Effective Features and Machine Learning Methods for Human Recognition Based on Multi-biometric Systems

Inas Al-TAIE

A thesis submitted for the degree of

Doctor of Computer Science at the

Department of Computer Science and Electronic Engineering

University of Essex

March 2020

Abstract

Biometrics are fundamental to a wide range of technologies that require credible authentication approach to approve personal identification. This thesis aims to identify effective features and machine learning methods for human recognition based on multiple biometrics and produce the sufficient combination of single biometric systems suitable in specific applications for identification purposes. For example, banking systems which use multi-biometric authentication for login procedures and the police and criminal evidence applications. This thesis goes through general ideas of the most common biometrics that used for a personal identification and their application areas. It has been focusing on two well-known linear subspace-learning techniques that have become the most popular techniques for face recognition; PCA and LDA. Different face classification techniques have been presented including supervised and unsupervised learning methods. The research focuses on assessing vision system performance and different databases that are suitable for biometric research. This research has made a number of contributions. Firstly improving the Viola-Jones face detection performance by using the brightness channel in HSV And HSL color spaces. Distance similarity measures have been compared for PCA- and LDA-based face, ear and palm biometrics. The face and ear recognition performance using SVM based on PCA and SVM based on a combination of PCA and LDA techniques have been compared with the PCA and LDA techniques based on distance similarity measures. Face, ear, palmprint, eye, and hand biometric recognition has been applied using three Deep and Shallow Convolutional Neural Networks (GoogleNET, VGG16, ResNET). An implementation of a person identification system fusing different combinations of biometric modalities; face, ear, eye, hand, and palmprint at score level has

been examined. Comparison of these combinations has been employed to assess the performance. Finally, different sizes of training/testing set are examined to achieve high recognition performance.

Acknowledgements

At the forefront, I would like to express my gratitude to the University of Baghdad and the Ministry of Higher Education and Scientific Research of Iraq for offering me a scholarship to obtain the PhD degree, and providing me the opportunity to undertake this experience. I would like also to thank the University of Essex, UK at which I was able to carry out this research.

I am indebted to my supervisor, Dr. Adrian F. Clark, for his continuous support and guidance throughout the thesis stages. He has given his knowledge and put in effort at all times for the benefit of this thesis.

I would also like to acknowledge the immense impact and support that my family has given me. In particular, my mother whose encouragement was beyond description.

I am more than grateful to my husband, (Dr. Nassr Azeez) for his never-ending support throughout this research. Moreover, my special thank to my wonderful boys, Abdulazeez, Mohammed, and my new born baby Ryan for always making me smile and for their patience and understanding whilst doing my PhD thesis work.

Finally, my appreciation also goes out to my friends in Iraq and UK for their support and to all the rest of the family for believing in me and keeping me going.

...

Contents

List of Figures

List of Tables

Glossary

Correlation Is a statistical measure that indicates the extent to which two or more variables fluctuate together.

Gaussian Smoothing It is a two 2D convolution operator that used to 'blur' images and remove details and noise like mean filter but it uses a different kernel.

Canny Edge Detection It is a multi-step algorithm that can detect edges with noise suppressed at the same time.

Orientation A line between two points P_1 and P_2 has no given direction but has well defined orientation.

Image Orientation provides a way to specify a rotation to be applied to an image.

Gaber Filter It is a linear filter that used for texture analysis.

Zero Crossing Point It is the point when the sign of a mathematical function changes.

Convolution Is a mathematical operation on two function (f and g) to produce a third function that express how the shape of one is modified by the other.

FWER (Family Wise Error Rate) Is the probability of making one or more false discoveries of type I errors when performing multiple hypotheses tests.

Type I Error Falsely inferring the existence or reality of something that is in fact not real or does not in fact exist.

Haar Wavelet Is a sequence of rescaled squared shaped function. The main disadvantage, it is not continuous therefore it is not differentiable.

Continuous Function small change in the input results to arbitrary small change in the output.

A vector Space It is a collection of objects called vectors which may be added or multiplied (scaled) by numbers called scalers.

A Distance Function It is a function that defines a distance between each pair of elements of a set.

Inner product space It is a vector space with additional structure called innerproduct. This additional structure associated each pair of vectors in the space with a scaler quantity known as the innerproduct of the vectors.

Classification It is the task of approximating a mapping function (f) from input (x) to discrete output variables (y).

Regression It is the task of approximating a mapping function (f) from input (x) to a continuous output variables (y).

Loss Function measure the inconsistency between predicted value (y) and the actual label.

Softmax Function Calculates the probabilities distribution of an event over "n" different events.

Orthogonal Vectors Two vectors are orthogonal if they are perpendicular, i.e., they form a right angle.

A Right Angle It is an angle of exactly 90∘(degrees), corresponding to a quarter turn.

Chapter 1

Introduction

1.1 Human Biometrics

With the progress of technology and the development of systems which rely on its work to provide personal identification services, the need to find an effective system for discrimination of individuals is growing. To use an ATM or to access a bank account, a person needs to prove their identity. There are different means that provide users identification such as passports, access cards, PIN (the unique number that is given to each card) and passwords. Although all of these means provide satisfactory proof of personal identification, unfortunately they have many drawbacks: in particular, passwords, PIN, access cards can be lost or forgotten [1]. Hence, it has become imperative to find more efficient security systems to determine the identity of a person. Consequently, biometrics have become the focus of attention. The term biometric refers to the science in which an entity can be distinguished automatically depending on their physical or behavioural characteristics. Biometric technology can be used for both identification or verification/authentication purposes. Verification systems are concerned with validating of a person's claim. The identification system is used to identify a person which can be based on several human traits. There are certain criteria that must be taken into consideration to determine the possibility of using these traits in biometric systems: universality, uniqueness, permanence, collectab-

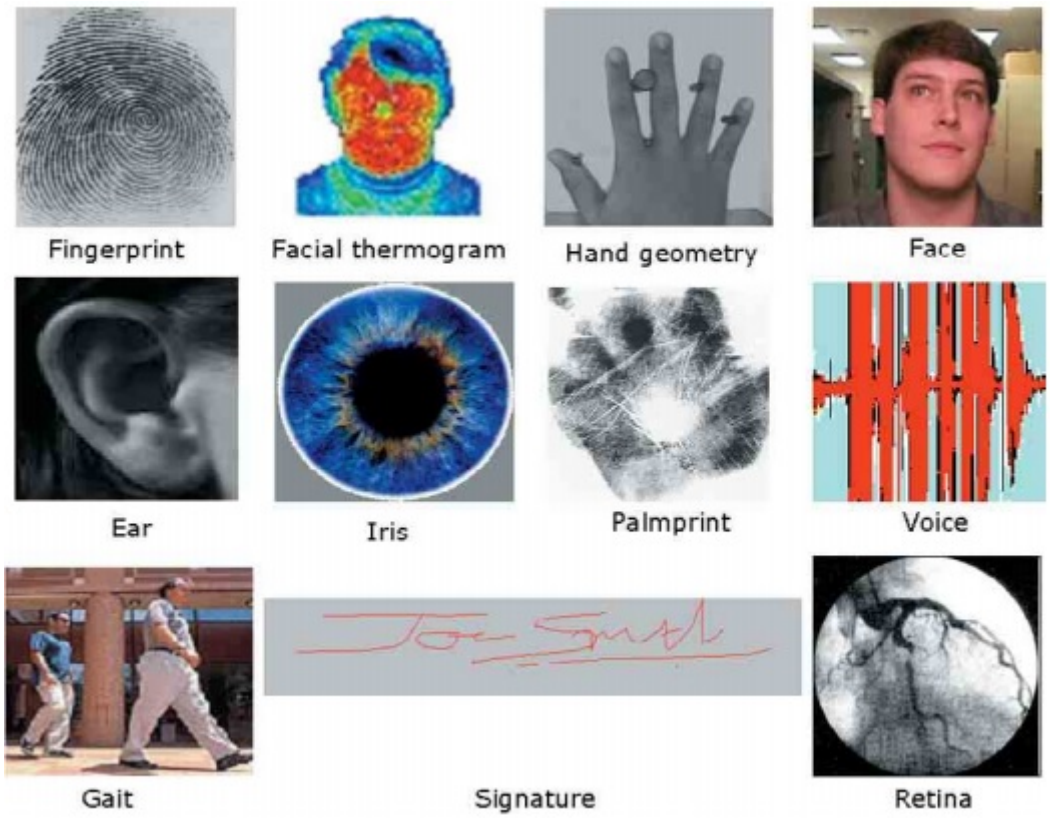


Figure 1.1: Examples of biometric traits used authentication of an individual [4].

ility, performance, acceptability, and circumvention [2]. The role of biometrics has become important in all types of security [3]. Figure 1.1 shows some of the biometrics used for person authentication. Biometric methodologies involve two techniques:

- **Physiological techniques**

  Attempt to determine the identity of a person depending on unique physiological traits of the user. Examples are discussed in the following paragraphs.

  Face Recognition Each face has a set of features, that distinguish it from other people's faces. In other words, key challenge of face recognition is how to learn these features in programs that are used for the purpose of identifying faces [5] [4]. Face recognition system has become one of the most popular and successful applications in computer vision [6]. However, it needs to exceed several problems. One of the main obstacles is how to identify a person whose facial picture may

not be frontal. Consequently, it is not easy to obtain a system capable to discriminate a person with a rotated face. The recognition result would be affected by the size of the image because some approach needs a standard size images and sometimes small size image is not clear enough for recognition. Another problem, a person may change dramatically within a short period of time due to glasses, makeup and head hair style. Another major problem for face recognition is lighting condition. Under different lighting condition, the same person may be appear quite different. In addition, facial expression will also cause a face varies. All of the mentioned problems will decrease the accuracy of such systems [7] [8].

Retina and Iris Retina and iris scanning depend on the unique physiological properties of the eye to identify a person. The retina consists of neural cells which are located in the back of the eye. Due to the sophisticated organisation of the capillaries which provide the retina with blood, each retina is unique to each person. Iris scanning uses the external rather than internal surface of the eye; It is reported to be the most accurate and effective all over the world. It is a fast process, a user can complete the process within just a few seconds [9]. Despite its reliability, There are concerns about hygiene and accessibility issues. If scanning device is shared and necessitates users to place their eyes on sockets used by others, it might become unhygienic unless completely cleaned out after each use. In terms of accessibility, iris scanning may be problematic for users with certain medical conditions. For example, diabetes could change the look of the eye over time. In addition, commercial iris scanning systems have proved unpopular due to the high cost [10] [11].

Fingerprint Fingerprint identification has one of the highest levels of accuracy. Its low-cost and high convenience makes them one of the common authentication methods. Fingerprint can be viewed as an oriented texture pattern. It uses the impressions made by the minute ridge formations or patterns found on the fingertips. The arrangement of ridge patterns for each person is unique, and the patterns of any one individual remain permanent, unchanged throughout

life [12]. Security systems based on fingerprint have some restrictions and limitations for example, a fingerprint may have some burns, cuts, and small temporary or permanent injuries. Moreover, fake fingers that made up from silicon and/or gelatin have the ability to hack fingerprint based recognition systems [13].

Hand Geometry This is based on the fact that each person has a different hand shape and with the time it does not change dramatically [14]. This system is highly implemented for public acceptance, ease of use and collection capability [15]. There are disadvantages of using the hand recognition device, in the very young and old age hand geometry and size changes which can cause the wrong result. In terms of hygiene issues, some users are reluctant to put their hand on the device for scanning where others have touched. The essential disadvantage of the hand recognition device is, it is considered fairly expensive [16] [17].

DNA (Dioxyribo Nucleic Acid) It is also known as DNA fingerprinting or DNA profiling. DNA recognition has been used to identify individuals by identifying the uniqueness in their DNA profiles. Studies show that a person's DNA is absolutely unique (except for identical twins) and is present in every cell [5]. Since every human is having DNA, it has high universality. It is considered to be highly stable due to the fact that DNA doesn't change over time. Although, DNA recognition has high performance and accuracy it is not usually accepted by the people because DNA could detect more than just the identity (critical information related to genetic of persons). The main disadvantages of using DNA, it needs several hours to be gained and the test is expensive to apply [18] [5].

- **Behavioural techniques**

Behavioral biometrics relate to a specific behaviour of a human. Examples include:

Signature One of the behavioural techniques that used to distinguish people is by their hand-written signature. An automatic authentication system not only based on this but also on the dynamic movement during writing [19]. The significant benefit of signature recognition is that it is highly resistant to impostors. For example, it is quite easy to form a signature but it is so difficult to copy the behavioral patterns inherent to the signing process [20]. Signature Recognition is well suited to high-value transactions. Signature recognition is widely accepted by users compared to the other biometric systems. On the downside, it is prone to high error rates, essentially when the behavioral features of signatures are unsteady [5] [19].

Voice Recognition This works by analysing the patterns of the air pressure and the waveforms that resulting when a person is talking. Although this technique has become more advanced in recent years and has low cost, it is considered unreliable due to low accuracy [21]. Cold may cause voice problems and recording tapes may be used to attack the system. Besides, the voice is the easiest to duplicate of all the biometric options [5].

Keystone Dynamics It is one of the most modern methods that is used in biometric systems. It analyses the way of typing keys or the access time that spent by a user to find keys. But this technique depends mostly on the user mood [3] [5]. Using keystroke dynamics as an authentication tool has some disadvantages. Firstly, typing patterns can be inconsistent due to the fact that cramped muscles and sweaty hands could significantly change a user's typing pattern. Besides, typing patterns change depending on the keyboard type being used, which leads to a complicated verification process [5].

Gait Recognition Since the last few decades, human biometric recognition using gait has become one of a challenging area for the researchers in the applications such as security and surveillance. Because of the ease of capturing the data, it has attracted the new researchers to explore it.

The most attractive characteristic of gait is its unobtrusiveness, it can be obtained from a distance and with no need to the prior consent of the observed subject. Most other biometrics such as iris, face, and hand geometry can be captured only by a physical contact or at a close distance from the observed subject [25]. There are some apparent limitations in gait capturing that making it difficult to identify and record all parameters that affect gait. Gait recognition has to depend on a video series taken in controlled or uncontrolled surrounding. Even if the accuracy of measuring certain gait parameters improves, it remains unknown if the knowledge of the parameters provides sufficient discrimination ability to enable largescale deployment of gait recognition technologies. Moreover, studies report that gait changes over time and it is affected by weight carrying and clothing conditions [23] [24]. For these reasons, gait is not expected to be used as a single means of identification of individuals. Instead, it is seen as a potentially valuable in a multi-modal biometric system [25].

1.2 Application Areas of Biometrics

Biometric technology can make a key contribution to a great number of applications. The following are some of these application areas [26] [27].

Justice/Law Enforcement Law enforcement agencies always in a great need on accurate and quick identification of individuals specially suspects and criminals and it can be efficiently achieved with the use of biometrics. Error in identification can result in a wrongly identified people and innocent people getting hurt or losing their lives. Over a long time, police agencies used different means to identify criminals for example, AFIS (An Automated Fingerprint Identification System), face, palm, vein, voice, iris and DNA recognition. Nowadays, law enforcement trend to apply multi-modal biometrics: face, fingerprint and hand-shape recognition to improve performance [27].

Border Control Systems Every nation focuses to secure their borders from illegal entry of individuals or goods. This action put direct pressure on border agencies to provide a safe and secure border environment for the movement of travellers worldwide. Border control biometrics is considered an important area of application for the biometric technology [28]. biometric identification is an attractive choice to achieve robust border control security systems due to the fact that no two individuals have the same traits such as facial patterns, fingerprints, iris patterns, and DNA. Biometric identification systems have the ability to identify people more quickly and accurately than traditional checking like passports and visas which might result in identification errors easily and is time consuming. Conversely, biometric identification can accurately identify any person within seconds by depending on physiological or behavioural traits which cannot be lost, forgotten, stolen. Different means used to recognize an individual's identity, for example iris scanners, face recognition, and fingerprint [28]. Nowadays, biometric identification systems are assessed to be much more reliable for border control due to their dependence on multiple biometric traits for identification rather than a single biometric trait. The integration of two or more factors of biometric authentication systems help to meet strict performance requirements put by border control security conscious customers [29].

Access Control Information security essentially ensures the privately, safety and availability of information. It basically supplies the necessary protection to information, systems and infrastructures from different possible threats. Physical and logical access controls are imperative protection schemes in information security [30] [31]. Physical access controls guarantee only the authorized individuals to access buildings or rooms including, for example documents filling and IT infrastructures. On the other hand, logical access controls save the network facilities, computers, and information systems from unauthorized access. Recent advances in biometric technologies have proliferated the applications of biometric systems into the physical and logical access control domain.

One of the fundamental uses for biometrics technology is providing access control for restrictive facilit-

ies or equipments. Moreover, it can be applied to regulate access to computer networks and transport systems in order to allow only authorized individuals from accessing these protected resources. Various means used to recognize people identities such as irises, fingerprints, voice patterns, hand geometries [32].

Banking Systems/Financial Transactions Biometric is considered one of the most common uses in banking systems and financial transactions. With recent implementations of on-line and mobile payments, user names and passwords are no longer enough to gain mobile financial service customers more protection. Hence, it became necessary to provide high security and convenient protection systems by using biometrics technology. Various biometric modalities such as fingerprint, voice recognition, finger vein, palm vein or iris have been used in financial services authentication [39] [40].

Airport Security With the increasing demand for air transportation across countries, it is required that airport services issues and operational security have to be improved by using biometric-based identifiers in order to reduce passengers secure threats. Unauthorized access tops the list of airport security threats. Using Biometric solutions is expected to increase due to it is more reliable and cost effective. Different biometrics have been used to verify identities especially face, iris, retina, and fingerprint with differing levels of success [33] [34].

1.3 Problem Statement

A wide range of systems require authoritative personal recognition techniques to determine the identity of an individual. Practically, biometrics applications increased dramatically in many fields, such as security systems, access to computer networks, on-line banking services, border control [35] [36]. Single biometric security systems have some restrictions and limitations. For example, iris recognition suffers from some problems like camera distance, eyelashes occlusion and lenses. Face features are

unstable and change during time, and twins may have identical facial features. A fingerprint may have some burns, cuts, and small temporary or permanent injuries. Moreover, fake fingers that made up from silicon and/or gelatin have the ability to hack fingerprint based recognition systems. Cold may cause voice problems and recording tapes may be used to attack the system [37]. DNA needs several hours to be gained. Besides, DNA includes critical information related to genetic of persons and the test is expensive to apply [38]. Gait is not stable because it is sensitive to body weight. Signature is not universal and changes with time. Overview of biometrics showed that it is impossible to find the best single biometric suitable for all applications, populations and technologies. Therefore, the integration of biometric modalities can solve single modal system limitations to achieve the desired performance requirements. Multi-biometrics is an interesting research topic, it is used to recognizing individuals for security purposes to boost security levels. Multi-biometric systems distinct over single biometric systems by addressing the issue of non universality and noisy data [38]. Multi-biometric systems can facilitate the indexing of large-scale biometric databases. Moreover, it becomes hard for an impostor to spoof all the biometric characteristics of an authorized enrolled individual. In other words, a person can spoof one biometric but it is currently not possible to spoof many at same time. In addition, multi-biometric recognition systems have advantages in the continuous monitoring or tracking of a person when a single trait is not suitable in use or if part of biometric sources become unavailable of a failed. Consequently, this thesis aims to find a sufficient combination of single biometric suitable in specific applications for identification and/or verification purposes. For example, banking systems which use multi-biometric authentication for login procedures and the police and criminal evidence applications.

1.4 Contributions

This research has made a number of contributions, highlighted in the following series of bullet points.

- Improving the Viola-Jones face detection performance by using the brightness channel in HSV And HLS Colour Spaces, chapter 4.

- Distance similarity measures have been compared for PCA- and LDA-based face, ear and palm biometrics, chapter 5.
- The face and ear recognition performance using SVM based on PCA and SVM based on a combination of PCA and LDA techniques have been compared with the PCA and LDA techniques depend on distance similarity measures, chapter 5.
- Face, ear, palmprint, eye, and hand biometric recognition have been applied using three deep Convolutional Neural Networks (GoogleNET, VGG16, ResNET) and Shallow Convolutional Neural Networks, chapter 6.
- Multi-biometric recognition systems have been applied involving a different combination of single biometric traits: face, ear, palmprint, eye, and hand. Comparison among these combinations has been employed to assess the performance, chapter 7.

1.5 Publications

There have been seven published conference papers generated from parts of the thesis contributions. First 4 publications directly relevant to my research, last 3 are in a different application area but the techniques that used have contributed. The publications are listed in the following section:

1. Inas Al-Taie, Nassr Azeez, Arwa Basbrain, and Adrian Clark. "The Effect of Distance Similarity Measures on the Performance of Face, Ear and Palm Biometric Systems." In Digital Image Computing: Techniques and Applications (DICTA), 2017 International Conference on, pp. 1−7. IEEE, 2017.

   Distance or similarity measures are essential components used by distance-based recognition techniques. Since the Euclidean distance function is the most widely used distance metric in PCA and LDA recognition systems, no empirical study examines the recognition performance based on these two methods by using different distance functions, especially for biometric authentication

domain problems. The aim of this paper is to investigate whether the distance function can affect the PCA and LDA performance over different biometrics datasets. This paper helps the researcher to identify suitable distance measures for datasets. Our experiments are based on three different types of biometrics datasets containing a face, ear and palmprint data with four different distance functions including Euclidean, Manhattan, Mahanoblis and Cosine similarity distance are used during PCA and LDA classification individually. The presence of statistically significant performance differences is assessed using McNemar's Test.

2. Inas Al-Taie, Adrian Clark, and Nassr Azeez. "Improving the Viola-Jones Face Detection Performance by using the Brightness Channel in HSV and HLS Colour Spaces." In Irish Machine Vision and Image Processing (IMVIP), 2017 International Conference on, PP. 178 — 185. IPRCS, 2017.

   The algorithm due to Viola and Jones is the de facto standard way for locating faces in images. Although in widespread use, it does have shortcomings: it is effective on only frontal images, performance decreases with changes to the head size in an image, and it is sensitive to illumination. Generally, the algorithm is applied to grey-scale images in which the value of each pixel represents the intensity of light at this pixel. Illumination plays a significant role in the effectiveness of Viola-Jones. This paper investigates whether using the V and L channels, from HSV and HSL color spaces respectively, have an effect on the performance of Viola-Jones face detection. The presence of statistically significant performance differences is assessed using McNemar's Test. It is found that using the V-channel of HSV color space yields significantly fewer incorrect classifications than when the algorithm is applied to grey-scale images or to the L channel.

3. Inas Al-Taie, Adrian Clark, and Nassr Azeez. "Similarity Measures and the Performance of Biometric Systems." In Irish Machine Vision and Image Processing (IMVIP), 2017 International Conference on, PP. 115 — 122. IPRCS, 2017.

   Distance similarity measures are core components used by distance-based classification algorithms. This paper investigates whether the way in which similarity is measured can affect the perform-

ance of PCA- and LDA-based recognition systems using face, ear and palmprint biometric datasets. Four distance functions were considered: Euclidean, Manhattan, Mahanobolis distances and a Cosine similarity measure. The presence of statistically-significant performance differences was assessed using McNemar's test. It was found that all distance measures considered identified LDA as significantly outperforming PCA but that no individual similarity measure was more reliable than the others, leading to the conclusion that the content of the database used has an effect on the similarity measure.

4. Inas Al-Taie, Nassr Azeez, Wafa Yahya, Arwa Basbrain, Adrian Clark. "Biometric Recognition Systems Based on SVM pca and SVM pca, lda Techniques." 2019 Third World Conference on Smart Trends in Systems Security and Sustainablity (WorldS4). IEEE, 2019.

   In this work, the performance of face and ear recognition using SVM based on PCA and SVM based on a combination of PCA and LDA techniques have been compared with the PCA and LDA techniques depend on distance similarity measures. No previous study examines the recognition performance based on these techniques, especially for biometric authentication domain problems. The experimental work shows that the recognition based on SVM_{pca} and $SVM_{pca,lda}$ techniques can achieve a better performance than using PCA based on distance similarity measures. While, no significant differences were found using SVM_{pca}, $SVM_{pca,lda}$ and LDA based on distance similarity measures.

5. Basbrain, Arwa M., Inas Al-Taie, Nassr Azeez, John Q. Gan, and Adrian Clark. "Shallow convolutional neural network for eyeglasses detection in facial images." In Computer Science and Electronic Engineering (CEEC), 2017, pp. 157 − 161. IEEE, 2017.

   Automatic eyeglasses detection plays a main role in different facial analysis systems. To improve the robustness of these systems and cope with real-world applications, a high-speed eyeglasses detector that can achieve high accuracy is needed. This paper presents an effective and efficient method for eyeglasses detection in facial images based on extracting deep features from a well-

designed shallow convolutional neural network (CNN). The main contribution of this paper is to address the two essential aspects of CNN: the size of the training dataset required and the depth of the network architecture. As a result, a significantly more accurate shallow CNN architecture, Shallow-GlassNet, is obtained. The results have demonstrated the superior performance of the proposed framework which achieves a mean accuracy of 99.73%.

6. Nassr Azeez, Inas Al-Taie, Wafa Yahya, Arwa Basbrain, Adrian Clark. "Regional Agricultural Land Texture Classification Based on GLCMs, SVM and Decision Tree Induction Techniques." In Computer Science and Electronic Engineering (CEEC), 2018.

   Texture is a natural characteristic of all surfaces, which describes the visual patterns, and each has homogenization properties. In this paper, our concern is with the Regional Agricultural Land texture classification using grey level cooccurrence matrices (GLCMs). Texture discrimination is performed to partition a textured image into areas, each related to a homogeneous texture where samples of four different textures are extracted from the image. For each patch, the four features of the GLCM matrices namely, dissimilarity, correlation, angular second moment, and homogeneity are computed. Finally, for texture classification we have used two well-known methods: Support Vector Machine and decision tree induction. The results show that these texture features have high discrimination accuracy and classification using support vector machines gives better results as compared to the decision tree induction classifier.

7. Nassr Azeez, Inas Al-Taie, Wafa Yahya, Arwa Basbrain, Adrian Clark. "Regional Agricultural Land Classification Based on Random Forest (RF), Decision Tree, and SVMs Techniques." Fourth International Congress on Information and Communication Technology. Springer, Singapore, 2020.

   Land cover observation using remote sensing data requires robust classification techniques which give the accurate complex land cover mapping. Scientists and researchers have made great efforts in improving classification accuracy considerably. The objective of this paper is to present results obtained with the Random Forest (RF) classifier and to compare its performance with the Sup-

port Vector Machines (SVMs) technique. The mentioned techniques are applied over a complex mediterranean landscape image with five of land cover categories and low inter-class separability. Results showed that the random forest classifier performance outperforms the SVMs technique performance in terms of classification accuracy and training time with an overall accuracy of 92.7% while the SVMs accuracy is 85.0%

1.6 Thesis Structure

The thesis structure is summarized in terms of chapters: The first part chapter 2 presents the relevant fundamentals of the most common single biometric systems and literature review about some of the human biometrics traits that been used for a person identification/vitrification. The second part describes the basis of multi-biometric systems, reviews the most popular fusion strategies that been used in multi-biometric systems, and finally introduces some of the challenges in multi-biometric system design.

Chapter 3 sheds light on some of the ethical issues in biometric technologies. This chapter also reviews existing databases that been used for biometric research. Furthermore, performance evaluation techniques commonly been used in computer vision research are reviewed.

Chapter 4 presents the Viola-Jones face detection framework and the fundamental aspects related to this technique. This chapter also examines how the well-established Viola-Jones algorithm can be made more effective by applying it to the brightness channel of a transformed color image rather than the conventional grey-scale image.

Chapter 5 introduces conventional statistical techniques that been used for biometric classification; supervised learning and unsupervised learning. In this chapter four similarity measures: Manhattan, Euclidean, Cosine similarity, and Mahalanobis distance have been compared for PCA- and LDA-based face, ear, and palm biometrics. Furthermore, the face and ear recognition performance using SVM based

on PCA and SVM based on a combination of PCA and LDA techniques have been compared with the PCA and LDA techniques using distance similarity measures.

Chapter 6 presents the relevant fundamentals of convolutional neural networks (CNNs) that used for biometric recognition. Biometric recognition using three deep convolutional neural networks (Google-NET, VGG16, ResNET50) and shallow convolutional neural networks have been applied to five different databases: face, ear, palmprint, eye, and hand. Caffe is a very widely used framework for deep learning. In this chapter, we explain how to get started with CAFFE and the steps that should be set.

In chapter 7, the experimental work of multiple biometric recognition has been introduced. Multiple databases have been generated which consist of different single databases; face, ear, eye, hand, and palmprint. Different combination of single biometric systems that based on score fusion has been examined to achieve high recognition performance.

Finally, chapter 8 presents the conclusions drawn from this research and makes suggestions for further work.

Chapter 2

Single and Multiple Biometrics

2.1 Introduction

The human biometrics system is a common research subject in computer vision. Many studies have conducted to identify or verify individuals using their body's characteristics: physiological or behavioural. The first part of this chapter presents the relevant fundamentals of the most common single biometric systems and literature review about some of the human biometrics traits that used for a person identification/vitrification. The second part describes the basis of multi-biometric systems, reviews the most popular fusion strategies that used in multi-biometric systems, and finally introduces some of the challenges in multi-biometric system design.

2.2 Face Recognition

Face recognition is one of the most important research problem spanning various fields and disciplines. Over the last ten years, it has become one of the most popular and successful applications in computer vision [7] [6]. In additional to having many practical applications such as access control, bank card identification, information security, law enforcement, and surveillance, face recognition is an essential

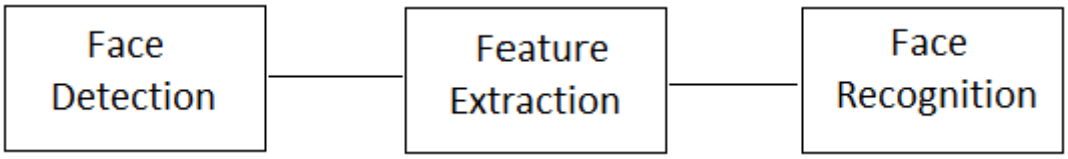


Figure 2.1: Face recognition process [44].

human behaviour that is important for effective interactions among people [41] [42]. The human face has a complex and dynamic structure with characteristics that can change over time. Humans have the ability to recognize faces, machine learning is now being developed to do this task. Scientists try to understand the architecture of the human face when building or improving face recognition systems. Human face recognition system utilizes data obtained from senses, such as visual, auditory, and tactile. These data are used individually or together to memorize and store faces. Handling sizeable data and memorizing many faces are a key issue for a machine recognition system [43] [7].

Facial recognition is a biometric application that is used to identify/verify a certain individual in a digital image by analysing and matching patterns. In general, a face recognition system involves three step process: face detection, feature extraction and face recognition Figure 2.1 presents a generic face recognition system.

In general, pattern recognition can be classified into four categories: model matching, statistical approach, syntactic approach, and neural networks [6]. The model matching category establishes several models for each label class and compares these models with the test pattern to obtain a suitable decision. The statistical approach extracts information from training data and employs different kinds of machine learning techniques for dimension reduction and recognition [6]. The syntactic approach, also known as rule-based pattern recognition, is based on human knowledge or some physical rules. The term knowledge here refers to the rules that the recognition system employs to perform specific actions. Finally, the neural networks is a framework build on a recognition unit called perceptron. With various numbers of perceptrons, layers of them, and optimization criteria, the neural networks can be applied to a wide range of recognition tasks.

2.2.1 Face Localization

The first and the most important step of face recognition is face detection (face localization). That means, when faces are located accurately the recognition step will be less complicated [7]. Face detection is the process of locating and extracting human faces in a digital image. In other words, face detection determines whether or not there are any faces in an image and if present, where they are. The expected outputs of face detection step are patches delimiting each face in the input image [45]. In order to make face recognition system more easy and robust, face alignment subsequently is applied to justify the scales and orientations of patches [46]. A variety of techniques have been proposed to determine a face location in an image, ranging from simple edge-based techniques to composite high-level techniques utilizing advanced pattern recognition approaches [45]. In 2002, Yang et al. [47] grouped the various methods into five categories: knowledge-based techniques, feature invariant methods, template matching approaches, appearance-based methods, and part-based methods. Knowledge-based techniques use predefined basics to locate a face based on human knowledge; feature invariant methods attempt to find the features of face structure which are resilient to pose and lighting variations; template matching approaches employ pre-stored face templates to judge weather an image is a face or not; appearance-based methods depend on learning face models from a set of representative training face images to perform detection [48]. Part-based methods based on the idea of using detected parts to represent human faces.

- Knowledge-based techniques: These rule-based techniques encode human knowledge of what forms a typical face. Usually, the rules save the relationships between facial features. These techniques are designed fundamentally for face localization, which attempts to locate the image place of a single face. In this subsection, two examples rule-based technique based on hierarchical knowledge-based method and vertical / horizontal projection are introduced [6].

Hierarchical knowledge-based method [49] This method is formed of the multi-resolution hierarchy of images and particular rules defined at every image level. The hierarchy is composed by image sub-sampling. The face detection process starts from the highest layer in the hierarchy which has the lowest resolution and then extracts possible face candidate depend on the general look of faces. After that the middle and bottom layers hold rule of more details for example, the facial features alignment and verify each face candidate.

Horizontal/vertical projection [50] Depending on the observations that human mouths and eyes have the lowest intensity among other regions of faces, this method performs the horizontal and vertical projection on the test image and locates minimums as facial feature candidates which together form a face candidate. Finally, each face candidate is validated by further detection rules such as eyebrow and nostrils. This method is hard to perform on complicated background images and can not be used on images with multiple faces.

- Feature invariant approaches: These algorithms attempt to build structural features that remain even when the pose, viewpoint, or lighting conditions change, and then use these features to locate faces. To distinguish from the knowledge-based methods, the feature invariant approaches attempt to extract features and find face candidates and then verify each candidate by spatial relations among these features, while the knowledge-based methods usually use information of the whole image to detect faces [51] [6].

Locating faces by colour information It is known that the amount of melanin compound in human skin determines its colour. To some extent, there is number of hues and saturation that describe skin-like pixels [52]. The general idea of this technique is to establish a decision rule to identify between skin and non-skin pixels in an image by applying a skin pixel classifier. In 2002, Hsu et al. [53] used color information for skin-color detection to extract human face regions. To deal with various illumination conditions, they extracted the 5% brightest pixels and

Figure 2.2: The locations of the missing features are estimated from two feature points. The ellipses show the areas which with high probability include the missing features [56].

employed their mean color for lighting compensation. However, there are many drawbacks with this method. Firstly, there are other components, such as wood, have somewhat similar hues to skin [54]. Secondly, the skin colour is affected by illumination.

In 2017, Lakshmipriya and Krishnaveni [55] used YCbCr color model to detect face skin regions. Viola Jones algorithm is employed to locate the face region, then YCbCr color model is used to extract the face skin regions alone from the background. This technique effectively detects the face skin regions of various sizes, poses and expressions under various environmental illumination conditions.

Face detection based on random labelled graph matching A probabilistic technique to locate a human face in a cluttered scene has been developed by Leung et al., using on local feature detectors and random graph matching [56]. Their motivation is to represent face detection as a search problem to find the arrangement of specific features that is most likely to be a face region. The first step, a set of local feature detectors, is performed to the image to capture candidate locations for facial features such as nose and eyes. Since the facial feature detectors are not completely reliable, the spatial arrangement of the facial features must be used also to locate face; see Figure 2.2.

- Template matching methods: In this category, several standard face patterns are stored to characterize the face as one unit or the features of the face separately. The correlations between an input face image and the stored face pattern are calculated for face localization and detection. The following paragraph summarizes a face detection method built on deformable template matching, where the faces template is deformable according to some rules and limitations.

  Adaptive appearance model In the traditional template matching methods, the deformation constraints are determined depending on user defined rules [57] . These constraints seek for some specific properties or prior knowledge, while not all the patterns have these properties. In addition, the traditional methods are fundamentally used for shape or boundary matching not used for texture matching. Cootes et al. [58] proposed the active shape model (ASM) that exploits information from training data to output the deformable constraints. They used the PCA technique [59] to learn the probable variation of object shapes. The active shape model can deal with variation of shape but not texture variation. Following their works, Edwards et al. proposed a technique that can deal with both boundary and texture matching by first matching an ASM to the shape features in the image, then using a separate eigenface model (texture model based on the PCA) to reconstruct the texture in a shape-normalized frame. This technique is not guaranteed to give an ideal fit of the appearance (boundary and texture) model to the image because some errors in boundary matching model can result in a shape-normalized texture map that can not be reconstructed correctly using PCA model. To direct the shape matching shape and texture simultaneously, Cootes et al. proposed the active appearance model (AAM) [60] [61].

- Appearance-based methods: In contrast to model matching methods, the models are learned from a set of training images which capture the representative variability of facial appearance. The appearance-based methods consider all regions of the face for face detection not only the facial feature points.

Example-based learning for view-based human face detection Given a window size, the appearance-based method scans the image and analyses each covered region. Sung et al. [62] selected the window size of 19×19 for training and represented each extracted patch by a 381-dimensional vector [63]. A face mask is employed to disregard pixels close to the boundaries of the window which may involve background pixels, and decrease the vector into 283 dimensions. The Gaussian mixture model is used in order to best capture the distribution of the face samples [47].

Locating faces using the Viola-Jones algorithm In 2001, Paul Viola and Michael Jones proposed the first object detection framework that characterized by high detection accuracy and low computation time, faster than any previous approach by about 15 times. In general, the idea of this technique is to collect a large set of resized of face and non-face images and train a classifier to discriminate them. There are three main factors with this framework.The first factor is using an adaptive representation for images called 'Integral Image' which allows very quickly feature computation. The second factor is applied an efficient classifier which depend on using the AdaBoost learning algorithm which reduce the numbers of features used by the detector. The third factor is combining classifiers in a 'cascade' in such a way that background regions of the image are discarded quickly while spending more calculation on face regions. The algorithm consists of the following stages:

- Haar Features Selection
- Creating Integral Image
- Adaboost Training algorithm
- Cascaded Classifiers

Due to the fact that Viola-Jones face detection framework may fail to detect faces under various environmental conditions. Sanghun and Chulhee [64], 2016, proposed a method that is based

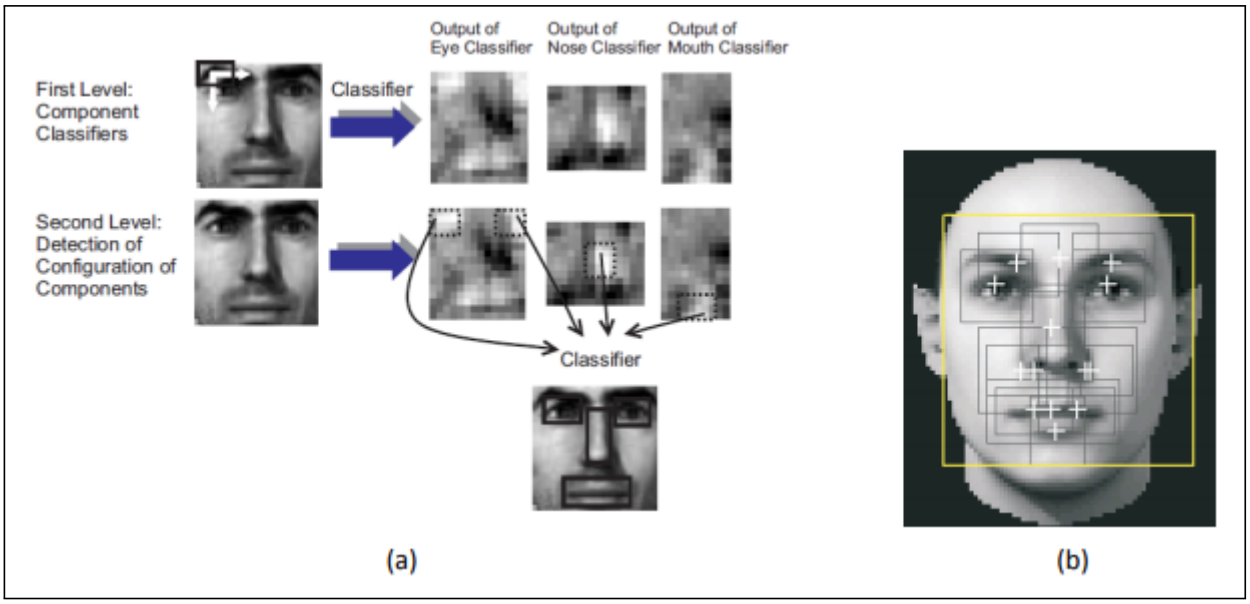


Figure 2.3: (a) Represents the first and the second level of the component-based classifier using four components. (b) Represents the fourteen learned components which are denoted by the black boxes with the corresponding center marked by crosses [78]

on using the difference of Gaussian (DoG) filter. They combined an original image and its DoG-filtered image to reduce illumination effects and thus enhance the face detection performance. They applied a skin color validation process after face candidates were gained using the YCbCr color space. The experimental results showed that the proposed methods reduced more than 50% of false positives.

