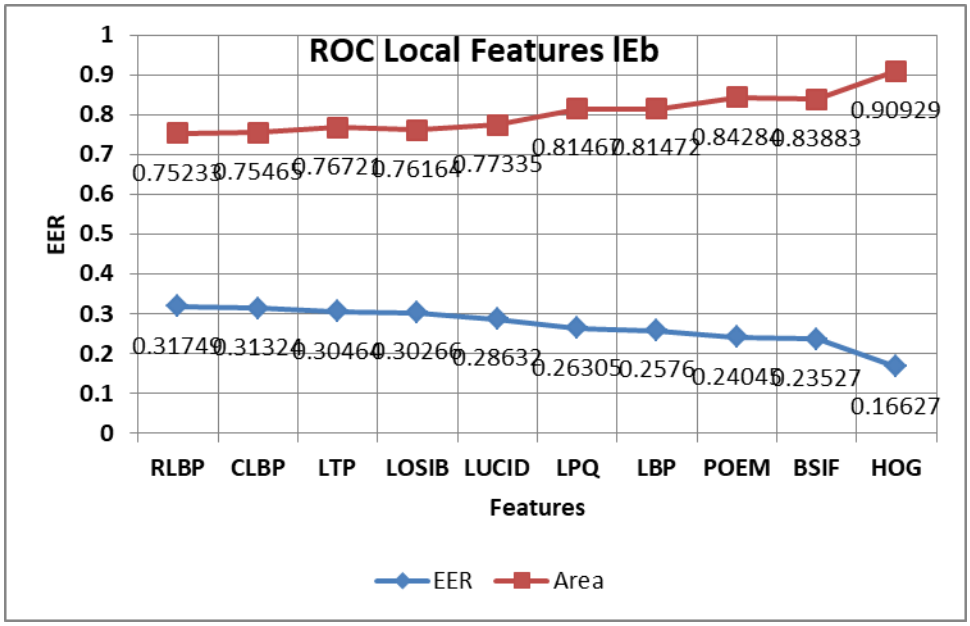


Figure 50: EER and AUC values for these ten descriptors for non-learning eyebrow-based user authentication applied on ROI consisting of left eyebrow.

descriptor obtained the best result with AUC of 0.88, and EER of 19.4%. HOG descriptor obtained the best result with AUC of 0.90, and EER of 16.6% even for left ROI. Further, a fusion of HOG for left and right eyebrow region reduced the EER to 15.32% at score level. As a part of future work, a fusion of local and global descriptors will be evaluated at feature and score level for further performance enhancement.

Figure 52 shows the evaluation of right and left eyebrows with GIST feature using non-learning eyebrow-based for user authentication. As it can seen (Figure 52 (A)), an EER of 13.92% has been obtained from left eyebrows. while, an EER of 13.75% has been obtained from right eyebrow (Figure 52 (B)).

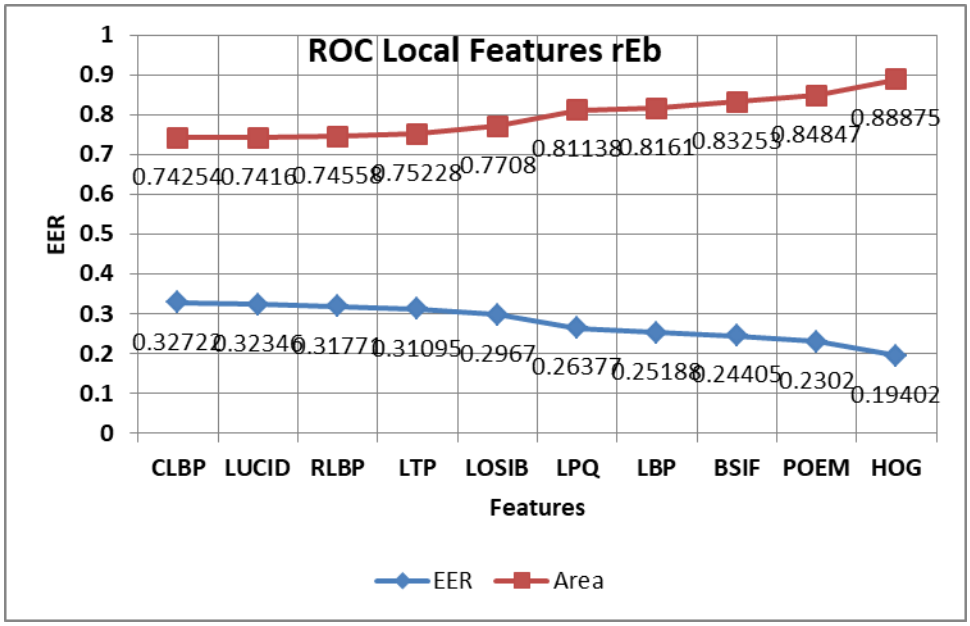


Figure 51: EER and AUC values for these ten descriptors for non-learning eyebrow-based user authentication applied on ROI consisting of right eyebrow.

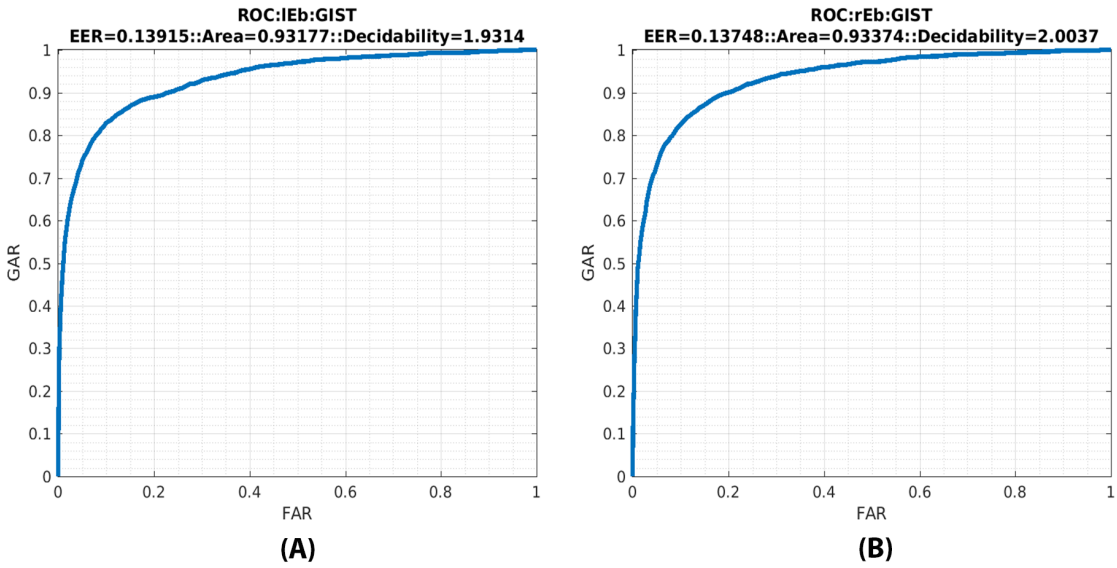


Figure 52: ROC of global (GIST) descriptors for non-learning eyebrow-based user authentication applied on ROI consisting of (A) left and (B) right eyebrows.

4.4.3 Learning Eyebrows-based User Authentication

The most experiments (Experiment (2) in section 4.4.1) of eyebrows-based user authentication evaluated in different light settings showed that the deep feature (VGG16) descriptor out-performed the other global (GIST) descriptor and local (HOG) descriptor in most experiments of eyeglasses presence.

Table (12) shows Equal Error Rate (EER) Area Under Curve (AUC), and decidability (d´) obtained for eyebrow ROI across indoor and outdoor lighting conditions for iPhone without presence of eyeglasses (Experiment (1) in section 4.4.1). The best EER of 1.26%, AUC of 0.9972, and decidability of 4.2358 was obtained using a fusion of left and right eyebrow ROIs in indoor lighting condition (In:In) using a fusion of descriptors. While, the worst EER of 6.39%, AUC of 0.9798, decidability of 3.1317 were obtained for the fused eyebrow in indoor versus outdoor settings (In:Out) using fused of the descriptor.

Figure (53) shows the EER evaluation with different light condition for eyebrow without eyeglasses existence. In this case, the GIST descriptor performs better than other descriptors in most light setting cases.

Table (13) shows EER obtained for eyebrow ROI across indoor and outdoor lighting conditions for iPhone in presence of eyeglasses (Experiment (2) in section 4.4.1). The best EER of 0.63%, AUC of 0.9942, and decidability of 5.6802 were obtained with the fusion of left and right eyebrow ROIs in indoor light condition (In:In) using the fused descriptor. While, the worst ERR of 12.25%, AUC of 0.9323, and decidability of 2.4168 were obtained using fused eyebrow in indoor versus outdoor setting (In:Out) using the fused descriptor.

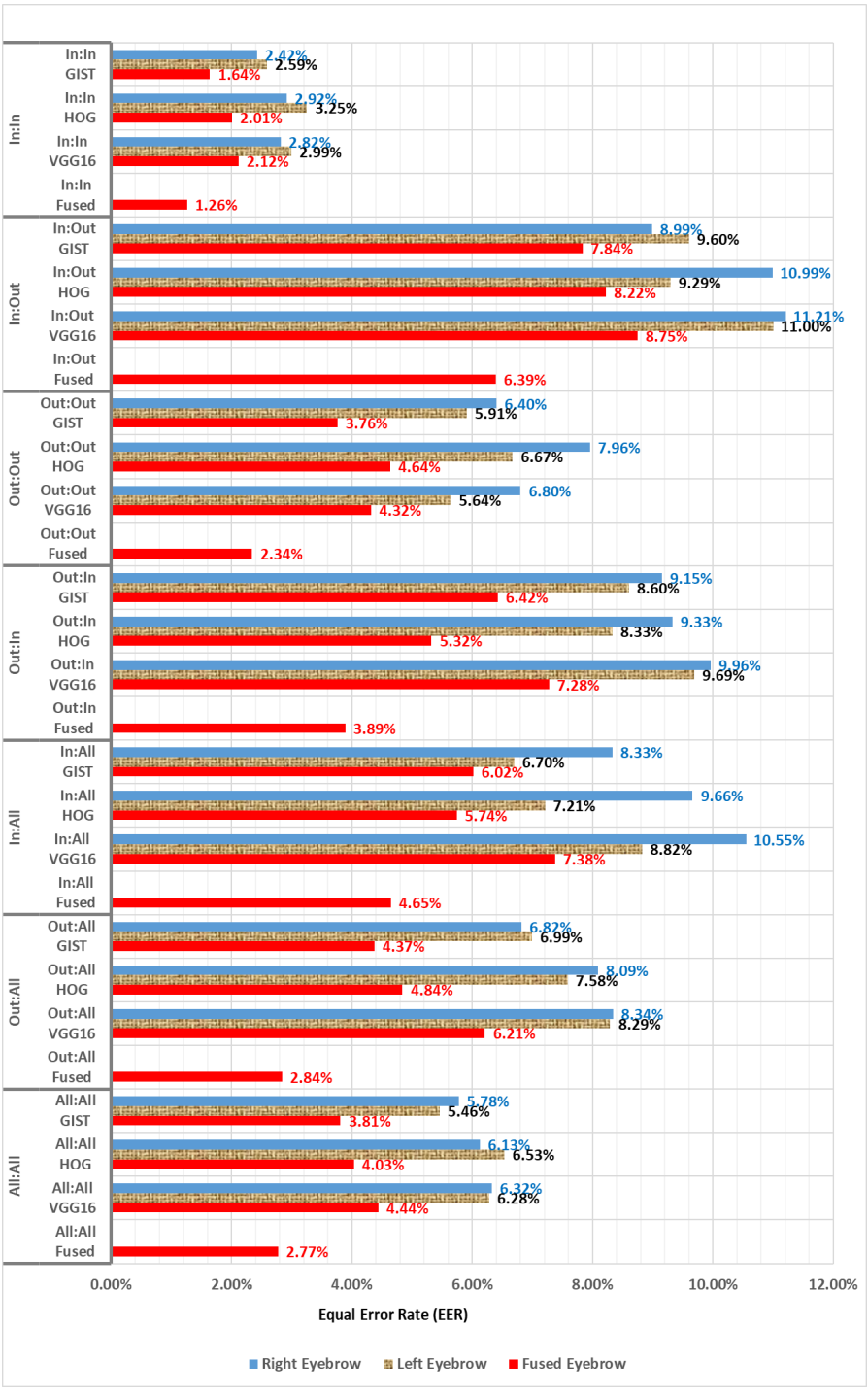


Figure 53: Test evaluation using comparison of EER with different light conditions for VISOB database, iPhone device, eyebrow without eyeglasses existence.