- Part-based methods: With the development of the graphical model framework the part-based method lately attracts more attention [65]. This approach based on the idea of using detected parts to represent human faces.

Component-based face detection based on the SVM classifier Bernd et al. [66] suggested the face detection technique that involving of a two-level hierarchy of support vector machine (SVM) classifiers. On the first level, components classifiers detect a face components independently. On the second level, a unique classifier checks if there is a matching between the geometrical configuration of the detected components in the image and a geometrical model of a face. Figure 2.3 shows the process of their technique.

2.2.2 Feature Extraction

After the face detection step, human face patches are extracted from images. Using these patches directly for face recognition have some drawbacks. First, usually each patch contains more than 1000 pixels, which are too huge to build a robust recognition system. Second, face patches may be taken from various camera alignments, with various face expressions and illuminations. To overcome these drawbacks, features extraction is performed to achieve information packing, dimension reduction, salience extraction, and noise removing. After the features extraction step, a face patch is usually transformed into a vector with fixed dimension or a set of points and their corresponding locations [6]. In other words, feature extraction aim to find a certain representation of the image data that can highlight relevant information [67] [68]. Commonly, a face image is represented by a high dimensional vector including pixel values holistic representation) or a set of vectors in which each vector summarizes the content of a local region using a high level transformation (local representation) [69]. Facial features extraction can be classified into two major methods.

Appearance-Based Methods During the past twenty years, appearance-based methods draw the most attention against other methods. Appearance-based methods are also known as holistic-based methods, which attempt to identify faces using global representations of the entire image rather than local facial features [70] [71]. Global facial information is mainly represented by a number of features that are derived from the pixel information of face images, these features capture the variance among different individual faces and used to identify unique individuals. In other words, appearance-based methods employ whole information of a face patch and implement some transformations on this patch to obtain a compact representation (feature vectors) for recognition [72] [73]. Eigenfaces (performed by the PCA) and fisherfaces (performed by the LDA) are the most famous examples of appearance-Based Methods

Model-Based Methods Model-based face recognition methods attempt to construct a human face model that capture facial variations. previous knowledge of the human face is substantially utilized to

design the model. For example, model based matching captures the features of distance and relative position from the placement of internal facial elements [74] [70]. Three various extraction methods are distinct; feature template based methods,structural matching methods, and generic methods based on lines,edges, and curves [73].

2.2.3 Face Recognition

The next step after the features extracted and selected is to classify the image. A wide variety of classification methods are used for face recognition. Sometimes two or more classifiers are integrated to achieve better results. One way or another, classifiers have an important influence in face recognition. Usually, classification methods involve some learning; supervised, unsupervised or semi-supervised. The key difference between these methods is weather the label of each training sample is known or unknown. Supervised learning (label is known) attempt to learn the model of the relation between the feature vectors and their corresponding labels during the learning process. Conversely, unsupervised learning (label is unknown) attempt to learn the distribution of the possible categories of feature vectors in the training data set. Semi-supervised learning lies between unsupervised and supervised learning which means only part of the training data has labels. In 2008, Mazanec et al. [75] presented the results of face recognition using PCA, LDA, and SVMs methods. In addition, they proposed a combination of PCA and LDA methods with SVMs and they used different classifiers to match the probe image to the database images. They used these classifiers in the form of simple metrics (Cosine, Mahalinobis, and LdaSoft) and more complex SVMs. The experimental results indicated that using LDA with ldaSoft achieved the highest face recognition rate. Experiments showed that using PCA with SVM and LDA with SVM introduced a better recognition rate than PCA and LDA methods. In 2013, Sodhi and Lal [76] discussed different steps involved in face recognition using PCA and LDA methods. They mentioned that using some combination of the standard distance similarity measures such as City Block, Euclidean, and Mahalanobis might outperform the individual distance similarity measures. In 2014, Rozario et al. [77] performed a quantitative accuracy analysis of four face recognition methods: PCA, ICA (Independent

Component Analysis), LDA, and SVMs under different environmental conditions such as illumination, alignment and pose variations. The experimental results indicated that using SVMs method gave better accuracy in comparison with the other methods.

2.3 Ear Detection and Recognition

Due to the fact that the ear shape is unique and permanent (the appearance of ear does not change extremely during a human life). Additionally, it is not affected by the change of expression such as the face. And more, ear images can be taken without a person's knowledge and therefore it is suitable for security system. Hence, the researcher's interest in ear recognition system has grown in recent year [79]. In general, the idea of ear recognition is to represent an input image to a features set and then compares this against the other feature sets of different images to define its identity. Ear recognition involves four stages. The first stage is ear detection (determine the ear's position in an image). The second stage is feature extraction. In this stage, the image is represented as a vector that represent the discriminatory information in the ear image. Then, the system compares the extracted features from the input ear image against those stored in the database to determine the ear's identity. In matching stage, scores are generated to indicate the similarity to other ear images. Lastly, the system uses these score to introduce a final decision [80] [81].

2.3.1 Ear Detection

Ear detection from side face images is an essential step in different application including biometric recognition technology. But accurate ear detection for a real-time application is a challenging function due to the fact that the appearance of ear images can vary under various illumination and viewing [82]. There are few reported techniques for correctly ear detection from side face images. One of the earliest techniques is to detect the ear contour using Canny edge maps, it was developed by John F. Canny in 1986 [83] [84]. In order to detect and crop the ear, Yan and Bowyer in 2005 illustrated approach

including a two-line landmark. The first line along the border between the face and the ear, the second one from the top to the bottom of the ear [85]. L. Yuan and Z.C. Mu [86] proposed a technique that is based on skin-color and contour information. This technique detects ear by estimating the ear location and by improving the ear position using contour information. In this technique presumes the shape of the ear is elliptical and fits an ellipse to the edges to obtain the accurate location of the ear. In 2007, Yan and Bowyer proposed another technique to detect the ear pit using Gaussian smoothing and curvature estimation and then applying an active algorithm to extract contour of the ear [87]. This technique fails when the ear pit is invisible. Most of the techniques that have been mentioned above detect the ear from a small part of the side face around the ear. But sometimes it is required, in particular for non intrusive application to localize the ear from a whole portion of the side face image. Also, the above approaches fail when the quality of input images is poor. Almost of the mentioned approaches are not fast enough to be implemented in real time systems. To overcome such problems, S. Islam et al. in 2008 proposed a fully automatic ear detection approach based on Haar features and cascaded AdaBoost provided by the OpenCV library for object detection. The experimental work which applied on a test set of 203 face images showed that all the ears were detected accurately with a low false positive rate as they mentioned in their paper [88]. In 2009, S.Prakash et al. [89] suggested an automatic ear detection technique based on template matching, in which an ear template gains a appropriate size by resizing the created template. The resizing process is based on the size of the skin region of profile face image and it works well when profile face contains only facial parts. In 2010, a technique proposed by A.H. Cummins et al. [90] which is known as image ray transform. It is depend on an analogy to light rays.The accuracy is reported to be 99.6% for enrolment to 252 images of the XM2VTS database.

2.3.2 Ear Recognition

In 1997, Mark Burge and Wilhelm Burger described the first attempt to recognize a human ear automatically. A mathematical graph model was used to represent and match the features(curves and edges) within a 2D ear image [91]. After two years, Belen Moreno et al. demonstrated an automated ear

recognition system depended on different features (ear shape and wrinkles) [92]. Since then, various feature extraction and matching schemes have been employed for ear recognition, based on computer vision and image processing algorithms [93]. In 2007, M. Ali et al. described a new ear recognition method based on using wavelet transform to extract features from the image. Results achieved are up to 94.3% [94]. Victor et al. used PCA technique to both ear and face recognition. The conclusion indicated that the performance of face recognition was better than ear [95]. However, K. Chang et al. carried out later an experiment and conclude that no significant difference was noticed between face and ear biometrics, when they used PCA technique. They proposed that the reason for the conflict, there is low control over lighting, earrings and hair within the data set. Furthermore, they reported a recognition rate of 90.9% [96] [97].

2.4 Hand Shape Recognition

In recent years, hand recognition has become an active topic for researchers in view of its potential use in human-computer interaction, camera surveillance and identification of individuals [98]. Hand shape biometrics determine the identity of a person based on determining a number of measurements derived from it and matching of various hand geometric features such as: finger lengths, width, hand size, and hand contour [99] [100]. Since the early 1970s, hand geometry-based recognition systems have been available. In 1988, Sidlauskas described a 3D hand profile systems which have been used for hand geometry recognition [101]. In 2000, Sanches and Raul described a technique based on using Gaussian mixture modelling (GMM), a statistical pattern recognition technique, which is usually applied to speaker identification [102].

2.4.1 Hand Detection

Accurate hand detection in images remains a challenging issue due to the changes of hand appearance. Hand detection is the first and essential step in most hand recognition systems. Different methods have

been proposed to detect a hand [103] [98]. There are two major categories [104]. Methods in the first category called appearance based methods which depend on the hand appearance itself. In other words, these methods exploit hands features only without using any features from other surrounding body parts. In 2000, Zhu et al. [105] generated two models; a hand color model and a background color model for an image. They classified each pixel in the image into either hand pixel or background pixel according to these models. Ong et al. [106] presented a novel unsupervised method to train a detector which is characterized by efficiency and robust. It detects the presence of human hands and also classifies the shape of the hand. In 2010, a hand posture classification method proposed by Tran et al. [107] including of two steps. The first step tries to detect skin regions using a very fast color segmentation algorithm based on thresholding technique. The second step aims at classifying each skin regions into one of hand posture classes based on Cascaded Adaboost technique. In the second category, methods utilize the information supplied by surrounding parts. The methods deal with the hand shape as a deformable shape, and then use shape matching algorithms to detect hands. In 2006, Athitsos et al. [108] proposed a method including shape models that used to represent shapes of variable structure. The detection algorithm aims at searching for items of such shapes in images by finding optimal concurrences between image features and shape models. Coughlan et al. [109] presented a novel template that detects the edges of an open hand without the user initialization. Zhang et al. [104] suggested a method that evaluates the four hand features; color, gradient norm, temporal motion and motion residue and then they explore the probability of these features to build a robust hand detector. In 2012, Nguyen et al. presented a hand detection approach relied on Viola-Jones detector [110]. This approach detects internal portion of the hand regardless of background based on local features of this internal portion. The set of these features has been called as Internal Features. In case of Haar-like features, it is know as Internal Haar-like features. In addition, they proposed a hand detection framework that combines several individual hand posture detectors. The Experimental results indicate that "the proposed method outperforms the conventional method based on Viola-Jones detector with the same computation time" [98].

2.5 Palm Print Recognition

During the last fifteen years, many various problems related to palmprint recognition and systems securities have been addressed. Palmprint is one of the reliable modality since it possess many features such as principal lines, wrinkles, singular points, and ridges. Also palmprint modality is unique for each person and universal. Palmprint recognition is employed in law enforcement, civil applications, and different applications where access control is fundamental. Palmprint recognition techniques have been assorted into two main groups [111]. The first approach is based on low resolution features where only principal lines, wrinkles, and texture are extracted and the second approach is based on high resolution features where in addition to principal lines and wrinkles, other discriminant features such as ridges, minutiae, and singular points can be extracted [112]. In 2004, C.C. Han [113] proposed a specific verification technology by using two hand-based features, the hand geometry and the palmprint, these two features are simultaneously grabbed by the CCD camera-based devices. Geometrical features of the hands are employed to verify the identity. The samples with confused hand shapes should be to re-check by the palmprint features. Jiaa et al., 2008, have proposed palmprint verification depend on robust line orientation code. They used modified finite Radon transform for feature extraction that extracts orientation feature. The line matching technique has been used for matching of a test image with training image which is build on pixel-to-area algorithm. In 2009, Zhang et al. proposed a dynamic selection scheme by using global texture feature measurement and the detection of local interesting points. The experimental results shows that palmprint patterns can be well described by textures [112]. A. Kong and D. Zhang, 2011, have presented a novel feature extraction scheme (the Competitive Coding Scheme) for palmprint identification. In this scheme, the orientation information from the palm lines has been extracted and stored in the Competitive Code. Dai and Zhou,2012, introduces high resolution method for palmprint recognition using multiple features extraction [114]. In 2012, high resolution palmprint recognition system has been proposed by Cappelli et al. [115] which is based on minutiae extraction. Pre-processing is performed by segmentation of an image from its background. Local frequencies and

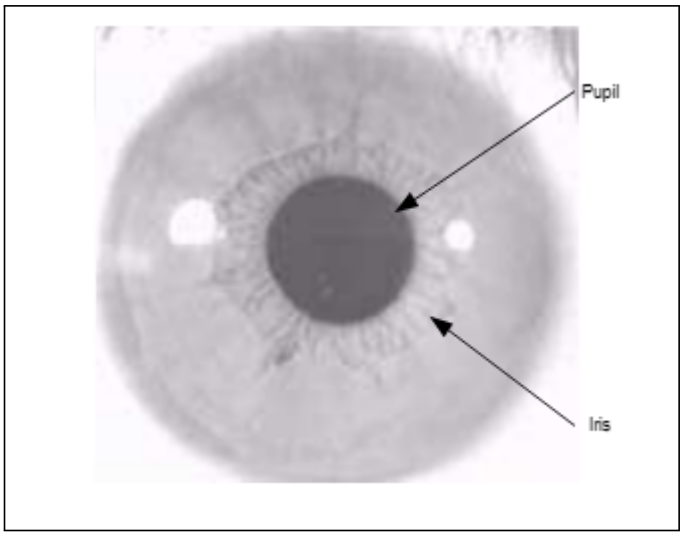


Figure 2.4: An image of human iris [117]

local orientations are estimated to enhance the quality of image. To extract the minutiae, Gabor filters with features contextual filtering approach is applied.

2.6 Iris Print Recognition

For the last few decades, iris patterns have attracted lots of attention in biometric technology because they have stable and unique features for personal identification [116]. The structure of iris is consisted of many layers, the outer surface is made of epithelial cell which are thickly coloured. See Figure 2.4 Efforts to devise authoritative mechanical means for biometric personal discrimination have a long history. In 1993, Daugman developed the feature extraction method based on information from a set of 2D Gabor filters [117]. In this work, a 256 byte code generated by quantizing the angle of the local phase according to the outputs of both the real and imaginary parts of the filtered image, then comparing the percentage of non identical bits between a pair of iris representations via XOR operator. Boles and Boashash, 1998, [118] presented a new method for recognizing the iris of the human eye by calculating zero-crossings of the wavelet transform at various resolution levels over concentric circles on the iris. The resulting $1D$ signals are compared with model features by applying different dissimilarity functions. In 2001, Lim et al. [117] proposed an iris recognition system which contains a compact representation

scheme for iris patterns by using the 2D wavelet transform. In 2014, H. Rai and A. Radav [119] presented a novel method for iris patterns recognition by using a combination of two techniques; support vector machine and Hamming distance. They selected the zigzag collarette region of the iris for iris feature extraction because it captures the important regions of iris complex pattern. The proposed technique effective and reliable with a recognition rate of 99.91% on CASIA . S. Umer et al.,2017, proposed a novel cancelable iris recognition system which based on the performance of using different feature learning techniques. They used Sparse Representation Coding followed by Spatial Pyramid Mapping technique for feature computation from iris pattern. The results show the effectiveness and robustness of the proposed technique [120].

2.7 Fingerprint Recognition

The fingerprint is a physiological biometric trait to identify individuals and verify their identity. Fingerprint recognition is one of the most popular technique employed in biometrics because of their uniqueness and consistency during a lifetime [121]. Fingerprint identification uses the impressions made by the minute ridge patterns found on the fingertips whereas no two persons have totally the same arrangement of ridge formations. According to FBI, fingerprints vary even for ten fingers of a same individual. In fingerprint recognition, features of interest can be divided into local and global features. local features or Minutia points are the unique traits of fingerprint ridges that are employed for positive identification while global Features are the traits that any person can see with the naked eye, such as pattern area, core point, delta, type lines, and basic ridge patterns [122]. There are different algorithms and techniques which achieved accurate results for the fingerprint recognition system. Jiang et al., 2015, presented a new design of the minutiae based fingerprint matching technique that particularly created for GPU based massively parallel architectures. The parallel design achieves speed-up ratios of up to 15 with single GPU compared to multi-threaded CPU applications [123]. In 2016, C. Yuan et al. proposed a new software-based liveness detection technique by using multi-scale local phase

quantity (LPQ) and principal component analysis (PCA). They used multi-scale LPQ to construct feature vectors of a fingerprint. PCA technique is introduced to reduce the feature vectors dimensionality and gain more effective features. Lastly, a training model is obtained using support vector machine classifier (SVM), and the liveness of a fingerprint is detected depending on the training model. Experimental results show that the proposed method can detect the liveness of individuals' fingerprints and achieve high recognition reliability [124]. D. Peralta et al., 2017, proposed a generic decomposition methodology for minutiae-based matching methods that separates the calculation of the matching scores into lower level steps that can be performed in parallel in a elastic manner [125].

2.8 Body Shape

Athough, body-shape-based biometrics is not popular as a stand alone system, it is considered helpful for development of multi-biometric systems [126]. In 2003, Godil et al. investigated the advantage of using static human body distances as a biometric for human identification. In their work the 3D landmark data from the CAESAR database is used. In this database each individual is represented with a simple biometric comprising of distances between fixed connection body locations [127]. In 2002, Collins et al. captured human body shape cues such as body height and width, as well as gait cues such as the amount of arm swing and stride length. They tested their technique on the CMU motion of body (MoBo) database [128]. This technique used "a template- matching based feature extraction and nearest-neighbour classifier" [126].

2.9 Multi Biometric Systems

Although biometric systems have priority over the traditional security techniques such as key, smart card or password [129], they also are subject to many limitations such as noise in sensed data and spoof attacks. One of the solutions to these problems is by implementing multi biometric systems, that

means many sources of biometric information are used to identify a person [130].

2.9.1 Multi Biometric System Categories

Depending on the nature of the biometric information sources, a multi biometric system can be classified into different categories; multi-sensor, multi-algorithm, multi-sample, multi-instance and multi-modal systems. The multi biometric systems categories are depicted as in Figure 2.5

Multi-sensor systems Multi-sensor systems utilize various sensors to capture single biometric trait of a subject. In 2004, Lee et al. reported the example of this system where multiple 2D cameras are used to capture the image of an individual [132]. Subsequently, Kong et al. (2005) [131] applied an infrared sensor and visible-light sensor to capture the image of a person's face while Rowe and Nixon [134] and Pan et al. [133], a multi spectral camera has been used to capture images of iris, face or finger. The application of multi-sensors in the scientific researches is able to improve the recognition ability of the biometric systems [135].

Multi-algorithm systems These systems integrate the output of multiple methods for example, feature extraction or/and classification algorithms for the same biometrics information [136]. In other words, the acquired data by more than one algorithm helps to enhance the performance. In addition, it is cost effective because utilization of new sensor is not required. However, this system has a drawback due to various feature extraction and matching modules which can result in complexity of system computation. In 2003, Lu et al. [137] have combined three different feature extraction techniques which are Principle Discriminate Analysis (PCA), Linear Discriminate Analysis (LDA) and Independent Component Analysis (ICA) to improve a face recognition system. Another researcher has also combined multiple algorithms; Iterative Closet Point (ICP), PCA and LDA to implement 3D face recognition [138]. In Imran et al. [139], three subspace algorithms such as PCA, FLD and ICA are applied for palm print and face separately to determine the best algorithm performance. The result displays that the ICA technique performs well for

both individual modalities.

Multi-sample systems Multi-sample systems employ multiple samples derived from the same biometrics captured by a single sensor. The same algorithm operates each of the samples and the singular results are fused to gain an overall recognition results. This system has been studied in Chang et al. [140] for face recognition in which 2D face image has been used as a baseline to compare the performance of multi-sample 2D and 3D face in speech recognition. In 2007, Samad et al. have proposed multi-sample technique to UMACE filter classifier by integrating scores from various samples from lipreading features and spectrographic features [141].

Multi-instance systems In this system, the biometric data extracted from the multiple instances of the same individual trait. For example, the left and right finger and iris of a same individual is proposed in Jang et al. [142] and Prabhakar and Jain [143], respectively.

Multi-modal systems Multi-modal systems combine the evidence of multiple biometric traits which can come from a variety of modalities [144] to extract the biometric data of an individual. The multi-modal system is considered as a reliable system due to the presence of multiple independent biometrics. However, the drawback of this system is the substantial cost due to the requirement of many sensors. In 1993, Brunelli and Falagivna [145] reported the example of this system where a person identification system using face and speech is presented. This research showed that the system performance has been improved [146] by combining three biometrics frontal face, face profile and voice using sum rule combination scheme. Another research which used a combination of face, fingerprint and finger vein has been presented in Hong et al. [147] while Ramli et al. [148], and Lip and Ramli [149] combined the speech signal and lipreading image as a second modality to assist the performance of the single modal system in the multibiometric systems.

2.9.2 Biometrics and Fusion Strategies

A biometric system depends on gaining a personal identification by using a single biometric identifier so it is incapable to achieve the desired acquirement. The achieved results indicate that using multiple biometrics to identify a person is more forcible than using only one. Recent attention has focused on multi modal biometrics in particular, in late 90s. In 1998, Hong and Jain integrated a PCA based face and a minutiae based fingerprint identification system [150]. In 2000, a commercial multimodal approach was developed by Frischholz and Dieckmann [151]. This approach combined lip motion and face images with the voice from an audio signal to achieve a person verification. In 2003, Fierrez-Aguilar and Ortega-Garcia proposed a multi-modal approach using face, minutiae-based fingerprint and an online signature verification system [152]. Ross and Jain (2003) combined face, fingerprint and hand geometry at the matching score level [153]. In the same year, Kumar et al. presented multimodal verification system using hand images. This system combined hand geometry and palm image at the feature and match score [154]. In 2004, a multimodal biometric system was developed by Toh et al. [155] using vioce, hand geometry and fingerprint at match-score-level fusion. Camlikaya et al. (2008), proposed a multimodal biometric authentication system that combine fingerprint and voice mod [156]. Shahin et al. (2008) [157] presented a multi modal system using hand veins, hand geometry and fingerprint to achieve high security. In 2009, Chandran et al. [158] integrated fingerprint and iris to improve the performance. Poinsot et al. (2009) [159] proposed palm and face multi modal biometrics using Gabor filter for features extraction. Soltane et al. (2010), demonstrated a human verification method that integrated face and speech data in order to overcome the drawbacks of single biometric verification [160]. A feature fusion system of palm print and face characteristics using a simple fusion method presented by GU Bokade et al. (2013) [161]. Since feature set includes relevant information to the captured biometric evidence, fusion at feature level expected to provide more precise results as compared to other fusion techniques. In 2013, MD Dhameliya et al. [162] fused the two biometric systems palmprint and fingerprint at feature level to develop a multimodal biometric system.

Features are extracted using Gabor filtering. The recognition rate gained was 87%. M Gogoi and DK Bhattacharyya, 2014, presented an effective technique for decision level fusion of fingerprint and iris biometrics [163].

In 2014, Benaliouche and Touahria [164] investigated the performance of iris and fingerprint multimodal recognition system based on three different techniques: weighted sum rule, classical sum rule, and fuzzy logic technique. The scores from these two biometric traits are fused at the matching score and the decision levels. The experimental results showed that the recognition performance based on the fuzzy logic technique for the matching scores combinations at the decision level outperformed the other techniques followed by the classical weighted sum rule.

In 2018,Javad et al. present a multibiometric system based on two types of biometric modalities, palmprint and fingerprint based on a score fusion [165]. In 2017, Reddy and Bindu present a multibiometric recognition system using three models of biometrics face, iris and fingerprint. The fusion is implemented at the matching-score level. The experimental results showed that the designed system achieves a high recognition rate [166]. In 2018, khodadoust et al. present a multibiometric system based on two types of biometric modalities, palmprint and fingerprint based on a score fusion [165].

Figure 2.5 shows different types of multibiometric systems. One of the essential issues that must be taken into consideration in the design of a multibiometrics system is to define the type of information that need to be fused. The term fusion in biometrics system refers to the integration of information from multiple biometric sources [168]. Researchs shown that using fusion of multiple biometric evidences, improve the authentication performance [136]. Fusion scheme in biometric systems can be performed at several levels; sensor level, feature level, score level and decision level fusion [167] [84]. The most popular fusion scheme performs matching score level where the scores obtained by the person matchers are integrated [169].

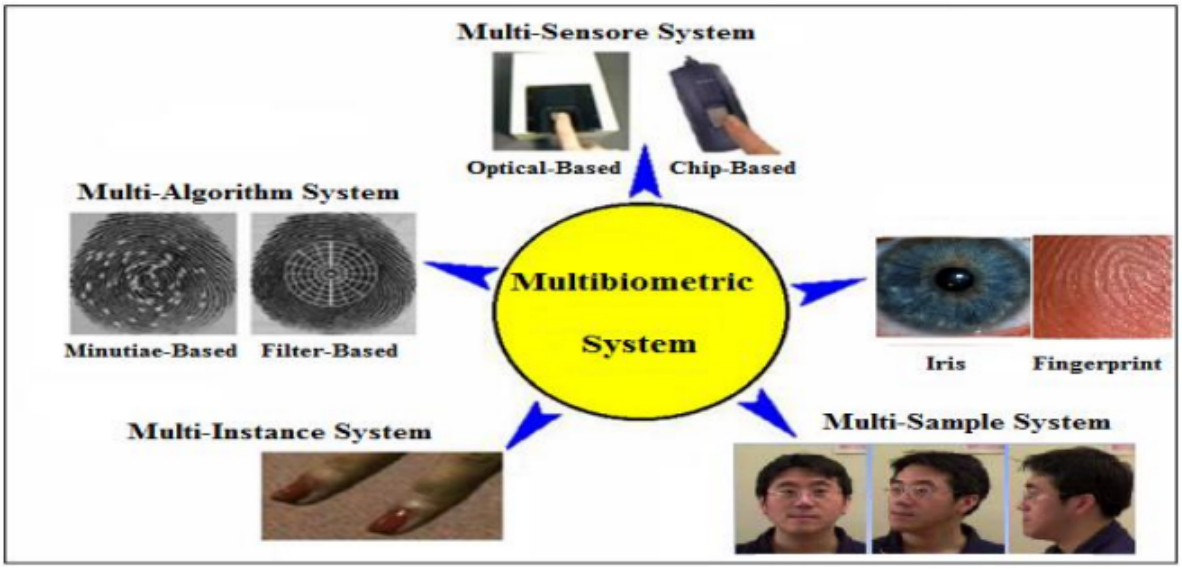


Figure 2.5: Different types of multi-biometric [173]

Sensor Level Fusion Sensor fusion is the integration of data derived from different sources; radar, sonar and other acoustic, TV cameras, thermal imaging camera. So, the resulting information would be more accurate than when these sources used individually [84] [170].

Feature Level Fusion In feature level fusion, different feature sets that have been extracted from multiple biometric sources are combined into a single feature set. The main benefit of feature level fusion is to detect the correlation values between these features and identify a distinct set of features that can improve the accuracy of recognition [170] [84].

Score Level Fusion Match score refers to the measurement of the similarity between the input and template feature vectors. In score level fusion, different match scores output by individual biometric matchers are combined into a new match score in order to reach at a final recognition decision.

Decision Level Fusion In a multi-biometric system, fusion is implement at the decision level, abstract level, when only the decisions output by different biometric matchers are obtainable [171] [172]. Fusion can be performed at any level in a biometric system. Figure 2.6 shows various multi-biometric fusion levels.

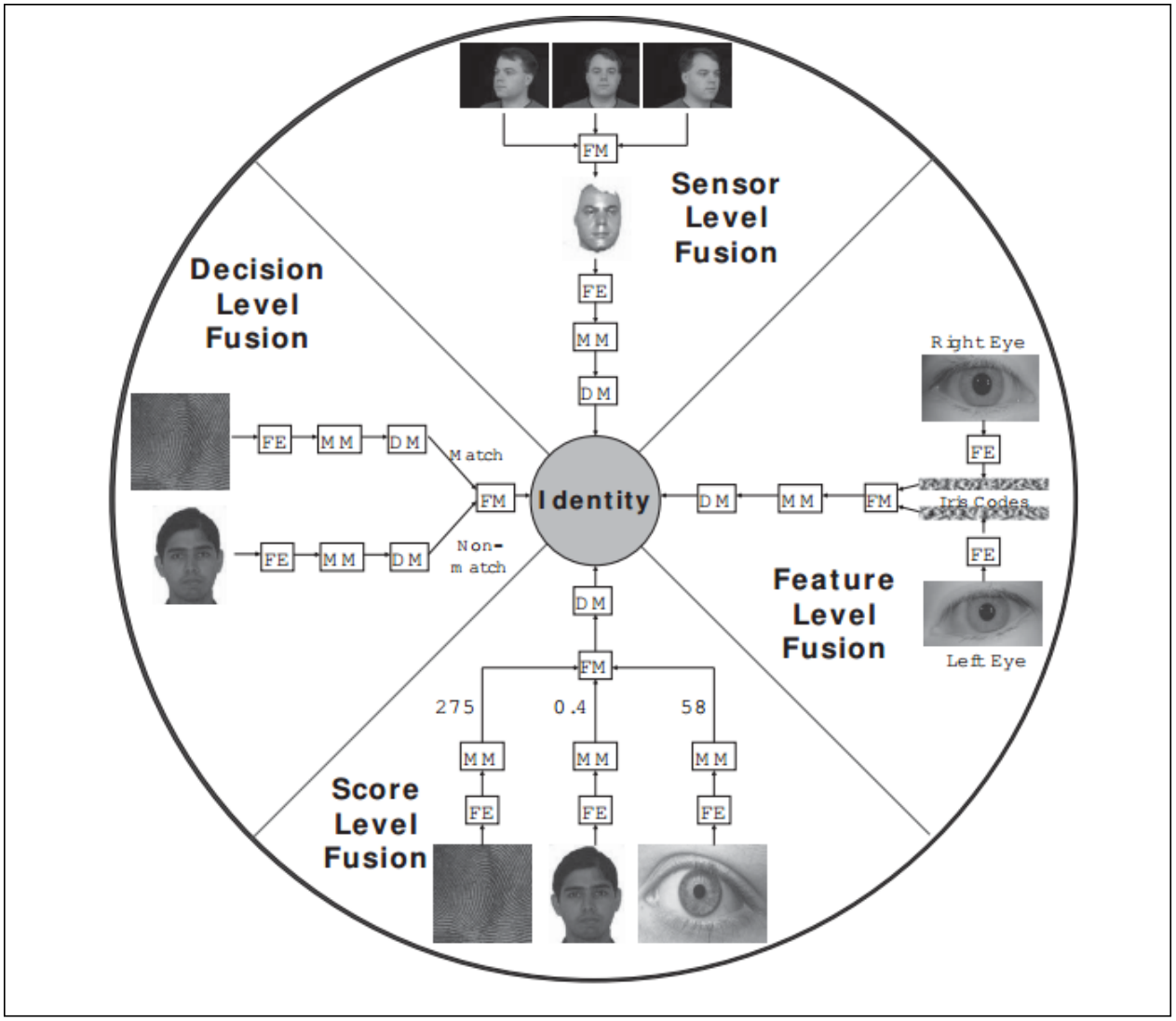


Figure 2.6: Multi-biometric fusion levels. FE: feature extraction module; MM: matching module; DM: decision-making module; FM: fusion module [168].

2.9.3 Challenges in Multibiometric System Design

Multi-biometric system designing is considered as a challenging problem because it depends on different factors such as information sources, acquisition and processing construction, information fusion strategy. Consequently, it is hard to choose the ideal sources of information and the ideal fusion strategy for a specific application. The following are some of these factors [174] [168].

Heterogeneity of Information Sources Integration of information is considered to be an effective stage in multi-biometric systems because the information amount that available to the fusion model decreases from the sensor level to the decision level. However, the incompatibility of the acquired information lead to the inability to fuse information at the sensor or feature level.

Fusion Complexity In some cases, even if the acquired information sources are compatible (for example: two impressions of the same finger, impression of two different fingers for the same person), the complication of the fusion algorithm may reduce the advantages of the fusion process. For instance, the sensor level or feature level fusion consists of extra processing which may be complex such as designing a new algorithm to achieve the fused data matching process. Furthermore, the raw data resulting from the sensor and the extracted features are usually affected by different types of noise for example, background chaos in a face image and faked minutiae in a fingerprint minutiae set. Hence, the use of fusion process at the sensor and feature level may not lead to improve performance.

Varied Discriminative Ability Each biometric source can provide different amount of discriminatory information. Consider a multi-biometric system which has two matchers A and B, the former has high accuracy compared to the later. If a simple fusion rule is employed that allocates equal weights to the information derived from both matchers, the accuracy of the multi-biometric system at the most will be less than the accuracy of the matcher A. Furthermore, there are some multi-biometric systems use weak biometric traits like height, gender, ethnicity which have lower discriminatory information compared

to other biometric identifiers such as face, fingerprint and iris. Hence, it is fundamental to measure the amount of discriminatory information in each source individually and allocate appropriate weights to the sources depending on their information content.

Correlation Between Sources In general, the different biometric sources in many multi-biometric systems may not be independent. However, there are a number of systems depend on correlated information sources. Examples of such systems, include:

- systems using traits of physically related. For example, speech and lip movement of a person.
- multiple matchers performing on the same biometric raw data or extracted feature representation. For example, multiple face matchers which applied on the same face image.
- systems using multiple samples of the same trait. For example, the impressions of two right finger of the same person.

In general, researchers indicate that biometric systems based on independent evidences can be expected to improve system accuracy further compared to biometric systems using correlated sources fusion. However, the impact of the biometric sources correlation on the fusion performance is partially were known. Therefore, information fusion in biometric systems is still an active research area.

2.10 Conclusions

The first part of this chapter presents the relevant fundamentals of the most common single biometric systems and literature review about some of the human biometrics traits that been used for person identification. The second part describes the basis of multi-biometric systems, reviews the most popular fusion strategies that been used in multi-biometric systems, and finally introduces some of the challenges in multi-biometric system design. Next chapter will shed light on some of the ethical issues in biometric technologies and review existing databases used for biometric research. Furthermore, performance

evaluation techniques commonly used in vision research will be reviewed.

Chapter 3

Biometric Databases and Assessing Vision System Performance

3.1 Introduction

Biometric data is commonly used in systems that aim to identify a specific user through unique traits. Computer image processing is considered one form of biometric analysis that utilizes biometric data. In most biometric analysis systems, there is request for a big amount of biometric data. This data must be stored and secured from unauthorized access especially when biometric data is stored in centralised databases. These systems build on complex algorithms that sort data in somehow that will obtain an identifying result in a given application. Developers use unique features from one person to another in order to achieve effective biometric identification. This chapter sheds light on some of the ethical issues in biometric technologies. It also reviews existing databases that used for biometric research. Furthermore, performance evaluation techniques commonly used in vision research are reviewed.

3.2 Ethical Issues in Biometric Technologies

Because biometric technologies are build on measurements of physiological or behavioural traits of the individual body add to collection and storage of individual data, they raise a host of ethical concerns which are related to the protection of individual values such as personal liberty, privacy, equity, and dignity. The civil liberty communities argue that the technology weakens the human rights for personal privacy and anonymity [175]. It is parasitical and has the ability to produce serious impact on individual freedom and democratic rights. But due to many considerations and threats around the world, such as threat of terrorism, security, crime prevention and identity theft and fraud, it has become essential to have the capability to preserve human identity for later identification and verification [176].

3.3 Databases for Biometric Research

3.3.1 The Database of Faces AT&T

The AT& T Database of Faces [177]contains a set of face images taken at the lab between 1992 and 1994. There are 40 distinct subjects of ten different images each. The images for some subjects were taken at different times, varying the lighting, facial expressions, and facial details. All the images were taken against a dark background and all the subjects in an frontal, upright position. Examples of images from the AT& T face Database are reproduced in Figure 3.1.

3.3.2 Caltech Face Database

Caltech Frontal face dataset [178] is collected by Markus Weber at California Institute of Technology. The database consists of 450 Jpeg format face images of 896×592 pixels. It has 27 unique subjects with different backgrounds,expressions, and lightening. Examples of images from the Caltech face Database are reproduced in Figure 3.2.

Figure 3.1: Samples of AT& T face database [177].

Figure 3.2: Samples of Caltech face database [178].

Figure 3.3: Example face images in CASIA-FaceV5 [179].

3.3.3 CASIA-FaceV5

CASIA Face Image Database Version 5.0 (or CASIA-FaceV5) [179] contains 2,500 color facial images of 500 subjects. The face images of CASIA-FaceV5 have been taken using Logitech USB camera in one session. All face images are 16 bit color BMP files and the image resolution is 640 × 480. Typical intra-class variations include illumination, pose, expression, eye-glasses, imaging distance.