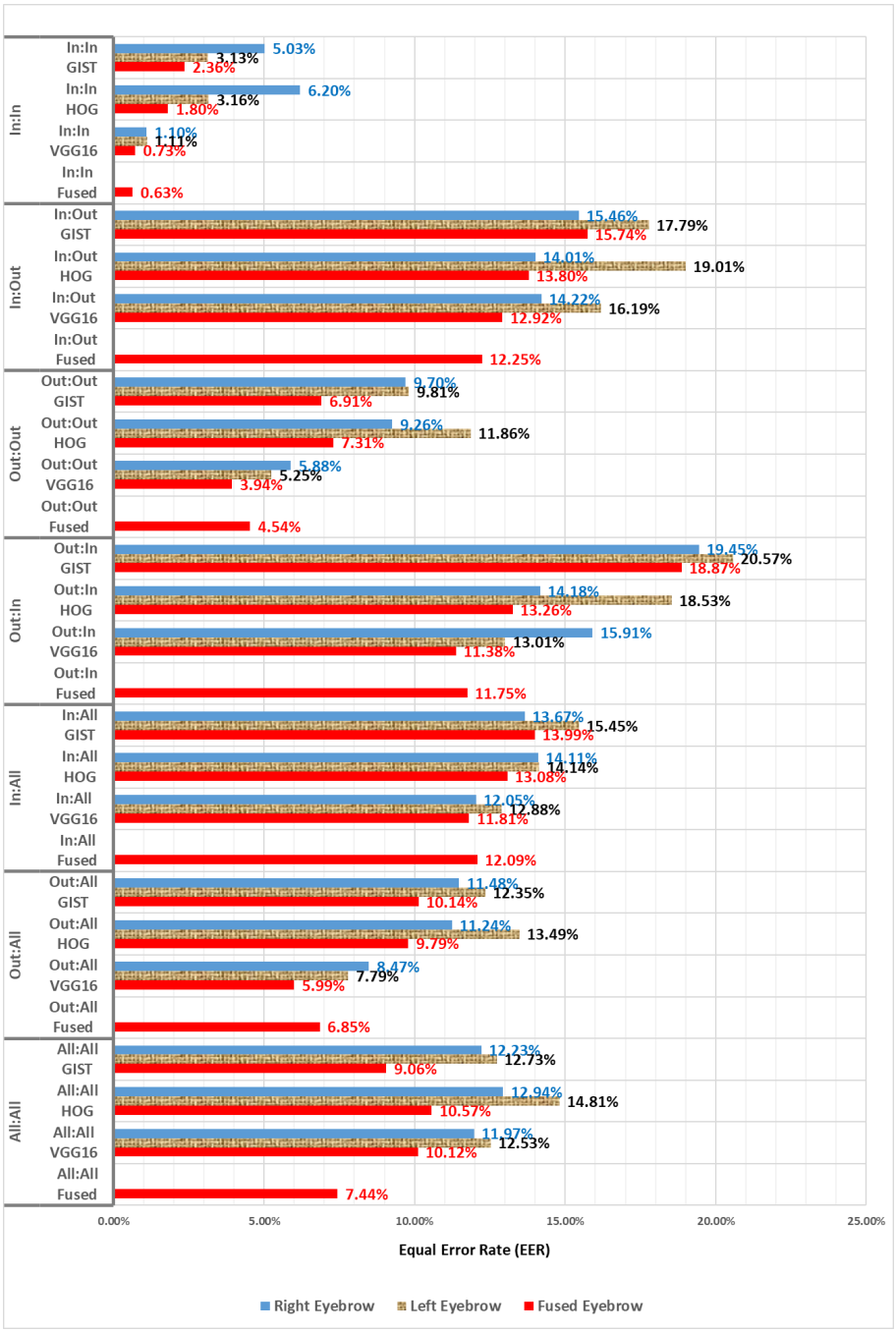


Figure 54: Test evaluation using comparison of EER with different light conditions for VISOB database, iPhone device, eyebrow with eyeglasses existence.

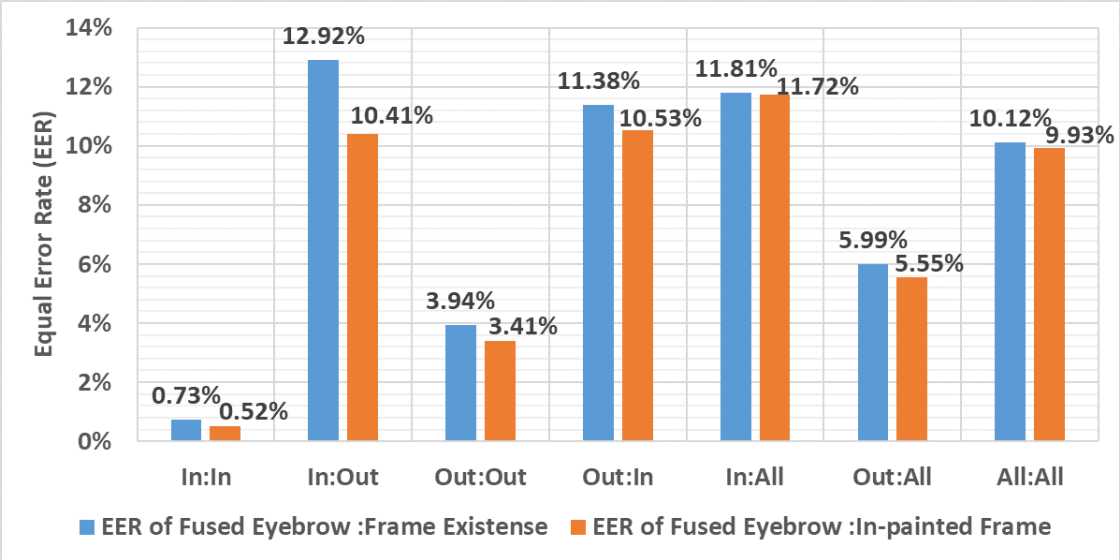


Figure 55: Test evaluation using comparison of EER with different light conditions for VISOB database, iPhone device, eyebrow with eyeglasses removal using in-painting.

Figure (54) shows the EER evaluation with different light condition for eyebrow with eyeglasses existence. In this case, VGG16 descriptor performs better than other descriptors in most light settings (except for All versus All (All:All) light setting).

Figure 55 show example of ocular region with eyeglasses and after segmentation and removal. Table 14 shows EER enhancement of 2.51% for indoor vs. outdoor (In:Out) light settings, and the worst EER enhancement of 0.09% for indoor vs All (In:All) lighting conditions were obtained for eyebrow-based authentication after eyeglasses segmentation and removal using convolutional-deconvolutional approach followed by in-painting. The convolutional-deconvolutional approach is used because of its superiority over cascaded CNNs-based approach for eyeglasses segmentation.

Table (15) shows EER obtained for eyebrow ROI across indoor and outdoor lighting conditions for OPPO without the presence of eyeglasses (Experiment (2) in section 4.4.1). The best EER of 2.03%, AUC of 0.9947, and decidability of 4.1548 were obtained with the fusion of left and right eyebrow ROIs in all light conditions (All:All) using the fused descriptor. While, the worst EER of 4.27%, AUC of 0.9866, and decidability of 3.4608 were obtained using fused eyebrow in indoor versus outdoor settings (In:All) using fused descriptor.

Figure (56) shows the EER evaluation with different light condition for eyebrow without eyeglasses existence. In this case, VGG16 descriptor performs better than other descriptors in most light setting cases.

Table (16) shows EER obtained for eyebrow ROI across indoor and outdoor lighting conditions for OPPO without the presence of eyeglasses (Experiment (2) in section 4.4.1). The best EER of 2.71%, AUC of 0.9975, and decidability of 4.1271 were obtained with the fusion of left and right eyebrow ROIs in indoor light conditions (In:In) using the fused descriptor. While, the worst ERR of 13.01%, AUC of 0.9283, and decidability of 2.1298 were obtained using fused eyebrow in indoor versus outdoor setting (In:Out) using the fused descriptor. Overall, VGG16 outperformed all the other descriptors (GIST and HOG) in all the light settings in the presence of eyeglasses.

Figure (57) shows the EER evaluation with different light condition for eyebrow with eyeglasses existence. In this case, VGG16 descriptor performs better than other descriptors in most light setting cases.

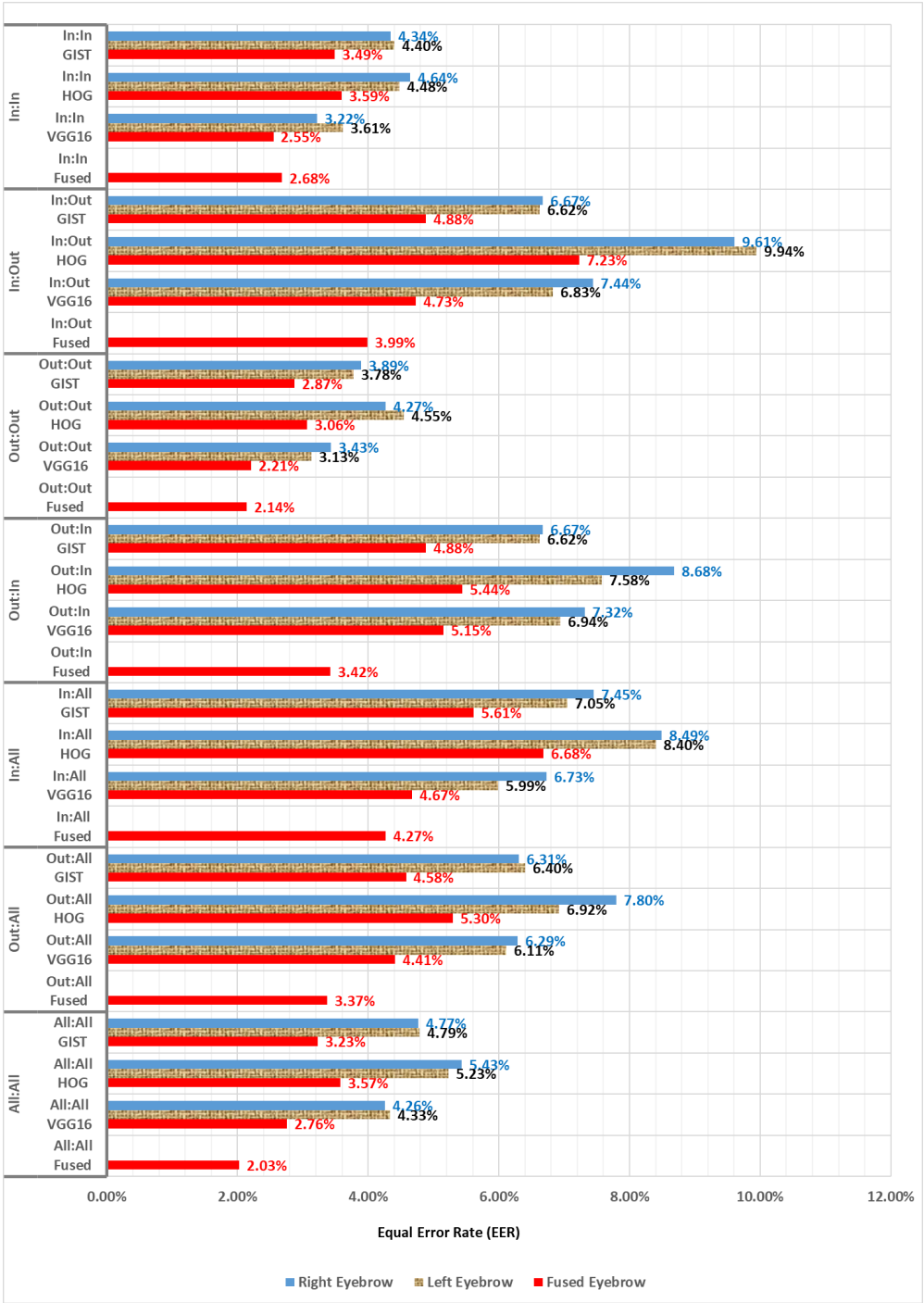


Figure 56: Test evaluation using comparison of EER with different light conditions for VISOB database, OPPO device, eyebrow without eyeglasses existence.