3.3.4 AMI Ear Database

AMI Ear Database [180] was created by Esther Gonzalez during her PhD in Computer Science. The images have been taken in an indoor environment. The database consists of 100 different subjects in the age range of 19-65 years. For each individual, seven images (six right ear images and one left ear image) were taken. All the images were taken using a Nikon D100 camera, under the same lighting conditions, with the subject placed seated at a distance of about 2 meters from the camera and looking at some previously fixed marks. It used a 135 mm focal length for six of the seven images and 200 mm focal length for the image we called ZOOM. Five of the captured images were right side profile (right ear) with the individual facing forward (FRONT), looking up and down (UP, DOWN) and looking

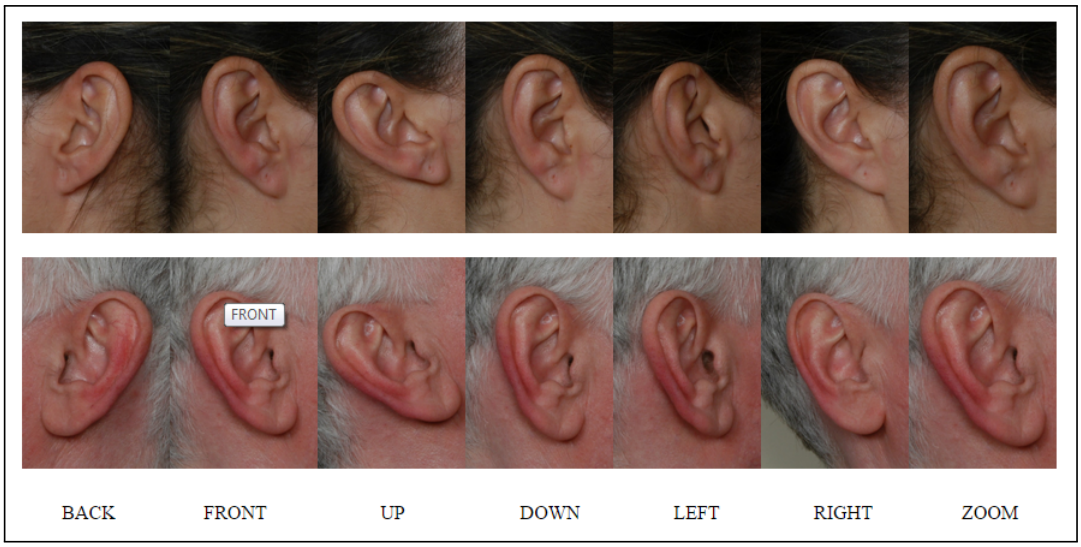


Figure 3.4: Samples of AMI ear database [180].

left and right (LEFT, RIGHT). The sixth image of right profile was taken with the subject also facing forward but with a different camera focal lenght (ZOOM). Last image (BACK) was a left side profile (left ear), with the subject facing forward and with the same camera focal length than the previous five images. The database of 700 images has been sequentially numbered for every subject with an integer identification number. The resolution of these images is 492 × 702 pixels and all these images are available in jpeg format. Few examples of images from the AMI Ear Database are reproduced in Figure 3.4.

3.3.5 IIT Delhi Ear Database

The IIT Delhi ear image database [181] is collected from the 121 different subjects and each subject has at least three ear images. All the images are acquired from a distance using simple imaging setup. The images have been taken in an indoor environment. The currently available database is acquired from the 121 different subjects and each subject has at least three ear images. All the subjects in the database are in the age group 14-58 years. The database of 471 images has been sequentially numbered for every user with an integer identification/number. The resolution of these images is 272 × 204 pixels and all these images are available in jpeg format. In addition to the original images, this database also provide the automatically normalized and cropped ear images of size 50 × 180 pixels. Recently, a larger

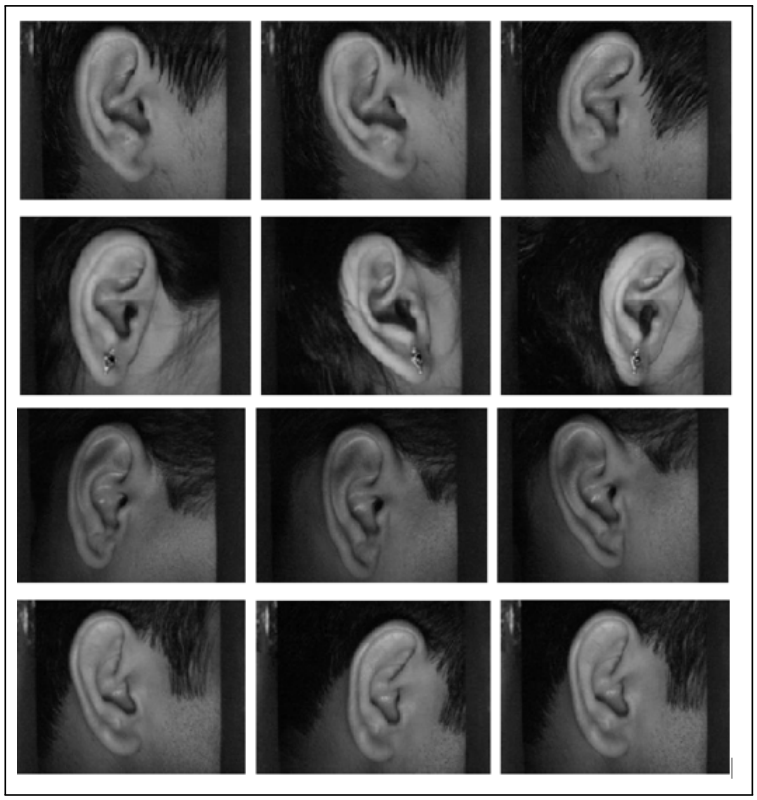


Figure 3.5: Samples of Delhi ear database [181].

version of ear database (automatically cropped and normalized) from 212 users with 754 ear images is also integrated and made available on request.The sample images from the IIT Delhi ear database are reproduced in Figure 3.5.

3.3.6 Cambridge Handshape Database

The size of the data set is about 1GB. The data set consists of 900 image sequences of 9 gesture classes which are defined by 3 primitive hand shapes and 3 primitive motions, see Figure 3.6. Therefore, the target task for this data set is to classify different shapes as well as different motions at a time [181] [182]. Each class contains 100 image sequences (5 different illuminations x 10 arbitrary motions x 2 subjects). Each sequence was recorded in front of a fixed camera having roughly isolated gestures in space and time. See Figure 3.7 for typical sample sequences of the 9 classes and Figure 3.8 for 5 different illumination prototypes.

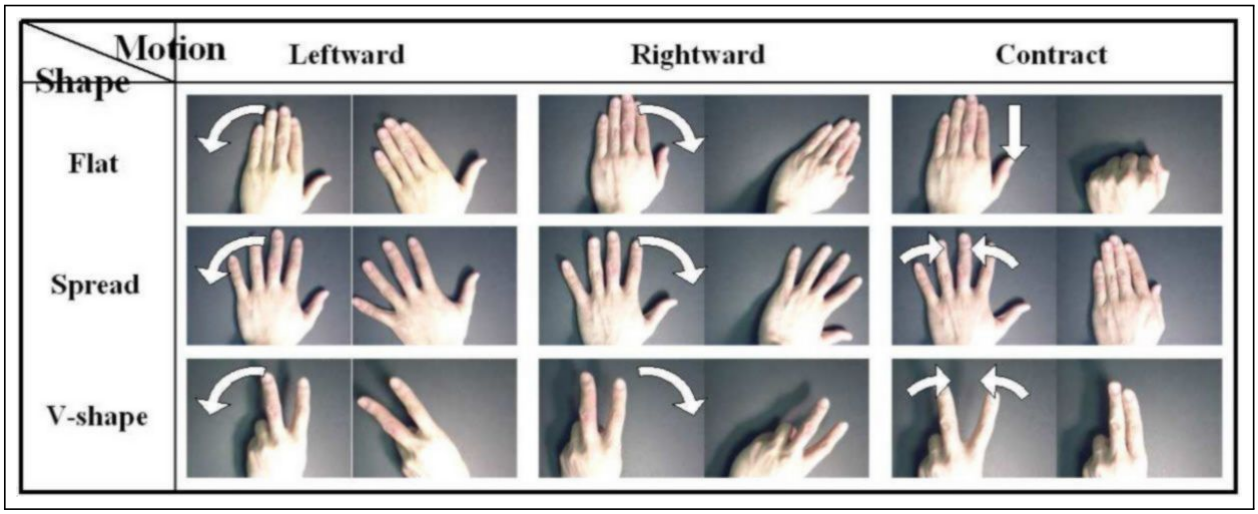


Figure 3.6: Hand-Gesture database. 9 different gesture classes are generated by 3 different primitive shapes and motions [181] [182].

3.3.7 CASIA-IrisV4

CASIA-IrisV4 is an extension of CASIA-IrisV3 including six subsets; CASIA-Iris-Interval, CASIA-Iris-Lamp, CASIA-Iris-Twins, CASIA-Iris-Distance, CASIA-Iris-Thousand and CASIA-Iris-Syn. CASIA-IrisV4 consists of 54,601 iris images from more than 1,800 genuine subjects in addition to 1,000 virtual subjects. All the iris images are 8 bit gray-level JPEG files, collected under near infrared illumination. The six data sets were collected or synthesized at different times. Each image in CASIA-IrisV4 has a unique file name which denotes some useful properties associated with the image for example, left/right/double, subset category, subject ID, image ID, class ID. Figure 3.9 shows some examples of CASIA-IrisV4 images [183].

3.3.8 Groups Images Dataset

The database contains 5,080 of people images such as wedding, bride, groom, portrait, group shot, group photo, group portrait or family portrait. The search for images is repeated for 270 different days, 100 images are returned for any given image capture day at maximum. Each face is labeled as being in one of seven age categories: 0-2, 3-7, 8-12, 13-19, 20-36, 37-65, and 66+. In addition, all faces are labeled with gender also [184]. Figure 3.10 shows examples of these images.

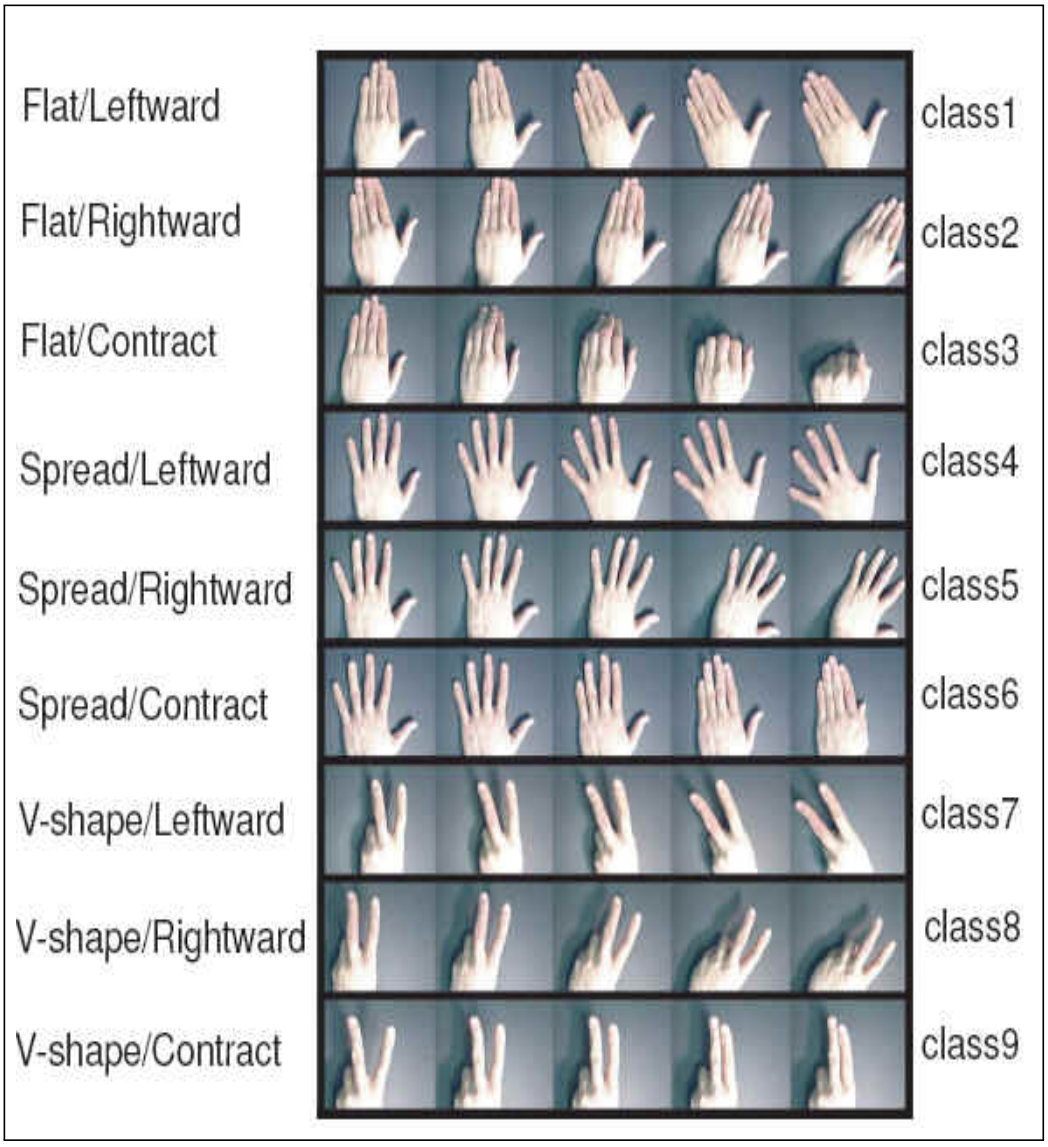


Figure 3.7: Sample sequences of the 9 gesture classes.

Figure 3.8: 5 different illumination conditions in the database.

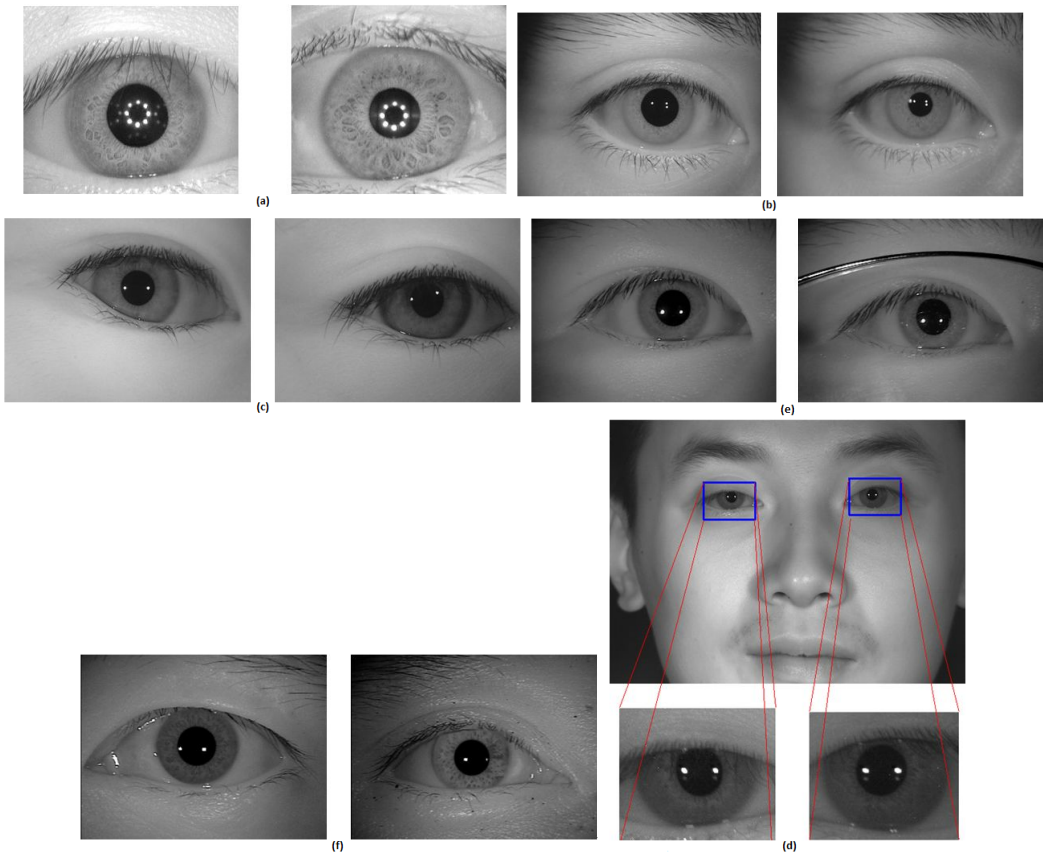


Figure 3.9: Examples of CASIA-IrisV4 subsets. (a) Represents CASIA-Iris-Interval, (b) represents CASIA-Iris-Lamp, (c) represents CASIA-Iris-Twins, (d) represents CASIA-Iris-Distance, (e) represents CASIA-Iris-Thousand and finally (f) represents CASIA-Iris-Syn [183].

Figure 3.10: Examples of groups images dataset [184].

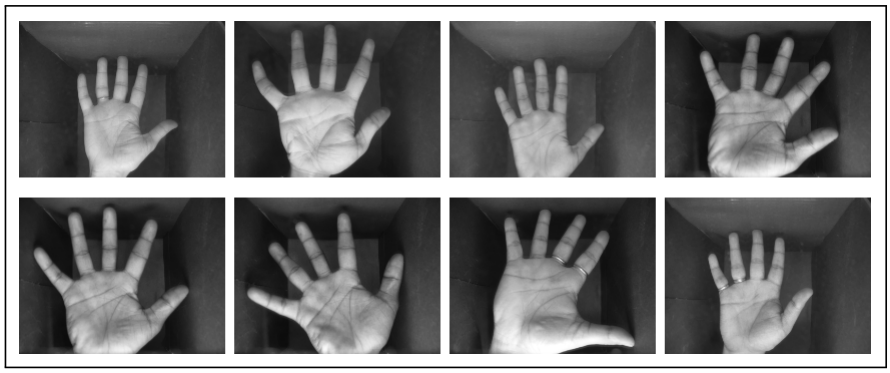


Figure 3.11: Samples of IIT Delhi palmprint database

3.3.9 IIT Delhi Palmprint Database V1.0

The IIT Delhi palmprint image database consists of the hand images acquired in the IIT Delhi campus in 2006. All the database images are collected in the indoor environment from 235 users of age 12-57 years. All the images are available in bitmap format and resolution of 800×600 pixels. Seven images of the left and right hand from each subject are acquired in varying hand pose variations. The acquired images have been numbered for every user with an identification number in sequence. In addition to the original images, there are plamprint images of 150×150 pixel which cropped and normalized automatically [185] [186]. Figure 3.11 shows examples of this database.

3.3.10 Southampton MultiBiometric Tunnel Database

The southampton multi-biometric tunnel database [187] is multiview synchronised video of subjects walking in a constrained environment. The environment is designed with airports and other high throughput environments in mind. It is able to acquire a variety of non-contact biometrics in a non-intrusive manner. The system uses eight synchronised IEEE1394 cameras to capture gait and additional cameras to capture images from the face and one ear, as an individual walks through the tunnel. The tunnel acquires data automatically as a subject walks through it and is designed for the collection of very large gait datasets [188].

3.3.11 West Virginia University MultiBiometric Datasets

Since 2008, West Virginia University(WVU) has performed large, medium, and small scale biometric data collection that includes two releases. First release of the dataset collection [189] includes image and sound files for six biometric modalities: iris, face, voice, fingerprint, hand Geometry, and palmprint. Second release of the dataset collection includes image and video files for the modalities: iris, face, face video and voice, fingerprint, hand geometry, and palmprint. Both releases of the dataset include soft biometrics such as height and weight, for subjects of different age, ethnicity and gender with variable number of sessions/subject.

3.3.12 The MultiBiometric databases Used

We generate multiple databases which consist of combinations of single databases: face, ear, eye, hand, and palmprint. Each of these databases consists of 50 distinct subjects with $10, 25, 40$ images each. Different splits between training/testing sets are examined to achieve high recognition performance. The following points have been taken into consideration for the multi-biometric databases that have been used.

- Collection of our own database would require institutional ethical approval.
- Can perhaps have too much information about each individual.
- Some are difficult to collect such as iris and fingerprint.
- The size of the database would be fairly small and perhaps not diverse in gender or race.
- By mixing individual biometrics, we can explore a much larger 'space' than those from single people– though we accept that we may explore combinations that are unrealistic.
- individual databases used are freely available, so other researchers could also download them and reproduce our findings.

3.4 Assessing Vision System Performance

Characterize and assess the performance of techniques that used in computer vision is crucial to the development of discipline. The performance of a system is assessed by running the program on a large number of inputs and counting the cases number of correct and incorrect results which produced by the the program. The performing of each individual test can yield one of four possible results:

- True positive (TP): Obtains when a test should give a positive result does so.
- True negative (TN): Obtains when a test should give a negative result does so.
- False positive(FP): Obtains when a test should give a positive result but actually gives a negative result.
- False negative(FN): Obtains when a test should give a negative result but actually gives a positive result.

3.4.1 Assessing an Individual Algorithm

Researchers have introduced various procedures to present the data graphically. The following describe three of these methods.

The Receiver Operating Characteristic Curve (ROC)

ROC curves often used to present results for problems of binary decision in machine learning. An ROC curve is a graphical plot that clarify the performance of a system by plotting the true-positive rate (TPR) against the false-positive rate (FPR) as its discriminant threshold is varied. The TPR determines the number of correct positive results that produced among the total positive samples prepared during the test. On the other hand, the FPR determines the number of incorrect positive results that produced among total negative samples prepared during the test. The ROC curves in Figure 3.12 have been interpreted as follows.

Figure 3.12: Examples of ROC curves [192]

- the curve nearest to the top left-hand corner of the plot represents the most accurate test.
- the curve nearest to a 45° diagonal represents the worst accurate test.
- the area under the curve represents a measure of the test accuracy.
- the plot depicts relative trade-off between the true positive and the false positive rate: any increase in true positive rate will be associated to an increase in false positive.

The accuracy of a test is measured by the area under its ROC curve. To estimate area under the curves, different methods have been proposed. Two methods are commonly used; a parametric method based on using a maximum likelihood estimator to fit the data points to a smooth curve, a non-parametric method based on fitting trapezoids under the curve. Both of these methods are accessible as computer programs and give an estimation to area under the curves and standard error to compare various tests or the same test in different samples [190] [192]. Standard error can be referred on these plots. After the mean of a set of observations is calculated, some indication of how accurate the estimated mean is likely to be to the parametric ("true") mean must be given. One way to do this is by using confidence limits. The assessed confidence limits can then be plotted as error bars or error ellipses around the points.

Precision-Recall and Related Curves

Precision-Recall (PR) curves commonly used in Information Retrieval [191]. Precision is a measure of quality, whereas recall is a measure of quantity. In simple terms, high precision implies that an algorithm returned basically more relevant results than irrelevant results, whilst high recall implies that an algorithm returned basically most of the relevant results.In a classification process, the precision for a class is defined as the number of true positives (the number of instances that are classified correctly as belonging to the positive class) divided by the sum of true positives and false positives (the number of instances that classified incorrectly as belonging to the class). Recall is defined as the number of true positives divided by the sum of true positives and false negatives. In PR curves, true-positive rate (TPR) and false-positive rate (FPR) are frequently plotted versus each other to find operating point trade-off.

$$precision = \frac{TP}{TP + FP} \tag{3.1}$$

$$recall = \frac{TP}{TP + FN} \tag{3.2}$$

In general, the ROC and precision-recall curves are useful in evaluating how various parameters applied to an algorithm influence performance.

Confusion Matrices

Confusion matrix also known as an error matrix or a matching matrix. It is a specific table that gives visualization of an algorithm performance. Columns of the matrix represent the items in a predicted class whilst rows represent the items in an actual class [193]. Using confusion matrices allow to see if the system is confusing two classes. In another words, mislabelling one as another. As it is shown in Figure 3.13, the primary diagonal values of the table represent the number of instances that are classified

		actual									
		0	1	2	3	4	5	6	7	8	9
expected	0	20	0	0	6	0	0	1	0	10	0
	1	0	25	0	0	0	0	0	6	0	0
	2	0	0	31	0	0	0	0	0	0	0
	3	0	0	0	21	0	0	0	0	10	0
	4	0	0	0	0	31	0	0	0	0	0
	5	0	0	0	0	0	22	0	0	9	0
	6	1	0	0	1	0	2	23	0	3	1
	7	0	8	0	0	0	0	0	23	0	0
	8	4	0	1	3	2	1	3	0	13	4
	9	0	0	0	2	0	0	0	3	1	27

Figure 3.13: An Example of confusion matrix [192].

correctly, while off-diagonal values represent the number of mis-classifications. Through observation, it is clarified that small numbers along the primary diagonal represent cases in which the performance of the classification is poor. For instance, the actual digit ’6’ has been mis-classified as ’3’ once, as ’5’ two times and as ’8’ three times, while the actual digit ’4’ never been misclassified [192].

3.4.2 Comparing Two Algorithms

Using ROC or Precision-Recall Curves

The common way to compare algorithms is by their ROC or precision-recall curves. Often ROC curves to be compared intersect each other. And then it becomes difficult to indicate the points accuracy in the curve by using error bars or tells if there is any significant differences in performance or not. Hence, it is hard to the user to choose which curve that represents the best method for their application. See Figure 3.14.

McNemar’s Test For Class Labels

McNemar’s test is a statical type of comparison that works on a pair of algorithms to explore where one processing succeed and the other failed [194]. Table 3.1 shows 2×2 truth table for two algorithms A

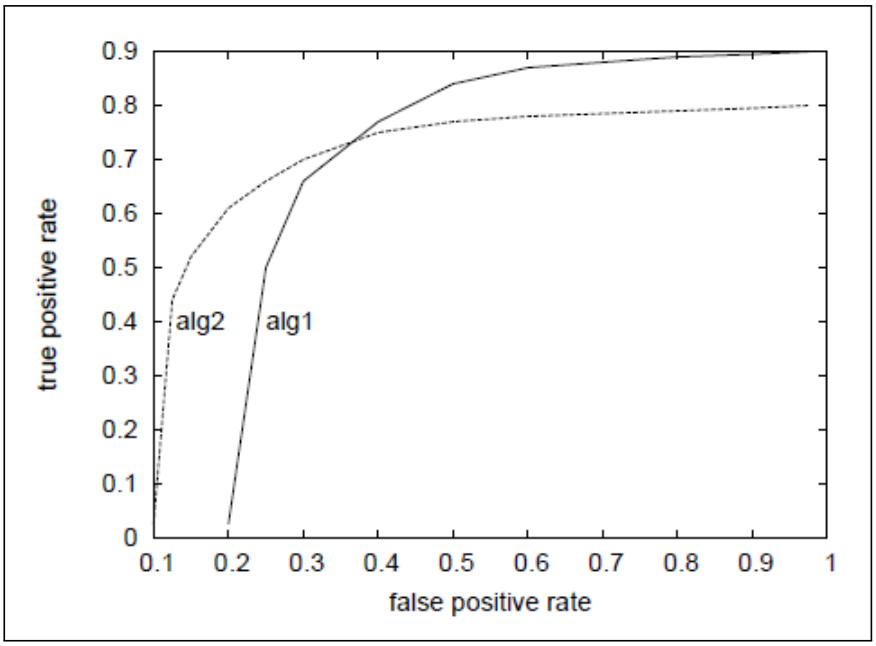


Figure 3.14: Crossing ROC curves [192]

and B, where N_{sf} refers to the number of tests for which algorithm A succeeded and algorithm B failed, N_{fs} refers to the number of tests for which algorithm A failed and algorithm B succeeded, and so on. McNemar's test involves computation of the so-called Z-score using:

$$Z = \sqrt{x^2} \rightarrow \frac{|N_{sf} - N_{fs}| - 1}{\sqrt{N_{sf} + N_{fs}}} \tag{3.3}$$

$$x^2 = \frac{(|N_{sf} - N_{fs}| - 1)^2}{N_{sf} + N_{fs}} \tag{3.4}$$

where the -1 is a continuity correction. Confidence limits can be associated with the value of Z. That means, if the result form algorithms A is similar to algorithm B then Z will be near zero. Otherwise, the Z value will increase when the two algorithms results diverge [195] [192].

Table 3.1: Truth table for McNemar's test

	Algorithm B failed	Algorithm B succeeded
algorithm A failed	N_{ff}	N_{fs}
algorithm A succeeded	N_{sf}	N_{ss}

Many evaluating approaches take no account of the number of tests that have been performed: the size

of the dataset may be sufficiently small that any difference in performance could have arisen purely by chance. Instead, McNemar's test takes this into account.

3.4.3 Comparing Several Algorithms

McNemar's test compares only pairs algorithms. When several algorithms need to be compared, it becomes necessary to find other ways that making multiple pairwise comparisons. Some caution is needed when comparing several methods. In particular, using such comparisons tends to boost the family-wise error rate [194] [192]. More precisely, it is possible for each comparison to give a specific result by chance. Consequently, performing several such comparisons increases the probability of this happening. If a result probability arising by chance in any single comparison is α, so the probability of not gaining such an error is $1-\alpha$. Different methods used to reduce the α value, one of these method called Bonferroni correction which is a modification made to α when more than two statistical tests are applied simultaneously on a dataset [195].

3.5 Conclusions

This chapter sheds light on some of the ethical issues in biometric technologies. This chapter also reviews existing databases that used for biometric research. Furthermore, performance evaluation techniques commonly used in vision research are reviewed. Next chapter will present the Viola-Jones face framework and the fundamental aspects related to this technique. Also examine the well-established Viola-Jones algorithm can be made more effective by applying it to the brightness channel of a transformed colour image rather than the conventional grey-scale image obtained from a colour one.

Chapter 4

Improving The Viola-Jones Face Detection Performance by Using The Brightness Channel in HSV And HLS Colour Space

4.1 Introduction

The first and arguably the most important step of face recognition is face detection (sometimes called face location) because knowing where faces are located makes the recognition phase less complicated [7]. Viola and Jones proposed an object detection framework characterized by high detection accuracy and low computation time, and applied it to the problem of face detection. Their technique involves collecting a large set of face and non-face images and training a classifier to discriminate them. Although the Viola-Jones technique is considered as the most successful face location algorithm available which is used within almost every camera and smartphone, it does has shortcomings; it misses faces (false negatives) and identifies non-face regions as being faces (false positives), it is effective on only frontal images, performance decreases with changes to the head size and it is sensitive to illumination. Generally, the algorithm is applied to grey-scale images in which the value of each pixel represents the

intensity of light at this pixel. Illumination plays a surprisingly significant role in the effectiveness of Viola-Jones performance. There are some color spaces such as HSV and HLS which separate the image intensity (luma), from the color information (chroma). In computer vision, leaving the color components alone and using the intensity component only is useful in some applications for example, robustness to removing shadows or lighting changes. No study examines Viola-Jones detection performance using different color spaces to overcome illumination problem. Therefore, the aim of this chapter is to explore whether these the Viola-Jones failure rates can be reduced by the simple expedient of manipulating the colour space of an image before presenting it to Viola-Jones. The remainder of this chapter is structured as follows. Section 4.2 describes the Viola-Jones face recognition algorithm, describing its major stages in some detail. Section 4.3 outlines the most common failure modes of the technique followed by colour space in section 4.4. Then section 4.5 presents the results of experiments that demonstrate how the effectiveness of the algorithm can be improved by first manipulating the colour space of an image (Viola Jones has been employed for face and hand detection) . Section 4.6 draws conclusions.

4.2 Face Detection Using The Viola-Jones Approach

Before the advent of the Viola-Jones algorithm, the principal way of identifying faces was by colour, but that approach was easily confused by other regions of a skin or by other similarly-coloured image features. Unlike previous approaches, Viola-Jones operates on grey-scale images. The Viola-Jones approach involves the reduction of any input image to a fixed size, for example 24×24 pixels. After that, all possible Haar features need to be extracted from the image. There are 162,336 likely features, even for small images, and this explains why the learning algorithm is a slow process. Most of the calculated features are irrelevant, a few of these features will help to detect the face. The learning process tries to find which set of features that reduce the error rate, the number of mis-classification, in the classification of the face and non-face image regions. Individually, each feature is considered as a weak classifier. Therefore, all the weak classifiers are combined in such away to gain a strong classifier. This is known

as Adaboost learning algorithm. Viola and Jones found that about 6,000 features classified all faces correctly. Despite the fact that the time needed to compute 6,000 features is less than the time taken for 162,336, it remains possible to make this process more efficient. Viola and Jones introduce the concept of a classifier cascade which contains a quick test to check whether a region specifically does not include a face. In this case, it means there is no need to calculate the other Haar features [196].

4.2.1 Haar-like Features

Viola and Jones adapted the Haar wavelets idea and advanced the so-called Haar-like features. A Haar-like feature is defined as the difference value of rectangular image regions. This difference is then used to classify subsections of an image. In Viola-Jones object detection framework, a window of a desired size is moved over the original image, the Haar-like feature is calculated for each subsection of the input image. Then the difference is compared to a threshold in order to classify non-objects from objects. Because a Haar-like feature is a weak classifier, a huge number of Haar-like features are basic to describe an object with acceptable accuracy. For this reason, the Haar-like features used by Viola-Jones object detection framework are organized in a classifier cascade to build a strong learner or classifier.

Using integral image, an intermediate representation for the image, is the key interest of a Haar-like feature over most other features which led to a quick features calculation. In other words, a Haar-like feature of any size can be calculated in a fixed interval of time. Viola-Jones approach used three kinds of features as shown in Figure 4.1:

- two-rectangle feature calculates the difference between the sum of the pixels inside two rectangular regions.
- A three-rectangle feature is defined as the difference between the sum within two outside rectangles subtracted from the sum of the pixels within a center rectangle
- a four-rectangle feature is defined as the difference between diagonal pairs of rectangles.

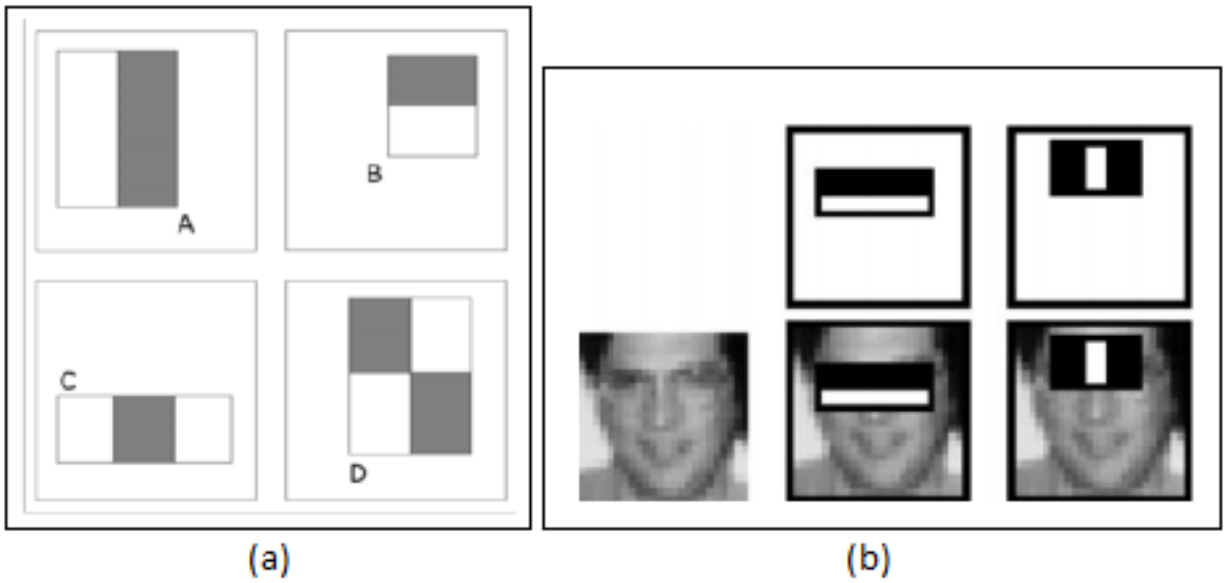


Figure 4.1: (a) Feature types used by Viola and Jones (b) Haar features that look similar to the eye region and the bridge upper nose region is applied on a face [196].

Most Haar features are irrelevant but a few will help to detect the face. The learning process tries to find which set of features reduces the error rate, the number of mis-classifications, in the classification of the face and non-face image regions. Individually, each feature is considered as a weak classifier; all the weak classifiers are combined in such away to gain a strong classifier [205].

4.2.2 Integral Image

Viola-Jones approach characterized by using an intermediate representation for the original which is know as an integral image. The integral image is defined as a two-dimensional table (Summed Area Table) with the same size as the original image. Making use of the so-called *integral image* representation allows the computation of the sums of rectangular regions in constant time, so the computation of the Haar features is rapid. The value of a point (x, y) within an integral image is computed as the sum of all pixels placed in the above-left region of the original image, including the value at the point (x, y) itself. This allows one to calculate the sum of a rectangular region at any position or size in the image to be achieved using only four values of the integral image. The pixel value at the point (x, y) in the

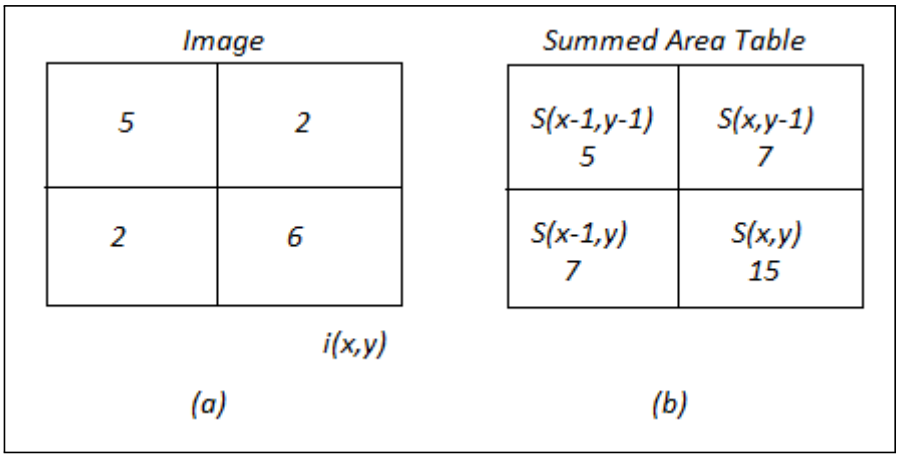


Figure 4.2: (a) The original image (b) The summed area table.

integral image is simply computed by:

$$s(x,y) = i(x,y) + s(x-1,y) + s(x,y-1) - s(x-1,y-1) \tag{4.1}$$

The pixel value $i(x,y)$ in the original image is added to the values directly above and left to this pixel at $s(x-1,y)$ and $s(x,y-1)$ from the Summed Area Table. Then, the value exactly top-left of $i(x,y)$ from the Summed Area Table at $s(x-1,y-1)$ is subtracted. Figure 4.2 illustrates the summed area table.

4.2.3 Adaboost Training and Feature Selection

The technique known as *adaptive boosting* or *AdaBoost* is a common approach that can be used to combine multiple weak classifiers into one strong classifier to enhance performance. With the 24 × 24-pixel windows used in Viola-Jones, there are 162,336 possible rectangle features, and computing them all is time-consuming. Adaboost was used to establish those that contribute to distinguishing face and non-face regions. A training set of labelled images is prepared by scaling all images to 24 × 24 pixels. So, each image has index $l, l = 1 \ldots L$. A corresponding value y_l is assigned for each image [196]: $y_l = 1$ for faces and $y_l = 0$ for non-faces. Then, some weights are initialized:

$$w_{1,l} = \frac{1}{2p_-}, \frac{1}{2p_+} \tag{4.2}$$

where p_- and p_+ are the number of negative (images without faces) and positive (images with faces) images in the images set. For $i = 1 \dots L$, the following steps are performed:

1. Normalize the weights as following

$$\frac{w_{i,l}}{\sum_{j=1}^{n} w_{i,j}} \quad \rightarrow \quad w_{i,l} \tag{4.3}$$

2. For each feature j, train a classifier h_j which is restricted to use a single feature. The classifier's error rate ε_j is evaluated with respect to $w_{i,j}$ as

$$\varepsilon_j = \sum_{l=0}^{L-1} w_{i,l} |h_j(x_l) - y_l| \tag{4.4}$$

3. Choose the classifier, h_i with the lowest error ε_i. Update the weights:

$$\begin{aligned} w_{i+1,l} &= w_{i,l} \beta_i^{1-\varepsilon_i} \\ \beta_i &= \tfrac{\varepsilon_i}{1-\varepsilon} \end{aligned} \tag{4.5}$$

The final strong classifier is:

$$h(x) = \begin{cases} 1, & \text{if } \sum_{i=0}^{I-1} \alpha_i h_i(x_l) \geq \frac{1}{2} \sum_{i=0}^{I-1} \alpha_i \\ 0, & \text{otherwise} \end{cases} \tag{4.6}$$

where $\alpha_i = \log \frac{1}{\beta_i}$

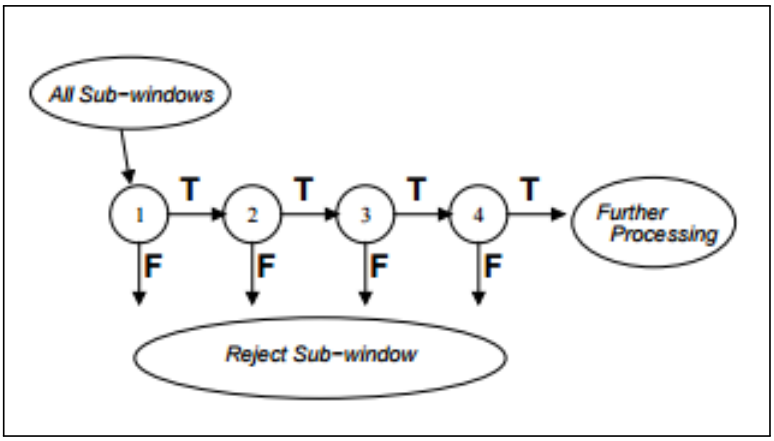


Figure 4.3: Detection cascade scheme [198]

4.2.4 Cascaded Classifiers

Cascading depends on the idea of using a sequence of several classifiers to increase the performance of detection and to reduce the time of computation [196]. In other words, all the collected information produced from a given classifier is used for the next classifier in the cascade as additional information.The first cascading classifier is used by Viola and Jones face detector in 2001. The estimation of the strong classifiers introduced by the learning process can be implemented quickly, but not fast enough to execute in real-time. For this reason, the need arises to arrange these strong classifiers in a cascade, so each classifier is trained only on the chosen samples which introduced by the previous classifiers. In cascading, all the features are clustered into several stages where each stage includes a classifier of a certain number of features. The task of each stage is to decide whether a given sub-window is a face or not. if a given sub-window fails in any of the stages, it is immediately neglected as not a face [110]. A simple framework for cascade training is shown in Figure 4.3.

4.3 Failure Modes of the Viola-Jones Algorithm

As alluded to in section 4.1, the Viola-Jones algorithm does not find all faces in all images, and can identify non-face regions as being faces. The following paragraphs identify the most common failure modes.

Figure 4.4: Head rotated horizontally [199]

Figure 4.5: Face images demonstrating the range of illumination variation [199]

4.3.1 Head Pose

The Viola-Jones algorithm works only for subject looking at the camera, However, humans do not always face the camera — indeed, good portrait photographs rarely have the subject posed full face, looking at the photographer (Figure 4.4) [199] [200]. This is arguably the most common failure mode of the algorithm.

4.3.2 Illumination

Illumination plays a surprisingly significant role in the effectiveness of Viola-Jones. For example, a back-lit subject photographed on a bright day means that their face is largely in shadow and hence exhibits poor contrast. Viola-Jones is much less effective in such cases (see Figure 4.5).

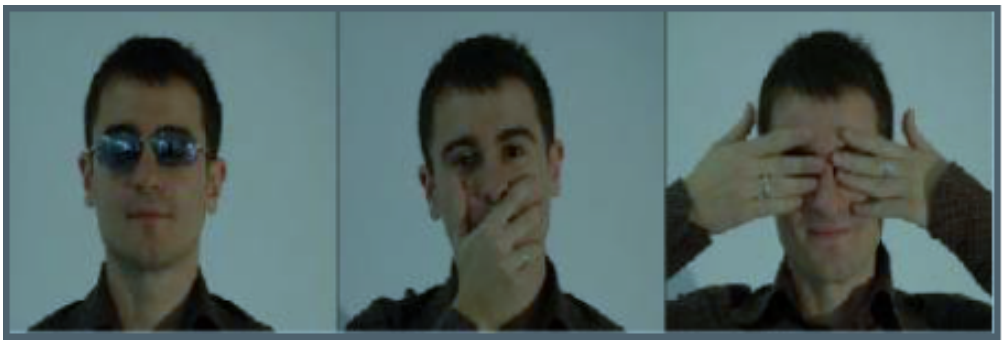


Figure 4.6: Simple occluded face images [199]

Figure 4.7: Facial expression of a person [199]

4.3.3 Obstruction in Front of The Face

The Haar features used to classify a detected object as a face or not depend on the variation of intensity of the different parts of a face. The presence of an occluding object in front of a face prevents the determination of the required feature that needed to detect it. Figure 4.6 shows some obstructions that cause Viola-Jones to fail.

4.3.4 Facial Expression

A discrepancy may occur between a recorded face and the current facial expression due to variation in the facial expression. Figure 4.7 shows different facial expressions for a person from a University of Minnesota database, most of which cause Viola-Jones to fail.

4.4 Colour Space

A *colour space* is a specific mathematical model of representing a set of colours. The most common colour model, the one used by most computer displays and digital cameras, is red–green–blue (RGB). An RGB color space can be described as all possible colors, which can be made from three colors for red, green and blue. In other words, each pixel of an image is allocated a range of 0 to 255 intensity values of RGB components. However, other colour spaces abound, with acronyms such as YIQ, YUV, YC_bC_r and CMYK [201] [202]. This work has investigated two specific colour spaces related to an artist's notion of hue, saturation and brightness, namely HSV and HLS [203], both of which can be obtained by a simple transformation of RGB. The principal difference between HLS and HSV is the calculation of the brightness component (L or V), which determines the distribution and the range of the brightness and the saturation (S) [201] [203]. HSV (Hue, Saturation, Value) and HSL (Hue, Saturation, Lightness) are alternative ways to represent the RGB color model, designed by computer graphics researchers in the 1970s to be aligned with the way human vision recognizes color-making attributes. The HSV (Hue, Saturation, Value) model, defines a color space in terms of three components:

- Hue refers to the color type (such as red, blue, or yellow), ranges from $0-360$
- Saturation refers to the "vibrancy" or "purity" of the color, ranges from $0-100$. In other words, the lower the saturation of a color, the more "grayness" is present
- Value refers to the brightness of the color, ranges from $0-100$. Brightness is the relative lightness or darkness of a particular color, from black (no brightness) to white (full brightness)

HSL stands for (hue, saturation, lightness/luminance), and is often also called HLS, is similar to HSV model with "lightness" replacing "brightness". The difference between these two models is that a perfectly light color in HSL is pure white; but a perfectly bright color in HSV is similar to shining a white light on a colored object. More specifically, shining a bright white source on a red target causes the target to still appear red, just brighter. Otherwise, shining a low light on a red target causes the target

to appear dark and less bright.

4.5 The Experimental Work

In the experimental work, Viola-Jones has been applied to face and hand databases to explore whether these the Viola-Jones failure rates can be reduced by the simple expedient of manipulating the colour space of an image before presenting it to Viola-Jones.

In general, databases used for face recognition typically contain a single face, so there are no scene components that could potentially confuse Viola-Jones. Hence, this work has assimilated a database of group photographs, each of which contains more than one face; the total number of faces in all the images is 2250. Different lightings have been cast on each image. Each of these images has been transformed into three different single-channel representations: grey-scale, lightness (L of HSL) and value (V of HSV). In each case, all of the images were transformed into the relevant colour space before being passed through the Viola-Jones algorithm and all the detected faces recorded. These detected faces were then compared with the ground truth of all the faces in all the images and used to determine whether the face had been found or not, and to establish whether non-face regions had been identified as faces (false positives). Figures 4.8, 4.9 show some of the experimental results.

In experiments of these kinds, it is common to quote the numbers true positives *etc*; but despite their widespread use these numbers actually have little meaning. In establishing whether there is a genuine, statistically-significant performance difference in applying Viola-Jones to images obtained from different colour spaces, a more meaningful approach is to employ a statistical test. The best available test is the McNemar's test, as defined in chapter 3. Table 4.1 shows the Z values (equations 3.3, 3.4) of Viola-Jones applied on grey, H and L.

color spaces	N_{sf}	N_{fs}	Z
V and grey	146	57	6.176
L and grey	82	44	3.296
V and L	45	19	4.238

Table 4.1: Z-values of Viola-Jones applied to detect faces on grey, H and L for face detection.

Figure 4.8: The Viola-Jones approach applied to different colour-space channels. (a) output when applied to V-channel image; (b) output when applied to L-channel image(c); output when applied to grey-scale image.

Figure 4.9: Viola-Jones applied to different images with different lightness +10, +20, -10, -20 percentage. (a) Viola-Jones applied on V-channel image; (b) output when applied to L-channel image; (c) output when applied to grey-scale image.

Results obtained from the comparisons using McNemar's test of the original Viola-Jones algorithm which is based on the grey-scale image with the V-channel of HSV space and the L-channel of HLS space are shown in table 4.1. We can see that, the number of cases in which the Viola-Jones based on V-channel algorithm succeeds and the Viola-Jones based on grey algorithm fails, $N_{sf} = 146$ and vice versa, $N_{fs} = 57$. In the other hand, The number of cases in which the Viola-Jones based on L-channel algorithm succeeds and the Viola-Jones based on grey algorithm fails, $N_{sf} = 82$ and vice versa, $N_{fs} = 44$. It is normal to look for $Z > 1.96$, which means that the results from the algorithms would be expected to differ by chance only one time in 20. Consequently, both of these algorithms are well above $z_{crit} = 1.96$, so one can conclude that with a probability > 0.95, the Viola-Jones algorithm is more effective if the image is first transformed to the HSV or HLS color space and the appropriate brightness channel of that space (V, L channels) is used. The table also indicate that the Viola-Jones based on V-channel of HSV color space is more effective than the Viola-Jones based on L-channel of HLS color space where $N_{sf} = 45$ and $N_{fs} = 19$, though whether or not that the difference is statistically significant would involve performing McNemar's test between those sets of results.

The hand data set that been used in the experimental work consists of 900 image. Different lightings have been cast on each image and then transformed into three different channel representations: grey-scale, lightness (L of HSL) and value (V of HSV) before being passed through the Viola-Jones algorithm. The detected hands were then compared with the ground truth of all the hands in all the images and used to determine whether the hand had been found or not.

Results obtained from the comparisons using McNemar's test of the Viola-Jones algorithm to detect hands applied on the grey-scale image, the V-channel, and the L-channel are shown in table 4.2. Obviously, the number of cases in which the Viola-Jones based on V-channel algorithm succeeds and the Viola-Jones based on grey algorithm fails, $N_{sf} = 96$ and vice versa, $N_{fs} = 43$. In the other hand, The number of cases in which the Viola-Jones based on L-channel algorithm succeeds and the Viola-Jones based on grey algorithm fails, $N_{sf} = 71$ and vice versa, $N_{fs} = 35$. The table also indicates that the

Viola-Jones based on V-channel of HSV color space is more effective than the Viola-Jones based on L-channel of HLS color space where $N_{sf} = 55$ and $N_{fs} = 21$. The experimental results has indicated that using the V-channel of HSV is the best for hand detection followed by the L-channel of HLS and the grey-scale image in order depending on the $z_v alues$ that been achieved. Table 4.2 shows the Z values (equations 3.3, 3.4) of Viola-Jones applied to detect hands on grey, H and L.

color spaces	N_{sf}	N_{fs}	Z
V and grey	96	43	4.410
L and grey	71	35	3.399
V and L	55	21	3.785

Table 4.2: Z-values of Viola-Jones applied on grey, H and L for hand detection.

4.6 Conclusions

Due to the fact that Viola-Jones face detection framework may fail to detect faces under various environmental conditions. Therefore, using some color spaces such as HSV and HLS which separate the image intensity (luma), from the color information (chroma) might be useful. This work has shown that the well-established Viola-Jones algorithm used for both face and hand detection can be made more effective and less sensitive to illumination by applying it to the brightness channel of a transformed colour image, rather than the conventional grey-scale image obtained from a colour one. Moreover, the experimental results has indicated that using the V-channel of HSV is the best followed by the L-channel of HLS and the grey-scale image in order. This means that the simple, fast transformation from RGB pixels to HSV or HLS as a precursor to using conventional Viola-Jones face detection and hand detection should yield significantly fewer false positives and negatives, a significant improvement in its effectiveness. Next chapter will introduce conventional statistical techniques that used for biometric classification.

Chapter 5

Biometric Recognition using Conventional Statistical Techniques

5.1 Introduction

Classification is the process to categorize data into several categories, based on their similarities. There are three main classification techniques that used for classification; supervised learning, unsupervised learning and semi-supervised learning [197]. This chapter introduces two major methods that used for biometric image classification; supervised learning and unsupervised learning. Also, some other related methods that used for classification have been discussed such as, SVM, PCA, and LDA. Four distance similarity measures: Manhattan, Euclidean, Cosine similarity, and Mahalanobis distance have been compared for PCA- and LDA-based face, ear, and palm biometrics. Furthermore, the face, ear and palm recognition performance using SVM based on PCA and SVM based on a combination of PCA and LDA techniques have been compared with the PCA and LDA techniques depend on distance similarity measures.

5.2 Supervised and Unsupervised Learning

In supervised learning, pre-labelled objects (samples) are required to determine an accurate classification of an image. The task here is to build a classifier based on these labeled objects; a model which is able to predict the class of a new object. This can be done by using different techniques for example, Support Vector Machines (SVMs), Random Forest Classification, and Convolutional Neural Networks (CNNs). Supervised learning process involving three main steps:

- build a model: Choose a machine learning algorithm
- Learning (training): Learn the model using the training data.
- Testing: Test the model using unseen test data to assess the model accuracy.

$$Accuracy = \frac{Number\ of\ correct\ classifications}{Total\ number\ of\ test\ cases} \tag{5.1}$$

The unsupervised learning task is inferring a function to describe and explain key features from unlabeled data. the unsupervised classification outcomes (groupings of pixels that have common characteristics) are based on the software analysis of an image without providing sample classes. The most common task within unsupervised learning is clustering that groups the given samples of unknown classes into clusters such that the subjects in a group will be similar (or related) to one another and different from (or unrelated to) the objects in other groups, each represented as a vector $\mathbf{x} = [x_1, \cdots, x_N]^T$ in the N-dimensional feature space, into a set of clusters [204]. Today many different unsupervised classification techniques are commonly used in different applications. The two most used techniques are the ISODATA and the K-mean clustering algorithm. In general, both of these two algorithms are an iterative procedure. Figure 5.1 shows the classification steps. The first step in the classification process is to define the classes which are related to the property of the image and it should be defined clearly. Then, features of these classes are extracted where features of each class differ from one to another. Various classification techniques like supervised and unsupervised learning are used according to the data to

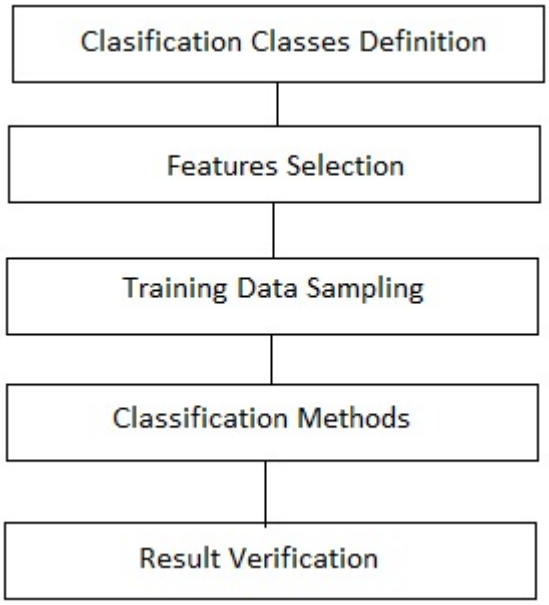


Figure 5.1: The classification steps

obtain correct decision rule which is necessary to sample the training data. The last step is choosing an appropriate classification method like SVM, PCA and linear discriminate analysis (LDA)to gain decision.

Classification methods that are based on matching represent each class by an initial pattern vector. An unknown pattern is allocated to the class to which it is closest. Using the minimum distance classifier is the simplest approach to find the closest class, as its name implies, it determines the distances between the unknown and each of the initial vectors and chooses the smallest distance to make a decision.

5.3 Distance and Similarity Measures

Similarity is usually described as the degree of how much alike two data are. In other words, it is a distance between the features of two objects. If this distance is large it will be the low degree of similarity whereas small distance will be the high degree of similarity [206]. The most common techniques that used to determine the similarity between objects are:

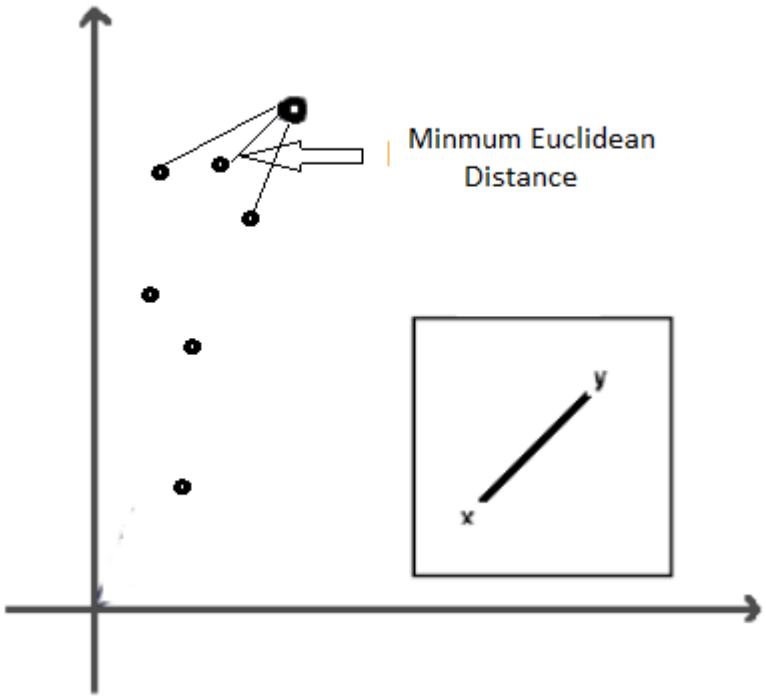


Figure 5.2: The Euclidean distance.

Euclidean Distance It is also known as simply distance. The Euclidean distance between two feature vectors x and y is simply the path length connecting them [207] as determined by Pythagoras's theorem:

$$\text{Eculidean Distance}(x,y) = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{5.2}$$

For this to be meaningful, the individual components of the feature vector need to be of the same type. For example, if one component is a radius and the other an angle, the Euclidean distance is not a sensible similarity measure. Figure 5.2 illustrates the euclidean distance.

Manhattan Distance The Manhattan distance between two points is the sum of the absolute differences of their coordinates; the x-coordinates and y-coordinates [206].

$$\text{Manhatten Distance}(x,y) = \sum_{i=1}^{k}|x_i - y_i| \tag{5.3}$$

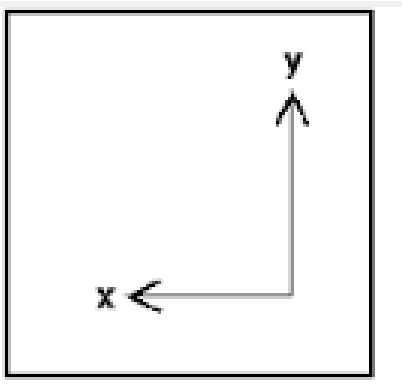


Figure 5.3: The Manhatten distance.

Again, the individual components of the feature vector need to be of a consistent type for this to make sense. See Figure 5.3.

Minkowski Distance The Minkowski distance is a combination metric form of Euclidean distance and Manhattan distance.

$$d^{MKD}(i,j) = \lambda\sqrt{\sum_{k=0}^{n-1}|y_{i,k} - y_{j,k}|^{\lambda}} \tag{5.4}$$

d^{MKD} is the Minkowski distance between the data record i and j, k the index of a variable, n the total number of variables y and λ the order of the Minkowski metric [185].

if $\lambda = 2$ then $d^{MKD}(i,j)$ is Euclidean distance.

if $\lambda = 1$ then $d^{MKD}(i,j)$ is Manhattan distance

Cosine Similarity This essentially measures the angle between a pair of feature vectors via a normalized inner product. The cosine of 0° is 1, and for any other angle is less than 1. Two vectors with the same orientation have a cosine similarity of 1, two vectors at 90° have a similarity of 0, and two opposed vectors have a similarity of -1 [185]. See Figure 5.4.

$$\text{sim}(x,y) \equiv \cos(\theta) = \frac{x.y}{|x||y|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} \tag{5.5}$$

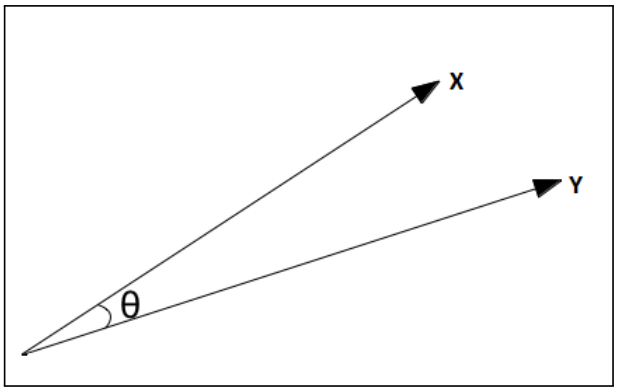


Figure 5.4: The Cosine distance.

Mahalanobis Distance: The Mahalanobis distance between two vectors _x and _y can be calculated as follows:

$$d_{Mahalanonbis}(x,y) = \sqrt{\sum_{i=1}^{N} \frac{(x_i - y_i)^2}{S_i^2}} \tag{5.6}$$

where, S_i is the standard deviation of the x_i and y_i over the sample set [208] [209].

5.4 Principal Component Analysis Algorithm (PCA)

PCA is a well-established statistical pattern recognition technique for data reduction and feature extraction [210]. It works by identifying the principal types of variation of an input data set and then calculating a basis set that maximizes variance along orthogonal directions in feature space. This decomposition is optimal in the linear least squares sense. An individual feature vector can be calculated as a weighted sum of the basis set [211], so those weights describe a feature vector completely. In the context of face recognition, the PCA approach is usually termed *eigenfaces* [212], relying on a low-dimensional representation of face images.Faces images which are similar in overall formation will be clustered in feature space and therefore can be characterized by a low dimensional subspace.

An eigenface (basis vector) will not generally correspond to the variation of a common facial feature such as eyes, or mouth but rather to the variation present in the input to the eigen decomposition.

The face recognition process will categorize a probe image containing a face as known or unknown, depending on how close the feature vector calculated from the probe image matches those of subject enrolled in the system [213].

Computation of Eigenfaces

In 2-dimension, a face image with size $N \times N$can also be represented as one dimensional vector of dimension N^2. For instance, face image from ORL database with size 112×92 can be represented as a vector of dimension 10,304, or a point in a 10,304 dimensional space. The images map to a set of points in this huge space. Faces images which are similar in overall formation will not be randomly distributed in this huge image space and therefore can be characterized by a low dimensional subspace. The PCA method tries to find the vectors that best computation for the face images distribution within the whole image space. These vectors are known as eigenfaces, which define the subspace of face images [210].
step1: Obtain face images $\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_M$ (training faces) face images must be centered and of the same size. It is very essential, the face images must be normalized (centered and of the same size) [214]

step2: represent every image $\mathbf{I}_i$ as a vecter $\boldsymbol{\Gamma_i}$

step3: Compute the average face vector $\boldsymbol{\Psi}$

$$\boldsymbol{\Psi} = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{\Gamma}_i \tag{5.7}$$

step4: Subtract the mean face:

$$\boldsymbol{\Phi}_i = \boldsymbol{\Gamma}_i - \boldsymbol{\Psi} \tag{5.8}$$

step5: Compute the covariance matrix **C**:

$$\mathbf{C} = \frac{1}{M}\sum_{n=1}^{M}\boldsymbol{\Phi}_n\boldsymbol{\Phi}_n^T = \mathbf{A}\mathbf{A}^T \rightarrow (N^2 \times N^2 matrix) \tag{5.9}$$

Where

$$\mathbf{A} = [\mathbf{\Phi}_1 \mathbf{\Phi}_2 .... \mathbf{\Phi}_M] \rightarrow (N^2 \times Mmatrix) \tag{5.10}$$

step6: Compute the eigenvectors $\mathbf{u}_i$ of $\mathbf{AA}^T$

The matrix $\mathbf{AA}^T$ is not practical because it is very large!

step6.1: Consider the matrix $\mathbf{A}^T\mathbf{A}$ ($M \times M$ matrix)

step6.2: Compute the eigenvectors $\mathbf{V}_i$ of $\mathbf{A}^T\mathbf{A}$

$$\mathbf{A}^T\mathbf{A}\mathbf{v}_i = \mu_i \mathbf{v}_i \tag{5.11}$$

Now what is the relation between $\mathbf{u}_i$and $\mathbf{v}_i$?

$\mathbf{A}^T\mathbf{A}\mathbf{v}_i = \mathbf{m}\mathbf{u}_i\mathbf{v}_i \rightarrow \mathbf{A}\mathbf{A}^T\mathbf{A}\mathbf{v}_i = \mu_i\mathbf{A}\mathbf{v}_i \rightarrow \mathbf{C}\mathbf{A}\mathbf{v}_i = \mu_i\mathbf{A}\mathbf{v}_i$ or $\mathbf{C}\mathbf{u}_i = \mu_i\mathbf{u}_i$ where $\mathbf{u}_i = \mathbf{A}\mathbf{v}_i$

Thus, $\mathbf{AA}^T$ and $\mathbf{A}^T\mathbf{A}$ have the same eigenvalues. That means, their eigenvectors are related as: $\mathbf{u}_i = \mathbf{A}\mathbf{v}_i$

step6.3: Compute the M best eigenvectors of $\mathbf{AA}^T$: $\mathbf{u}_i = \mathbf{A}\mathbf{v}_i$

It should be mentioned that $\mathbf{AA}^T$ may have up to N^2 eigenvalues and eigenvectors while $\mathbf{A}^T\mathbf{A}$ may have up to M eigenvalues and eigenvectors

step7: Keep only K eigenvectors which are corresponding to the K largest eigenvalues.

Face Recognition Using Eigenfaces

- Given an unknown face image Γ (centered and of the same size as the training faces)

step1: Normalize

$$\mathbf{\Gamma} : \mathbf{\Phi} = \mathbf{\Gamma} - \mathbf{\Psi} \tag{5.12}$$

step2: Project on the eigenspace

$$\hat{\mathbf{\Phi}} = \sum_{i=1}^{K} \mathbf{w}_i \mathbf{u}_i \quad (\mathbf{w}_i = \mathbf{u}_i^T \mathbf{\Phi}) \tag{5.13}$$

step3: Represent $\mathbf{\Phi}$ as :

$$\mathbf{\Omega} = \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_K \end{bmatrix} \tag{5.14}$$

step4: find

$$e_r = min_l ||\mathbf{\Omega} - \mathbf{\Omega}^l|| \tag{5.15}$$

step5: if $e_r < T_r$, then Φ is recognized as face l from the training set.

The distance e_r is called distance within the face space (difs) [214]

5.5 Linear Discriminant Analysis Algorithm (LDA)

LDA, sometimes termed Fisher Discriminant Analysis (FDA), also makes use of projection into a lower dimensional feature space. The important distinction compared to PCA, which decomposes purely on the basis of variation in the input data, is that LDA also involves the class to which the data belong, meaning that more discriminating information is retained [215]. Consider a situation where the significant variation is due to an external light source: the most significant eigenvector from PCA will encode this variation even though it provides no discriminating information, reducing the effectiveness of correct classification. LDA clusters the same classes tightly together, while different classes are placed as well away from each other [216].

As is mentioned before, PCA approach projects faces onto a lower dimensional sub-space with no distinc-

tion between inter- and intra-class variabilities. Consequently, it is optimal for representation but not for discrimination. While LDA find a subspace which maximizes the ratio of inter-class and intra-class variability. Therefore, it is more efficient for discrimination between data. Suppose that each sample belongs to one of the classes $\{Ci\}_{i=1}^{L}$. $[w_1, w_2, \therefore, w_n]$ is a set of eigenvectors of the sample covariance matrix related to the n largest eigenvalues [217]. The objective function is as follows:

$$\max_{w} \frac{\mathbf{W}^T \mathbf{S_b} \mathbf{W}}{\mathbf{W}^T \mathbf{S_w} \mathbf{W}} \tag{5.16}$$

Where,

$$\mathbf{S_w} = \sum_{j=1}^{L} \sum_{x_i \in C_j} (\mathbf{x}_i - \mathbf{m}_j)(\mathbf{x}_i - \mathbf{m}_j)^T \tag{5.17}$$

$$\mathbf{S_b} = \sum_{i=1}^{L} \mathbf{M}_i (\mathbf{m}_i - m)(\mathbf{m}_i - m)^T \tag{5.18}$$

where m_i denotes the mean of samples in C_i class, M_i is the number of samples in this class, and m is the mean of all the samples. $\mathbf{S_w}$ and $\mathbf{S_b}$ are the within-class and between-class scatter matrices of the training samples. thus, the goal of LDA is to find a projection axis that maximizes the between-class scatter and at the same time minimizes the within-class scatter.

Comparison between Eigen and Fisher Algorithms Eigenfaces and Fisherface are global methods that used as a preprocessing step for pattern recognition and machine learning problems. The desired outcome of both techniques is to reduce the dimension of the dataset with minimal information loss [218]. There are some similarities between the two techniques; both make use of projection of image into a feature space with a difference in the method used for projection and extract common features of the training images. The Eigenface method employs Principal Component Analysis to linearly project the image space into a low dimensional feature space. While, the Fisherface method employs Fisher Linear Discriminant Analysis (LDA) which is more efficient for the classification of different classes [219].

The Fisherface method maximizes the ratio of between-class scatter to that of within-class scatter. Consequently, it works better than PCA for purpose of discrimination. LDA is more effective especially when facial images have major variations in facial expression and illumination. With a large dataset that having multiple classes, it's better to use LDA because class separability is an important factor in that case [216].

5.6 Support Vector Machines (SVMs)

SVMs is a supervised learning technique with associated learning algorithms that analyse data that used for classification and regression analysis. Recently, SVMs has received a growing interest in the area of classification due to providing efficient and powerful classification algorithms which have the ability to deal with high-dimensional input features [220]. Originally, SVMs was formulated for linear two-class classification with margin, where margin refers to the minimal distance between the separating hyperplane and the closest data points. In other words, SVMs learning machine is defined by an optimal separating hyperplane, where the margin between the positive and negative examples is maximal [221]. The important feature of this technique is that the solution is based on those data points, which are located at the margin. These points are called support vectors. See Figure 5.5. The SVMs machine has the ability to find a solution for problems that can not be linearly separated by applying a non-linear transformation for the original input space into a high dimensional feature space, which means that a maximal margin classifier (an optimal separating hyperplane) with respect to the training dataset can be founded.

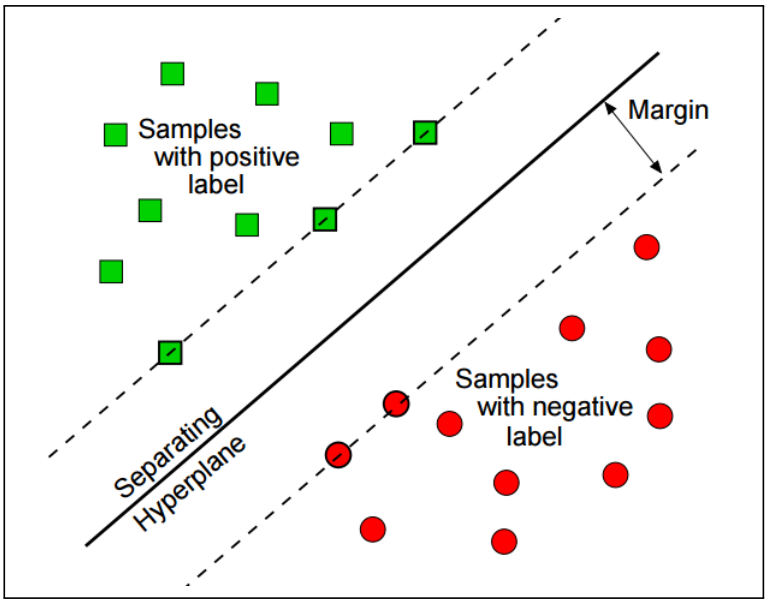


Figure 5.5: The optimal separating hyperplane [222]

5.7 The Experimental Work

5.7.1 Similarity Measures and The Performance of Biometric Systems

Distance similarity measures are core components used by distance-based classification algorithms. Since the Euclidean distance function is the most widely used distance metric in PCA and LDA recognition systems, no study examines the classification performance of these two methods by using different distance functions, especially for biometric authentication domain problems. The aim of this section is to investigate whether the distance function can affect the PCA and LDA performance over different biometrics datasets. This project experiment help other researchers to identify suitable distance measures for datasets. Ideally, the performance of a system should be independent of the choice of the similarity measure and this work ascertains whether this is the case. A statistical test, McNemar's test, is used for identifying performance differences.

Establishing the Eigenfaces Basis

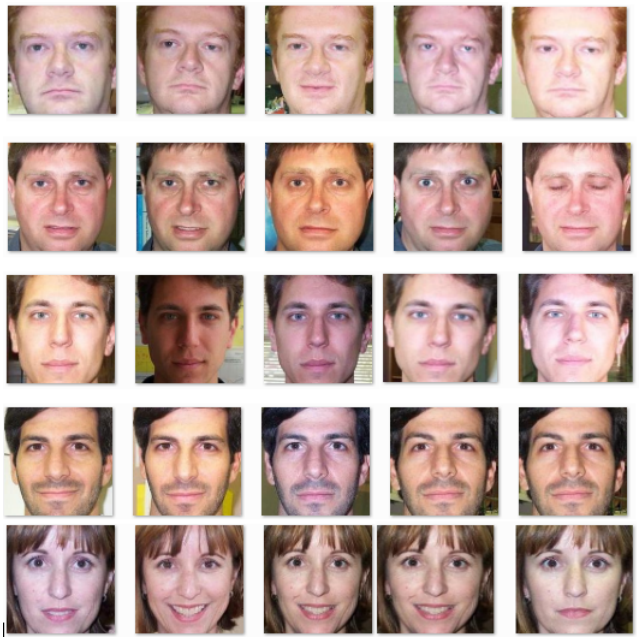
Steps

1. First of all, we have to obtain a training set of M face images $\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_M$. They should be: face-wise aligned, with eyes in the same level and faces of the same size, normalized so that every pixel has a value between 0 and 255 , and of the same $N \times N$ size.

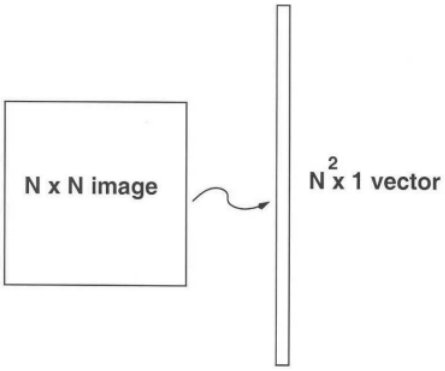
   we want to obtain a set $\{\mathbf{I}_1, \mathbf{I}_2, ..., \mathbf{I}_M\}$, where

$$\mathbf{I}_i = \begin{bmatrix} a^i_{1,1} & a^i_{1,2} & \cdots & a^i_{1,N} \\ a^i_{2,1} & a^i_{2,2} & \cdots & a^i_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ a^i_{N,1} & a^i_{N,2} & \cdots & a^i_{N,N} \end{bmatrix}_{NXN}$$

   In our example M = 25 as shown below:



2. Once we have that, we should change the representation of a face image $\mathbf{I}_i$ from a $N \times N$ matrix, to a Γ_i point in N^2—dimensional space. We concatenate all the rows of the matrix $\mathbf{I}_i$ into one big vector of dimension N^2.



$$\mathbf{\Gamma}_i = \begin{bmatrix} p^i_{1,1} \\ p^i_{1,2} \\ \vdots \\ p^i_{1,N} \\ p^i_{2,1} \\ p^i_{2,2} \\ \vdots \\ p^i_{2,N} \\ \vdots \\ p^i_{N,1} \\ p^i_{N,2} \\ \vdots \\ p^i_{N,N} \end{bmatrix}_{N^2X1}$$

where $i = 1, 2, ..., M$

3. Since we are much more interested in the characteristic features of those faces, we subtract everything that is common between them (the average face $\mathbf{\Psi}$). Figure 5.6 shows The average face of the images. The average face of the images can be defined as

$$\mathbf{\Psi} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{\Gamma}_i$$

Figure 5.6: The average face of the images

then each face differs from the average by the vector

$$\mathbf{\Phi}_i = \mathbf{\Gamma}_i - \mathbf{\Psi}$$

4. Now we should attempt to find a set of M orthonormal vectors which best describe the distribution of our data. Firstly, we need to compute the covariance matrix $\mathbf{C}$:
$\mathbf{C} = \frac{1}{M}\sum_{n=1}^{M} \mathbf{\Phi}_n \mathbf{\Phi}_n^T = \mathbf{A}\mathbf{A}^T$, where $\mathbf{A} = [\mathbf{\Phi}_1 \mathbf{\Phi}_2 .... \mathbf{\Phi}_M]$. Then find the eigenvectors $\mathbf{u}_k$ and eigenvalues λ_k of $\mathbf{C}$.

5. Compute the M best eigenvectors of $\mathbf{A}\mathbf{A}^T$: $\mathbf{u}_i = \mathbf{A}\mathbf{v}_i$ by computing the eigenvectors of $\mathbf{A}^T\mathbf{A}$. keep only K eigenvectors (corresponding to the K largest eigenvalues). Figure 5.7 shows the first 20 Eigenfaces. See chapter 2 for more detail.

6. The features extracted by the Eigenfaces algorithm can be reconstructed back into an image and display it. We will need to normalise the image to ensure it displays correctly as the reconstruction might give pixel values bigger than 1 or smaller than 0. Figure 5.8 shows the Eigenfaces reconstruction images.