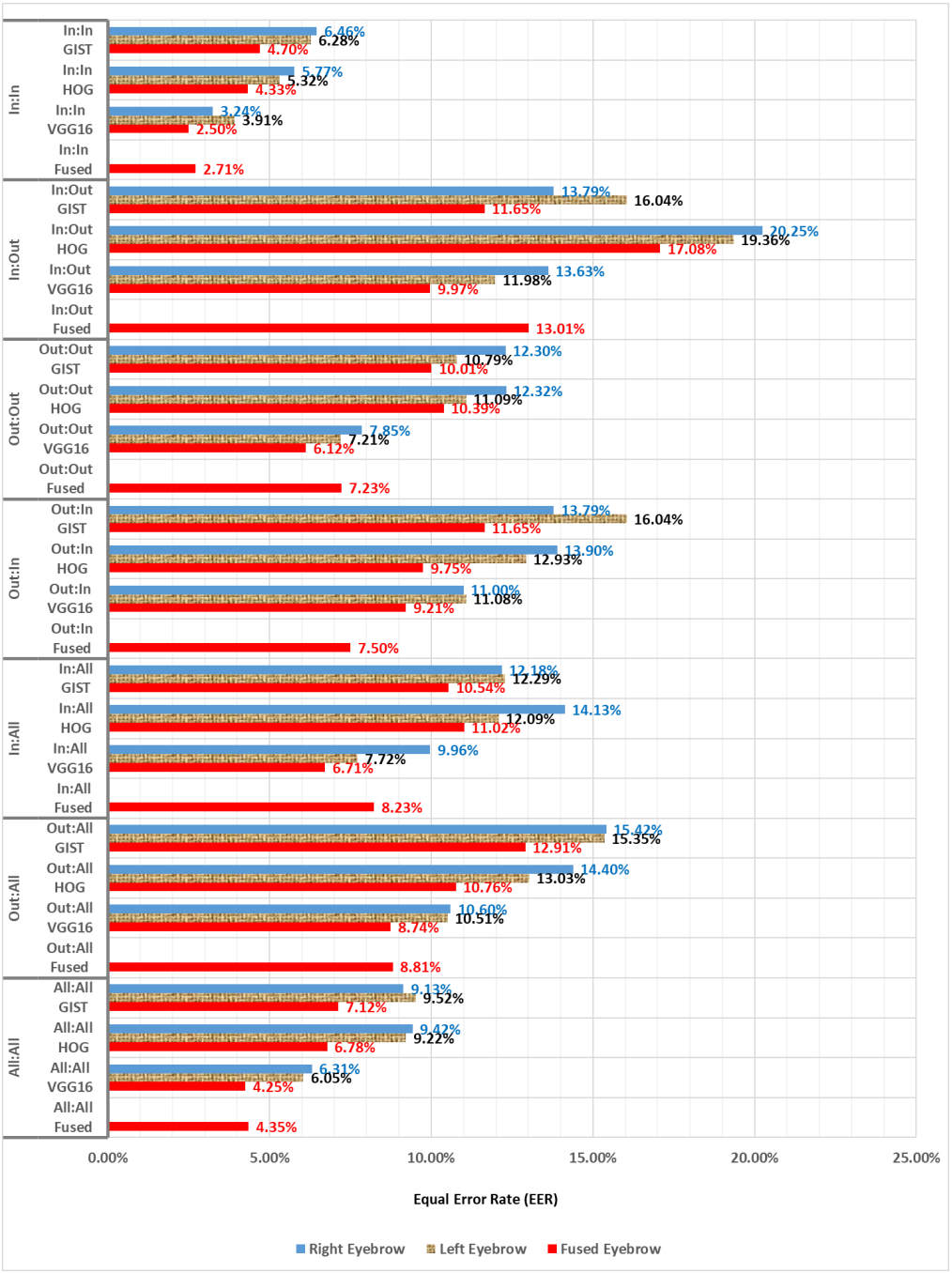


Figure 57: Test evaluation using comparison of EER with different light conditions for VISOB database, OPPO device, eyebrow with eyeglasses existence.

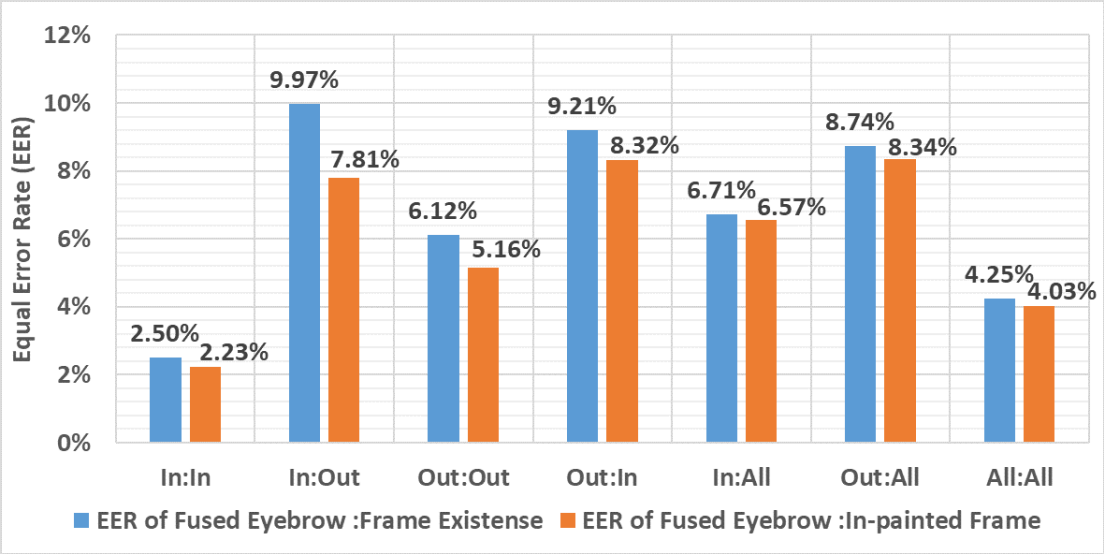


Figure 58: Test evaluation using comparison of EER with different light conditions for VISOB database, iPhone device, eyebrow with eyeglasses removal using in-painting.

Figure 58 show example of ocular region with eyeglasses and after segmentation and removal. Table 17 shows EER enhancement of 2.16% for indoor vs. outdoor (In:Out) light settings, and the worst EER enhancement of 0.14% for indoor vs All (In:All) lighting conditions were obtained for eyebrow-based authentication after eyeglasses segmentation and removal using convolutional-deconvolutional approach followed by in-painting. The convolutional-deconvolutional approach is used because of its superiority over cascaded CNNs-based approach for eyeglasses segmentation.

4.4.4 Results of Eyeglasses-based User Authentication

Similar to previous experiments, the results have been compared using the score level fusion of all the descriptors (GIST, HOG, VGG16) for left and right eyebrows across different light settings (Experiment (3) in section 4.4.1).

Table (18) shows EER obtained for eyebrow ROI across indoor and outdoor lighting conditions for iPhone (Experiment (2) in section 4.4.1). The best EER of 3.4400%, AUC of 0.9880, and decidability of 3.9726 were obtained with the fusion of left and right eyebrow ROIs in indoor light condition (In:In) using fused descriptors. While, the worst ERR of 17.550%, AUC of 0.8763, and decidability of 1.8797 were obtained using fused eyebrow in indoor versus outdoor settings (In:Out) using fused descriptors. Overall, the deep (VGG16) the outperformed other descriptors (GIST, and HOG) in most light settings.

Figure (59) shows the EER evaluation of VISOB iPhone device with different light condition for eyeglasses. In this case, VGG16 descriptor performs better than other descriptors in most light setting cases.

Table (19) shows EER obtained for eyebrow ROI across indoor and outdoor lighting conditions for OPPO (Experiment (2) in section 4.4.1). The best EER of 7.690%, AUC of 0.9616, and decidability of 2.9524 were obtained with the fusion of left and right eyebrow ROIs in indoor light conditions (In:In) using the fused descriptor. While, the worst EER of 18.230%, AUC of 0.8815, and decidability of 1.8268 were obtained using fused eyebrow in outdoor versus indoor settings (Out:In). Overall, the deep (VGG16) descriptor outperformed other descriptors (GIST, and HOG) in all light settings.

Figure (60) shows the EER evaluation of VISOB OPPO device with different light condition for eyeglasses. In this case, VGG16 descriptor performs better than other descriptors in most light setting cases.

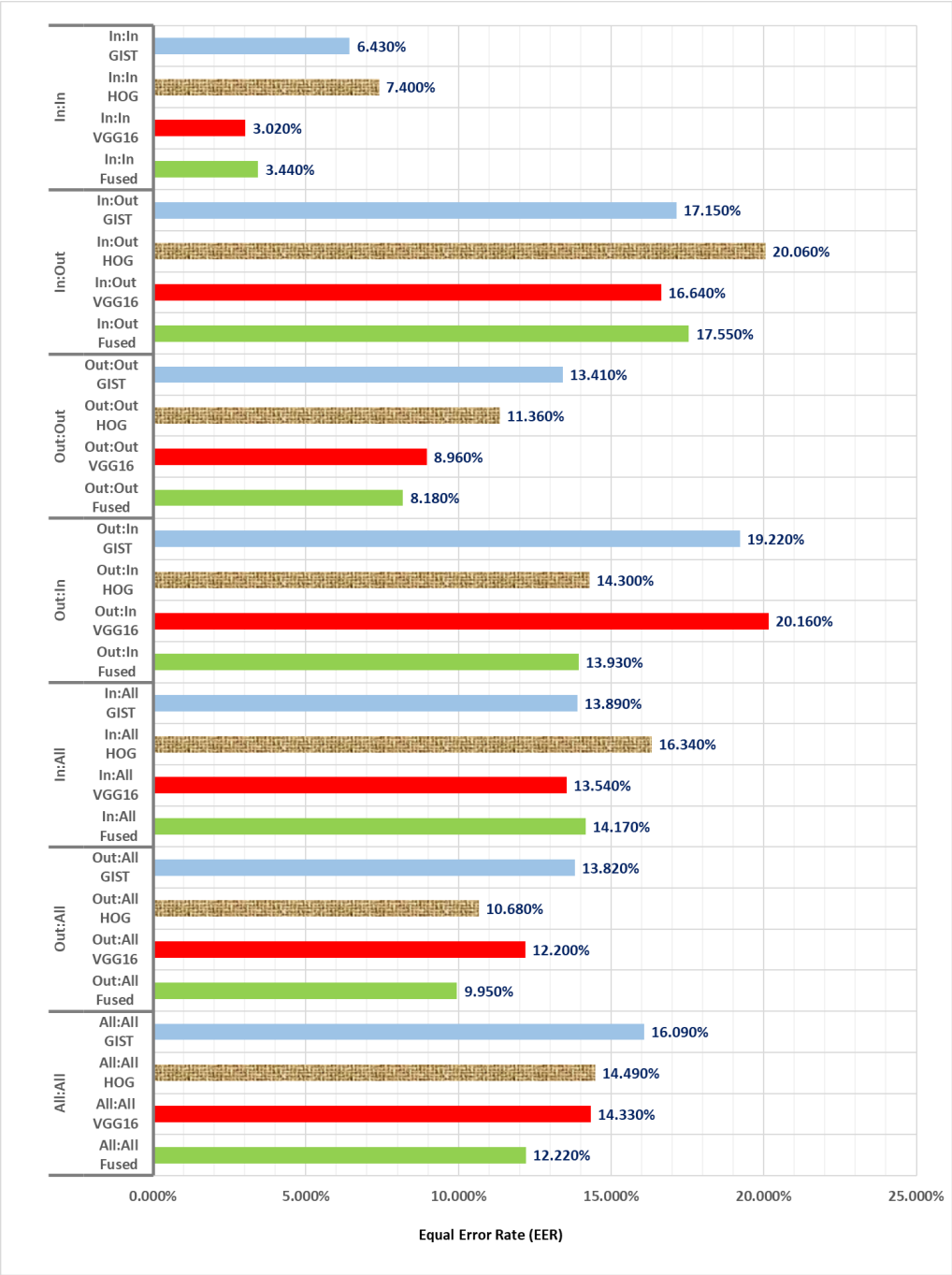


Figure 59: Test evaluation using comparison of EER with different light conditions for VISOB database, OPPO device, eyebrow with eyeglasses existence.