7. Once the eigenfaces are created, a new face image $\mathbf{\Gamma}$ can be transformed into it's eigenfaces components by a simple operation: $\mathbf{\Phi} = \mathbf{\Omega} = \begin{bmatrix} w_1 \\ w_2 \\ \cdots \\ w_R \end{bmatrix}_{R\times 1}$, where $R \leq M$
The weights $w_i \in \Omega$ describe the contribution of each eigenface in representing the input face

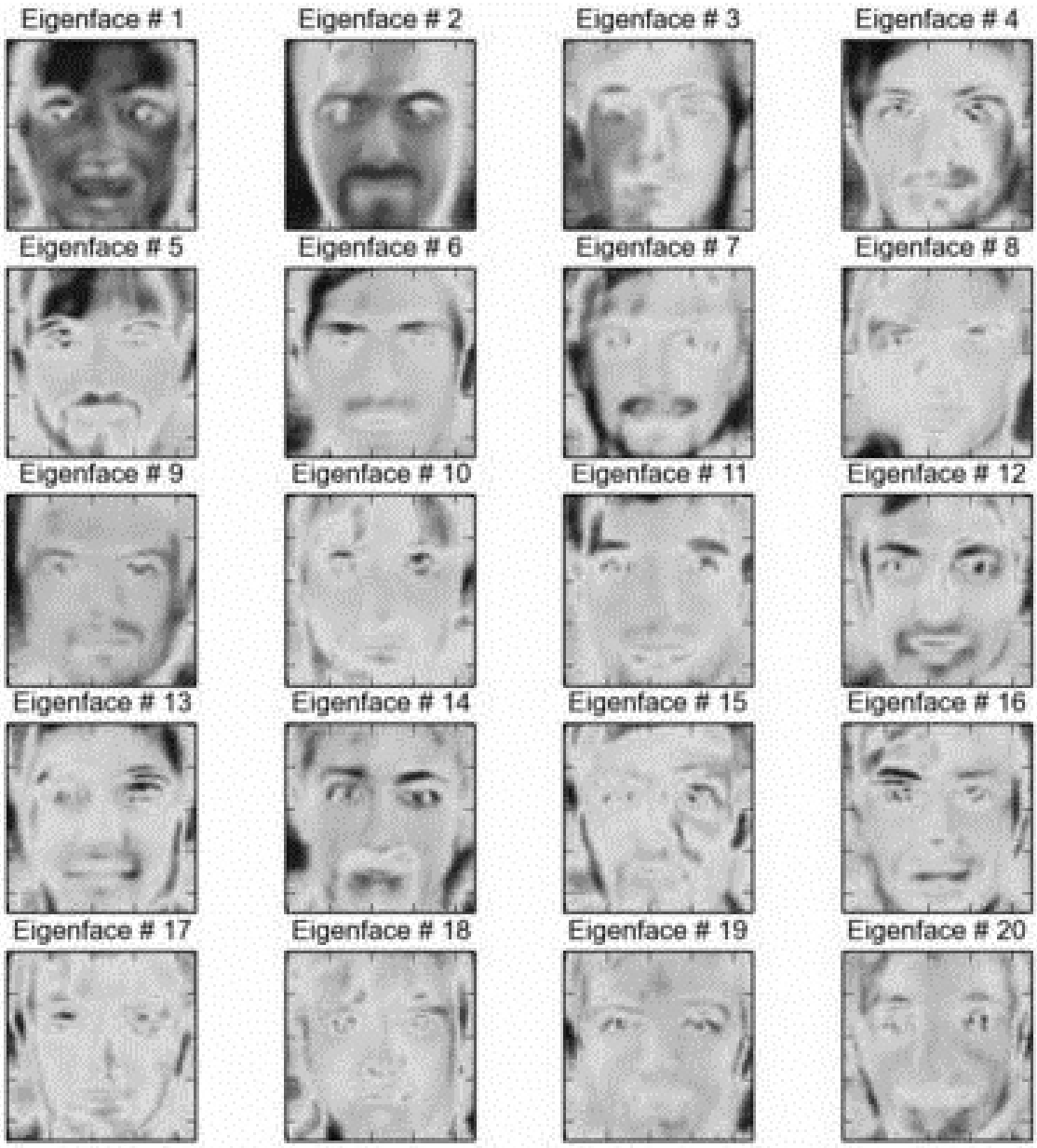


Figure 5.7: The first 20 Eigenfaces

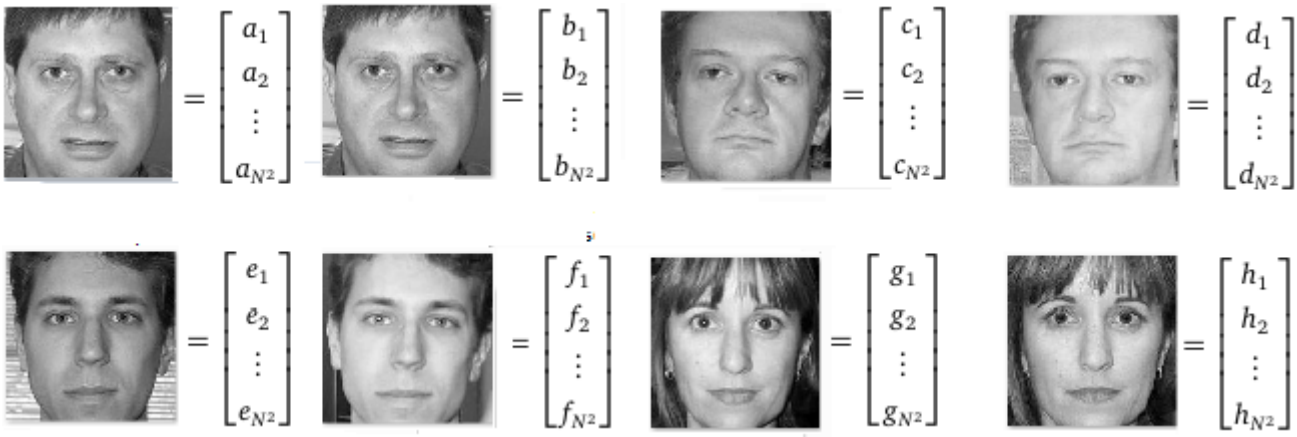
Figure 5.8: Eigenfaces reconstruction images

image. We can use this vector for face recognition by finding the smallest distance ϵ_{rec} between the input face and training faces weight vectors, i.e. by calculating $\epsilon_{rec} = \min ||\Omega - \Omega_i||$, where Ω_i represent the input unknown image. If $\epsilon_{rec} \prec \theta_{rec}$, where θ_{rec} is a threshold chosen heuristically, then we can say that the input image is recognized as the image with which it gives the lowest score.

Establishing the Fisherface Basis

Steps

1. As it mentioned before, we need to obtain a training set of M face images. We should change the representation of a face image $\mathbf{I}_i$ from a $N \times N$ matrix, to a $\mathbf{\Gamma}_i$ point in N^2—dimensional space.



2. Compute the average of all faces:

$$\vec{M} = \frac{1}{M} \begin{bmatrix} a_1 & +b_1 & +\cdots & +h_1 \\ a_2 & +b_2 & +\cdots & +h_2 \\ \vdots & \vdots & & \vdots \\ a_{N^2} & +b_{N^2} & +\cdots & +h_{N^2} \end{bmatrix}$$

In this example M=8

3. Compute the average face for each person:

$$\vec{x} = \frac{1}{2}\begin{bmatrix} a_1 & + & b_1 \\ a_2 & + & b_2 \\ \vdots & & \vdots \\ a_{N^2} & + & b_{N^2} \end{bmatrix} \qquad \vec{y} = \frac{1}{2}\begin{bmatrix} c_1 & + & d_1 \\ c_2 & + & d_2 \\ \vdots & & \vdots \\ c_{N^2} & + & d_{N^2} \end{bmatrix}$$

$$\vec{z} = \frac{1}{2}\begin{bmatrix} e_1 & + & f_1 \\ e_2 & + & f_2 \\ \vdots & & \vdots \\ e_{N^2} & + & f_{N^2} \end{bmatrix} \qquad \vec{w} = \frac{1}{2}\begin{bmatrix} g_1 & + & h_1 \\ g_2 & + & h_2 \\ \vdots & & \vdots \\ g_{N^2} & + & h_{N^2} \end{bmatrix}$$

4. Subtract them from the training faces:

$$\vec{a}_m = \begin{bmatrix} a_1 & - & x_1 \\ a_2 & - & x_2 \\ \vdots & & \vdots \\ a_{N^2} & - & x_{N^2} \end{bmatrix}, \vec{b}_m = \begin{bmatrix} b_1 & - & x_1 \\ b_2 & - & x_2 \\ \vdots & & \vdots \\ b_{N^2} & - & x_{N^2} \end{bmatrix}, \vec{c}_m = \begin{bmatrix} c_1 & - & y_1 \\ c_2 & - & y_2 \\ \vdots & & \vdots \\ c_{N^2} & - & y_{N^2} \end{bmatrix}, \vec{d}_m = \begin{bmatrix} d_1 & - & y_1 \\ d_2 & - & y_2 \\ \vdots & & \vdots \\ d_{N^2} & - & y_{N^2} \end{bmatrix}$$

$$\vec{e}_m = \begin{bmatrix} e_1 & - & z_1 \\ e_2 & - & z_2 \\ \vdots & & \vdots \\ e_{N^2} & - & z_{N^2} \end{bmatrix}, \vec{f}_m = \begin{bmatrix} f_1 & - & z_1 \\ f_2 & - & z_2 \\ \vdots & & \vdots \\ f_{N^2} & - & z_{N^2} \end{bmatrix}, \vec{g}_m = \begin{bmatrix} g_1 & - & w_1 \\ g_2 & - & w_2 \\ \vdots & & \vdots \\ g_{N^2} & - & w_{N^2} \end{bmatrix}, \vec{h}_m = \begin{bmatrix} h_1 & - & w_1 \\ h_2 & - & w_2 \\ \vdots & & \vdots \\ h_{N^2} & - & w_{N^2} \end{bmatrix}$$

5. We build scatter matrices S_1, S_2, S_3, S_4

$$\mathbf{S}_1 = (\vec{a}_m \vec{a}_m^T + \vec{b}_m \vec{b}_m^T), \mathbf{S}_2 = (\vec{c}_m \vec{c}_m^T + \vec{d}_m \vec{d}_m^T)$$

$$\mathbf{S}_3 = (\vec{e}_m \vec{e}_m^T + \vec{f}_m \vec{f}_m^T), \mathbf{S}_4 = (\vec{g}_m \vec{g}_m^T + \vec{h}_m \vec{h}_m^T)$$

And then calculate the within-class scatter matrix S_w

$$\mathbf{S_W} = \mathbf{S}_1 + \mathbf{S}_2 + \mathbf{S}_3 + \mathbf{S}_4$$

Then, calculate the between class-scatter matrix

$$\mathbf{S_B} = 2(\vec{x}-\vec{m})(\vec{x}-\vec{m})^T + 2(\vec{y}-\vec{m})(\vec{y}-\vec{m})^T + 2(\vec{z}-\vec{m})(\vec{z}-\vec{m})^T + 2(\vec{w}-\vec{m})(\vec{w}-\vec{m})^T$$

6. Find orthonormal vectors (W) that best account distribution of the face images within the entire space.

$$\mathbf{W_{opt}} = argmax_w \frac{\mathbf{W}^T\mathbf{S_B}\mathbf{W}}{\mathbf{W}^T\mathbf{S_W}\mathbf{W}}$$

$$= [w_1, w_2, ..., w_m]$$

7. If $\mathbf{S_W}$ is non-singular ($M \geq N^2$) Then, the columns of $\mathbf{W}$: $[w_1, w_2, ..., w_m]$ are eigenvectors of $\mathbf{S_W}^{-1}\mathbf{S_B}$

   i. Compute $\mathbf{S_W}^{-1}$,

   ii. Multiply the matrices $\mathbf{S_W}^{-1}\mathbf{S_B}$

   iii. Compute the eigenvectors $[\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m]$

   If $\mathbf{S_W}$ is singular ($M < N^2$):

   i. Apply PCA first to reduce the dimension of faces from N^2 to M

$$\mathbf{W_{pca}} = argmax_W |\mathbf{W}^T\mathbf{S_T}\mathbf{W}|$$

   Where, $\mathbf{S_T}$(Total Scatter Matrix) $= \sum_{N=1}^{N} (\mathbf{x}_k - \vec{M})(\mathbf{x}_k - \vec{M})^T$

   ii. Apply FLD on the output

$$\mathbf{W_{fld}} = argmax_w \frac{|\mathbf{W}^T\mathbf{w_{pca}}^T\mathbf{S_B}\mathbf{w_{pca}}\mathbf{W}|}{|\mathbf{W}^T\mathbf{w_{pca}}^T\mathbf{S_W}\mathbf{w_{pca}}\mathbf{W}|}$$

iii. Compute eigenvectors

$$\mathbf{w_{opt}} = \mathbf{w_{fld}}\mathbf{w_{pca}} = [w_1, w_2, ..., w_m]$$

Figures 5.9 5.10 show an example of our fisherfaces and its reconstruction images, respectively:

8. Project eigenvectors (faces) onto the LDA subspace as described:

$$\vec{X}_{LDA} = W^T\vec{X} \quad , \quad \vec{Y}_{LDA} = W^T\vec{Y}$$

$$\vec{Z}_{LDA} = W^T\vec{Z} \quad , \quad \vec{w}_{LDA} = W^T\vec{w}$$

9. Apply a nearest-neighbour classifier to determine the similarity between the query input face and the database faces.

Face database: For this work, Caltech Faces Database (Front images) has been used [223]. Some 433 labelled face images were used for training and 100 for testing. Figure 5.11 represents face recognition system based on PCA and LDA respectively.

Ear database: IIT Delhi Ear Database has been used in this work [181]. The database has been split into 397 training images and 100 testing images. Figure 5.12 represent ear recognition system based on PCA and LDA, respectively .

Palm database: The palm database consists of 288 images with 50 subjects. Some 288 images were used for training and 100 for testing. Figure 5.13 represent palm recognition system based on PCA and LDA respectively [224] [186].

The results found when running PCA and LDA on face, ear, and palm datasets with different distance measures are mentioned in Table 5.1:

Table 5.2 showing the *Z_value* to the LDA and PCA techniques based on four similarity distance measures for matching. It is interesting to note that in all cases, the fisherfaces technique outperformed the

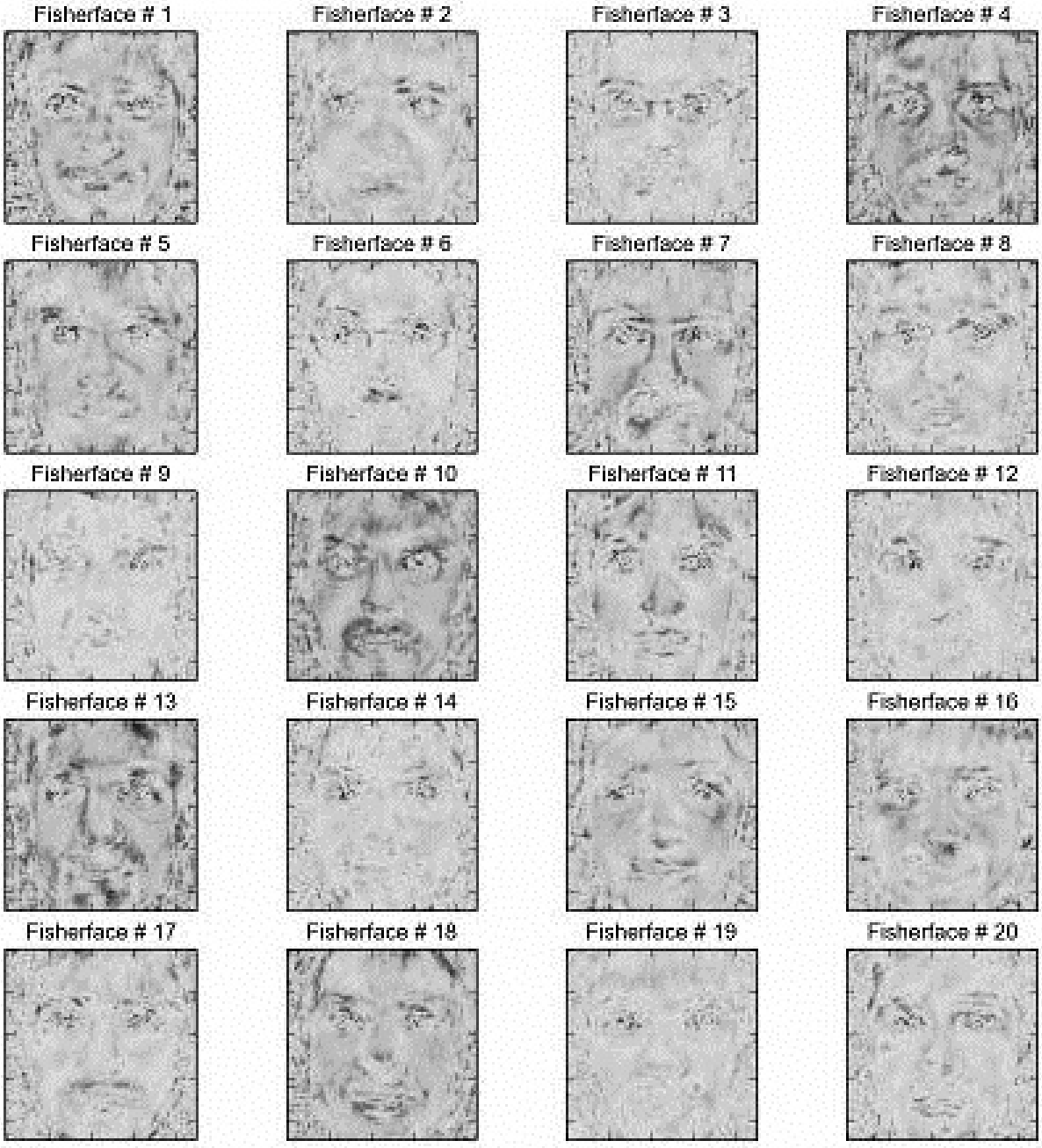


Figure 5.9: The first 20 Fisherfaces

Figure 5.10: Fisherfaces reconstruction images

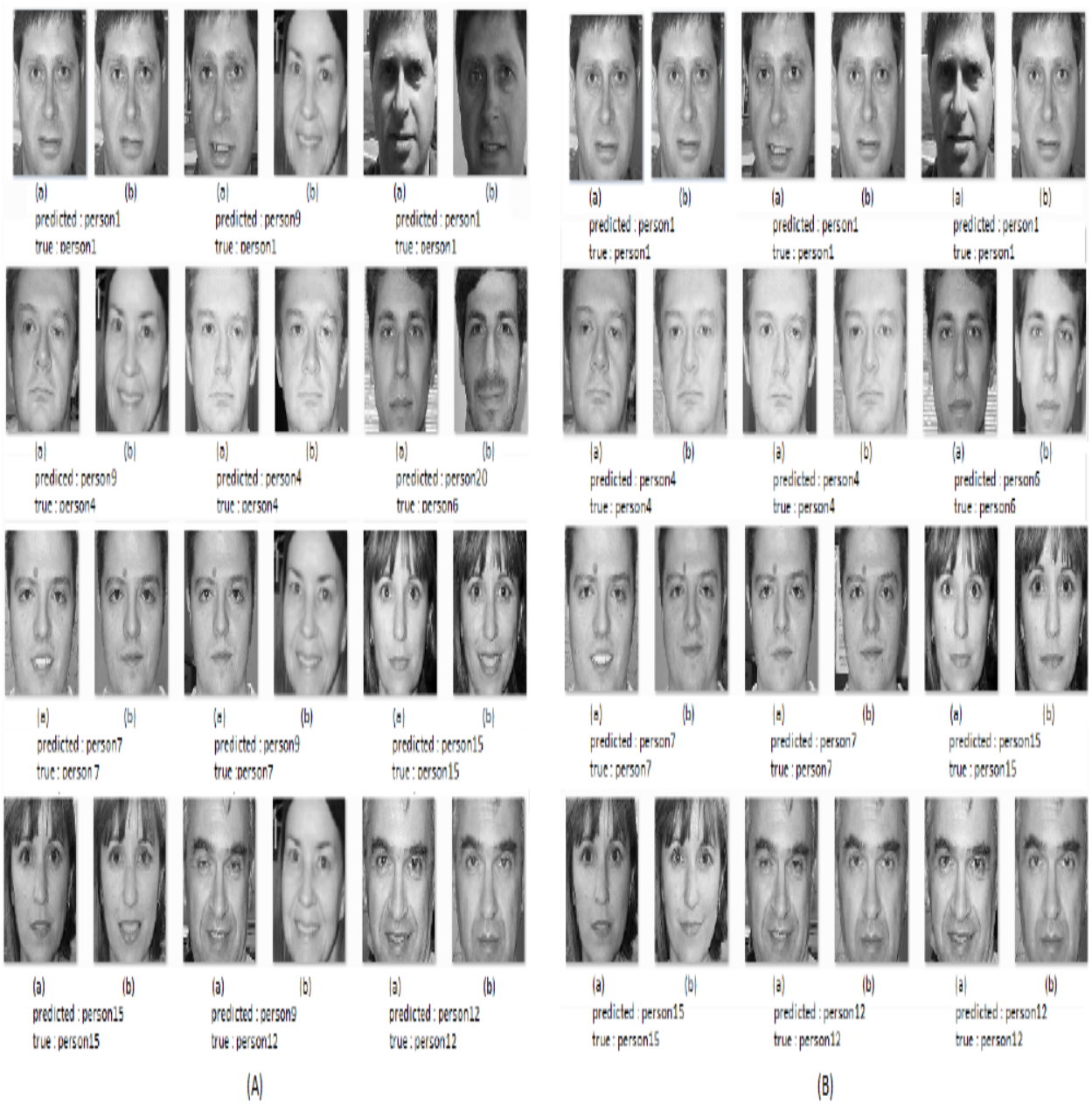


Figure 5.11: The output of face recognition using PCA and LDA. (a) shows probe images and (b) the closest match in the database using Euclidean distance.

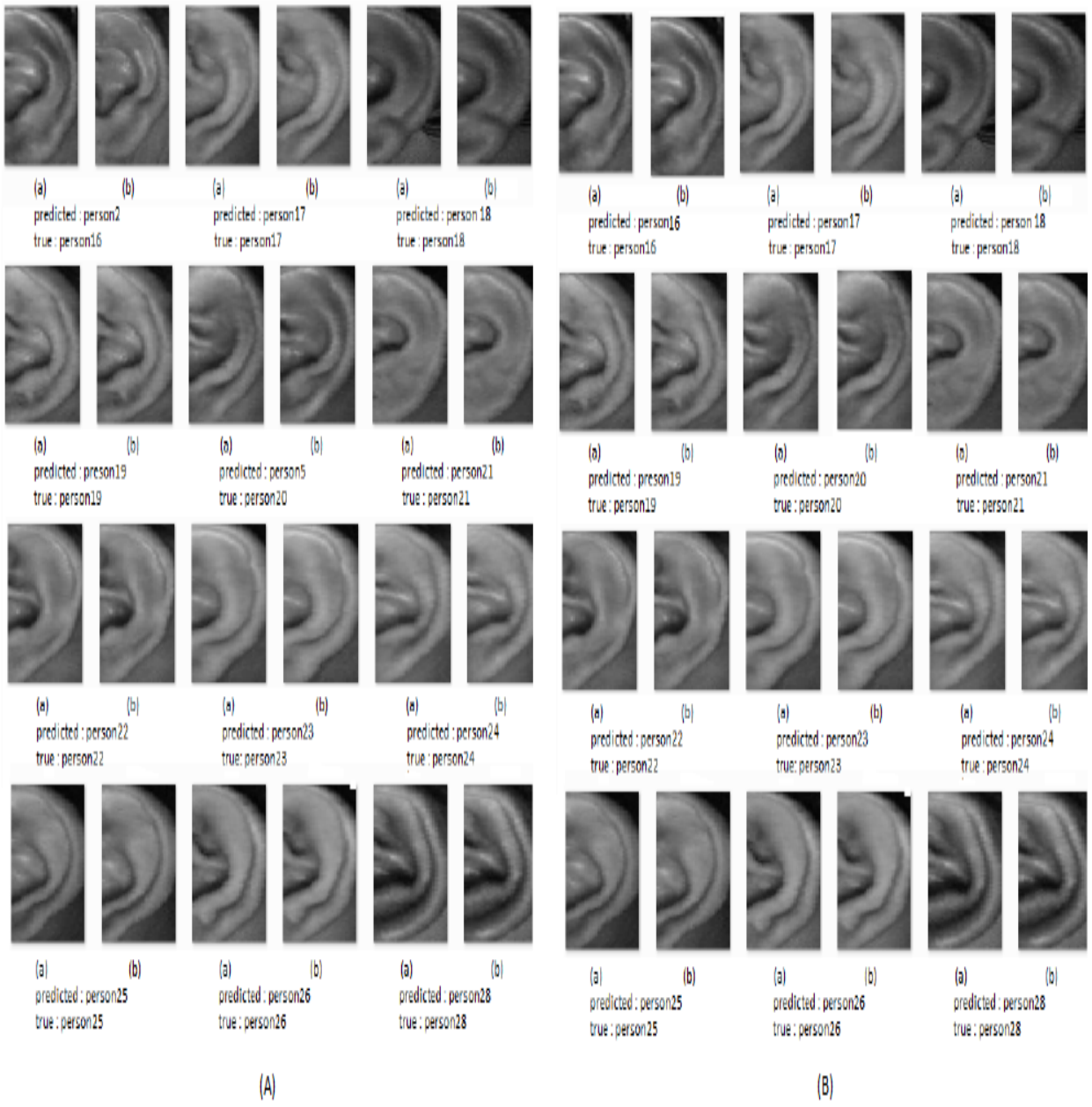


Figure 5.12: Ear recognition using PCA. (a) shows probe images and (b) the closest match in the database using Euclidean distance.

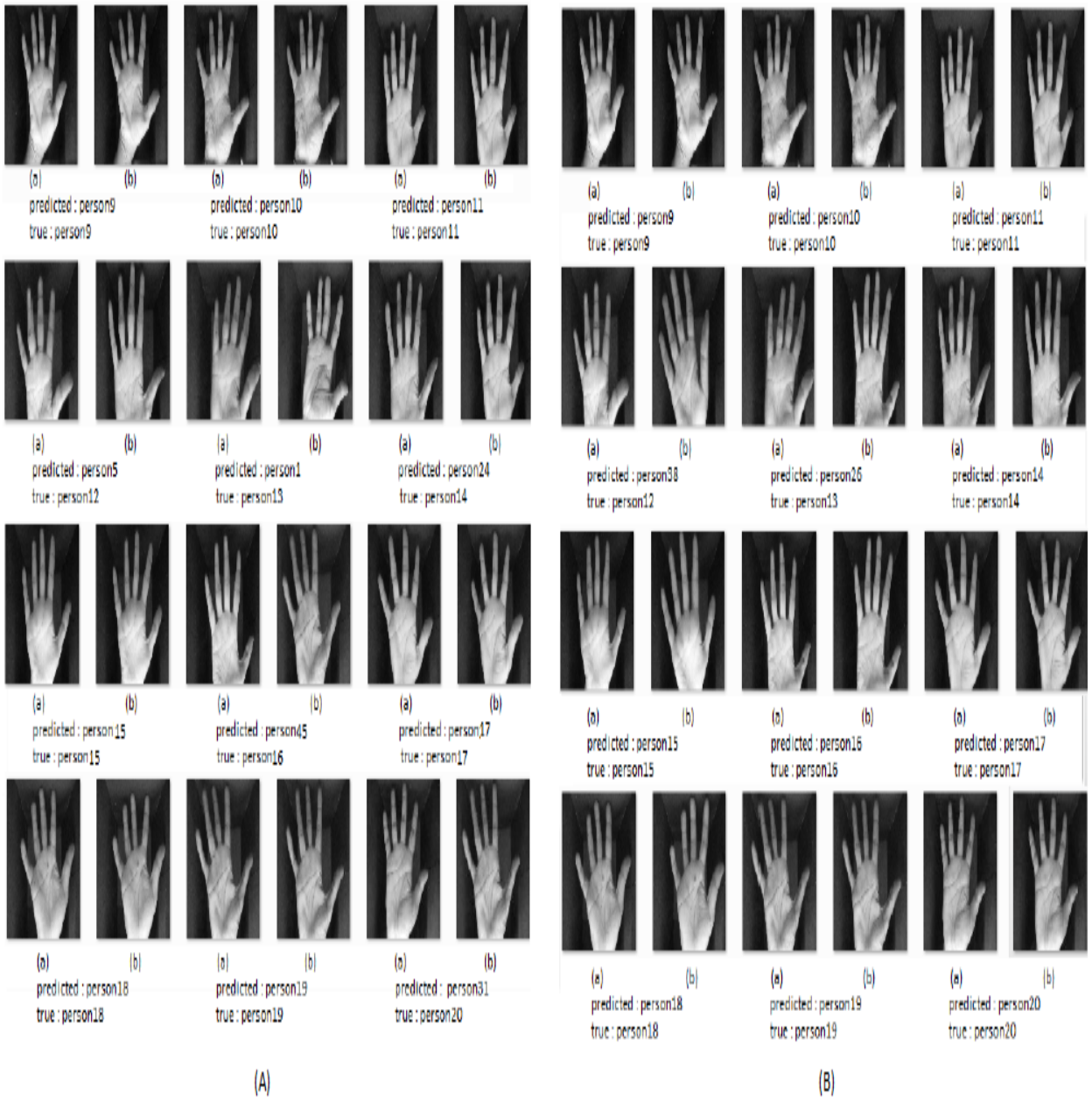


Figure 5.13: Palm recognition using PCA. (a) shows probe images and (b) the closest match in the database using Euclidean distance.

eigenfaces technique regardless of the similarity distance measure used for classification

PCA algorithm

LDA algorithm	Biometric	Euclidean		Manhattan		Mahalanobis		Cosine	
		N_{sf}	N_{fs}	N_{sf}	N_{fs}	N_{sf}	N_{fs}	N_{sf}	N_{fs}
	Face	37	**1**	19	**0**	58	**0**	37	**3**
	Ear	59	**0**	12	**3**	46	**4**	41	**1**
	Palm	28	**4**	22	**4**	80	**0**	36	**0**

Table 5.1: Counts of the discrepancies between the LDA and PCA outputs using the various distance measures

PCA algorithm

LDA algorithm	Biometric	Euclidean	Manhattan	Mahalanobis	Cosine
	Face	5.678	4.130	7.485	5.218
	Ear	7.551	2.066	5.798	6.018
	Palm	4.066	3.333	8.832	5.833

Table 5.2: The *Z_value* between the LDA and PCA techniques.

Tables 5.3, 5.4, 5.5, 5.6, 5.7 and 5.8 illustrate the 'True/False' matching and *z_value* (3.3 and 3.4 for face, ear and palm recognition based on PCA and LDA using different distance similarity measures. It is clear that using the eigenfaces technique based on the Manhattan similarity distance classifier outperformed the other distance classifier in the majority of cases. No significant differences were found between the mentioned classifiers when fisherfaces has been used for recognition.

Face Recognition based on PCA

Algorithm	Manhattan			Mahalanobis			Cosine		
	N_{sf}	N_{fs}	z_value	N_{sf}	N_{fs}	z_value	N_{sf}	N_{fs}	z_value
Euclidean	2	19	**3.710**	43	23	**2.338**	21	19	**0.158**
Manhattan				51	13	**4.625**	46	10	**4.677**
Mahalanobis							19	37	**2.271**

Table 5.3: The 'Trure/False' matching and *z_value* for face recognition based on PCA and distance similarity measures.

Face Recognition based on LDA

Algorithm	Manhattan			Mahalanobis			Cosine		
	N_{sf}	N_{fs}	z_value	N_{sf}	N_{fs}	z_value	N_{sf}	N_{fs}	z_value
Euclidean	0	1	**0**	0	0	**0**	4	0	**1.5**
Manhattan				0	0	**0**	5	0	**1.788**
Mahalanobis							5	0	**1.788**

Table 5.4: The 'Trure/False' matching and *z_value* for face recognition based on LDA and distance similarity measures.

Ear Recognition based on PCA

	Manhattan			Mahalanobis			Cosine		
Algorithm	N_{sf}	N_{fs}	z_value	N_{sf}	N_{fs}	z_value	N_{sf}	N_{fs}	z_value
Euclidean	2	51	**6.593**	18	21	**0.320**	10	33	**3.355**
Manhattan				49	2	**6.441**	28	3	**4.311**
Mahalanobis							7	29	**3.5**

Table 5.5: The 'Trure/False' matching and *z_value* for ear recognition based on PCA and distance similarity measures.

Ear Recognition based on LDA

	Manhattan			Mahalanobis			Cosine		
Algorithm	N_{sf}	N_{fs}	z_value	N_{sf}	N_{fs}	z_value	N_{sf}	N_{fs}	z_value
Euclidean	2	0	**0.707**	16	0	**3.75**	4	2	**0.408**
Manhattan				14	0	**3.475**	2	6	**1.06**
Mahalanobis							0	18	**4.008**

Table 5.6: The 'Trure/False' matching and *z_value* for ear recognition based on LDA and distance similarity measures.

Palm Recognition based on PCA

	Manhattan			Mahalanobis			Cosine		
Algorithm	N_{sf}	N_{fs}	z_value	N_{sf}	N_{fs}	z_value	N_{sf}	N_{fs}	z_value
Euclidean	0	3	**1.154**	27	1	**4.725**	4	1	**0.894**
Manhattan				30	1	**5.029**	7	1	**1.768**
Mahalanobis							1	24	**4.4**

Table 5.7: The 'Trure/False' matching and *z_value* for palm recognition based on PCA and distance similarity measures.

Palm Recognition based on LDA

	Manhattan			Mahalanobis			Cosine		
Algorithm	N_{sf}	N_{fs}	z_value	N_{sf}	N_{fs}	z_value	N_{sf}	N_{fs}	z_value
Euclidean	0	1	**0**	0	3	**1.154**	0	3	**1.154**
Manhattan				0	2	**0.707**	0	3	**1.154**
Mahalanobis							0	1	**0**

Table 5.8: The'Trure/False' matching and *z_value* for palm recognition based on LDA and distance similarity measures.

5.7.2 Biometric Recognition System Based on SVM_pca and SVM_pca,lda Techniques

PCA is one of the most common feature extraction and data representation technique which is used mainly for dimensionality reduction in the fields of pattern recognition, computer vision and etc. However, this method is usually affected by light illumination [225]. PCA is based on linear projection of

an image space to a low dimension feature space which represents the eigen vectors of covariance matrix that corresponds to the direction of Principal Components of the original data [220]. Meanwhile, LDA is another popular dimensionality reduction technique which based on projecting the data onto a lower-dimensional vector space, thus achieving maximum discrimination by maximizing the ratio of the between-class distance to the within-class distance [226].

This experiment deals with the face and ear recognition systems. PCA and a combination of LDA and PCA are used for feature extraction. Then SVMs is implemented for classification to identify the similarity between the probe images from the same individual and different individuals after all the images in the database are represented with relevant features. The experimental results show that recognition based on SVM_pca has achieved a better performance than the SVM_pca,lda technique. In addition, the recognition performance based on SVM_pca outperformed the recognition performance using PCA based on distance similarity measures. While, No significant differences were found using SVM_pca,lda and LDA based on distance similarity measures. See tables 5.9, 5.10, 5.11 and 5.12.

Algorithm	Face Recognition	
	Accuracy	Error Rate
SVM_pca	0.964	0.035
SVM_pca,lda	0.928	0.071

Table 5.9: Face recognition performance based on SVM_pca and SVM_pca,lda clasifiers.

	Distance Similarity Measures							
Face	Euclidean		Manhattan		Mahalanobis		Cosine	
Algorithm	Accuracy	e_rate	Accuracy	e_rate	Accuracy	e_rate	Accuracy	e_rate
PCA	0.665	0.334	0.812	0.187	0.388	0.61	0.718	0.281
LDA	0.959	0.040	0.964	0.035	0.964	0.035	0.937	0.062

Table 5.10: Face recognition based on PCA and LDA techniques using distance similarity measures.

Algorithm	Ear Recognition	
	Accuracy	Error Rate
SVM_pca	0.75	0.25
SVM_pca,lda	0.74	0.26

Table 5.11: Ear recognition performance based on SVM_pca and SVM_pca,lda clasifiers.

Distance Similarity Measures

Ear	Euclidean		Manhattan		Mahalanobis		Cosine	
Algorithm	Accuracy	e_rate	Accuracy	e_rate	Accuracy	e_rate	Accuracy	e_rate
PCA	0.31	0.69	0.68	0.32	0.32	0.68	0.53	0.47
LDA	0.89	0.11	0.87	0.13	0.73	0.27	0.91	0.09

Table 5.12: Ear recognition based on PCA and LDA techniques using distance similarity measures.

5.8 Conclusions

In this chapter, two experiments have been performed. In the first experiment, four distance similarity measures: Manhattan, Euclidean, Cosine similarity and Mahalanobis distance have been compared for PCA- and LDA-based face, ear, and palm biometrics. The comparative experiment shows that using the LDA technique outperformed the PCA technique in all cases in handling variation in lighting and expression regardless of the distance similarity measure used for classification. While, using the PCA technique based on the Manhattan similarity distance classifier outperformed the other distance classifier in the majority of cases. No significant differences were found between the above mentioned distance measures when LDA has been used for recognition. The experiment shows that LDA works better with large dataset having multiple classes. Whereas, PCA performs better in case where number of samples per class is less.

In the second experiment, the face and ear recognition performance using SVM based on PCA and SVM based on a combination of PCA and LDA techniques have been compared with the PCA and LDA techniques depend on distance similarity measures. From the experimental results we can conclude that recognition based on SVM_pca has achieved a better performance than the SVM_pca,lda technique. In addition, the recognition performance based on SVM_pca outperformed the recognition performance using PCA based on distance similarity measures. While, No significant differences were found using SVM_pca,lda and LDA based on distance similarity measures.

The next chapter will present the relevant fundamentals of convolutional neural networks (CNNs) that used for biometric recognition. Biometric recognition using three deep and three shallow convolutional

neural networks (GoogleNET, VGG16, ResNET50) have been applied to five different databases: face, ear, palmprint, iris, and hand.

Chapter 6

Biometric Recognition Using CNNs

6.1 Introduction

In machine learning, a convolutional neural network (CNN or ConvNets) is a category of feed-forward artificial neural network that has proven highly effective in various areas such as classification and image recognition. One of the first convolutional neural networks (CNNs) was LeNet by Yann LeCun in 1998 which boost the field of deep learning [227] [228]. There have been many new architectures proposed which considered as improvements over the LeNet, but all depend on the base concepts from the LeNet. CNN integrates feature extraction with classification, it receives the input data and produces the final results of classification without any additional process. CNNs can be considered as a robust automatic feature extractor which can handle big training samples and the features are learned automatically [229].

This chapter presents the relevant fundamentals of convolutional neural networks that been used for biometric recognition. Biometric recognition using three deep convolutional neural networks (GoogleNET, VGG16, ResNET50) and shallow convolutional neural networks have been applied to five different databases: face, ear, palmprint, iris, and hand. Caffe is a very widely used framework for deep learning. In this chapter, we explain how to get started with CAFFE and the steps that should be set

6.1.1 The Convolution Steps

There are four main stages which considered as the basic for every Convolutional Network:

- Convolution
- Non Linearity (ReLU)
- Pooling or Sub Sampling
- Classification (Fully Connected Layer)

These operations are the basic building blocks of every Convolutional Neural [230] as shown in Figure 6.1:

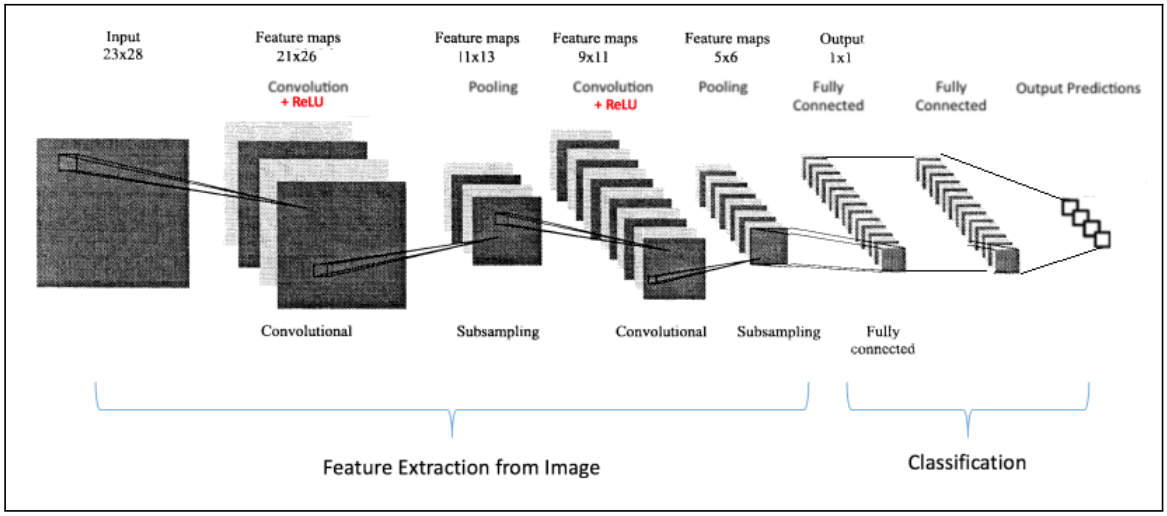


Figure 6.1: A typical convolutional network

ConvNet derives its name from the 'convolution' operator. The main purpose of Convolution in a ConvNet is extracting features from the input image data and preserving the spatial relationship between pixels. This process is achieved by learning image features using small squares of input image data. Every image can be considered as a pixel values matrix. Consider a 5 x 5 image whose pixel values are only 0 and 1. Figure 6.2 shows the convolution operation.