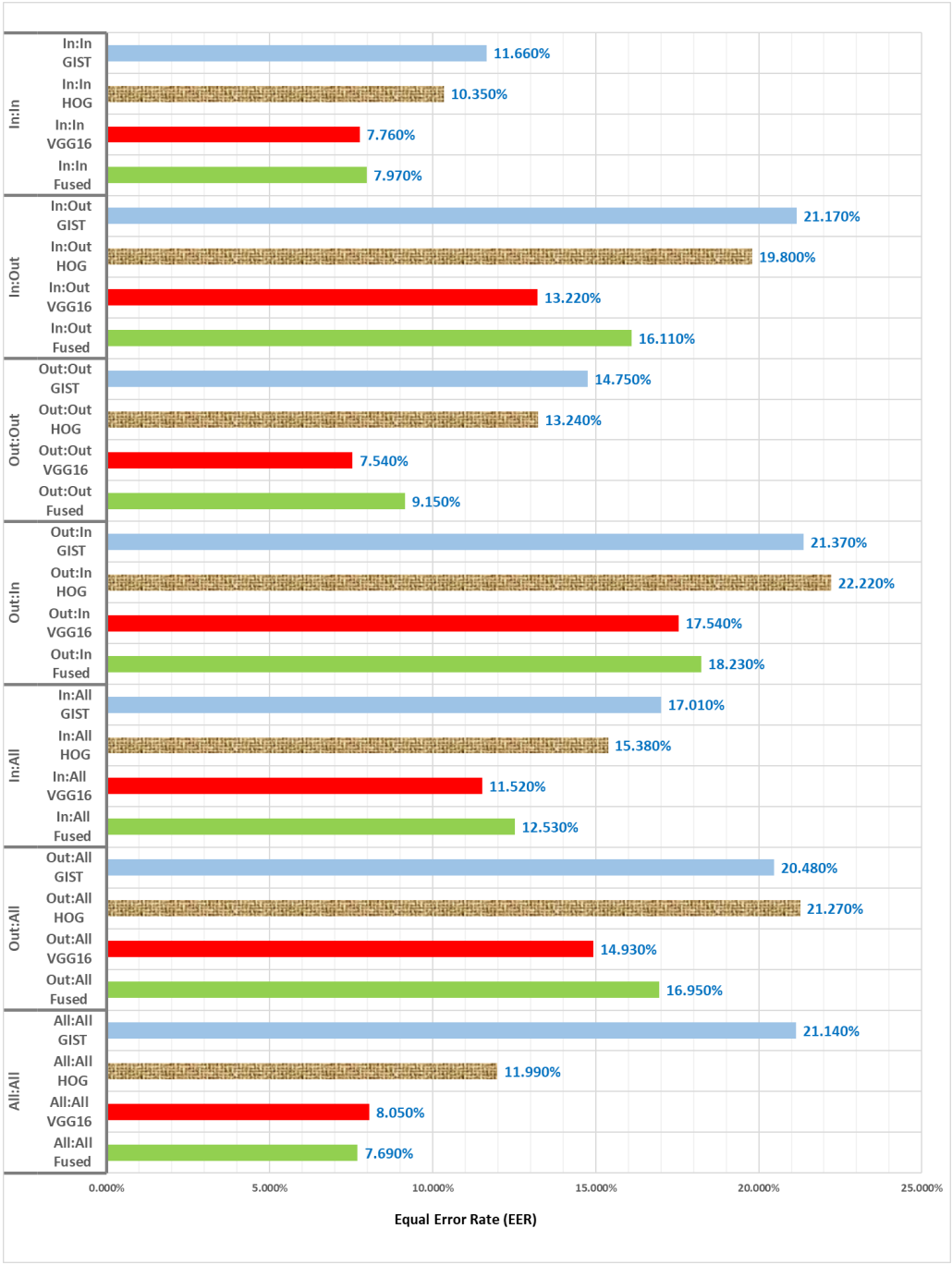


Figure 60: Test evaluation using comparison of EER with different light conditions for VISOB database, OPPO device, eyebrow with eyeglasses existence.

4.4.5 Results of Modalities Fusion

A score level fusion has been applied onto two soft biometric traits namely, fused eyebrow modality (score level fusion of left and right eyebrow) in the presence of eyeglasses with eyeglasses modality. The deep features (VGG16) have been fused for ROIs (eyebrow and eyeglasses) since the VGG16 feature outperformed all other used global (GIST) and local (HOG) descriptor for eyeglasses and eyebrows in the presence of eyeglasses.

Table (20) shows EER obtained for fused modalities (fused eyebrows, and eyeglasses) across indoor and outdoor lighting conditions for iPhone (Experiment (2) in section 4.4.1). The best EER of 0.720%, AUC of 0.9944, and decidability of 6.6212 were obtained using the fusion of modalities in indoor light conditions (In:In) using fused (VGG16) descriptor. While, the worst ERR of 14.350%, AUC of 0.9268, and decidability of 2.4422 were obtained using the fusion of modalities in indoor versus outdoor setting (In:Out) using fused (VGG16) descriptor for both modalities. Overall, the fused (VGG16) for both modalities outer-performed all eyeglasses scores, and most of the fused eyebrow scores.

Figure (60) shows the EER evaluation of VISOB iPhone device with a different light condition for fusion of VGG16 for eyebrows and eyeglasses. In this case, the fused descriptor performs better than other descriptors in most light setting cases.

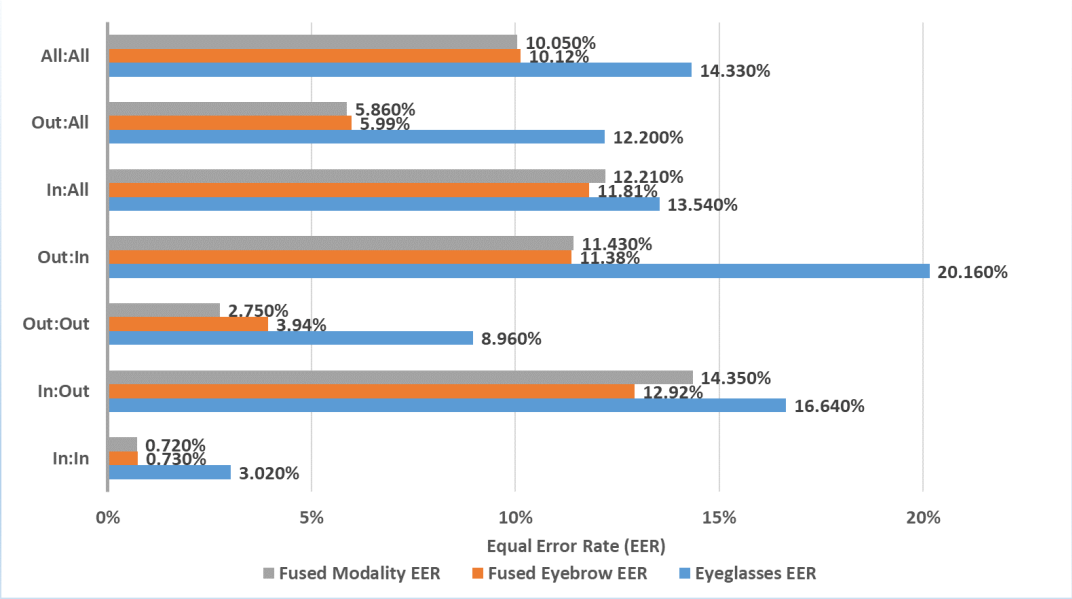


Figure 61: Test evaluation using comparison of EER with different light conditions for VISOB database, OPPO device, eyebrow with eyeglasses existence.

Table (21) shows EER obtained for fused modalities (fused eyebrows, and eyeglasses) across indoor and outdoor lighting conditions for OPPO (Experiment (4) in section 4.4.1). The best EER of 2.570%, AUC of 0.9940, and decidability of 4.1481 were obtained with the fusion of modalities in indoor light condition (In:In) using fused (VGG16) descriptor. While, the worst ERR of 10.340%, AUC of 0.9574, and decidability of 2.4876 were obtained using fused eyebrow in outdoor versus indoor settings (Out:In) using fused (VGG16) descriptor. Overall, the fused (VGG16) for both of modalities outer-performed all eyeglasses results, and fused eyebrow results in different light conditions.

Figure (60) shows the EER evaluation of VISOB OPPO device with different light conditions for fusion of VGG16 for eyebrows and eyeglasses. In this case, the fused descriptor performs better than other descriptors in most light setting cases.

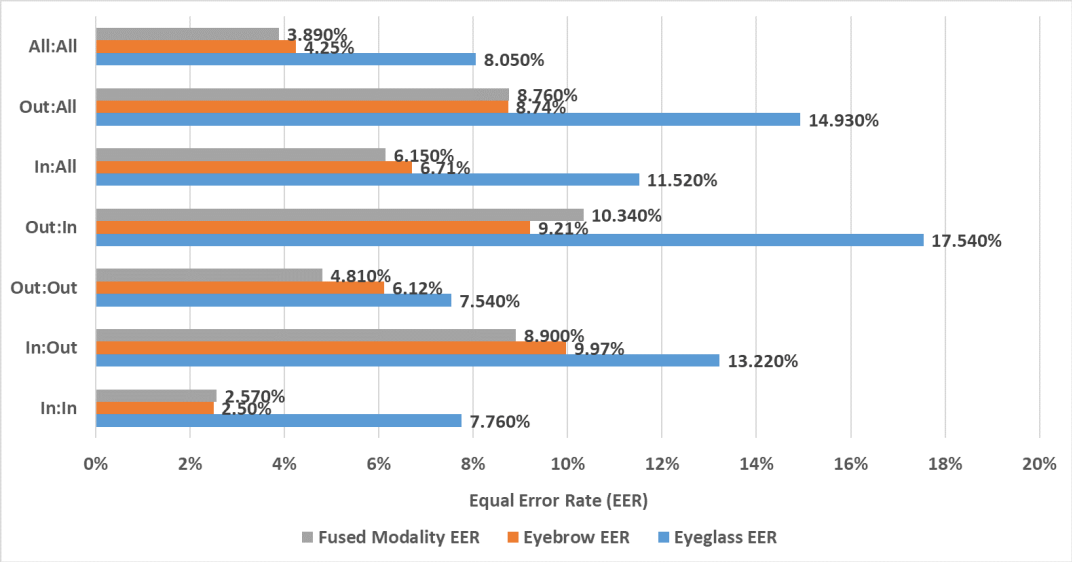


Figure 62: Test evaluation using comparison of EER with different light conditions for VISOB database, OPPO device, eyebrow with eyeglasses existence.

Table 11: Summary of the previous work on eyebrow-based User Authentication.

Ref.	**Method Used**	**Dataset**	**Subjects**	**Images**	**Accuracy**
[106]	WHT, DCT, and DFT with LBP	FRGC	4003	16012	FAR= 8.2%
[111]	Fast Fourier Transform	BJUT + colored FERET	109 + 989	1118 + 1978	97-98%
[113]	Hidden Markov Model	In-house small	27	54	92.60%
[114]	Discrete Fourier Transform	In-house small	32	64	93.75%
[40]	GO, LBP, and SIFT	In-house small	30	899	80%
[107]	Active Shape Model	MBGC + AR	574 + 126	2341 + 4000	85%-76%
[116]	Biorthogonal wavelet	In-house small	50	100	FAR= 29.58%
[108]	Sparsity Preserving Project	In-house small	32	320	92.50%
[109]	GIST, and CLBP	UBIRIS	261	11102	70.82%

Table 12: EER, AUC, and d´ values for eyebrow-based user authentication across indoor and outdoor light conditions without presence of eyeglasses for VISOB-iPhone.

Experiment	**Feature**	**Right Eyebrow**			**Left Eyebrow**			**Fused Eyebrow**		
		EER	**AUC**	**d'**	**EER**	**AUC**	**d'**	**EER**	**AUC**	**d'**
In:In	GIST	2.42%	0.9938	3.6766	2.59%	0.9951	3.6903	1.64%	0.9965	4.2358
	HOG	2.92%	0.9918	3.9876	3.25%	0.9911	3.8626	2.01%	0.9939	4.5062
	VGG16	2.82%	0.9929	3.9422	2.99%	0.9935	3.9208	2.12%	0.9961	4.4025
	Fused							**1.26%**	**0.9972**	**4.9968**
In:Out	GIST	8.99%	0.9601	2.6285	9.60%	0.9646	2.6287	7.84%	0.9733	2.9247
	HOG	10.99%	0.9493	2.6382	9.29%	0.9621	2.6912	8.22%	0.9671	2.9166
	VGG16	11.21%	0.9587	2.4917	11.00%	0.9577	2.5294	8.75%	0.9728	2.6741
	Fused							**6.39%**	**0.9798**	**3.1317**
Out:Out	GIST	6.40%	0.9798	2.9353	5.91%	0.9837	2.9329	3.76%	0.9917	3.3695
	HOG	7.96%	0.9672	2.7716	6.67%	0.9802	2.8816	4.64%	0.9889	3.2353
	VGG16	6.80%	0.9769	2.8553	5.64%	0.9853	2.9274	4.32%	0.9900	3.1850
	Fused							2.34%	0.9954	3.7193
Out:In	GIST	9.15%	0.9641	2.4818	8.60%	0.9658	2.5112	6.42%	0.9783	2.8303
	HOG	9.33%	0.9639	2.5000	8.33%	0.9721	2.5521	5.32%	0.9846	2.8805
	VGG16	9.96%	0.9625	2.3721	9.69%	0.9692	2.3655	7.28%	0.9816	2.5910
	Fused							3.89%	0.9914	3.1347
In:All	GIST	8.33%	0.9653	2.9205	6.70%	0.978	2.9897	6.02%	0.9828	3.3179
	HOG	9.66%	0.9622	2.9655	7.21%	0.9759	3.0613	5.74%	0.9809	3.3409
	VGG16	10.55%	0.9639	2.8245	8.82%	0.9691	2.8856	7.38%	0.9789	3.0671
	Fused							4.65%	0.9883	3.6186
Out:All	GIST	6.82%	0.9792	2.7446	6.99%	0.9783	2.7330	4.37%	0.9895	3.1290
	HOG	8.09%	0.9691	2.6502	7.58%	0.9769	2.6989	4.84%	0.9872	3.0523
	VGG16	8.34%	0.9741	2.6093	8.29%	0.9756	2.6025	6.21%	0.9858	2.8482
	Fused							2.84%	0.9947	3.4193
All:All	GIST	5.78%	0.9835	2.7765	5.46%	0.9806	2.6850	3.81%	0.9878	3.6934
	HOG	6.13%	0.9806	2.7759	6.53%	0.9783	2.7383	4.03%	0.9795	3.1389
	VGG16	6.32%	0.9771	2.6213	6.28%	0.9788	2.6408	4.44%	0.9847	2.8694
	Fused							2.77%	0.9917	3.4292

Table 13: EER, AUC, and d´ values for eyebrow-based user authentication across indoor and outdoor light conditions with presence of eyeglasses for VISOB-iPhone.

Experiment	Feature	Right Eyebrow			Left Eyebrow			Fused Eyebrow		
		EER	**AUC**	**d'**	**EER**	**AUC**	**d'**	**EER**	**AUC**	**d'**
In:In	GIST	5.03%	0.9868	3.0836	3.13%	0.9888	3.1882	2.36%	0.9922	3.5949
	HOG	6.20%	0.9831	3.4292	3.16%	0.992	3.6807	1.80%	0.9944	4.1310
	VGG16	1.10%	0.9952	5.5194	1.11%	0.9942	5.7510	0.73%	0.9948	6.7217
	Fused							**0.63%**	**0.9942**	**5.6802**
In:Out	GIST	15.46%	0.9139	2.0461	17.79%	0.8879	1.7418	15.74%	0.9136	2.0638
	HOG	14.01%	0.9255	2.3137	19.01%	0.8857	1.8831	13.80%	0.924	2.2925
	VGG16	14.22%	0.9362	2.3933	16.19%	0.9184	2.3173	12.92%	0.9406	2.4681
	Fused							**12.25%**	**0.9323**	**2.4168**
Out:Out	GIST	9.70%	0.95	2.2686	9.81%	0.9571	2.2337	6.91%	0.9667	2.5498
	HOG	9.26%	0.9541	2.4138	11.86%	0.9442	2.3140	7.31%	0.9671	2.7534
	VGG16	5.88%	0.9801	2.9365	5.25%	0.9802	3.0922	3.94%	0.9869	3.3221
	Fused							4.54%	0.982	3.2965
Out:In	GIST	19.45%	0.9067	1.8068	20.57%	0.8307	1.3273	18.87%	0.8790	1.7566
	HOG	14.18%	0.9282	2.0033	18.53%	0.9032	1.8859	13.26%	0.9473	2.2608
	VGG16	15.91%	0.9237	1.9054	13.01%	0.9481	2.1345	11.38%	0.9583	2.1658
	Fused							11.75%	0.953	2.3183
In:All	GIST	13.67%	0.8977	1.8218	15.45%	0.8879	1.6773	13.99%	0.8968	1.8851
	HOG	14.11%	0.9157	2.2228	14.14%	0.9124	2.0517	13.08%	0.9202	2.3164
	VGG16	12.05%	0.9205	2.4448	12.88%	0.9263	2.5180	11.81%	0.9246	2.5962
	Fused							12.09%	0.9085	2.3364
Out:All	GIST	11.48%	0.9449	2.1808	12.35%	0.9208	1.9272	10.14%	0.944	2.3222
	HOG	11.24%	0.9508	2.3285	13.49%	0.9323	2.1823	9.79%	0.9642	2.6209
	VGG16	8.47%	0.9675	2.6267	7.79%	0.9761	2.8047	5.99%	0.9821	2.9577
	Fused							6.85%	0.9761	2.9879
All:All	GIST	12.23%	0.9376	1.7723	12.73%	0.9391	1.8303	9.06%	0.9546	2.0379
	HOG	12.94%	0.9367	1.9527	14.81%	0.9151	1.7286	10.57%	0.9449	2.0960
	VGG16	11.97%	0.9442	2.0796	12.53%	0.9391	2.1188	10.12%	0.9546	2.2468
	Fused							7.44%	0.9647	2.3468

Table 14: EER for eyebrows-based user authentication across indoor and outdoor light conditions in presence of eyeglasses and applying eyeglasses in-painting for iPhone device in VISOB dataset.

Experiment	**Feature**	**EER of Fused Eyebrow**		
		Frame Existense	**In-painted Frame**	**Difference**
In:In	VGG16	0.73%	0.52%	0.21%
In:Out	VGG16	12.92%	10.41%	2.51%
Out:Out	VGG16	3.94%	3.41%	0.53%
Out:In	VGG16	11.38%	10.53%	0.85%
In:All	VGG16	11.81%	11.72%	0.09%
Out:All	VGG16	5.99%	5.55%	0.44%
All:All	VGG16	10.12%	9.93%	0.19%

Table 15: EER, AUC, and d´ values for eyebrow-based user authentication across indoor and outdoor light conditions without presence of eyeglasses for VISOB-OPPO.

Experiment	**Feature**	**Right Eyebrow**			**Left Eyebrow**			**Fused Eyebrow**		
		EER	**AUC**	**d'**	**EER**	**AUC**	**d'**	**EER**	**AUC**	**d'**
In:In	GIST	4.34%	0.9880	3.5123	4.40%	0.9871	3.4888	3.49%	0.9913	3.9256
	HOG	4.64%	0.9835	2.7737	4.48%	0.9825	3.7951	3.59%	0.9866	4.1968
	VGG16	3.22%	0.9923	3.9642	3.61%	0.9914	4.0124	2.55%	0.9952	4.3544
	Fused							2.68%	0.9939	4.7011
In:Out	GIST	6.67%	0.9775	2.9150	6.62%	0.9794	2.8830	4.88%	0.9872	3.2781
	HOG	9.61%	0.9585	2.5159	9.94%	0.9602	2.5565	7.23%	0.9742	2.7746
	VGG16	7.44%	0.9751	2.5135	6.83%	0.9776	2.6464	4.73%	0.9867	2.7693
	Fused							3.99%	0.9884	3.1950
Out:Out	GIST	3.89%	0.9869	3.3700	3.78%	0.9875	3.3592	2.87%	0.9907	3.7935
	HOG	4.27%	0.9855	3.5299	4.55%	0.9837	3.5181	3.06%	0.9897	3.9398
	VGG16	3.43%	0.9916	3.6569	3.13%	0.9912	3.7002	2.21%	0.9946	4.0440
	Fused							2.14%	0.9932	4.4441
Out:In	GIST	6.67%	0.9775	2.9150	6.62%	0.9794	2.8830	4.88%	0.9872	3.2781
	HOG	8.68%	0.9685	2.7306	7.58%	0.9687	2.7589	5.44%	0.9792	3.0603
	VGG16	7.32%	0.9705	2.7022	6.94%	0.9762	2.7753	5.15%	0.9835	2.9947
	Fused							3.42%	0.9887	3.5476
In:All	GIST	7.45%	0.9723	2.8249	7.05%	0.9742	2.8564	5.61%	0.9810	3.1509
	HOG	8.49%	0.9653	2.8013	8.40%	0.9660	2.8515	6.68%	0.9752	3.0769
	VGG16	6.73%	0.9775	2.8455	5.99%	0.9828	2.9703	4.67%	0.9879	3.1103
	Fused							**4.27%**	**0.9866**	**3.4608**
Out:All	GIST	6.31%	0.9775	2.9552	6.40%	0.9793	2.9592	4.58%	0.9858	3.3240
	HOG	7.80%	0.9713	2.8514	6.92%	0.9728	2.9051	5.30%	0.9816	3.1928
	VGG16	6.29%	0.9789	2.8671	6.11%	0.9806	2.9406	4.41%	0.9875	3.1488
	Fused							3.37%	0.9892	3.6404
All:All	GIST	4.77%	0.9863	3.1350	4.79%	0.9872	3.1669	3.23%	0.9916	3.6073
	HOG	5.43%	0.9835	3.2198	5.23%	0.9828	3.2285	3.57%	0.9894	3.6516
	VGG16	4.26%	0.9894	3.2894	4.33%	0.9898	3.3246	2.76%	0.9946	3.6348
	Fused							**2.03%**	**0.9947**	**4.1548**

Table 16: EER, AUC, and d´ values for eyebrow-based user authentication across indoor and outdoor light conditions with presence of eyeglasses for VISOB-OPPO.