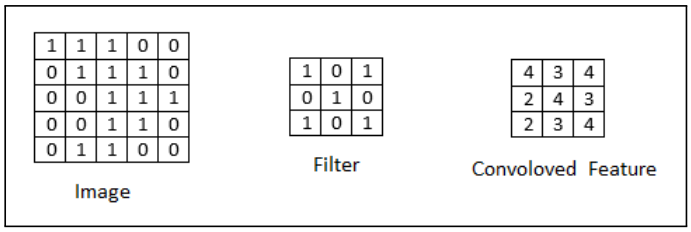


Figure 6.2: The convolution operation

The convolved feature or called 'Feature Map' is calculated, by sliding the filter over the image and computing the dot product. The convolution of another filter, over the same image gives a different feature map. In other words, different filters generate different feature maps from the same original image.

In practice, a CNN learns the values of these filters on its own during the training process. Its still need to specify some parameters such as the number of filters, filter size, architecture of the network) before the training process. In other words, The number of filters means the more image features get extracted and therefore the better network performance at recognizing patterns in unknown images. The size of the Feature Map (Convolved Feature) is controlled by three parameters that we need to decide before the convolution step is performed:

Depth. Refers to the number of filters that used for the convolution operation

stride. Refers to the number of pixels by which the filter matrix been slide over the input matrix. When the stride is 1 then the filters jump one pixel at a time. When the stride is 2, then the filters move 2 pixels at a time. So, having a larger stride will give smaller feature maps.

Zero-padding. Refers to the process of adding zeros around the border of the input matrix. A good feature of zero padding is that it allows to control the size of the feature maps [231] [232].

6.1.2 Introducing Non Linearity (ReLU)

An additional operation called ReLU used after every Convolution operation. ReLU is an operation applied per pixel and replaces all negative pixel values in the feature map by zero. The purpose of the ReLU operation is to introduce non-linearity in ConvNet, since most of the real – world data used with ConvNet to learn would be non-linear (Convolution is a linear operation – element wise matrix multiplication and addition, so it is important to account for non-linearity by introducing a non-linear function like ReLU) [233].

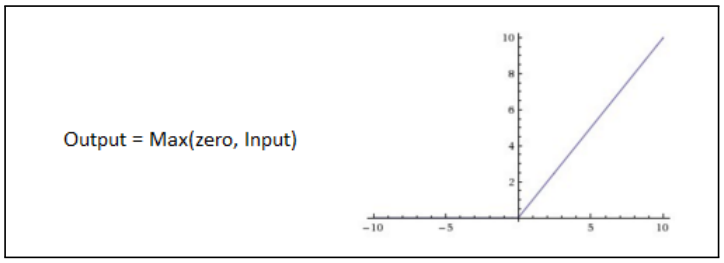


Figure 6.3: the ReLU operation.

6.1.3 The Pooling Step

In addition to convolution operations, pooling operations considered as another main building block in CNNs, it is also called subsampling. Pooling operations aim to reduce the dimensionality of each feature map but reserve the most important information by using some functions to summarize subregion such as Max, Average and Sum function [234] [231]. The main functions of Pooling are:

- makes the input representations smaller and more manageable.
- reduces the number of parameters and computations in the network.
- makes the network invariant to slight distortions and translations in the input image. In other words, a small distortion within the input image will not change the output of Pooling – since the maximum/average value is taken in a local neighborhood.
- helps to arrive at an almost scale invariant representation of the input image, This is very powerful

to detect objects in an image [235].

6.1.4 Fully Connected Layer

The term 'Fully Connected' refers to that every neuron in any layer is connected to every neuron on the adjacent layer. The outcome from the convolutional and pooling layers considered as high-level features of the input image. The main purpose of the Fully Connected layer is to use these high-level features for classifying the input image into different classes based on the training dataset. Most of the features resultant from convolutional and pooling layers may be good for the classification process, but using combinations of those features might be better. As is shown in Figure 6.1, the main purpose of convolution and pooling layers is extracting features from the input image while fully connected layer acting as a classifier [233].

6.1.5 The ConvNets Training Process

The overall training process of the Convolution Network may be summarized as below:

Step 1 Initialize all filters and parameters/weights with random values

Step 2 The network takes a training image as input, applies the forward propagation step (convolution, ReLU and pooling operations along with forwarding propagation in the Fully Connected layer) and finds the output probabilities for each class.

Step 3 Calculate the total error at the output layer:

$$\text{Total Error} = \sum \frac{1}{2}\text{target probability} - \text{output probability} \tag{6.1}$$

Step 4 Use Backpropagation to calculate the gradients of the error with respect to all weights in the network and use gradient descent to update all filter values/weights and parameter values to

minimize the output error.

- The weights are adjusted in proportion to their contribution to the total error.
- This means that the network has learnt to classify a particular image correctly by adjusting its weights/filters such that the output error is reduced.
- Parameters such as the number of filters, filter sizes, architecture of the network have all been fixed before Step 1 and do not change during training process — only the values of the filter matrix and connection weights get updated.

Step 5 Repeat steps $2-4$ with all images in the training set.

The above steps train the ConvNet — this essentially means that all the weights and parameters of the ConvNet have now been optimized to correctly classify images from the training set [236].

6.2 The Shallow CNN Architecture

Shallow-Deep Network (SDN) is a generic modification to deep neural networks (DNNs) for introducing internal classifiers (ICs) to overcome the weakness that occurs when a DNN can reach correct predictions before its final layer. Figure 6.4 shows the Shallow-Deep Networks. Each internal classifier involves of a single fully connected layer that comes after a feature reduction layer. The feature reduction layer picks the large output of a network's internal layers and reduces its size. By using the reduced output, The fully connected layer introduces the internal prediction [242].

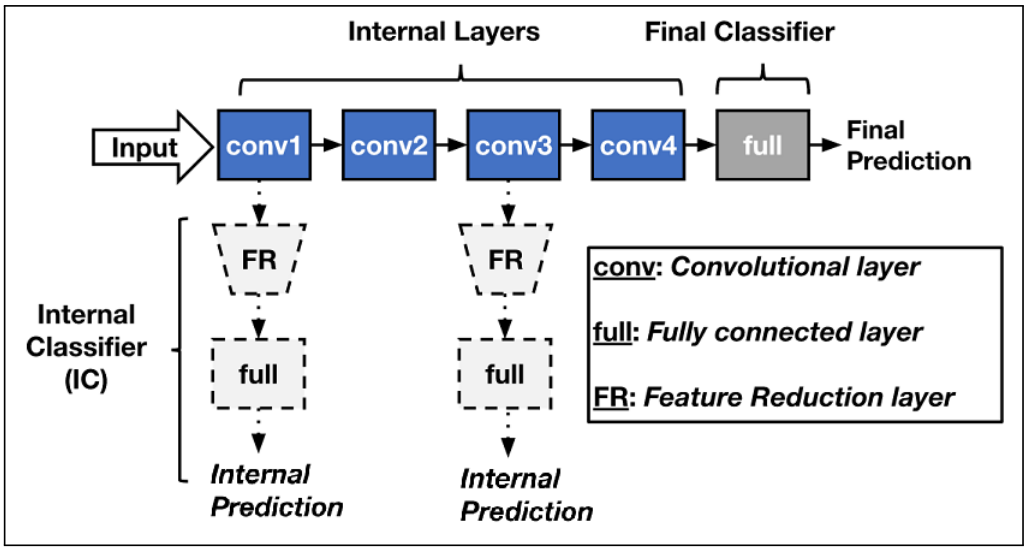


Figure 6.4: The Shallow-Deep Networks [242]

6.3 The CNN Architectures

Many architectures have been proposed to achieve high and efficient image classification and recognition.

6.3.1 VGG16

VGG16 is a convolutional neural network, it is 16 layers deep and has the ability to classify images into 1000 object categories. The network has 41 layers. There are 16 layers with learnable weights: 13 convolutional layers, and 3 fully connected layers. The input to convolution layer is of fixed size 224×224 RGB image. The image is passed through a stack of convolutional layers, where the filters are used with a very small receptive field 3×3. The convolution stride is fixed to 1 pixel. Spatial pooling is carried out by five max-pooling layers and finally Max-pooling is performed over a 2×2 pixel window, with stride 2. Figure 6.5 represents the 16 layers deep of VGG16 network and figure 6.6 represents the VGG16 layers [238].

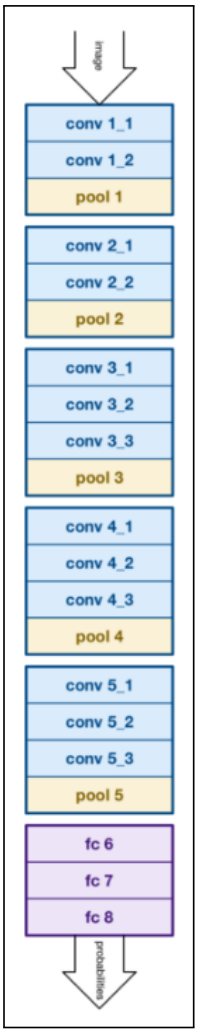


Figure 6.5: The 16 Layers Deep of VGG16 Network

1	'input'	Image Input	224x224x3 images with 'zerocenter' normalization
2	'conv1_1'	Convolution	64 3x3x3 convolutions with stride [1 1] and padding [1 1 1 1]
3	'relu1_1'	ReLU	ReLU
4	'conv1_2'	Convolution	64 3x3x64 convolutions with stride [1 1] and padding [1 1 1 1]
5	'relu1_2'	ReLU	ReLU
6	'pool1'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
7	'conv2_1'	Convolution	128 3x3x64 convolutions with stride [1 1] and padding [1 1 1 1]
8	'relu2_1'	ReLU	ReLU
9	'conv2_2'	Convolution	128 3x3x128 convolutions with stride [1 1] and padding [1 1 1 1]
10	'relu2_2'	ReLU	ReLU
11	'pool2'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
12	'conv3_1'	Convolution	256 3x3x128 convolutions with stride [1 1] and padding [1 1 1 1]
13	'relu3_1'	ReLU	ReLU
14	'conv3_2'	Convolution	256 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
15	'relu3_2'	ReLU	ReLU
16	'conv3_3'	Convolution	256 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
17	'relu3_3'	ReLU	ReLU
18	'pool3'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
19	'conv4_1'	Convolution	512 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
20	'relu4_1'	ReLU	ReLU
21	'conv4_2'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
22	'relu4_2'	ReLU	ReLU
23	'conv4_3'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
24	'relu4_3'	ReLU	ReLU
25	'pool4'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
26	'conv5_1'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
27	'relu5_1'	ReLU	ReLU
28	'conv5_2'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
29	'relu5_2'	ReLU	ReLU
30	'conv5_3'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
31	'relu5_3'	ReLU	ReLU
32	'pool5'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
33	'fc6'	Fully Connected	4096 fully connected layer
34	'relu6'	ReLU	ReLU
35	'drop6'	Dropout	50% dropout
36	'fc7'	Fully Connected	4096 fully connected layer
37	'relu7'	ReLU	ReLU
38	'drop7'	Dropout	50% dropout
39	'fc8'	Fully Connected	1000 fully connected layer
40	'prob'	Softmax	softmax
41	'output'	Classification Output	crossentropyex with 'tench' and 999 other classes

Figure 6.6: The VGG16 Network Layers

The proposed Shallow-VGG16, as shown in 6.7, contains 10 convolutional layers and 3 fully connected layers. In 6.7, the black shaded part represents the deleted layers in Shallow-VGG16

1	'input'	Image Input	224x224x3 images with 'zerocenter' normalization
2	'conv1_1'	Convolution	64 3x3x3 convolutions with stride [1 1] and padding [1 1 1 1]
3	'relu1_1'	ReLU	ReLU
4	'conv1_2'	Convolution	64 3x3x64 convolutions with stride [1 1] and padding [1 1 1 1]
5	'relu1_2'	ReLU	ReLU
6	'pool1'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
7	'conv2_1'	Convolution	128 3x3x64 convolutions with stride [1 1] and padding [1 1 1 1]
8	'relu2_1'	ReLU	ReLU
9	'conv2_2'	Convolution	128 3x3x128 convolutions with stride [1 1] and padding [1 1 1 1]
10	'relu2_2'	ReLU	ReLU
11	'pool2'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
12	'conv3_1'	Convolution	256 3x3x128 convolutions with stride [1 1] and padding [1 1 1 1]
13	'relu3_1'	ReLU	ReLU
14	'conv3_2'	Convolution	256 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
15	'relu3_2'	ReLU	ReLU
16	'conv3_3'	Convolution	256 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
17	'relu3_3'	ReLU	ReLU
18	'pool3'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
19	'conv4_1'	Convolution	512 3x3x256 convolutions with stride [1 1] and padding [1 1 1 1]
20	'relu4_1'	ReLU	ReLU
21	'conv4_2'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
22	'relu4_2'	ReLU	ReLU
23	'conv4_3'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
24	'relu4_3'	ReLU	ReLU
25	'pool4'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
26	'conv5_1'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
27	'relu5_1'	ReLU	ReLU
28	'conv5_2'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
29	'relu5_2'	ReLU	ReLU
30	'conv5_3'	Convolution	512 3x3x512 convolutions with stride [1 1] and padding [1 1 1 1]
31	'relu5_3'	ReLU	ReLU
32	'pool5'	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
33	'fc6'	Fully Connected	4096 fully connected layer
34	'relu6'	ReLU	ReLU
35	'drop6'	Dropout	50% dropout
36	'fc7'	Fully Connected	4096 fully connected layer
37	'relu7'	ReLU	ReLU
38	'drop7'	Dropout	50% dropout
39	'fc8'	Fully Connected	1000 fully connected layer
40	'prob'	Softmax	softmax
41	'output'	Classification Output	crossentropyex with 'tench' and 999 other classes

Figure 6.7: The Shallow VGG16 layers

6.3.2 ResNet50

ResNet50 is a convolutional neural network, the ResNet50 network is 50 layers deep and it has 171 layers. This network can classify images into 1000 object classes. Figure 6.8 shows the ResNet50 network layers.

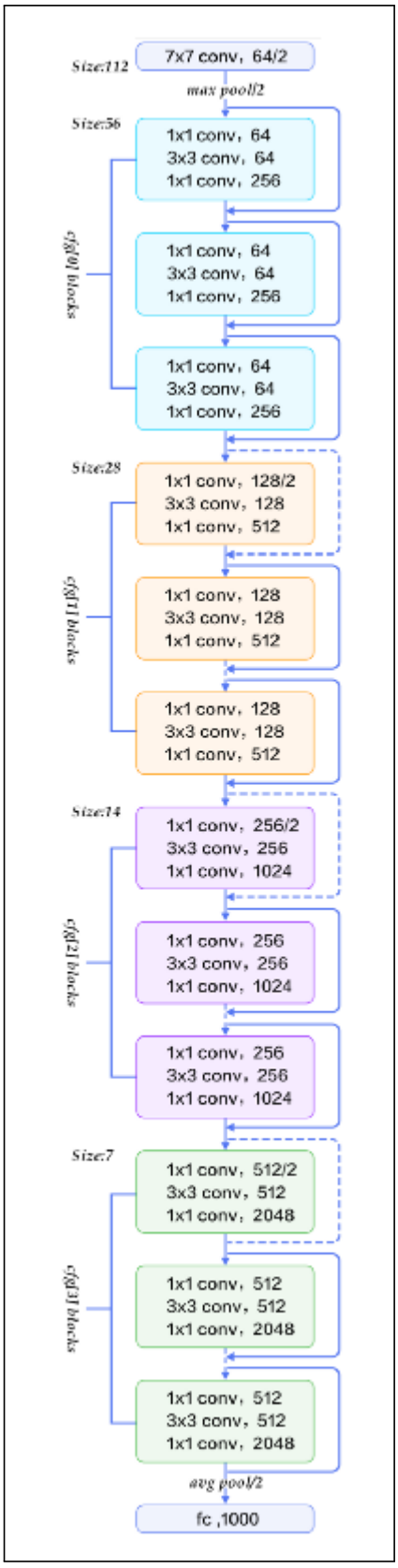


Figure 6.8: The ResNet Network Layers.

6.3.3 GoogleNet

GoogleNet is a convolutional neural network, it is 22 layers deep and it has 144 layers [240]. Figure 6.9 shows the GoogleNet network layers.

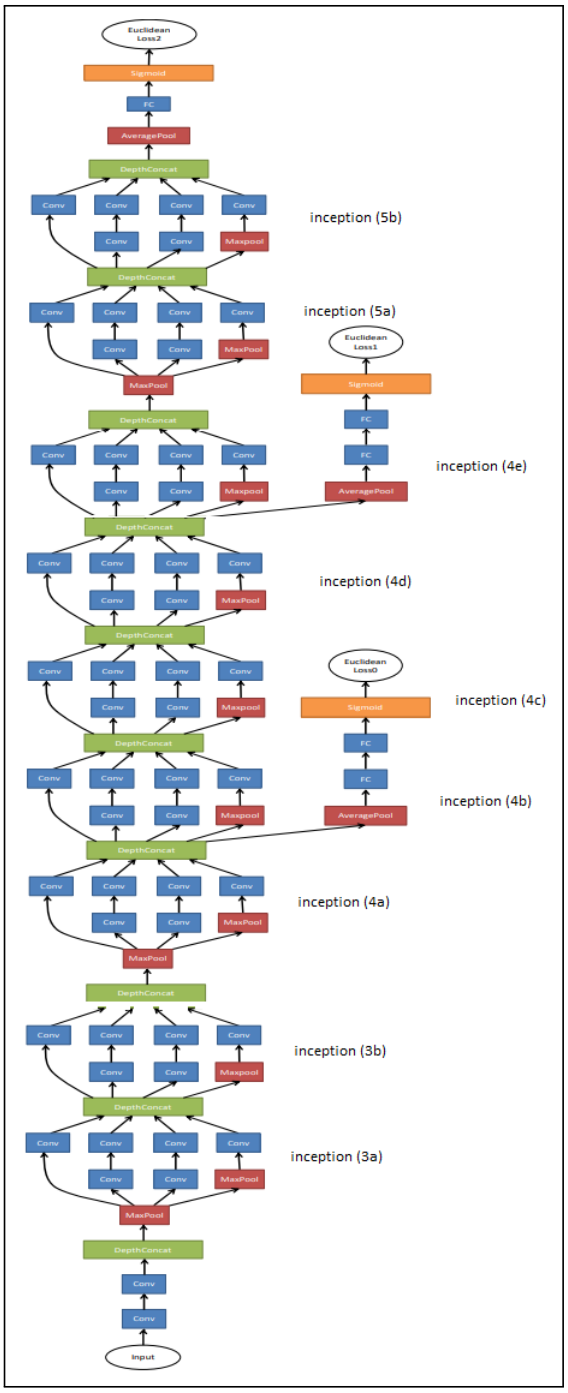


Figure 6.9: The GoogleNet Network Layers [241].

In the proposed Shallow-GoogleNet, as shown in Figure 6.10, the inception (5a) and the inception (5b)

are deleted from the architecture of GoogleNet. The black shaded part represents the deleted layers in Shallow-GoogleNet network.

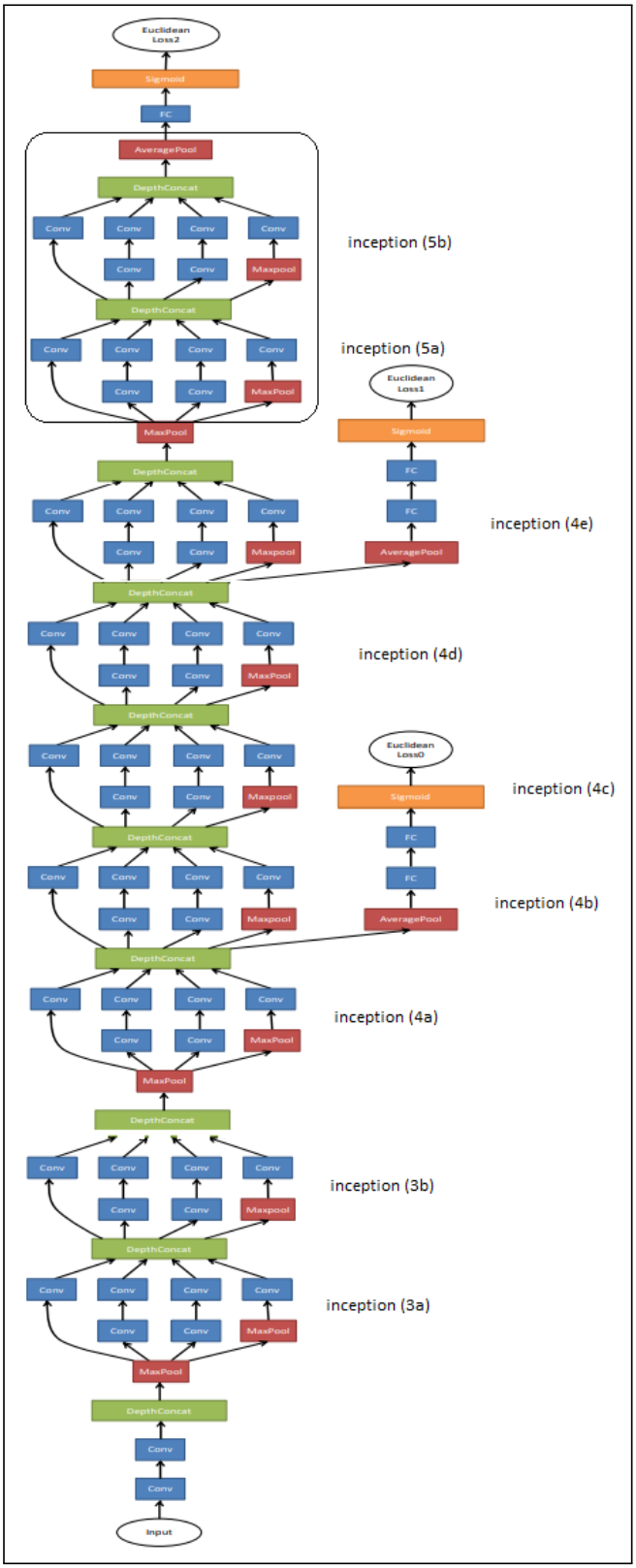


Figure 6.10: The Shallow GoogleNet Network Layers.

6.4 CAFFE

Caffe is a very widely deep learning framework developed by the Berkeley Vision and Learning Center (BVLC). To get started with CAFFE, there are many steps that should be set:

Get a desktop with a good GPU : Although most deep learning platforms can be run on CPU, it's in general much slower. Up to this point, building a PC with a decent GPU is the best option.

Install Caffe The official documentation of Caffe has detailed instruction on installing Caffe [237];

Data Preparation In this step, the images are stored in a format that can be used by Caffe. To create a Caffe dataset, two files: train.txt and val.txt will be needed to prepare. They will contain paths to images and class number from train and test data respectively.

Converting to LMDB Caffe is a high performance computing framework. Therefore, it is not necessary to use I/O file which slows down the computation time, To feed Caffe with large images dataset it's good to use LMDB format for the dataset. We need to change following things in the file $caffe/examples/imagenet/create\_imagenet.sh$:

- $EXAMPLE = examples/faces$: where we are going to store LMDB
- $DATA = data/faces/faces\_data$: folder with faces train.txt, val.txt
- $TRAIN\_DATA\_ROOT$: folder with train images
- VAL_DATA_ROOT: folder with test images
- $RESIZE = true$: we need to resize all images to same size and following piece of code:

```
echo "Creating train lmdb..."
GLOG_logtostderr=1 $TOOLS/convert_imageset \
 -- resize_height=$RESIZE_HEIGHT \
-- resize_width=$RESIZE_WIDTH \
-- shuffle \
$TRAIN_DATA_ROOT \
$DATA/train.txt \
$EXAMPLE/faces_train_lmdb
echo "Creating val lmdb..."
GLOG_logtostderr=1 $TOOLS/convert_imageset \
 -- resize_height=$RESIZE_HEIGHT \
 -- resize_width=$RESIZE_WIDTH \
 -- shuffle \
 $VAL_DATA_ROOT \
 $DATA/val.txt \
 $EXAMPLE/ faces_val_lmdb
```

Model definition In this step, a CNN architecture is chosen and its parameters are defined in a configuration file with extension .prototxt. Modify network architecture files in $caffe/models$. We should change bold things in the next snippet: size of images (from $RESIZE = true$) and paths to LMDB images ($\$EXAMPLE/faces\_train\_lmdb$, $\$EXAMPLE/faces\_val\_lmdb$) in file $train\_val.prototxt$. In our our work, three networks have been chosen; GoogleNet, RezNet50 and VGG16.

```
name: " VGGNet "
layer {
  name: "data"
  type: "Data"
  top: "data"
  top: "label"
  include {
    phase: TRAIN
  }
  transform_param {
    mirror: true
    crop_size: 224
    mean_value: 104
    mean_value: 117
    mean_value: 123
  }
  data_param {
    source: "examples/faces/faces_train_lmdb"
    batch_size: 32
    backend: LMDB
  }
}
```

```
layer {
  name: "data"
  type: "Data"
  top: "data"
  top: "label"
  include {
    phase: TEST
  }
  transform_param {
    mirror: false
    crop_size: 224
    mean_value: 104
    mean_value: 117
    mean_value: 123
  }
  data_param {
    source: "examples/facse/faces_val_lmdb"
    batch_size: 40
    backend: LMDB
  }
}
```

Solver definition The solver is responsible for model optimization. The solver parameters are defined in a configuration file with extension .prototxt.

Training and finetuning For training the network we need to run the following line:

```
./build/tools/caffe train -solver
models/VGGNet/quick_solver.prototxt -weights
models/VGGNet/VGGNet.caffemodel -gpu 0
```

The trained model is produced in a file with extension .caffemodel. After the training step, the .caffemodel trained model is used to make predictions of new unseen data.

6.5 The Experimental Work

The Caffe toolkit [237] on NVIDIA GeForce GTX 1080 GPU was applied to fine-tune the pre-trained GoogleNet, VGG16, and ResNet50 Deep CNN models. Face, ear, iris, hand, and palmprint are used, each of these databases consists of 50 distinct subjects with 40 images each with a total number of 2000. The database has been split into 1250 images for training, 500 for testing, and 250 for validation. Training dataset images used for fine-tuning are scaled to 224x224x3 to fit the CNN model input requirement. The cross-validation has been employed to find the optimal values for the parameters of each CNN to achieve high recognition performance. Two different sets of face, palm, ear, iris, and hand features are extracted using $inception5a-output$ and $inception5b-output$ layers for GoogleNet, pool4 and $conv5-2$ layers for VGG16, and $add-10$ and $add-11$ layers for ResNet50. Table 6.1 shows the recognition performance based on deep and shallow CNN and table 6.2 displays the training time that is needed for deep and Shallow CNN. The experimental results show that the recognition performance based on CNN outperformed the recognition performance based on shallow CNN in the majority of cases. For ear recognition based on CNN, using GoogleNet, VGG16 and ResNet50 achieves a high recognition performance. While ear recognition based on shallow GoogleNet shows a much better recognition performance as compared to shallow VGG16 and ResNet50. For iris recognition based on

CNN, using VGG16 increases the recognition performance as compared to GoogleNet and ResNet50. While iris recognition based on shallow GoogleNet shows a much better recognition performance as compared to VGG16 and ResNet50. For face recognition based on CNN, the three CNNs show a high recognition performance. While face recognition based on GoogleNet and ResNet50 shallow architecture outperformed the recognition performance based on shallow VGG16. For hand recognition based on CNN, the recognition performance based on ResNet50 outperformed the recognition performance based on GoogleNet and VGG16. While hand recognition based on shallow GoogleNet architecture shows a much better performance as compared to shallow VGG16 and ResNet50. For palm recognition, using GoogleNet, VGG16 and ResNet50 and shallow architecture achieve a high recognition performance. The experimental results show that using shallow GoogleNet is the appropriate network for face and palm recognition to reduce the recognition time and retain high recognition accuracy.

6.6 Conclusions

CNNs is a category of feedforward artificial neural network that has proven highly effective in various areas such as classification and image recognition. This chapter presents efficient and effective human biometrics recognition systems applied to five different databases: face, ear, palmprint, iris, and hand. The recognition process that used in the experimental work is based on using three different deep convolutional neural networks (GoogleNET, VGG16, ResNET50) and three shallow architecture which involve removing some layers in the CNNs. A comparison between these two architectures has been made to identify the appropriate architecture for each biometric to reduce the recognition time and retain high recognition accuracy. The experimental results show that the recognition performance based on deep CNN outperformed the recognition performance based on shallow CNN in the majority of cases due to the fact that at every layer, the network learns a new, more abstract representation of the input. While, shallow GoogleNet is the best network for face and palm recognition to reduce the recognition time and memory space to retain high recognition accuracy.

Table 6.1: Human biometric recognition based on CNN and shallow CNN

					CNN		Shallow CNN	
		Biometric	CNN Name	Layer	Validation	Test	Validation	Test
1	1	ear	googlenet	inception_5a-output inception_5b-output	98.0%	98.4%	98.4% 99.2%	98.0% 99.6%
	2	ear	vgg16	pool4 conv5_2	100.0%	100.0%	49.2% 33.2%	47.4% 32.8%
	3	ear	resnet50	add_10 add_11	100.0%	100.0%	70.0% 54.0%	68.8% 50.4%
2	4	eye	googlenet	inception_5a-output inception_5b-output	96.8%	97.6%	90.4% 93.2%	92.8% 90.6%
	5	eye	vgg16	pool4 conv5_2	100.0%	92.4%	65.6% 76.0%	71.4% 75.0%
	6	eye	resnet50	add_10 add_11	96.8%	99.2%	70.4% 73.6%	71.8% 65.8%
3	7	face	googlenet	inception_5a-output inception_5b-output	100.0%	99.6%	100.0% 100.0%	100.0% 100.0%
	8	face	vgg16	pool4 conv5_2	100.0%	100.0%	96.0% 98.0%	98.0% 100.0%
	9	face	resnet50	add_10 add_11	100.0%	100.0%	100.0% 96.0%	100.0% 96.6%
4	10	hand	googlenet	inception_5a-output inception_5b-output	92.8%	88.8%	92.4% 91.6%	87.2% 91.0%
	11	hand	vgg16	pool4 conv5_2	94.8%	91.6%	72.8% 69.6%	61.8% 64.4%
	12	hand	resnet50	add_10 add_11	98.0%	96.2%	82.0% 73.6%	77.0% 71.0%
5	13	palm	googlenet	inception_5a-output inception_5b-output	100.0%	100.0%	100.0% 100.0%	100.0% 100.0%
	14	palm	vgg16	pool4 conv5_2	100.0%	100.0%	100.0% 97.2%	100.0% 98.4%
	15	palm	resnet50	add_10 add_11	100.0%	100.0%	100.0% 98.0%	99.2% 98.8%

Table 6.2: Training time for deep and shallow CNN

					Training Time	
		Biometric	CNN Name	Layer	CNN	Shallow CNN
1	1	ear	googlenet	inception_5a-output	00:09:54	00:07:33
				inception_5b-output	00:11:13	00:09:45
	2	ear	vgg16	pool4	00:14:22	00:11:49
				conv5_2	00:15:55	00:13:51
	3	ear	resnet50	add_10	00:17:35	00:15:43
				add_11	00:20:11	00:18:52
2	4	eye	googlenet	inception_5a-output	00:10:55	00:08:34
				inception_5b-output	00:12:11	00:10:42
	5	eye	vgg16	pool4	00:13:51	00:11:33
				conv5_2	00:15:34	00:12:43
	6	eye	resnet50	add_10	00:17:24	00:15:40
				add_11	00:19:18	00:17:49
3	7	face	googlenet	inception_5a-output	00:13:09	00:11:49
				inception_5b-output	00:15:34	00:13:33
	8	face	vgg16	pool4	00:17:29	00:16:45
				conv5_2	00:19:51	00:17:53
	9	face	resnet50	add_10	00:20:31	00:18:53
				add_11	00:22:07	00:20:51
4	10	hand	googlenet	inception_5a-output	00:13:41	00:12:21
				inception_5b-output	00:15:44	00:13:33
	11	hand	vgg16	pool4	00:16:58	00:14:34
				conv5_2	00:18:31	00:15:54
	12	hand	resnet50	add_10	00:20:21	00:17:32
				add_11	00:21:54	00:19:32
5	13	palm	googlenet	inception_5a-output	00:12:33	00:11:30
				inception_5b-output	00:14:11	00:13:54
	14	palm	vgg16	pool4	00:16:21	00:15:21
				conv5_2	00:18:34	00:16:33
	15	palm	resnet50	add_10	00:20:21	00:18:34
				add_11	00:21:53	00:19:22

In the next chapter, the experimental work of multiple biometric recognition will be introduced. Multiple databases will be generated which consist of different single databases; face, ear, iris, hand, and palmprint. In addition, a different combination of single biometric systems that based on score fusion will be examined to achieve high recognition performance.

Chapter 7

Experiments in Biometric Recognition

Multimodal biometrics involves using of a combination of two or more biometric modalities in a human identification system. Multimodal biometrics provide additional information within different modalities to improve the recognition performance to overcome the drawbacks of single biometric systems. The experimental work described in this chapter is an implementation of a person identification system fusing different combinations of biometric modalities. The PCA and LDA classifiers are used to extract the features of the input face, ear, iris, hand, and palmprint images. A decision is made by matching the test image with the images registered in the database using Manhattan similarity measure: see chapter 5 for more details. The scores which are obtained from each biometric have been normalized, fused and then used for identification. A better result was obtained if the modalities are combined: Figure 7.1 represents the identification process based on PCA/LDA. In this experimental work, we generate multiple databases which consist of combinations of single databases: face, ear, iris, hand, and palmprint. Each of these databases consists of 50 distinct subjects with $10, 25, 40$ images each. Different splits between training/testing sets are examined to achieve high recognition performance.

7.1 The Identification Process

Three different experiments have been applied to examine the effect of the training/testing set size on the recognition performance. In the first experiment, each database consists of 50 distinct subjects with 10 images each with a total number of 500. The database has been split into 350 images for training and 150 for testing. In the second experiment, each of 50 distinct subjects has 25 images with a total number of 1250. The database has been split into 1000 images for training and 250 for testing. In the third experiment, each subject has 40 images with a total number of 2000. The database has been split into 1500 images for training and 500 for testing. A comparison between different combination of single biometric traits has been made. The practical work is summed up as follows. For face, ear, iris, palm, and hand images, each of size $p \times q$ pixels, all images in each dataset must be centred, of the same size and represented by a vector in $p.q$ dimensional space. Then, the average image is computed. PCA/LDA implementation is used to extract features of images and the information extracted from each image is represented by the distances corresponding to the distinct people. Once the eigenfaces are created, the weights for them are calculated and stored in. These weights describe the contribution of each eigenface in representing the input face. Figure 7.2 shows the training set of face images. Figure 7.3 shows the first 20 eigenfaces for ear dataset. The features extracted by the Eigenfaces algorithm can be reconstructed back into an image, figure 7.4 shows the reconstructed ear images. Given an unknown input image which is centred and of the same size as the training images. The weight and the distance of the input image is calculated and stored. The minimum distance classifier is used for classification. An acceptance or rejection decision is calculated by applying the Manhattan distance comparison: table 7.1 shows the Manhattan distance for faces, ears, palms, hand, and irises based on PCA and table 7.2 shows the Manhattan distances based on LDA. Figure 7.5 to Figure 7.7 present results for ear images. Figure 7.8 to Figure 7.10 present results for hand images. Figure 7.11 to Figure 7.13 present results for iris images. Figure 7.14 to Figure 7.16 present results for palm images. Before performing recognition based on fusion, all face, ear, palm, iris, and hand input images should be pre-processed and the stored

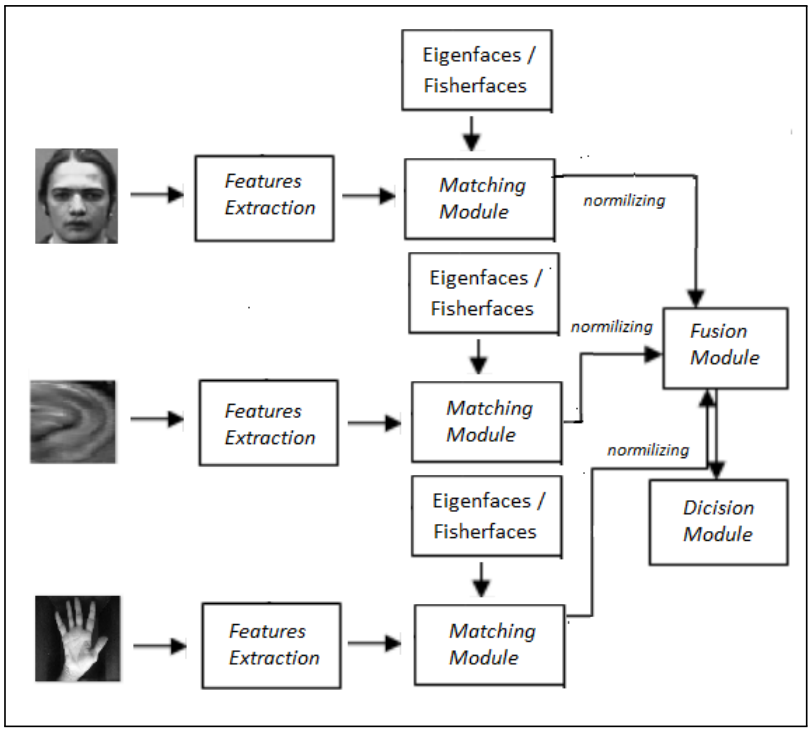


Figure 7.1: The identification process based on PCA and LDA

distances should be normalised.