Experiment	Feature	Right Eyebrow			Left Eyebrow			Fused Eyebrow		
		EER	**AUC**	**d'**	**EER**	**AUC**	**d'**	**EER**	**AUC**	**d'**
In:In	GIST	6.46%	0.9699	2.7389	6.28%	0.9723	2.7334	4.70%	0.9783	3.1315
	HOG	5.77%	0.9724	3.1460	5.32%	0.9746	3.1890	4.33%	0.9812	3.6074
	VGG16	3.24%	0.9901	3.7704	3.91%	0.9909	3.7861	2.50%	0.9931	4.1080
	Fused							**2.71%**	**0.9875**	**4.1271**
In:Out	GIST	13.79%	0.9252	2.0244	16.04%	0.9084	1.8642	11.65%	0.9415	2.2471
	HOG	20.25%	0.8667	1.5797	19.36%	0.8762	1.6844	17.08%	0.8934	1.7955
	VGG16	13.63%	0.9284	2.0789	11.98%	0.9437	2.1499	9.97%	0.9506	2.2548
	Fused							**13.01%**	**0.9283**	**2.1298**
Out:Out	GIST	12.30%	0.9384	2.1743	10.79%	0.9489	2.2496	10.01%	0.9549	2.4901
	HOG	12.32%	0.9315	2.2453	11.09%	0.9400	2.2960	10.39%	0.9504	2.5670
	VGG16	7.85%	0.9652	2.8290	7.21%	0.9697	2.9615	6.12%	0.9762	3.1384
	Fused							7.23%	0.9675	3.0416
Out:In	GIST	13.79%	0.9252	2.0244	16.04%	0.9084	1.8642	11.65%	0.9415	2.2471
	HOG	13.90%	0.9245	2.1448	12.93%	0.9297	2.0489	9.75%	0.9496	2.3935
	VGG16	11.00%	0.9539	2.3904	11.08%	0.9502	2.3263	9.21%	0.9687	2.5610
	Fused							7.50%	0.9618	2.7023
In:All	GIST	12.18%	0.9323	2.0814	12.29%	0.9304	2.0611	10.54%	0.9439	2.3237
	HOG	14.13%	0.9205	2.1618	12.09%	0.9332	2.2823	11.02%	0.9428	2.4662
	VGG16	9.96%	0.9546	2.5782	7.72%	0.9727	2.7160	6.71%	0.9751	2.8393
	Fused							8.23%	0.9608	2.8005
Out:All	GIST	15.42%	0.9127	1.8907	15.35%	0.9129	1.8608	12.91%	0.9330	2.1226
	HOG	14.40%	0.9163	2.0262	13.03%	0.9279	2.0186	10.76%	0.9426	2.2857
	VGG16	10.60%	0.9529	2.3267	10.51%	0.9531	2.3771	8.74%	0.9662	2.5331
	Fused							8.81%	0.9558	2.5776
All:All	GIST	9.13%	0.9626	2.3981	9.52%	0.9589	2.3494	7.12%	0.9719	2.7246
	HOG	9.42%	0.9532	2.4410	9.22%	0.9561	2.5006	6.78%	0.9676	2.8677
	VGG16	6.31%	0.9795	3.0028	6.05%	0.9835	3.0672	4.25%	0.9886	3.3470
	Fused							4.35%	0.9846	3.4583

Table 17: EER for eyebrows-based user authentication across indoor and outdoor light conditions in presence of eyeglasses and applying eyeglasses in-painting for OPPO device in VISOB dataset.

Experiment	**Feature**	**EER of Fused Eyebrow**		
		Frame Existense	**In-painted Frame**	**Difference**
In:In	VGG16	2.50%	2.23%	0.27%
In:Out	VGG16	9.97%	7.81%	2.16%
Out:Out	VGG16	6.12%	5.16%	0.96%
Out:In	VGG16	9.21%	8.32%	0.89%
In:All	VGG16	6.71%	6.57%	0.14%
Out:All	VGG16	8.74%	8.34%	0.40%
All:All	VGG16	4.25%	4.03%	0.22%

Table 18: EER, AUC, and d´ values for eyeglasses-based user authentication across indoor and outdoor light conditions for VISOB-iPhone.

Experiment	Descriptor	Eyeglasses		
		EER	**AUC**	**d'**
In:In	GIST	6.430%	0.9800	2.9439
	HOG	7.400%	0.9722	2.7629
	VGG16	3.020%	0.9911	3.9989
	Fused	**3.4400%**	**0.9880**	**3.9726**
In:Out	GIST	17.150%	0.8629	1.5220
	HOG	20.060%	0.8635	1.7119
	VGG16	16.640%	0.8951	2.1123
	Fused	**17.550%**	**0.8763**	**1.8797**
Out:Out	GIST	13.410%	0.9249	1.0433
	HOG	11.360%	0.9425	2.3432
	VGG16	8.960%	0.9671	2.7074
	Fused	8.180%	0.9607	2.8004
Out:In	GIST	19.220%	0.8906	1.6821
	HOG	14.300%	0.9348	2.0134
	VGG16	20.160%	0.8978	1.7973
	Fused	13.930%	0.9352	2.0535
In:All	GIST	13.890%	0.8954	1.3619
	HOG	16.340%	0.8844	1.6304
	VGG16	13.540%	0.9304	2.3381
	Fused	14.170%	0.9025	1.7878
Out:All	GIST	13.820%	0.9314	2.0214
	HOG	10.680%	0.9560	2.3284
	VGG16	12.200%	0.9497	2.5060
	Fused	9.950%	0.9625	2.6275
All:All	GIST	16.090%	0.9118	1.7776
	HOG	14.490%	0.9248	1.8995
	VGG16	14.330%	0.9299	2.0717
	Fused	12.220%	0.9403	2.1889

Table 19: EER, AUC, and d´ values for eyeglasses-based user authentication across indoor and outdoor light conditions for VISOB-OPPO.

Experiment	Descriptor	Eyeglasses		
		EER	**AUC**	**d'**
In:In	GIST	11.660%	0.9336	2.3766
	HOG	10.350%	0.9432	2.5122
	VGG16	7.760%	0.9719	3.0287
	Fused	7.970%	0.9578	3.0615
In:Out	GIST	21.170%	0.8471	1.4896
	HOG	19.800%	0.8663	1.6029
	VGG16	13.220%	0.9393	2.0549
	Fused	16.110%	0.8952	1.9123
Out:Out	GIST	14.750%	0.9094	1.9817
	HOG	13.240%	0.9139	2.0072
	VGG16	7.540%	0.9698	2.6935
	Fused	9.150%	0.946	2.6131
Out:In	GIST	21.370%	0.8523	1.5309
	HOG	22.220%	0.8503	1.5449
	VGG16	17.540%	0.8916	1.7757
	Fused	**18.230%**	**0.8815**	**1.8268**
In:All	GIST	17.010%	0.8921	1.8525
	HOG	15.380%	0.9058	1.9347
	VGG16	11.520%	0.9509	2.3634
	Fused	12.530%	0.9274	2.3183
Out:All	GIST	20.480%	0.8612	1.5927
	HOG	21.270%	0.8611	1.5995
	VGG16	14.930%	0.9142	1.9524
	Fused	16.950%	0.8954	1.9490
All:All	GIST	21.140%	0.9362	2.2408
	HOG	11.990%	0.9351	2.2634
	VGG16	8.050%	0.9700	2.7803
	Fused	**7.690%**	**0.9616**	**2.9524**

Table 20: EER, AUC, and d´ values for modality fusion of eyeglasses (VGG16) with eyebrow (VGG16) with presence of eyeglasses across indoor and outdoor light conditions for VISOB-iPhone.

Experiment	Eyeglass EER	Eyebrow EER	Fused Modality		
			EER	AUC	d'
In:In	3.020%	0.730%	**0.720**%	**0.9944**	**6.6212**
In:Out	16.640%	12.92%	**14.350**%	**0.9268**	**2.4422**
Out:Out	8.960%	3.94%	2.750%	0.9912	3.4583
Out:In	20.160%	11.38%	11.430%	0.9591	2.1505
In:All	13.540%	11.81%	12.210%	0.932	2.6212
Out:All	12.200%	5.99%	5.860%	0.9839	3.0574
All:All	14.330%	10.12%	10.050%	0.9566	2.3234

Table 21: EER, AUC, and d´ values for modality fusion of eyeglasses (VGG16) with eyebrow (VGG16) with presence of eyeglasses across indoor and outdoor light conditions for VISOB-OPPO.

Experiment	Eyeglass EER	Eyebrow EER	Fused Modality		
			EER	AUC	d'
In:In	7.760%	2.50%	**2.570**%	**0.9940**	**4.1481**
In:Out	13.220%	9.97%	8.900%	0.9652	2.3477
Out:Out	7.540%	6.12%	4.810%	0.9839	3.2726
Out:In	17.540%	9.21%	**10.340**%	**0.9574**	**2.4876**
In:All	11.520%	6.71%	6.150%	0.9803	2.8801
Out:All	14.930%	8.74%	8.760%	0.9638	2.5370
All:All	8.050%	4.25%	3.890%	0.9918	3.5240

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusions For Eyeglasses Detection

Eyeglasses detection represents an important step for glasses removal. In addition, eyeglasses cause real challenges for ocular and facial recognition due to glasses reflection, shadow, and frame occlusion.

Three schemes for prescription eyeglasses detection have been proposed. The first proposed scheme is a non-learning-based scheme consists of eye pair detection and eyeglasses detection. This scheme uses Viola-Jones to detect Region of Interest (ROI) followed by glass detection yielding an overall accuracy of 99.0% for FERET and 97.9% for VISOB datasets. Further, the factors detrimental to the performance of eyeglasses detection accuracy are mitigated as well. Experimental results on VISOB and FERET database prove the efficacy of the proposed approach.

The second scheme is the learning-based scheme which consists of three main steps (a) ROI detection, (b) Local Binary Pattern (LBP) and Histogram of Gradients (HOG) feature extraction and their fusion, and (c) applying classifiers such as Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and Linear Discriminant Analysis (LDA), and fusing the output of these classifiers. The latter obtained a best overall accuracy of about 99.3% on FERET and 100% on VISOB dataset. Moreover, SVM outperformed all other classifiers with an overall accuracy of 100%. Two level of fusion has

been used in this scheme. At first level, feature fusion has been applied to concatenate LPB features with HOG features. At the second level, the output of SVM, MLP, and LDA have been fused at the decision level for majority voting. The proposed fusion classifier outperforms existing learning-based schemes based on individual features such as LBP [81].

The third scheme is the deep learning scheme which described a comparative study for eyeglasses frame detection using different convolutional Neural Network (CNN) structures that have been applied on two different regions of interest (ROI), namely, Frame Bridge and extended ocular region. The best CNN model obtained an overall accuracy of 99.96% for ROI consisting of Frame Bridge. The main advantage of CNN models for eyeglasses detection over other classical approaches [47] (using feature extraction, descriptors fusion, and classification) is that the CNN model is structured both as feature extractor, using convolutional layers, and classifier, using fully connected layers, and both have trainable parameters that can have parameters adapted at each epoch. As a part of future work, eyeglasses segmentation will be applied using CNN models.

5.2 Conclusions For Eyeglasses Segmentation

Eyeglasses occlusion is considered as a real problem for facial and ocular recognition due to occlusion and glasses reflection. To mitigate this problem, an eyeglasses detection and segmentation approaches have been proposed to extract eyeglasses as a biometric trait. In this regard, two approaches have been proposed namely, cascaded Convolutional Neural Network (CNN), and Convolutional De-convolutional network.

The first segmentation scheme was cascaded convolutional Neural Network (CNN). This scheme consists of four main stages, namely region of interest (ROI) detection, cascaded CNN's for eyeglasses detection, weight generation, and glasses segmentation, followed by mathematical and binarization operations. Two CNN models have been proposed for this scheme which are CNN-01, and CNN02 models. CNN-01 model has been implemented to detect the existence of eyeglasses and generate the trained weights for the second CNN-02 model. Also, CNN-01 model implemented mainly with convolutional, and max-pooling layers for extracting a micro-structure feature, and fully connected with soft-max activation layer for eyeglasses detection. In addition, CNN-02 implemented with convolutional and up-sampling layers for eyeglasses segmentation. For this scheme, an experimental investigation on large scale VISOB dataset shows a 100% eyeglasses detection and 91% segmentation accuracy by our proposed approach.

The advantages of the proposed cascaded CNNs are (a) first CNN feeds the second CNN with emphasized eyeglasses, (b) instead of training two separate models, one cascaded model is trained for both detection and segmentation which is computationally efficient, and (c) eyeglasses segmentation could be obtained using trained weights obtained from CNN-01. Moreover, the microstructure features of convolutional layers showed a good performance in conjunction with fully connected layers for pixel-wise segmentation. As a part of future work, we plan to use larger scale databases and deeper CNN's for further performance enhancement.

The second segmentation scheme was a convolutional de-convolutional network which consisting of six main stages, namely ROI extraction, frame in-painting, CNN

model application, mask clean-up, segmentation and frame removal. This CNN model has been implemented with two convolutional layers, two de-convolutional, one max-pooling, one up-sampling layer, and one custom (lamda) layer. The last custom (Lamda) layer has one channel. The Lamda layer has been customized to satisfy four main operations which are summation of the sixteen heat maps, apply Gaussian low pass filter, apply Otsu's Binarization, find the loss metric using intersection over the union between the predicted binary mask and the reference target binary mask.

The proposed CNN model has been trained and evaluated using a dataset consists more than 4.1 million images, divided as 80% (3301280) training set, 15% (619040) validation set, and 5% (4269) testing set. The experimental results show a 97% segmentation accuracy on real eyeglasses images by our proposed approach.

To recap, the reported results show that the convolutional de-convolutional approach achieves better segmentation results over the cascaded approach. In other words, Convolution-deconvolution approach predicts the frame mask of eyeglasses better than cascaded approach; however, the cascaded approach involved eyeglasses detection which is required before segmentation.

5.3 Conclusions For User Authentication using Soft-Biometric

With the modern mobile technology revolution, biometrics has become a viable alternative to PINs and passwords for secure mobile access and transactions. Recent research has shown the potential of eyebrows as an independent biometric modality for recognition. Eyebrows leverage the benefit of being one-sixth of the facial region, and is

therefore computationally efficient and offer fast throughput. This work aims to exploit the eyebrow region for mobile user authentication and to show the impact factor such as eyeglasses existence on eyebrow-based user authentication. To this aim, Histogram of Oriented Gradients (HOG), a local descriptor, and GIST, a global descriptor, extracted from left and right eyebrow regions are evaluated.

Besides, we evaluated eyebrows biometrics for user authentication in smartphones. This is the first large-scale evaluation of eyebrow-based biometric on a mobile dataset in presence of eyeglasses. The experimental results suggest minimum EER of 3.23% on fusing the SVM output of the left and right eyebrow ROIs using GIST descriptors for OPPO mobile device VISOB database. Also, The experimental results suggest minimum EER of 1.08% on fusing the SVM output of the left and right eyebrow ROIs using HOG descriptors for FERET database. With the continuous advancement in the mobile hardware technology, the proposed approach can be used as a viable device-side application for user authentication and as a soft biometric trait. Also, this work shows the impact of eyeglasses on eyebrow ROI mitigated using eyeglasses segmentation and inpainting.

Local, global, and deep learning features have been used as image descriptors for user authentication under different light conditions. Also, the scores have been fused in three stages, namely, unit fusion (right, and left fusion), features fusion (local, global, and deep features fusion), and modality fusion (eyebrow and eyeglasses fusion). The reported results show that the deep feature outer performed the other features (local, and global fusion), and the feature fusion enhanced the authentication system in absence of eyeglasses. The best result of 0.63% EER using score level fusion of handcraft descriptors

(HOG, and GIST) with a deep VGG16 descriptor for eyebrows-based user authentication. Moreover, the best result of 3.44% EER using score level fusion of handcraft descriptors (HOG, and GIST) with a deep VGG16 descriptor for eyebrows-based user authentication.

Moreover, the deep (VGG16) descriptor outperforms other traditional handcrafted descriptors (GIST and HOG) for eyebrow-based with the presence of eyeglasses and eyeglasses-based user authentication. The fusion of eyebrows and eyeglasses modalities at score level using VGG16 feature descriptor showed enhanced performance over individual modalities.

5.4 Future Work

As a part of future work, prediction of eyeglasses frames using deep learning approach, I will use the full face to obtain the predicted frame with more deep and functional models (consequential CNN models).

For eyeglasses in-painting, I will use a more accurate in-painting process based on deep learning method rather than conventional method for eyeglasses removal .

As a part of future work in deep feature extraction, I will evaluate different pre-trained CNN models such as AlexNet, VGG19, ResNet-50, InceptionV3, and MobileNet. This evaluation should be applied to different biometric modalities such as faces, iris, and chin using different light conditions.

In addition, I will train to fuse the deep features (such as VGG or ResNet features) of primary modality (such face or iris) with the deep feature of soft-biometric modalities such as eyebrows or chin.

Also, as part of future work, I will also investigate soft-biometric prediction from ocular and facial images captured using mobile devices such as gender and age.

REFERENCE LIST

[1] Famous football stars and their look- alike. [Online]. Available: http://www.nairaland.com/2104767/famous-football-stars-look-alike

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[3] Deep cnn and supervision learning for visual recognition. [Online]. Available: https://blog.heuritech.com/2016/02/29/a-brief-report-of-the-heuritech-deep-learning-meetup-5/

[4] X. Han, T. Leung, Y. Jia, R. Sukthankar, and A. C. Berg, "Matchnet: Unifying feature and metric learning for patch-based matching," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. IEEE, 2015, pp. 3279–3286.

[5] B. Wu, H. Ai, and R. Liu, Eds., *Glasses Detection by Boosting Simple Wavelet Features*. ICPR , Cambridge , UK, Aug. 2004.

[6] A. A. Ross *et al.*, "Information fusion in fingerprint authentication," Ph.D. dissertation, Michigan State University. Department of Computer Science & Engineering, 2003.

[7] U. Uludag, "Secure biometric systems," Ph.D. dissertation, Michigan State University. Computer Science & Engineering, 2006.

[8] A. K. Jain and S. Z. Li, *Handbook of face recognition*. Springer, 2011.

[9] S. C. Draper, A. Khisti, E. Martinian, A. Vetro, and J. S. Yedidia, "Using distributed source coding to secure fingerprint biometrics," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 2. IEEE, 2007, pp. II–129.

[10] A. K. Jain, R. Bolle, and S. Pankanti, *Biometrics: personal identification in networked society: Book*. Springer Science & Business Media, 2006, vol. 479.

[11] C. Rathgeb, A. Uhl, and P. Wild, *Iris biometrics: from segmentation to template security: Book*. Springer Science & Business Media, 2012, vol. 59.

[12] A. Rattani and R. Derakhshani, "Ocular biometrics in the visible spectrum: A survey," *Image and Vision Computing*, vol. 59, pp. 1–16, 2017.

[13] A. Kong, D. Zhang, and M. Kamel, "A survey of palmprint recognition," *pattern recognition*, vol. 42, no. 7, pp. 1408–1418, 2009.

[14] C. Hegde, P. D. Shenoy, K. Venugopal, and L. M. Patnaik, "Authentication using finger knuckle prints," *signal, image and video processing*, vol. 7, no. 4, pp. 633–645, 2013.

[15] A. Ross, A. Jain, and K. Nandakumar, *Introduction to Biometrics: A Textbook*. Springer, 2011.

[16] L. Wang, *Behavioral Biometrics for Human Identification: Intelligent Applications: Intelligent Applications: A book*. IGI Global, 2009.

[17] R. Katiyar, V. K. Pathak, and K. Arya, "A study on existing gait biometrics approaches and challenges," *International Journal of Computer Science*, vol. 10, no. 1, pp. 135–144, 2013.

[18] G. Bailador, C. Sanchez-Avila, J. Guerra-Casanova, and A. de Santos Sierra, "Analysis of pattern recognition techniques for in-air signature biometrics," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2468–2478, 2011.

[19] A. K. Jain, P. Flynn, and A. A. Ross, *Handbook of biometrics*. Springer Science & Business Media, 2007.

[20] D. A. Reid and M. S. Nixon, "Using comparative human descriptions for soft biometrics," in *Biometrics (IJCB), 2011 International Joint Conference on*. IEEE, 2011, pp. 1–6.

[21] H. Ailisto, E. Vildjiounaite, M. Lindholm, S.-M. Mäkelä, and J. Peltola, "Soft biometrics: combining body weight and fat measurements with fingerprint biometrics," *Pattern Recognition Letters*, vol. 27, no. 5, pp. 325–334, 2006.

[22] D. A. Reid, M. S. Nixon, and S. V. Stevenage, "Soft biometrics; human identification using comparative descriptions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 6, pp. 1216–1228, 2014.

[23] S. Denman, C. Fookes, A. Bialkowski, and S. Sridharan, "Soft biometrics: unconstrained authentication in a surveillance environment," in *Digital Image Computing: Techniques and Applications, 2009. DICTA'09.* IEEE, 2009, pp. 196–203.

[24] Y. Gu, Y. Zhang, and Y. Zhang, "A novel biometric approach in human verification by photoplethysmographic signals," in *Information Technology Applications in Biomedicine, 2003. 4th International IEEE EMBS Special Topic Conference on*. IEEE, 2003, pp. 13–14.

[25] A. K. Jain, S. C. Dass, and K. Nandakumar, "Soft biometric traits for personal recognition systems," in *Biometric Authentication*. Springer, 2004, pp. 731–738.

[26] A. S. Mohammad and J. A. Al-Ani, "Towards ethnicity detection using learning based classifiers," in *Computer Science and Electronic Engineering (CEEC), 2017*. IEEE, 2017, pp. 219–224.

[27] S. Magnet, *When biometrics fail: Gender, race, and the technology of identity*. Duke University Press, 2011.

[28] F. Juefei-Xu and M. Savvides, "Can your eyebrows tell me who you are?" in *Signal Processing and Communication Systems (ICSPCS), 2011 5th International Conference on*. IEEE, 2011, pp. 1–8.

[29] T. H. N. Le, U. Prabhu, and M. Savvides, "A novel eyebrow segmentation and eyebrow shape-based identification," in *Biometrics (IJCB), 2014 IEEE International Joint Conference on*. IEEE, 2014, pp. 1–8.

[30] X. Yang, X. Xu, and C. Liu, "Eyebrow recognition based on sparsity preserving projections," in *Conference Anthology, IEEE*. IEEE, 2013, pp. 1–4.

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

[32] M.-H. Lin, S. Gao, K.-S. Huang, and J.-M. Wang, "Sequence-encoded multiple biometric template security system," May 21 2002, uS Patent 6,393,139.

[33] A. Rattani *et al.*, "Adaptive biometric system based on template update procedures, 2010," *PhD, University of Cagliari, Italy*.

[34] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–I.

[35] D. E. King, "Dlib-ml: A Machine Learning Toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009. [Online]. Available: http://jmlr.csail.mit.edu/papers/v10/king09a.html

[36] S. Yoon, *Fingerprint recognition: models and applications*. Michigan State University, 2014.

[37] A. Nagar, *Biometric template security*. Michigan State University. Computer Science, 2012.

[38] D. Dasgupta, A. Roy, and A. Nag, *Advances in User Authentication: book*. Springer, 2017.

[39] A. Ramchandra and R. Kumar, "Overview of face recognition system challenges," *International Journal Of Scientific & Technology Research*, vol. 2, no. 8, pp. 234–236, 2013.

[40] U. Park, A. Ross, and A. K. Jain, "Periocular biometrics in the visible spectrum: A feasibility study," in *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*. IEEE, 2009, pp. 1–6.

[41] C.-L. Lin and K.-C. Fan, "Biometric verification using thermal images of palm-dorsa vein patterns," *IEEE Transactions on Circuits and systems for Video Technology*, vol. 14, no. 2, pp. 199–213, 2004.

[42] B. Arbab-Zavar and M. S. Nixon, "Robust log-gabor filter for ear biometrics," in *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE, 2008, pp. 1–4.

[43] M. De Marsico, C. Galdi, M. Nappi, and D. Riccio, "Firme: Face and iris recognition for mobile engagement," *Image and Vision Computing*, vol. 32, no. 12, pp. 1161–1172, 2014.

[44] N. M. Nasrabadi, "Pattern recognition and machine learning," *Journal of electronic imaging*, vol. 16, no. 4, p. 049901, 2007.

[45] B. Schölkopf and A. J. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[46] A. A. Salah, "Machine learning for biometrics," in *Machine Learning: Concepts, Methodologies, Tools and Applications*. IGI Global, 2012, pp. 704–723.

[47] A. S. Mohammad, A. Rattani, and R. Derahkshani, "Eyeglasses detection based on learning and non-learning based classification schemes," in *Technologies for Homeland Security (HST), 2017 IEEE International Symposium on*. IEEE, 2017, pp. 1–5.

[48] R. Palaniappan and D. P. Mandic, "Biometrics from brain electrical activity: A machine learning approach," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 4, pp. 738–742, 2007.