Table 7.1: Manhatten distance for faces, ears, palms, hands, and irises based on PCA

Faces	Minimum Distance	Ears	Minimum Distance	Palms	Minimum Distance	Hands	Minimum Distance	Irises	Minimum Distance
f_{p1}	0.0327432	e_{p1}	0.0143803	pl_{p1}	0.0348488	h_{p1}	0.0962787	$iris_{p1}$	0.0165362
f_{p2}	0.0740500	e_{p2}	0.0185656	pl_{p2}	0.0635521	h_{p2}	0.1019694	$iris_{p2}$	0.0378175
f_{p3}	0.0922546	e_{p3}	0.0533082	pl_{p3}	0.1080658	h_{p3}	0.1077103	$iris_{p3}$	0.0388666
f_{p4}	0.1212889	e_{p4}	0.0567424	pl_{p4}	0.1119872	h_{p4}	0.1078077	$iris_{p4}$	0.0416875
f_{p5}	0.2824746	e_{p5}	0.0687295	pl_{p5}	0.1191199	h_{p5}	0.1237984	$iris_{p5}$	0.0486943
f_{p6}	0.2926958	e_{p6}	0.0805651	pl_{p6}	0.1265894	h_{p6}	0.1285813	$iris_{p6}$	0.0533318
f_{p7}	0.3404087	e_{p7}	0.0838133	pl_{p7}	0.1480250	h_{p7}	0.1374917	$iris_{p7}$	0.0588437
f_{p8}	0.3467941	e_{p8}	0.0943575	pl_{p8}	0.1587039	h_{p8}	0.14038849	$iris_{p8}$	0.0697203
f_{p9}	0.3541979	e_{p9}	0.0971861	pl_{p}	0.1599625	h_{p9}	0.1437796	$iris_{p9}$	0.0713442
f_{p10}	0.3558873	e_{p10}	0.1069176	pl_{p10}	0.1772897	h_{p10}	0.1440630	$iris_{p10}$	0.0718305

Table 7.2: Manhatten distance for faces, ears, palms, hands, and irises based on LDA

Faces	Minimum Distance	Ears	Minimum Distance	Palms	Minimum Distance	Hands	Minimum Distance	Irises	Minimum Distance
f_{p1}	0.0001188	e_{p1}	0.1901300	pl_{p1}	0.1364315	h_{p1}	0.0140181	$iris_{p1}$	0.0395210
f_{p2}	0.0001235	e_{p2}	0.1901300	pl_{p2}	0.2230689	h_{p2}	0.0710869	$iris_{p2}$	0.0935600
f_{p3}	0.0001430	e_{p3}	0.1901300	pl_{p3}	0.2625866	h_{p3}	0.1059091	$iris_{p3}$	0.0973866
f_{p4}	0.0001668	e_{p4}	0.9662956	pl_{p4}	0.3478251	h_{p4}	0.2204438	$iris_{p4}$	0.1062461
f_{p5}	0.0002138	e_{p5}	0.1966295	pl_{p5}	0.3526303	h_{p5}	0.1241496	$iris_{p5}$	0.1342058
f_{p6}	0.0044346	e_{p6}	0.1966295	pl_{p6}	0.3730244	h_{p6}	0.1373454	$iris_{p6}$	0.1371801
f_{p7}	0.0044632	e_{p7}	0.1966295	pl_{p7}	0.3766666	h_{p7}	0.1428376	$iris_{p7}$	0.1409095
f_{p8}	0.0044949	e_{p8}	0.1967776	pl_{p8}	0.3798038	h_{p8}	0.1504045	$iris_{p8}$	0.1558910
f_{p9}	0.0044957	e_{p9}	0.1977351	pl_{p}	0.3792590	h_{p9}	0.1731431	$iris_{p9}$	0.1726062
f_{p10}	0.0045300	e_{p10}	0.1998458	pl_{p10}	0.3852644	h_{p10}	0.1761673	$iris_{p10}$	0.1728823

For biometric recognition based on PCA; face, ear, palm, iris, and hand are tested individually with

Figure 7.2: Training set of face images

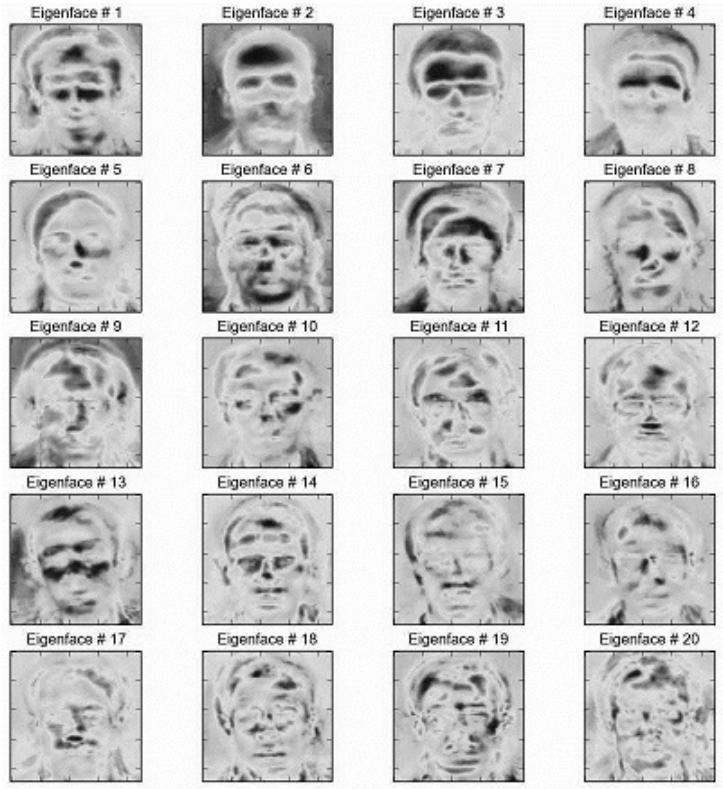


Figure 7.3: The first 20 eigenfaces for face dataset

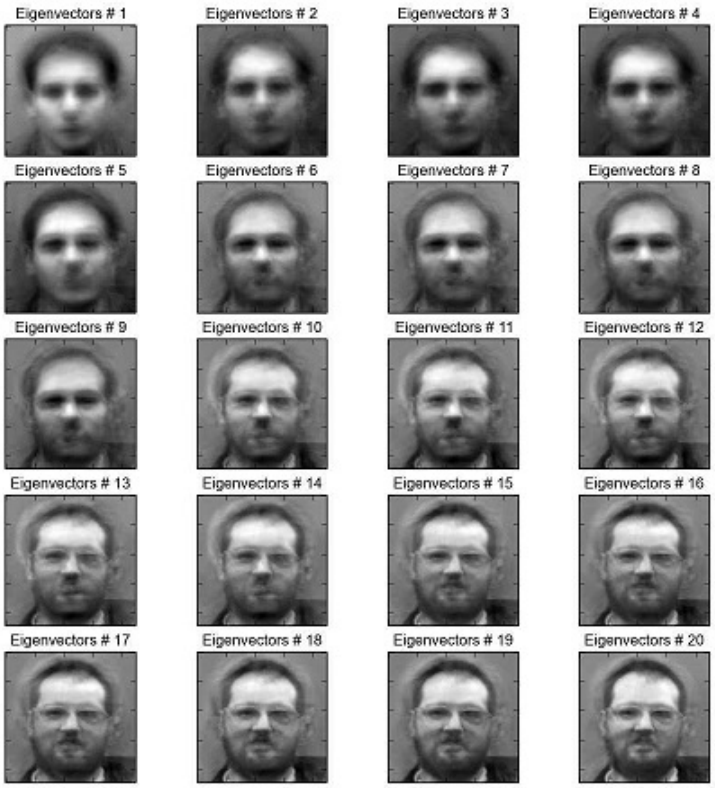


Figure 7.4: The reconstructed face images

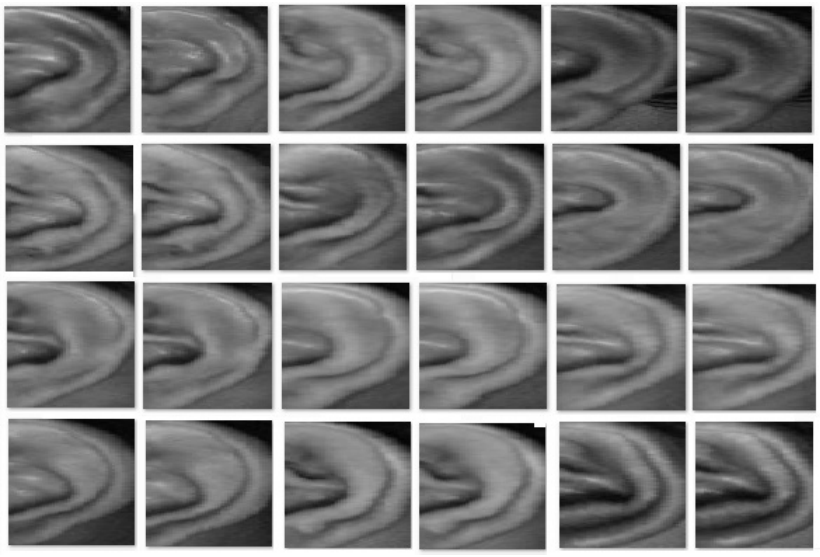


Figure 7.5: Training set of ear images

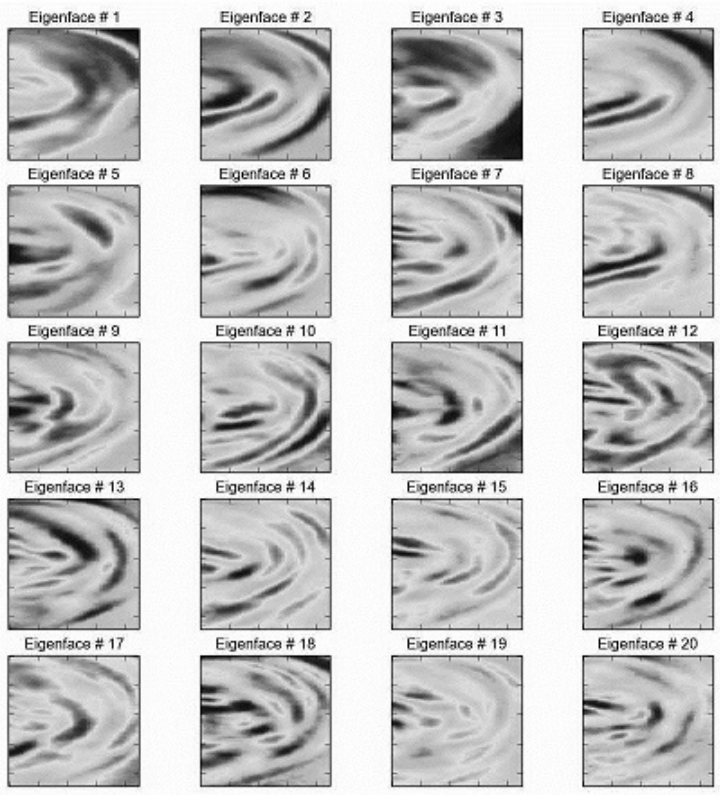


Figure 7.6: The first 20 eigenfaces for ear dataset

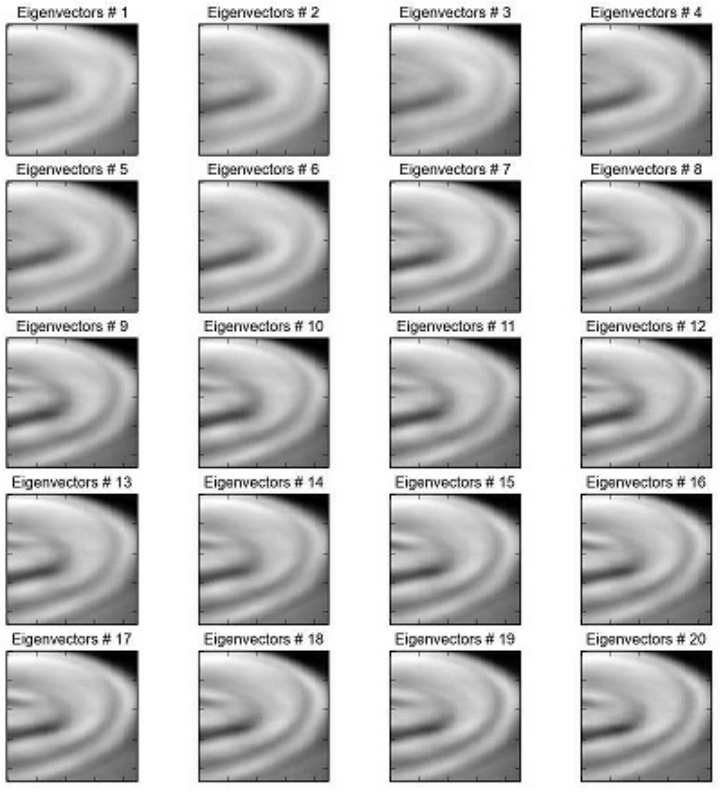


Figure 7.7: The reconstructed ear images

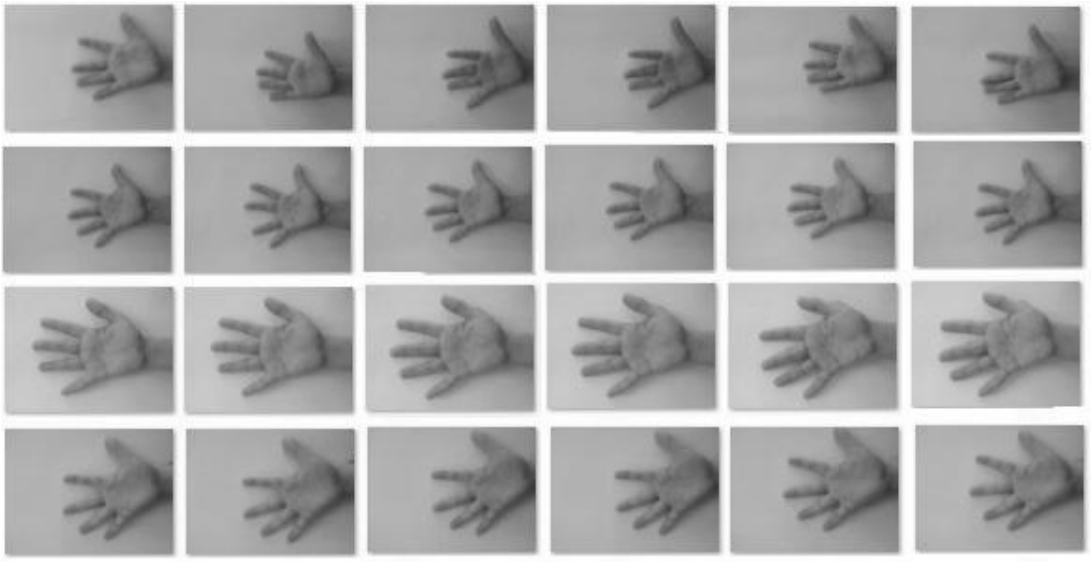


Figure 7.8: Training set of hand images

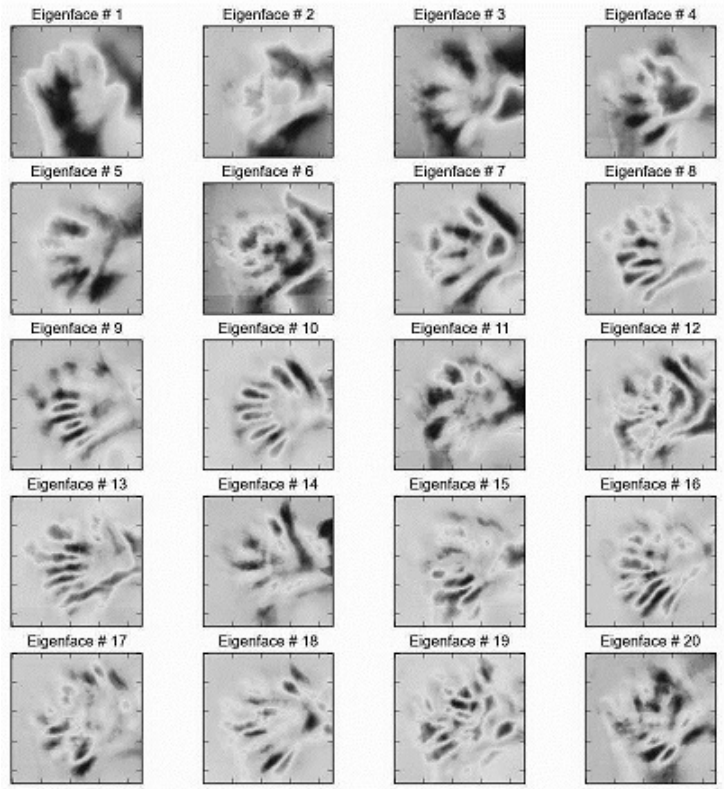


Figure 7.9: The first 20 eigenfaces for hand dataset

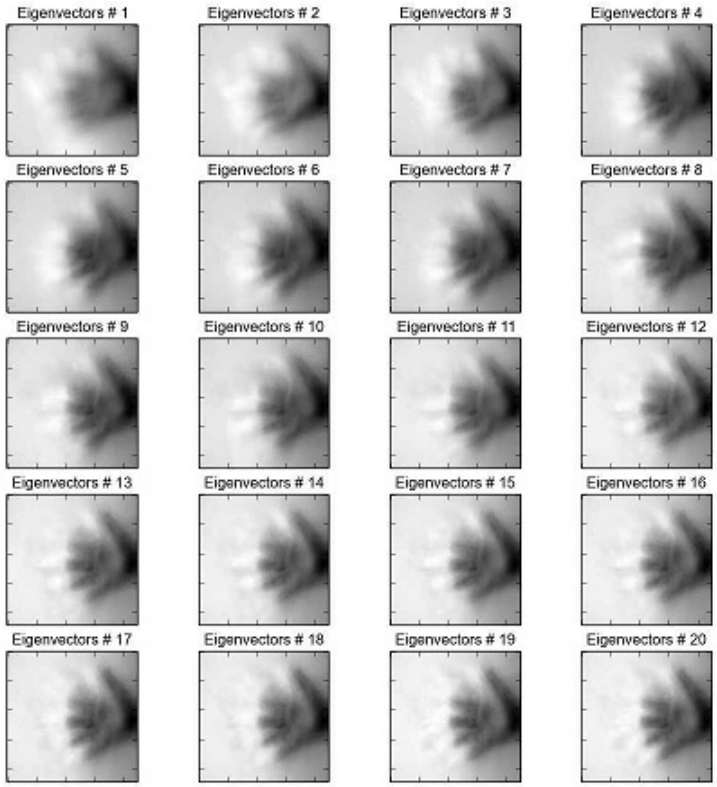


Figure 7.10: The reconstructed hand images

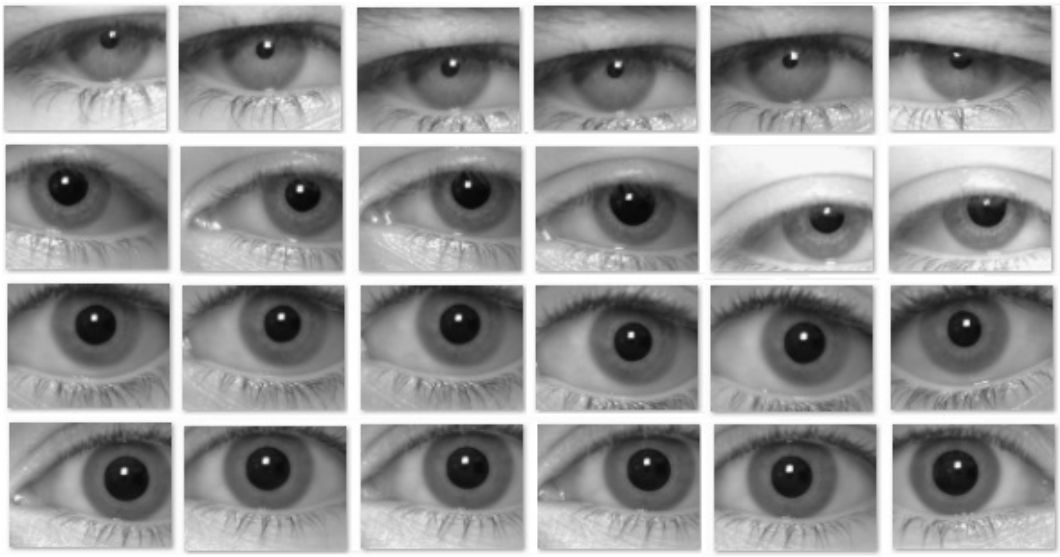


Figure 7.11: Training set of iris images

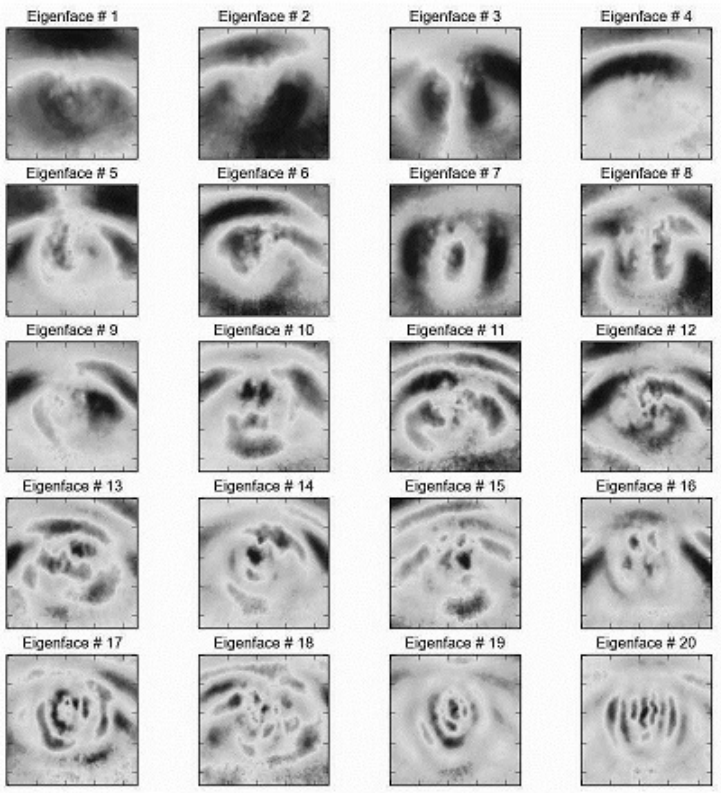


Figure 7.12: The first 20 eigenfaces for iris dataset

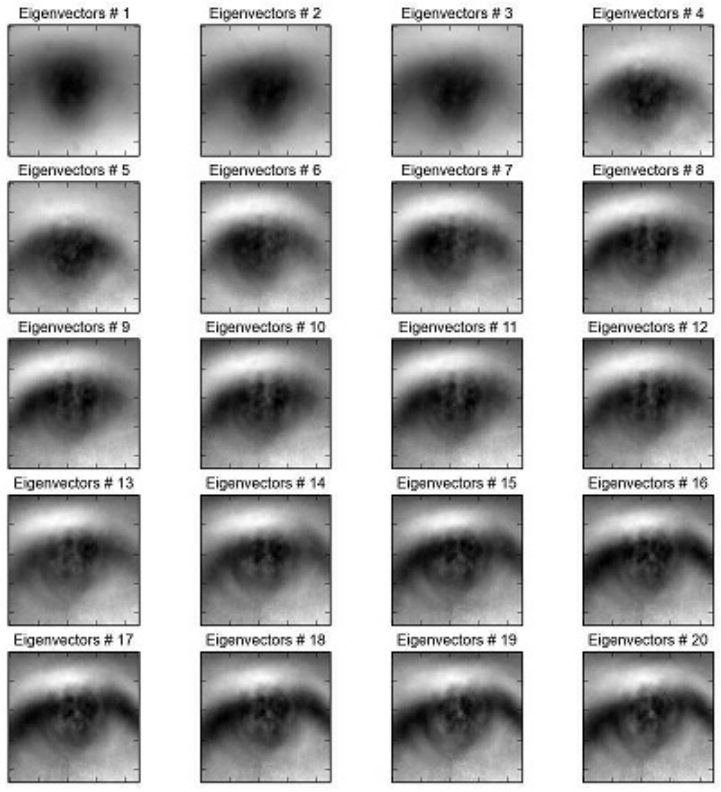


Figure 7.13: The reconstructed iris images

Figure 7.14: Training set of palm images

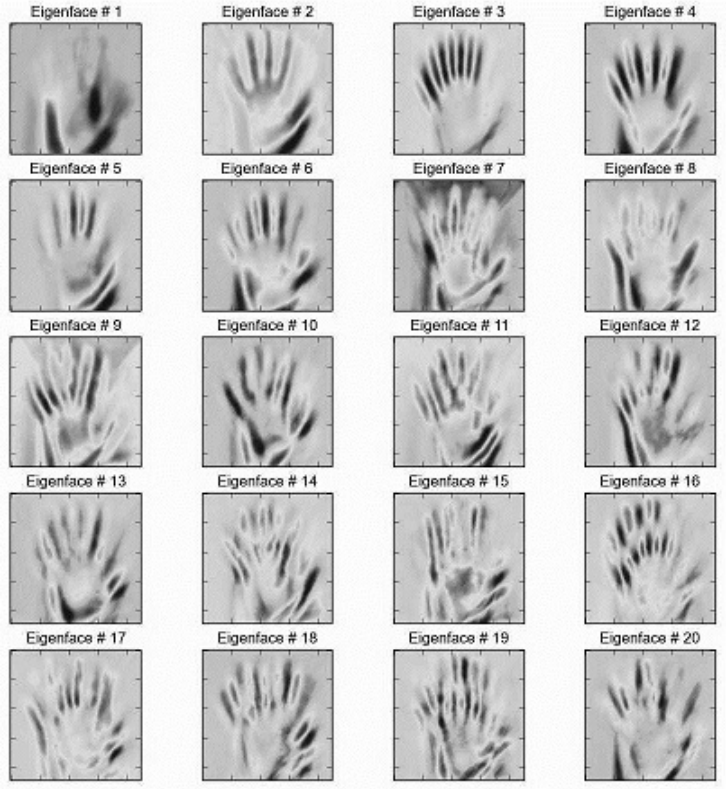


Figure 7.15: The first 20 eigenfaces for palm dataset

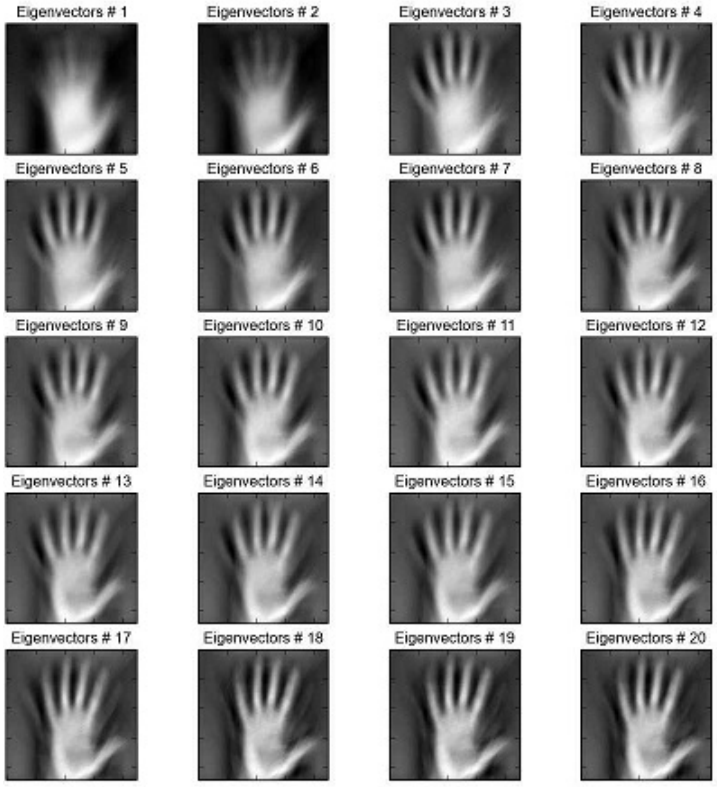


Figure 7.16: The reconstructed palm images

different size of the training/testing set size. Individual accuracy for face is found to be 0.86, 0.91, 0.99 for the dataset of 10, 25, 40 images for each person, respectively. Accuracy for ear 0.31, 0.64, 0.91, for palm 0.81, 0.91, 0.95, for iris 0.49, 0.64, 0.69 and for hand 0.77, 0.83, 0.93 as shown in table 7.3. The overall performance of the system has increased, showing accuracy for face, ear and hand of 0.96, 1.00, 1.00. The overall performance of the system has increased showing accuracy for face, ear and palm of 0.89, 0.97 and 1.00. Fusing face, hand and iris biometric modalities achieve an accuracy of 0.99, 1.00 and 1.00. The performance of the combined face, hand and palm modalities system has increased as well and achieves an accuracy of 0.99, 1.00, 1.00. Finally, the performance of the combined hand, palm and iris modalities system achieves an accuracy of 0.87, 0.94, 1.00. The comparative experiment shows that better result was obtained if the modalities are combined when the size of the training set has increased. For this experiment, the best results are obtained when we use the dataset that has 40 images for each person. Consequently, the size of the training set is a main factor to increase the recognition accuracy. In terms of reducing the storage space of memory and processing time, using fusing of two

Table 7.3: Individual and Combined Biometric Recognition Accuracy Based on PCA

Biometric	10 images each		25 images each		40 images each	
	Accuracy	$Error_{Rate}$	Accuracy	$Error_{Rate}$	Accuracy	$Error_{Rate}$
Face	0.86	0.14	0.91	0.09	0.99	0.01
Ear	0.31	0.69	0.64	0.36	0.91	0.09
Palm	0.81	0.19	0.91	0.09	0.95	0.05
Iris	0.49	0.51	0.64	0.36	0.69	0.31
Hand	0.77	0.23	0. 83	0.17	0.93	0.07
$Face + Ear + Hand$	0.96	0.04	1.00	0.00	1.00	0.00
$Face + Ear + Palm$	0.89	0.11	0.97	0.03	1.00	0.00
$Face + Hand + Iris$	0.95	0.05	1.00	0.00	1.00	0.00
$Face + Hand + Palm$	0.96	0.04	1.00	0.00	1.00	0.00
$Hand + Iris + Palm$	0.87	0.13	0.94	0.06	1.00	0.00
$Face + Ear$	0.83	0.17	0.85	0.15	0.97	0.03
$Face + Palm$	0.90	0.10	0.93	0.07	1.00	0.00
$Face + Hand$	0.88	0.12	0.93	0.07	0.95	0.05
$ear + hand$	0.67	0.33	0.76	0.24	0.92	0.08
$palm + hand$	0.90	0.10	0.93	0.07	0.95	0.05

biometric modalities achieve a good accuracy as compared to fusing of three biometric modalities in some cases.

For biometric recognition based on LDA; Individual accuracy for face is found to be 0.91, 0.99, 0.99 for dataset of 10, 25 and 40 images for each person, respectively. Accuracy for ear 0.73, 0.82 and 0.93, for palm 0.85, 0.94 and 0.99, for iris 0.65, 0.69 and 0.71 and for hand 0.81, 0.91, 0.93 as shown in table 7.4. The overall performance of the system has increased showing accuracy for face, ear and hand of 0.98, 1.00, 1.00. The overall performance of the system has increased, showing accuracy for face, ear and palm of 0.96, 1.00, 1.00. Fusing face, hand and iris biometric modalities achieve accuracy of 0.98, 1.00, 1.00. The performance of the combined face, hand and palm modalities system has increased as well and achieves an accuracy of 1.00, 1.00, 1.00. Finally, the performance of the combined hand, palm and iris modalities system achieves an accuracy of 0.92, 1.00, 1.00. The comparative experiment shows that individual biometric recognition accuracy that based on fisherfaces technique out performed the eigenfaces technique in most cases irrespective of the size of the training/testing set size and better results were obtained when the modalities were combined. This experiment shows that LDA based multimodal biometrics recognition gains good results as compared to the PCA based system even when we use the dataset that has 10 images for each person.

Table 7.4: Individual and Combined Biometric Recognition Accuracy Based on LDA

Biometric	10 images each		25 images each		40 images each	
	Accuracy	$Error_{Rate}$	Accuracy	$Error_{Rate}$	Accuracy	$Error_{Rate}$
Face	0.91	0.09	0.99	0.01	0.99	0.01
Ear	0.73	0.27	0.82	0.18	0.93	0.07
Palm	0.85	0.15	0.94	0.06	0.99	0.01
Iris	0.65	0.35	0.69	0.31	0.71	0.29
Hand	0.81	0.19	0.91	0.09	0.93	0.07
$Face + Ear + Hand$	0.98	0.02	1.00	0.00	1.00	0.00
$Face + Ear + Palm$	0.96	0.04	1.00	0.00	1.00	0.00
$Face + Hand + Iris$	0.98	0.02	1.00	0.00	1.00	0.00
$Face + Hand + Palm$	1.00	0.00	1.00	0.00	1.00	0.00
$Hand + Iris + Palm$	0.92	0.08	1.00	0.00	1.00	0.00
$Face + Ear$	0.87	0.13	0.94	0.06	0.97	0.02
$Face + Palm$	0.90	0.10	0.94	0.06	1.00	0.00
$Face + Hand$	0.92	0.08	0.94	0.06	0.97	0.03
$ear + hand$	0.82	0.18	0.85	0.15	0.95	0.05
$palm + hand$	0.93	0.07	0.95	0.05	0.99	0.01

7.2 Conclusions

Recently, many contributions have been achieved in the field of biometrics and investigations have been made in the multi-modal biometrics domain. Automatic individual identification is a fundamental task in computer vision applications. The proposed work is an implementation of an individual identification process fusing different combination of biometric modalities. The PCA and LDA classifiers are used to extract the features of the input biometrics images. A decision is made by matching the test image with the images registered in the database using Manhattan similarity measure. The scores which are obtained from each biometric at the score level have been normalized, fused and then used for identification. Multiple databases have been used which consist of different single databases; face, ear, iris, hand, and palmprint. Each of these databases consists of 50 distinct subject with $10, 25, 40$ images each, respectively. Different size of a training/testing set is examined to achieve high recognition performance. A better result was obtained if the modalities were combined. For biometric recognition based on PCA, the comparative experiment shows that better result was obtained if the modalities were combined and the size of the training set has increased. for this experiment, the best results are obtained when we use the dataset that has 40 images for each person. Consequently, the size of

the training set is a main factor to increase the recognition accuracy. For biometric recognition based on LDA, The comparative experiment shows that individual biometric recognition accuracy that based on fisherfaces technique outperformed the eigenfaces technique in most cases regardless of the size of the training/testing set size and better result was obtained when the modalities were combined. This experiment shows that LDA based multimodal biometrics recognition gains good results as compared to the PCA based system even when we use the dataset that has 10 images for each person.

Practically, the use of multi-biometrics information increases security levels. Consequently, it is expected to be employed dramatically in many fields such as banking systems/financial transactions, police and criminal evidence applications, and airport security, Time and Attendance Monitoring, Benefit Payment systems, and Prison visitor systems.

Regarding to banking systems, the authentication system used by most of the banks is dependent on traditional identification mechanisms such as passwords/PIN however, this authentication system could be attacked easily. Depending on the experimental work, we propose that using multi-biometric systems based on a combination of the mentioned biometrics can play a vital role to ensure secured banking services. In some cases related to banking robbery, it would be hard to capture a robber's face and ear because it usually are covered. Therefore, using multimodal biometric identification framework based on (body and gait) would be more useful. For on-line financial transactions, we suggest using a multiple biometric system based on (finger and palm print)

A more interesting way to use multi-model biometric recognition for surveillance. The obstacle with using surveillance camera footage is that individuals are only rarely looking at the camera directly, therefore the resulting images are hard to be processed by face recognition algorithms. Therefore, we suggest using multi-modal biometric identification framework based on (face,ear, hand) or (ear and hand) or (body and gate).

Using multi-model biometric recognition in the police and criminal evidence applications is so interesting. Imagine there is a witness to a crime, he saw a person running very quickly, but he was able to have a look at his face as he removed his balaclava. There might not be useful to use CCTV footage,

because of the mask. In the Uk police presents different faces to witness and the police will select the one that looks most like the person who did the crime. To increase security levels they can do the same, in principle, with hands etc and then ascertain whether witnesses choose the same person.

Chapter 8

Conclusions and Future Work

8.1 Thesis Contributions

The aim of the research was to identify effective features and machine learning methods for human recognition based on multiple biometrics and produce a sufficient combination of single biometric systems suitable in specific applications for identification purposes. For example, banking systems which use multi-biometric authentication for login procedures and the police and criminal evidence applications. At the start of the project, a human biometrics recognition based approach using feature matching seemed sensible. The key problem with using features was that it was unclear which detector and extractor to use, mainly because existing evaluations did not produce useful indicators of performance. A substantial part of the work reported in chapter 3 is concerned with approaches to evaluation that are statistically meaningful. Well-established tests such as McNemar's test are able to perform statistically meaningful comparisons and this work has established that they can be employed to compare the performances of feature detectors and extractors.

The results in chapter 4 suggested that the well-established Viola-Jones algorithm can be made more effective by applying it to the brightness channel of a transformed colour image, rather than the conventional grey-scale image obtained from a colour one. Moreover, the V-channel of HSV is more effective

than the L-channel of HLS. This means that the simple, fast transformation from RGB pixels to HSV as a precursor to using conventional Viola-Jones face detection should yield significantly fewer false positives and negatives, an improvement in its effectiveness.

Chapter 5 focused on the biometric recognition using conventional statistical techniques. Four similarity measures have been compared for PCA- and LDA-based face, ear and palm biometrics. Classification was performed using the nearest neighbour classier using Manhattan, Euclidean, Cosine similarity and Mahalanobis distance measures. The experimental results showed that both PCA and LDA perform well if presented with an image in the test set which is similar to an image in the training set. LDA shows a significantly better recognition performance in all cases as evidenced by McNemar's test, suggesting that it is better at handling variation in lighting and expression. However, the fact that the magnitudes of the similarity measures were not consistently rank-ordered shows that the choice of similarity measure can affect the conclusions drawn. There is clearly some feature or property of the content of datasets that affect the similarity measures. Two experiments have been examined in this part, both dealing with face and ear recognition systems. The first one depends on using the PCA technique for feature extraction. While the second depends on using a combination of Linear Discriminate Analysis and Principal Components Analysis to extract features. Finally, support vector machine algorithm is implemented for classification for both experiments to identify the similarity between the probe images from the same individual and different individuals after all the images in the database are represented with relevant features. No previous study examines the recognition performance based on these two methods, especially for biometric authentication domain problems. The experimental results using accuracy and the error rate indicated that the recognition based on SVM_{pca} and $SVM_{pca,lda}$ techniques can achieve a better performance than using PCA based on distance similarity measures. While no significant differences were found using SVM_{pca}, $SVM_{pca,lda}$ and LDA based on distance similarity measures.

Chapter 6 presented a human biometrics recognition systems (face, ear, eye, palmprint, hand) based on three deep convolutional neural networks (GoogleNET, VGG16, ResNET50) and shallow convolutional neural network by removing some layers in the network. CNNs is a category of feedforward artificial

neural network that has proven high effective in various areas such as classification and image recognition. Having established this, chapter 7 presented an implementation of a person identification system fusing different combinations of biometric modalities; face, ear, eye, hand, and palmprint. Different size of training/testing set are examined to achieve high recognition performance. The PCA and LDA classifiers are used to extract the features of the input face, ear, eye, hand ,and palmprint images. Decision is made by matching the test image with the images registered in the database using Manhattan similarity measure. The scores which are obtained from each biometric at the score level have been normalized, fused and then used for identification. Practically, the use of multi-biometrics information increase security levels. Consequently, it is expected to be employed dramatically in many fields such as banking systems/financial transactions, police and criminal evidence applications, and airport security.

8.2 Future Directions

The contributions from this research lay the groundwork for future research into the human biometrics recognition in order to explore a wide range of feature extraction and classification problems. Some specific issues should be considered for future research:

- Considering the shortcomings of performance evaluation metrics previously used to assess the performance of vision algorithms, see chapter 3 for more details, the application of statistical tests other than McNemar's test can be explored.

- Another important factor affecting algorithms' performances is the amount and type of data used. In future, the performance of feature operators, including Viola-Jones detector, can be performed on significantly larger amounts of data with more variety of imagery. The investigations in chapter 3 are limited to standard datasets, and capturing larger amount of image data and with an automated method of generating ground truth would clearly improve the situation.