[49] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[50] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[51] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution,

and fully connected crfs," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 834–848, 2018.

[52] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[53] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014, pp. 675–678.

[54] I. Guyon and A. Elisseeff, "An introduction to feature extraction," in *Feature extraction*. Springer, 2006, pp. 1–25.

[55] M. Nixon and A. Aguado, "Featureextractionandimageprocessing for computer vision." Academic Press, London, 2012.

[56] R. C. Gonzalez, "Digital image processing." Prentice hall, 2016.

[57] P. Soille, "Morphological image analysis: principles and applications." Springer Science & Business Media, 2013.

[58] K. Grauman and B. Leibe, "Visual object recognition," *Synthesis lectures on artificial intelligence and machine learning*, vol. 5, no. 2, pp. 1–181, 2011.

[59] S. A. Winder and M. Brown, "Learning local image descriptors," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[60] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

[61] T. Watanabe, S. Ito, and K. Yokoi, "Co-occurrence histograms of oriented gradients for pedestrian detection," in *Pacific-Rim Symposium on Image and Video Technology*. Springer, 2009, pp. 37–47.

[62] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[63] K. Zagoris, S. A. Chatzichristofis, and A. Arampatzis, "Bag-of-visual-words vs global image descriptors on two-stage multimodal retrieval," in *Proceedings of the 34th international ACM SIGIR conference on research and development in information Retrieval*. ACM, 2011, pp. 1251–1252.

[64] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision*, vol. 42, no. 3, pp. 145–175, 2001.

[65] Y. Liu and H. Zhang, "Visual loop closure detection with a compact image descriptor," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*. IEEE, 2012, pp. 1051–1056.

[66] I. Sikirić, K. Brkić, and S. Šegvić, "Classifying traffic scenes using the gist image descriptor," *arXiv preprint arXiv:1310.0316*, 2013.

[67] M. Douze, H. Jégou, H. Sandhawalia, L. Amsaleg, and C. Schmid, "Evaluation of gist descriptors for web-scale image search," in *Proceedings of the ACM International Conference on Image and Video Retrieval*. ACM, 2009, p. 19.

[68] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, "Context-based vision system for place and object recognition," in *null*. IEEE, 2003, p. 273.

[69] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*. Springer, 2006, pp. 404–417.

[70] M. Yang, K. Kpalma, and J. Ronsin, "A survey of shape feature extraction techniques," 2008.

[71] X. Bo, W. Junwen, L. Guangjie, and D. Yuewei, "Image copy-move forgery detection based on surf," in *Multimedia information networking and security (MINES), 2010 international conference on*. IEEE, 2010, pp. 889–892.

[72] A. Gordo, J. Almazán, J. Revaud, and D. Larlus, "Deep image retrieval: Learning global representations for image search," in *European Conference on Computer Vision*. Springer, 2016, pp. 241–257.

[73] N. Wang and D.-Y. Yeung, "Learning a deep compact image representation for visual tracking," in *Advances in neural information processing systems*, 2013, pp. 809–817.

[74] T. Xiao, Y. Xu, K. Yang, J. Zhang, Y. Peng, and Z. Zhang, "The application of two-level attention models in deep convolutional neural network for fine-grained image classification," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*. IEEE, 2015, pp. 842–850.

[75] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[76] C. Feichtenhofer, A. Pinz, and A. Zisserman, "Convolutional two-stream network fusion for video action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1933–1941.

[77] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[78] R. Blanco-Gonzalo and R. Sanchez-Reillo, "Biometrics on mobile devices," *Encyclopedia of Biometrics*, pp. 282–289, 2015.

[79] H. Saevanee and P. Bhatarakosol, "User authentication using combination of behavioral biometrics over the touchpad acting like touch screen of mobile device,"

Proceedings of the 2008 International Conference on Computer and Electrical Engineering, ICCEE 2008, pp. 82–86, 2008.

[80] J. Anil and L. P. Suresh, "Literature survey on face and face expression recognition," in *Circuit, Power and Computing Technologies (ICCPCT), 2016 International Conference on*. IEEE, 2016, pp. 1–6.

[81] A. FernÃ¡ndez, R. GarcÃa, R. Usamentiaga, and R. Casado, "Glasses detection on real images based on robust alignment," in *IEEE 3rd International Conference on Future Internet of Things and Cloud (FiCloud)*, May 2015, pp. 519–531.

[82] Z. Jing and R. Mariani, "Glasses detection and extraction by deformable contour," in *ICPR*, 2000.

[83] H. Wu, G. Yoshikawa, T. Shioyama, S. Lao, and M. Kawade, "Glasses frame detection with 3d hough transform," in *ICPR*, 2002.

[84] X. Jiang, M. Binkert, B. Achermann, and H. Bunke, "Towards detection of glasses in facial images," in *Proc. 14th ICPR*. Brisbane, 1998.

[85] C. Wu, C. Liu, H. Shum, Y. Xy, and Z. Zhang, "Automatic eyeglasses removal from face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 3, pp. 322–336, Mar. 2004. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2004.1262319

[86] Mohammad A Rattani A and D. R, "Eyeglasses Detection based on Learning and Non-learning based Classification Schemes," *IEEE International Symposium on Technologies for Homeland Security, 2017*, p. (accepted), 2017.

[87] A. S. Mohammad, A. Rattani, and R. Derakhshani, "Comparison of squeezed convolutional neural network models for eyeglasses detection in mobile environment," *Journal of Computing Sciences in Colleges*, vol. 33, no. 5, pp. 136–144, 2018.

[88] C. Du and G. Su, "Eyeglasses removal from facial images," *Pattern Recognition Letters*, vol. 26, no. 14, pp. 2215–2220, 2005.

[89] W. K. Wonga and H. Zhao, "Eyeglasses removal of thermal image based on visible information," *Information Fusion*, vol. 14, no. 2, pp. 163–176, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.inffus.2011.09.002

[90] W.-c. Cheng, H.-c. Liao, M.-h. Pan, and C.-c. Chen, "A Fatigue Detection System with Eyeglasses Removal," *Advanced Communication Technology (ICACT), 2013 15th International Conference on*, pp. 331–335, 2013.

[91] X. Jia and J. Guo, "Eyeglasses removal from facial image based on phase congruency," *Proceedings - 2010 3rd International Congress on Image and Signal Processing, CISP 2010*, vol. 4, no. 3, pp. 1859–1862, 2010.

[92] J. Heo, S. G. Kong, B. R. Abidi, and M. A. Abidi, "Fusion of Visual and Thermal Signatures with Eyeglass Removal for Robust Face Recognition," *2004 Conference on Computer Vision and Pattern Recognition Workshop*, no. July, pp. 122–128, 2004. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1384918

[93] Y.-K. W. Y.-K. Wang, J.-H. J. J.-H. Jang, L.-W. T. L.-W. Tsai, and K.-C. F. K.-C. Fan, "Improvement of Face Recognition by Eyeglass Removal," *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010 Sixth International Conference on*, 2010.

[94] J.-S. Park, Y. H. Oh, S. C. Ahn, and S.-W. Lee, "Glasses Removal from Facial Image Using Recursive PCA Reconstruction," *Lecture Notes in Computer Science / Audio- and Video-Based Biometric Person Authentication*, vol. 2688, no. 3, p. 1058, 2003.

[95] J. S. Park, Y. H. Oh, S. C. Ahn, and S. W. Lee, "Glasses removal from facial image using recursive error compensation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 5, pp. 805–811, 2005.

[96] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *International Journal of Computer Vision*, vol. 63, no. 2, pp. 153–161, 2005. [Online]. Available: http://dx.doi.org/10.1007/s11263-005-6644-8

[97] Zuiderveld and Karel, "Contrast limited adaptive histograph equalization," *Graphic Gems IV. San Diego: Academic Press Professional*, pp. 474 – 485., 1994.

[98] K. Kuman, T. Babu, N. vaishy, and K. Lavanya, "Stock market prediction by non-linear combination based on support vector machine regression model." *International Journal of Advanced Research in Computer Science*, vol. 7, no. 7, 2017.

[99] A. Mohammad, "Weather temperature forecasting using artificial neural network." *Journal of Engineering and Development and Al-Mustansyriah University*, vol. 15, no. 2, 2011.

[100] Y. Geng, J. Chen, R. Fu, G. Bao, and K. Pahlavan, "Enlighten wearable physiological monitoring systems: On-body rf characteristics based human motion classification using a support vector machine." in *IEEE transactions on mobile computing*, vol. 15(3), 2016, pp. 656–671.

[101] A. Mohammad, A. Hamoudi, and Y. Kareem, "Hand writing numbers detection using artificial neural networks," *Iraqi Journal of Computer and Communication and Control and System Engineering and University of Technology*, vol. 13, no. 2, 2013.

[102] D. Kinga and J. B. Adam, "A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[103] A. Rattani, R. Derakhshani, S. K. Saripalle, and V. Gottemukkula, "ICIP 2016 competition on mobile ocular biometric recognition," *Proceedings - International Conference on Image Processing, ICIP*, vol. 2016-Augus, pp. 320–324, 2016.

[104] http://www.nist.gov/itl/iad/ig/feret.cfm., 2011, [Date created: January 25, 2011 and Last updated: May 13, 2015].

[105] E. Vazquez-Fernandez and D. Gonzalez-Jimenez, "Face recognition for authentication on mobile devices," *Image and Vision Computing*, vol. 55, pp. 31 – 33, 2016.

[106] F. Juefei-Xu and M. Savvides, "Can your eyebrows tell me who you are?" *5th International Conference on Signal Processing and Communication Systems, ICSPCS'2011 - Proceedings*, pp. 0–7, 2011.

[107] T. H. N. Le, U. Prabhu, and M. Savvides, "A novel eyebrow segmentation and eyebrow shape-based identification," *IJCB 2014 - 2014 IEEE/IAPR International Joint Conference on Biometrics*, 2014.

[108] X. Yang, X. Xu, and C. Liu, "Eyebrow recognition based on sparsity preserving projections," *2013 IEEE Conference Anthology, ANTHOLOGY 2013*, no. 1, 2014.

[109] S. Bharadwaj, H. S. Bhatt, M. Vatsa, and R. Singh, "Periocular biometrics: When Iris recognition fails," *IEEE 4th International Conference on Biometrics: Theory, Applications and Systems, BTAS 2010*, no. c, pp. 2–7, 2010.

[110] V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello, "Continuous user authentication on mobile devices: Recent progress and remaining challenges," *IEEE Signal Processing Magazine*, vol. 33, no. 4, pp. 49–61, 2016.

[111] Y. Li, H. Li, and Z. Cai, "Human eyebrow recognition in the matching-recognizing framework," *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 170–181, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.cviu.2012.10.007

[112] M. Ichino and Y. Yamazaki, "Soft biometrics and its application to security and business," in *Biometrics and Kansei Engineering (ICBAKE), 2013 International Conference on*. IEEE, 2013, pp. 314–319.

[113] Y. Li and X. Li, "Hmm based eyebrow recognition," in *Intelligent Information Hiding and Multimedia Signal Processing, 2007. IIHMSP 2007. Third International Conference on*, vol. 1. IEEE, 2007, pp. 135–138.

[114] F. C. Li Yujian, "Eyebrow Recognition: a New Biometric Technique," in *Proceedings of the Ninth IASTED International Conference SIGNAL AND IMAGE PROCESSING*, 2007, pp. 206–210.

[115] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657–1663, 2010. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1774088.1774408http://ieeexplore.ieee.org/document/5339068/metrics

[116] C. Jun-bin, Y. Haitao, and D. Lili, "Eyebrows Identity Authentication Based on Wavelet Transform and Support Vector Machines," *Physics Procedia*, vol. 25, pp. 1337–1341, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1875389212006578

[117] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

VITA

Ahmad Saeed Mohammad was born in Iraq, Baghdad. He finished his bachelors degree in Computer Engineering, and his first master degree from University of Almustinsirya, Iraq-Baghdad (2002) and Anbar University, Iraq-Ramadi (2005), respectively. After that, He finished his second master degree in electrical engineering from University of Missouri-Kansas City (2015). At the time of writing this vita, he is current a Ph.D. candidate in electrical and computer engineering at University of Missouri-Kansas City.