- The two fundamental aspects of CNN are the size of the training dataset and the depth of the

network architecture. In chapter 6, a shallow convolutional neural network (CNN) design has been initialized and two different layer have been removed. For future research, different other layers can be removed from the depth of the deep neural network and testing its recognition performance on the validation/testing datasets.

- In chapter 7, a complete multi-biometric recognition system based on different combination of single biometric databases; face, ear, eye, hand, and palmprint has been examined by using score level fusion. In future, feature and decision level fusion can be explored and compared the system performance with the examined one.

- Two multi-biometric recognition systems based on convolutional neural network (CNN) design and shallow architecture can be applied and then compared the two systems performance.

Practically, biometrics applications increased dramatically in many fields to determine the identity of an individual. Theoretically, a single biometric identification might seem proficient but in fact there are considerable challenges and restrictions for example in face recognition the quality of the captured images might be affected by illumination conditions. In some cases, a camera can not be able to differentiate between the two users leading to incorrect matching. Another example when a scanner is unable to read dirty fingerprints clearly and this will lead to false database matches. Consequently, an enrolled user might be rejected whereas an impostor might be accepted falsely. Hence using a multi-biometric system will compensate the restrictions of a single biometric system. In other words, it becomes hard for an impostor to spoof all the biometric characteristics of an authorized enrolled individual because a person can spoof one biometric but it is currently not possible to spoof many at same time. Therefore, this work suggests that multiple biometric systems give a way to circumvent this potential problem.

Bibliography

[1] Shailaja, Dasari, and Phalguni Gupta. "A simple geometric approach for ear recognition." 9th International Conference on Information Technology (ICIT'06). IEEE, 2006.

[2] Le, Chien, and R. Jain. "A survey of biometrics security systems." EEUU. Washington University in St. Louis (2009).

[3] Anil, Ruud Bolle, and Sharath Pankanti. "Introduction to biometrics." Biometrics. Springer, Boston, MA, 1996. 1-41.

[4] Anil K., and Arun Ross. "Multibiometric systems." Communications of the ACM 47.1 (2004): 34-40.

[5] Anil K., Arun Ross, and Salil Prabhakar. "An introduction to biometric recognition." IEEE Transactions on circuits and systems for video technology 14.1 (2004).

[6] Wei-Lun. "Face recognition." GICE, National Taiwan University (2007).

[7] Zhao, Wenyi, et al. "Face recognition: A literature survey." ACM computing surveys (CSUR) 35.4 (2003): 399-458.

[8] TURK, MA. "Proc. IEEE Conf. Computer Vision and Pattern Recognition." CVPR'91. 1991.

[9] Daugman, John G. "High confidence visual recognition of persons by a test of statistical independence." IEEE transactions on pattern analysis and machine intelligence 15.11 (1993): 1148-1161.

[10] Daugman, John, and O. B. E. PhD. "University of Cambridge." How Iris Recognition Works (2004).

[11] Patel, Chandrakant D., Sanket Trivedi, and Sanjay Patel. "Biometrics in IRIS Technology: A Survey." International Journal of Scientific and Research Publications 2 (2012): 3-4.

[12] Jain, Anil, Arun Ross, and Salil Prabhakar. "Fingerprint matching using minutiae and texture features." Proceedings 2001 International Conference on Image Processing (Cat. No. 01CH37205). Vol. 3. IEEE, 2001.

[13] Jain, Anil K., et al. "An identity-authentication system using fingerprints." Proceedings of the IEEE 85.9 (1997): 1365-1388.

[14] Jain, Anil K., and Nicolae Duta. "Deformable matching of hand shapes for user verification." Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348). Vol. 2. IEEE, 1999.

[15] Ong, Michael Goh Kah, et al. "A single-sensor hand geometry and palmprint verification system." Proceedings of the 2003 ACM SIGMM workshop on Biometrics methods and applications. ACM, 2003.

[16] Garg, Pragati, Naveen Aggarwal, and Sanjeev Sofat. "Vision based hand gesture recognition." World Academy of Science, Engineering and Technology 49.1 (2009): 972-977.

[17] LaViola, Joseph. "A survey of hand posture and gesture recognition techniques and technology." Brown University, Providence, RI 29 (1999).

[18] Jeffreys, Alec J., Victoria Wilson, and Swee Lay Thein. "Individual-specific 'fingerprints' of human DNA." Nature 316.6023 (1985): 76.

[19] Gifford, M., and N. Edwards. "Trial of dynamic signature verification for a real-world identification solution." BT technology journal 23.2 (2005): 259-266.

[20] Qi, Yingyong, and Bobby R. Hunt. "A multiresolution approach to computer verification of handwritten signatures." IEEE Transactions on Image Processing 4.6 (1995): 870-874.

[21] Rabiner, L. "Theory and implementation of hidden Markov models." Fundamentals of speech recognition (1993).

[22] Katiyar, Rohit, Vinay Kumar Pathak, and K. V. Arya. "A study on existing gait biometrics approaches and challenges." International Journal of Computer Science Issues (IJCSI) 10.1 (2013): 135.

[23] Arora, Parul, and Smriti Srivastava. "Gait recognition using gait Gaussian image." 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN). IEEE, 2015.

[24] Sloman, Leon, et al. "Gait patterns of depressed patients and normal subjects." The American journal of psychiatry (1982).

[25] Katiyar, Rohit, Vinay Kumar Pathak, and K. V. Arya. "A study on existing gait biometrics approaches and challenges." International Journal of Computer Science Issues (IJCSI) 10.1 (2013): 135.

[26] Ashbourn, Julian. Biometrics: Advanced identity verification: The complete guide. Springer, 2014.

[27] Jain, Anil K., Patrick Flynn, and Arun A. Ross, eds. Handbook of biometrics. Springer Science and Business Media, 2007.

[28] National Research Council, and Whither Biometrics Committee. Biometric recognition: challenges and opportunities. National Academies Press, 2010..

[29] Juels, Ari, David Molnar, and David Wagner. "Security and Privacy Issues in E-passports." First International Conference on Security and Privacy for Emerging Areas in Communications Networks (SECURECOMM'05). IEEE, 2005.

[30] International Organization for Standardization, and International Electrotechnical Commission. Information Technology: Security Techniques: Code of Practice for Information Security Management. ISO/IEC, 2005.

[31] Jain, Anil K., Arun Ross, and Sharath Pankanti. "Biometrics: a tool for information security." IEEE transactions on information forensics and security 1.2 (2006): 125-143.

[32] Ahmad, Sharifah Mumtazah Syed, Borhanuddin Mohd Ali, and Wan Azizun Wan Adnan. "Technical issues and challenges of biometric applications as access control tools of information security." international journal of innovative computing, information and control 8.11 (2012): 7983-7999.

[33] Frederickson, H. George, and Todd R. LaPorte. "Airport security, high reliability, and the problem of rationality." Public Administration Review 62 (2002): 33-43.

[34] Jain, Anil K., et al. "Biometrics: a grand challenge." Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.. Vol. 2. IEEE, 2004.

[35] Gavrilova, Marina L., and Md Maruf Monwar. "Current Trends in Multimodal Biometric System—Rank Level Fusion." Pattern Recognition, Machine Intelligence and Biometrics. Springer, Berlin, Heidelberg, 2011. 657-673.

[36] Singhal, Rashmi, Narender Singh, and Payal Jain. "Towards an integrated biometric technique." International Journal of Computer Applications 42.13 (2012): 20-23.

[37] Karray, Fakhreddine, et al. "Multi modal biometric systems: A state of the art survey." Pattern Analysis and Machine Intelligence Laboratory, University of Waterloo, Waterloo, Canada 2007 (2007).

[38] El-Sayed, A. Y. M. A. N. "Multi-biometric systems: a state of the art survey and research directions." (IJACSA) International Journal of Advanced Computer Science and Applications. 2015. 6.

[39] Schaufele, David, and Guy Cihi. "Biometric-based systems and methods for identity verification." U.S. Patent Application No. 11/356,435.

[40] Rudolf Maarten, et al. "Business system and method using a distorted biometrics." U.S. Patent No. 7,120,607. 10 Oct. 2006.

[41] Tolba, A. S., A. H. El-Baz, and A. A. El-Harby. "Face recognition: A literature review." International Journal of Signal Processing 2.2 (2006): 88-103.

[42] Barnouti, Nawaf Hazim, Sinan Sameer Mahmood Al-Dabbagh, and Wael Esam Matti. "Face recognition: a literature review." International Journal of Applied Information Systems (IJAIS)–ISSN (2016): 2249-0868.

[43] Pornpanomchai, Chomtip, and Chittrapol Inkuna. "Human face recognition by euclidean distance and neural network." Second International Conference on Digital Image Processing. Vol. 7546. International Society for Optics and Photonics, 2010.

[44] De Carrera, Proyecto Fin, and Ion Marques. "Face recognition algorithms." Master's thesis in Computer Science, Universidad Euskal Herriko (2010).

[45] Hjelmas, Erik, and Boon Kee Low. "Face detection: A survey." Computer vision and image understanding 83.3 (2001): 236-274.

[46] Wagner, Andrew, et al. "Toward a practical face recognition system: Robust alignment and illumination by sparse representation." IEEE Transactions on Pattern Analysis and Machine Intelligence 34.2 (2011): 372-386.

[47] Yang, Ming-Hsuan, David J. Kriegman, and Narendra Ahuja. "Detecting faces in images: A survey." IEEE Transactions on pattern analysis and machine intelligence 24.1 (2002): 34-58.

[48] Zhang, Cha, and Zhengyou Zhang. "A survey of recent advances in face detection." (2010).

[49] Yang, Guangzheng, and Thomas S. Huang. "Human face detection in a complex background." Pattern recognition 27.1 (1994): 53-63.

[50] Kotropoulos, Constantine, and Ioannis Pitas. "Rule-based face detection in frontal views." 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing. Vol. 4. IEEE, 1997.

[51] Kakumanu, Praveen, Sokratis Makrogiannis, and Nikolaos Bourbakis. "A survey of skin-color modeling and detection methods." Pattern recognition 40.3 (2007): 1106-1122.

[52] A. F. Clark. CE316: Computer vision (lecture notes). Technical report, University of Essex, 2014.

[53] Hsu, Rein-Lien, Mohamed Abdel-Mottaleb, and Anil K. Jain. "Face detection in color images." IEEE transactions on pattern analysis and machine intelligence 24.5 (2002): 696-706.

[54] Tabassum, Mirza Rehenuma, et al. "Comparative study of statistical skin detection algorithms for sub-continental human images." arXiv preprint arXiv:1008.4206 (2010).

[55] Lakshmipriya, M., and K. Krishnaveni. "A Novel Technique to Detect Face Skin Regions using YCbCr Color Model."

[56] Leung, Thomas K., Michael C. Burl, and Pietro Perona. "Finding faces in cluttered scenes using random labeled graph matching." (1995): 637-644.

[57] Theodoridis, Sergios, and Konstantinos Koutroumbas. "Pattern recognition." (1998).

[58] Cootes, Timothy F., et al. "Active shape models-their training and application." Computer vision and image understanding 61.1 (1995): 38-59.

[59] Hastie, Trevor, et al. "The elements of statistical learning: data mining, inference and prediction." The Mathematical Intelligencer 27.2 (2005): 83-85.

[60] Edwards, Gareth J., Timothy F. Cootes, and Christopher J. Taylor. "Face recognition using active appearance models." European conference on computer vision. Springer, Berlin, Heidelberg, 1998.

[61] Cootes, Timothy F., Gareth J. Edwards, and Christopher J. Taylor. "Active appearance models." IEEE Transactions on Pattern Analysis and Machine Intelligence 6 (2001): 681-685.

[62] Sung, K-K., and Tomaso Poggio. "Example-based learning for view-based human face detection." IEEE Transactions on pattern analysis and machine intelligence 20.1 (1998): 39-51.

[63] Alpaydin, E. "Introduction to machine learning, 2nd edn. Adaptive computation and machine learning." (2010).

[64] Lee, Sanghun, and Chulhee Lee. "Illumination Normalization and Skin Color Validation for Robust Face Detection." Electronic Imaging 2016.19 (2016): 1-6.

[65] Felzenszwalb, Pedro F., et al. "Object detection with discriminatively trained part-based models." IEEE transactions on pattern analysis and machine intelligence 32.9 (2009): 1627-1645.

[66] Huang, Jennifer, Volker Blanz, and Bernd Heisele. "Face recognition using component-based SVM classification and morphable models." International Workshop on Support Vector Machines. Springer, Berlin, Heidelberg, 2002.

[67] Schroff, Florian, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[68] Klare, Brendan F., et al. "Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[69] Sarfraz, M. Saquib, Olaf Hellwich, and Zahid Riaz. "Feature extraction and representation for face recognition." Face recognition. IntechOpen, 2010.

[70] Jafri, Rabia, and R. Hamid. "Arabnia, A survey of face recognition techniques." Journal of information processing systems 5.2 (2009): 41-68.

[71] Lu, Xiaoguang. "Image analysis for face recognition." Personal notes (2003): 36.

[72] Shah, Deepali H., Dr JS Shah, and Dr Tejas V. Shah. "The Exploration of Face Recognition Techniques." International Journal of Application or Innovation in Engg. And Management (IJAIEM) Web Site: www. ijaiem. org Email: editor@ ijaiem. org, editorijaiem@ gmail. com 3.2 (2014).

[73] Parmar, Divyarajsinh N., and Brijesh B. Mehta. "Face recognition methods and applications." arXiv preprint arXiv:1403.0485 (2014).

[74] Kurmi, Uma Shankar, Dheeraj Agrawal, and R. K. Baghel. "Study of different face recognition algorithms and challenges." IJER, Volume 3 (2014): 112-115.

[75] Melišek, Ján Mazanec—Martin, and Miloš Oravec—Jarmila Pavlovicová. "Support vector machines, PCA and LDA in face recognition." J. Electr. Eng 59.203-209 (2008): 1.

[76] Sodhi, Kuldeep Singh, and Madam Lal. "Face recognition using PCA, LDA and various distance classifiers." Journal of Global Research in Computer Science 4.3 (2013): 30-35.

[77] Rozario, Liton Jude, et al. "Quantitative analysis of pca, ica, lda and svm in face recognition." World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering 8.9 (2014): 1613-1616.

[78] Heiselet, B., et al. "Component-based face detection." Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. Vol. 1. IEEE, 2001.

[79] Pflug, Anika, and Christoph Busch. "Ear biometrics: a survey of detection, feature extraction and recognition methods." IET biometrics 1.2 (2012): 114-129.

[80] Jain, Anil K., Arun A. Ross, and Karthik Nandakumar. Introduction to biometrics. Springer Science and Business Media, 2011.

[81] Abaza, Ayman. High performance image processing techniques in automated identification systems. West Virginia University, 2008.

[82] Iannarelli, A. V. "Forensic identification series: ear identification." Paramont, California 5 (1989).

[83] Islam, Syed MS, et al. "Biometric approaches of 2D-3D ear and face: A survey." Advances in computer and information sciences and engineering. Springer, Dordrecht, 2008. 509-514.

[84] AlMahafzah, Harbi, Mohammad Imran, and H. S. Sheshadri. "Multibiometric: Feature level fusion using FKP multi-instance biometric." arXiv preprint arXiv:1210.0818 (2012).

[85] Yan, Ping, and KevinW Bowyer. "Empirical evaluation of advanced ear biometrics." 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops. IEEE, 2005.

[86] Borse, Akshay Sanjay, et al. "EAR Recognition using Artificial Neural Networks." International Journal of Soft Computing and Engineering (IJSCE)', ISSN: 2277-3878.

[87] Yan, Ping, and Kevin W. Bowyer. "Biometric recognition using 3D ear shape." IEEE Transactions on pattern analysis and machine intelligence 29.8 (2007): 1297-1308.

[88] Islam, Syed MS, Mohammed Bennamoun, and Rowan Davies. "Fast and fully automatic ear detection using cascaded adaboost." 2008 IEEE Workshop on Applications of Computer Vision. IEEE, 2008.

[89] Prakash, Surya, Umarani Jayaraman, and Phalguni Gupta. "A skin-color and template based technique for automatic ear detection." 2009 Seventh International Conference on Advances in Pattern Recognition. IEEE, 2009.

[90] Cummings, Alastair H., Mark S. Nixon, and John N. Carter. "A novel ray analogy for enrolment of ear biometrics." 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS). IEEE, 2010.

[91] Burge, Mark, and Wilhelm Burger. "Ear biometrics for machine vision." 21st Workshop of the Austrian Association for Pattern Recognition. 1997.

[92] Moreno, Belé, Angel Sanchez, and José F. Vélez. "On the use of outer ear images for personal identification in security applications." Proceedings IEEE 33rd Annual 1999 International Carnahan Conference on Security Technology (Cat. No. 99CH36303). IEEE, 1999.

[93] Hurley, David J., Mark S. Nixon, and John N. Carter. "Force field feature extraction for ear biometrics." Computer Vision and Image Understanding 98.3 (2005): 491-512.

[94] Ali, M., M. Y. Javed, and A. Basit. "Ear recognition using wavelets." Proceedings of Image and Vision Computing New Zealand. 2007.

[95] Victor, Barnabas, Kevin Bowyer, and Sudeep Sarkar. "An evaluation of face and ear biometrics." Object recognition supported by user interaction for service robots. Vol. 1. IEEE, 2002.

[96] Chang, Kyong, et al. "Comparison and combination of ear and face images in appearance-based biometrics." IEEE Transactions on pattern analysis and machine intelligence 25.9 (2003): 1160-1165.

[97] Antakis, Spyridon. "A Survey on Ear Recognition." IEEE Conference on Neural Networks for Signal Processing IX. 2009.

[98] Nguyen, Van-toi, et al. "A method for hand detection based on Internal Haar-like features and Cascaded AdaBoost Classifier." The International Conference on Communications and Electronics (ICCE). 2012.

[99] Wu, Ying, and Thomas S. Huang. "Hand modeling, analysis and recognition." IEEE Signal Processing Magazine 18.3 (2001): 51-60.

[100] Duta, Nicolae. "A survey of biometric technology based on hand shape." Pattern Recognition 42.11 (2009): 2797-2806.

[101] Ross, Arun, Anil Jain, and S. Pankati. "A prototype hand geometry-based verification system." Proceedings of 2nd conference on audio and video based biometric person authentication. 1999.

[102] Sanchez-Reillo, Raul. "Hand geometry pattern recognition through gaussian mixture modelling." Proceedings 15th International Conference on Pattern Recognition. ICPR-2000. Vol. 2. IEEE, 2000.

[103] Yuan, Quan, Stan Sclaroff, and Vassilis Athitsos. "Automatic 2D hand tracking in video sequences." 2005 Seventh IEEE Workshops on Applications of Computer Vision (WACV/MOTION'05)-Volume 1. Vol. 1. IEEE, 2005.

[104] Zhang, Zhong, Rommel Alonzo, and Vassilis Athitsos. "Experiments with computer vision methods for hand detection." Proceedings of the 4th International Conference on PErvasive Technologies Related to Assistive Environments. ACM, 2011.

[105] Waibel, Xiaojin Zhu Jie Yang Alex. "Segmenting hands of arbitrary color." IEEE the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'000). Vol. 446. 2000.

[106] Ong, Eng-Jon, and Richard Bowden. "A boosted classifier tree for hand shape detection." Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.. IEEE, 2004.

[107] Tran, T. T. H., and T. T. M. Nguyen. "Invariant lighting hand posture classification." IEEE international conference on Progress in Informatics and Computing (PIC). 2010.

[108] Athitsos, Vassilis, et al. "Detecting instances of shape classes that exhibit variable structure." European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2006.

[109] Coughlan, James, et al. "Efficient deformable template detection and localization without user initialization." Computer Vision and Image Understanding 78.3 (2000): 303-319.

[110] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." CVPR (1) 1.511-518 (2001): 3.

[111] Dewangan, Dhaneshwar Prasad, and Abhishek Pandey. "A survey on security in palmprint recognitio: a biometric trait." Int. J. Adv. Res. Comput. Eng. Technol.(IJARCET) 1 (2012): 347.

[112] Sumalatha, K. A., and H. Harsha. "Biometric Palmprint Recognition System." A Review International Journal of Advanced Research in Computer Science and Software Engineering 4.1 (2014).

[113] Han, Chin-Chuan. "A hand-based personal authentication using a coarse-to-fine strategy." Image and Vision Computing 22.11 (2004): 909-918.

[114] Dai, Jifeng, Jianjiang Feng, and Jie Zhou. "Robust and efficient ridge-based palmprint matching." IEEE transactions on pattern analysis and machine intelligence 34.8 (2011): 1618-1632.

[115] Cappelli, Raffaele, Matteo Ferrara, and Dario Maio. "A fast and accurate palmprint recognition system based on minutiae." IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 42.3 (2012): 956-962.

[116] Azizi, Amir, and Hamid Reza Pourreza. "Efficient iris recognition through improvement of feature extraction and subset selection." arXiv preprint arXiv:0906.4789 (2009).

[117] Lim, Shinyoung, et al. "Efficient iris recognition through improvement of feature vector and classifier." ETRI journal 23.2 (2001): 61-70.

[118] Boles, Wageeh W., and Boualem Boashash. "A human identification technique using images of the iris and wavelet transform." IEEE transactions on signal processing 46.4 (1998): 1185-1188.

[119] Rai, Himanshu, and Anamika Yadav. "Iris recognition using combined support vector machine and Hamming distance approach." Expert systems with applications 41.2 (2014): 588-593.

[120] Umer, Saiyed, Bibhas Chandra Dhara, and Bhabatosh Chanda. "A novel cancelable iris recognition system based on feature learning techniques." Information Sciences 406 (2017): 102-118.

[121] Sharma, Rahul, Nidhi Mishra, and Sanjeev Kumar Yadav. "Fingerprint Recognition System and Tehniques: A Survey." International Journal of Scientific and Engineering Research 4.6 (2013): 1670.

[122] Thaiyalnayaki, K., S. Syed Abdul Karim, and P. Varsha Parmar. "Finger print recognition using discrete wavelet transform." International Journal of Computer Applications 1.24 (2010): 96-100.

[123] Miguel Lastra, Jesús Carabaño, Pablo D Gutiérrez, José M Benítez, and Francisco Herrera. Fast fingerprint identification using gpus. Information Sciences, 301:195–214, 2015.

[124] Yuan, Chengsheng, Xingming Sun, and Rui Lv. "Fingerprint liveness detection based on multi-scale LPQ and PCA." China Communications 13.7 (2016): 60-65.

[125] Peralta, Daniel, et al. "Minutiae-based fingerprint matching decomposition: methodology for big data frameworks." Information Sciences 408 (2017): 198-212.

[126] Wang, Liang, ed. Behavioral Biometrics for Human Identification: Intelligent Applications: Intelligent Applications. IGI Global, 2009.

[127] Godil, Afzal, Patrick Grother, and Sandy Ressler. "Human identification from body shape." Fourth International Conference on 3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings.. IEEE, 2003.

[128] Gross, Ralph, and Jianbo Shi. "The cmu motion of body (mobo) database." (2001).

[129] Prabhakar, Salil, Sharath Pankanti, and Anil K. Jain. "Biometric recognition: Security and privacy concerns." IEEE security and privacy 2 (2003): 33-42.

[130] Jain, Anil K., and Arun Ross. "Multibiometric systems." Communications of the ACM 47.1 (2004): 34-40.

[131] Kong, Seong G., et al. "Recent advances in visual and infrared face recognition—a review." Computer Vision and Image Understanding 97.1 (2005): 103-135.

[132] Lee, Jinho, et al. "Finding optimal views for 3D face shape modeling." Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004. Proceedings.. IEEE, 2004.

[133] Pan, Zhihong, et al. "Face recognition in hyperspectral images." IEEE Transactions on Pattern Analysis and Machine Intelligence 25.12 (2003): 1552-1560.

[134] Rowe, Robert K., and Kristin A. Nixon. "Fingerprint enhancement using a multispectral sensor." Biometric Technology for Human Identification II. Vol. 5779. International Society for Optics and Photonics, 2005.

[135] Jaafar, Haryati, and Dzati Athiar Ramli. "A review of multibiometric system with fusion strategies and weighting factor." International Journal of Computer Science Engineering (IJCSE) 2.4 (2013): 158-165.

[136] Abidi, Besma, and Mongi A. Abidi, eds. Face biometrics for personal identification: multi-sensory multi-modal systems. Springer Science and Business Media, 2007.

[137] Lu, Xiaoguang, Yunhong Wang, and Anil K. Jain. "Combining classifiers for face recognition." icme. Vol. 3. 2003.

[138] Chang, Kyong I., Kevin W. Bowyer, and Patrick J. Flynn. "Face recognition using 2D and 3D facial data." Workshop in Multidimonal User Authentication pp25-32. 2003.

[139] Imran, Mohammad, Ashok Rao, and G. Hemantha Kumar. "Multibiometric systems: A comparative study of multi-algorithmic and multimodal approaches." Procedia Computer Science 2 (2010): 207-212.

[140] Chang, Kyong I., Kevin W. Bowyer, and Patrick J. Flynn. "An evaluation of multimodal 2D+ 3D face biometrics." IEEE transactions on pattern analysis and machine intelligence 27.4 (2005): 619-624.

[141] Samad, Salina Abdul, Dzati Athiar Ramli, and Aini Hussain. "A multi-sample single-source model using spectrographic features for biometric authentication." 2007 6th International Conference on Information, Communications and Signal Processing. IEEE, 2007.

[142] Jang, Jain, et al. "Multi-unit iris recognition system by image check algorithm." International Conference on Biometric Authentication. Springer, Berlin, Heidelberg, 2004.

[143] Prabhakar, Salil, and Anil K. Jain. "Decision-level fusion in fingerprint verification." International Workshop on Multiple Classifier Systems. Springer, Berlin, Heidelberg, 2001.

[144] Jain, Anil, Karthik Nandakumar, and Arun Ross. "Score normalization in multimodal biometric systems." Pattern recognition 38.12 (2005): 2270-2285.

[145] Brunelli, Roberto, et al. "Automatic person recognition by using acoustic and geometric features." (1993).

[146] Kittler, Josef, et al. "On combining classifiers." IEEE transactions on pattern analysis and machine intelligence 20.3 (1998): 226-239.

[147] Hong, Lin, Anil K. Jain, and Sharath Pankanti. "Can multibiometrics improve performance?." Proceedings AutoID. Vol. 99. Citeseer, 1999.

[148] Ramli, Dzati Athiar, Salina Abdul Samad, and Aini Hussain. "A multibiometric speaker authentication system with SVM audio reliability indicator." IAENG International Journal of Computer Science 36.4 (2009): 313-321.

[149] Lip, Chia Chin, and Dzati Athiar Ramli. "Comparative study on feature, score and decision level fusion schemes for robust multibiometric systems." Frontiers in computer education. Springer, Berlin, Heidelberg, 2012. 941-948.

[150] Hong, Lin, and Anil Jain. "Integrating faces and fingerprints for personal identification." IEEE transactions on pattern analysis and machine intelligence 20.12 (1998): 1295-1307.

[151] Frischholz, Robert W., and Ulrich Dieckmann. "BiolD: a multimodal biometric identification system." Computer 33.2 (2000): 64-68.

[152] Fièrrez-Aguilar, Julian, et al. "A comparative evaluation of fusion strategies for multimodal biometric verification." International Conference on Audio-and Video-Based Biometric Person Authentication. Springer, Berlin, Heidelberg, 2003.

[153] Ross, Arun, and Anil Jain. "Information fusion in biometrics." Pattern recognition letters 24.13 (2003): 2115-2125.

[154] Kumar, Ajay, et al. "Personal verification using palmprint and hand geometry biometric." International Conference on Audio-and Video-Based Biometric Person Authentication. Springer, Berlin, Heidelberg, 2003.

[155] Toh, K-A., Xudong Jiang, and Wei-Yun Yau. "Exploiting global and local decisions for multimodal biometrics verification." IEEE Transactions on Signal Processing 52.10 (2004): 3059-3072.

[156] Camlikaya, Eren, Alisher Kholmatov, and Berrin Yanikoglu. "Multi-biometric templates using fingerprint and voice." Biometric technology for human identification V. Vol. 6944. International Soci-

ety for Optics and Photonics, 2008.

[157] Shahin, M. K., A. M. Badawi, and M. E. Rasmy. "A multimodal hand vein, hand geometry, and fingerprint prototype design for high security biometrics." 2008 Cairo International Biomedical Engineering Conference. IEEE, 2008.

[158] Chellin, George. "Performance analysis of multimodal biometric system authentication." (2009).

[159] Poinsot, Audrey, Fan Yang, and Michel Paindavoine. "Small sample biometric recognition based on palmprint and face fusion." 2009 Fourth International Multi-Conference on Computing in the Global Information Technology. IEEE, 2009.

[160] Soltane, Mohamed, Noureddine Doghmane, and Noureddine Guersi. "Face and speech based multi-modal biometric authentication." International Journal of Advanced Science and Technology 21.6 (2010): 41-56.

[161] Bokade, Gayatri Umakant, and Ashok M. Sapkal. "Feature level fusion of palm and face for secure recognition." International Journal of Computer and Electrical Engineering 4.2 (2012): 157.

[162] Dhameliya, Mitul D., and Jitendra P. Chaudhari. "A multimodal biometric recognition system based on fusion of palmprint and fingerprint." International journal of Engineering trends and technology 4.5 (2013): 1908-1911.

[163] Gogoi, Minakshi, and Dhruba Kr Bhattacharyya. "Fusion of fingerprint and iris biometrics using binary ant colony optimization." Proceedings of the Third International Conference on Soft Computing for Problem Solving. Springer, New Delhi, 2014.

[164] Benaliouche, Houda, and Mohamed Touahria. "Comparative study of multimodal biometric recognition by fusion of iris and fingerprint." The Scientific World Journal 2014 (2014).

[165] Khodadoust, Javad, et al. "Design and implementation of a multibiometric system based on hand's traits." Expert Systems with Applications 97 (2018): 303-314.

[166] M.Chandrasekhar Reddy and Ch.Hima Bindu. Biometric quality on finger, face and iris identification. In International Journal and Magazine of Engineering, Technology, Mangement and Research, pages 545–551. A peer Reviewed Open Access International Jounrnal, 2017.

[167] Vatsa, Mayank, et al. "Belief function theory based biometric match score fusion: Case studies in multi-instance and multi-unit iris verification." 2009 Seventh International Conference on Advances in Pattern Recognition. IEEE, 2009.

[168] Nandakumar, Karthik. Multibiometric systems: Fusion strategies and template security. MICHIGAN STATE UNIV EAST LANSING DEPT OF COMPUTER SCIENCE/ENGINEERING, 2008.

[169] Jain, Anil K., and Arun Ross. "Multibiometric systems." Communications of the ACM 47.1 (2004): 34-40.

[170] Ross, Arun A., Karthik Nandakumar, and Anil K. Jain. Handbook of multibiometrics: human recognition systems. Springer, 2006.

[171] Lam, Louisa, and S. Y. Suen. "Application of majority voting to pattern recognition: an analysis of its behavior and performance." IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 27.5 (1997): 553-568.

[172] Yang, Bian, et al. "Decision level fusion of fingerprint minutiae based pseudonymous identifiers." 2011 International Conference on Hand-Based Biometrics. IEEE, 2011.

[173] AlMahafzah, Harbi, and Maen Zaid AlRwashdeh. "A survey of multibiometric systems." arXiv preprint arXiv:1210.0829 (2012).

[174] Kiltter, Josef, and Norman Poh. "Multibiometrics for identity authentication: Issues, benefits and challenges." 2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems. IEEE, 2008.

[175] Sutrop, Margit. "Ethical issues in governing biometric technologies." International Conference on Ethics and Policy of Biometrics. Springer, Berlin, Heidelberg, 2010.

[176] Alterman, Anton. "“A piece of yourself”: Ethical issues in biometric identification." Ethics and information technology 5.3 (2003): 139-150.

[177] ORL database of faces. At & t laboratories. U.K. [Online] Available: http://www.uk.research.att.com/facedatabase.html, 1992-1994.

[178] Caltech Face Database, Face Recognition Homepage-databases. [Online] Available: http://www.face-rec.org/databases.

[179] CASIA Face Image Database Version 5.0. [Online] Available: https://github.com/joyhuang-note/dataset-collection/issues/5.

[180] AMI database by Esther Gonzalez, Luis Alvarez and Luis Mazorra is licensed under a Creative Commons Reconocimiento-NonCommercial-SinObraDerivada 3.0 Unported License.

[181] Kumar, Ajay, and Chenye Wu. "Automated human identification using ear imaging." Pattern Recognition 45.3 (2012): 956-968.

[182] Kim, Tae-Kyun, Shu-Fai Wong, and Roberto Cipolla. "Tensor canonical correlation analysis for action classification." 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007.

[183] Tan, T., and Z. Sun. "CASIA-IrisV4,” Chinese Academy of Sciences Institute of Automation." (2005).

[184] A. Gallagher and T. Chen. Understanding images of groups of people. In Proc. CVPR, 2009.

[185] Bora, Mr, et al. "Effect of different distance measures on the performance of K-means algorithm: an experimental study in Matlab." arXiv preprint arXiv:1405.7471 (2014).

[186] Kumar, Ajay, and Sumit Shekhar. "Personal identification using multibiometrics rank-level fusion." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 41.5 (2010): 743-752.

[187] Nixon, Mark. "Southampton MultiBiometric Tunnel Database." (2006).

[188] Seely, Richard D., et al. "The university of southampton multi-biometric tunnel and introducing a novel 3d gait dataset." 2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems. IEEE, 2008.

[189] WestVirginiaUniversit. West virginia university biometric data sets, 2008.

[190] Fawcett, Tom. "An introduction to ROC analysis." Pattern recognition letters 27.8 (2006): 861-874.

[191] Davis, Jesse, and Mark Goadrich. "The relationship between Precision-Recall and ROC curves." Proceedings of the 23rd international conference on Machine learning. ACM, 2006.

[192] A. Clark and C. Clark. Performance characterization in computer vision. Technical report, University of Essex, 2003.

[193] Powers, David Martin. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." (2011).

[194] Kanwal, Nadia. Low-level image features and navigation systems for visually impaired people. Diss. University of Essex, 2013.

[195] Yimyam, Panitnat, and Adrian F. Clark. "Agricultural produce grading by computer vision using genetic programming." 2012 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2012.

[196] Viola, Paul, and Michael J. Jones. "Robust real-time face detection." International journal of computer vision 57.2 (2004): 137-154.

[197] Jain, Maneela, and Pushpendra Singh Tomar. "Review of image classification methods and techniques." International journal of engineering research and technology 2.8 (2013).

[198] Jones, Michael, and Paul Viola. "Fast multi-view face detection." Mitsubishi Electric Research Lab TR-20003-96 3.14 (2003): 2.

[199] Chaudhari, Monali, Shanta Sondur, and Gauresh Vanjare. "A review on Face Detection and study of Viola Jones method." International Journal of Computer Trends and Technology (IJCTT) 25.1 (2015): 54-61.

[200] Sharifara, Ali, Mohd Shafry Mohd Rahim, and Yasaman Anisi. "A general review of human face detection including a study of neural networks and Haar feature-based cascade classifier in face detection." 2014 International Symposium on Biometrics and Security Technologies (ISBAST). IEEE, 2014.

[201] Sural, Shamik, Gang Qian, and Sakti Pramanik. "Segmentation and histogram generation using the HSV color space for image retrieval." Proceedings. International Conference on Image Processing. Vol. 2. IEEE, 2002.

[202] Ganesan, P., et al. "Satellite image segmentation based on YCbCr color space." Indian Journal of Science and Technology 8.1 (2015): 35.

[203] Setiawan, Nurul Arif, et al. "Gaussian mixture model in improved hls color space for human silhouette extraction." International Conference on Artificial Reality and Telexistence. Springer, Berlin, Heidelberg, 2006.

[204] Ghahramani, Z., et al. "Chapter:'Unsupervised Learning'in Advanced Lectures on Machine Learning." Springer (2004).

[205] Jones, Michael J., and Paul Viola. "Method and system for object detection in digital images." U.S. Patent No. 7,099,510. 29 Aug. 2006.

[206] Cha, Sung-Hyuk. "Comprehensive survey on distance/similarity measures between probability density functions." City 1.2 (2007): 1.

[207] Greenacre, Michael, and Raul Primicerio. "Measures of distance between samples: Euclidean." Fundacion BBVA Publication (December 2013) (2008): 978-84.

[208] Galeano, Pedro, Esdras Joseph, and Rosa E. Lillo. "The Mahalanobis distance for functional data with applications to classification." Technometrics 57.2 (2015): 281-291.

[209] De Maesschalck, Roy, Delphine Jouan-Rimbaud, and Désiré L. Massart. "The mahalanobis distance." Chemometrics and intelligent laboratory systems 50.1 (2000): 1-18.

[210] Eleyan, Alaa, and Hasan Demirel. Pca and lda based neural networks for human face recognition. Vol. 558. INTECH Open Access Publisher, 2007.

[211] Pissarenko, Dimitri. "Eigenface-based facial recognition." December 1st (2002).

[212] Naz, Erum, Umar Farooq, and Tabbasum Naz. "Analysis of principal component analysis-based and fisher discriminant analysis-based face recognition algorithms." 2006 International Conference on Emerging Technologies. IEEE, 2006.

[213] Singh, Abhishek, and Saurabh Kumar. Face recognition using pca and eigen face approach. Diss. 2012.

[214] Turk, Matthew, and Alex Pentland. "Eigenfaces for recognition." Journal of cognitive neuroscience 3.1 (1991): 71-86.

[215] Wagner, Philipp. "Face recognition with python." Tersedia dalam: www. bytefish. de [diakses pada 16 Februari 2015] (2012).

[216] Martínez, Aleix M., and Avinash C. Kak. "Pca versus lda." IEEE transactions on pattern analysis and machine intelligence 23.2 (2001): 228-233.

[217] Zhang, Cheng-Yuan, and Qiu-Qi Ruan. "Face Recognition Using L-Fisherfaces." Journal of Information Science and Engineering 26.4 (2010).

[218] MartõÂnez, Aleix M. "Member, ieee, and avinash c. kak,“pca versus lda”." IEEE Transactions on Pattern Analysis and machine intelligence 23.2 (2001).

[219] Lyu, Michael. "Face Recognition Algorithms Review Term Paper-December 2001."

[220] Salh, Thair A., and Mustafa Z. Nayef. "Face Recognition System based on Wavelet, PCA-LDA and SVM." Computer Engineering and Intelligent Systems Journal 4.3 (2013): 26-31.

[221] Weston, Jason. "Support Vector Machine (and Statistical Learning Theory) Tutorial. 4 Independence Way." Princeton, USA.

[222] Markowetz, Florian, Lutz Edler, and Martin Vingron. "Support vector machines for protein fold class prediction." Biometrical Journal: Journal of Mathematical Methods in Biosciences 45.3 (2003): 377-389.

[223] Grgic, Mislav, and Kresimir Delac. "Face recognition homepage." Zagreb, Croatia (www. face-rec. org/databases) 324 (2013).

[224] Kumar, Ajay. "Incorporating cohort information for reliable palmprint authentication." 2008 Sixth Indian Conference on Computer Vision, Graphics and Image Processing. IEEE, 2008.

[225] Mishra, Annapurna, Monorama Swain, and Bodhisattva Dash. "An Approach to Face Recognition of 2-D Images Using Eigen Faces and PCA." Signal and Image Processing 3.2 (2012): 143.

[226] Ahmed, Hythem, Jedra Mohamed, and Zahid Noureddine. "Face recognition systems using relevance weighted two dimensional linear discriminant analysis algorithm." Journal of Signal and Information Processing 3.1 (2012): 130-135.

[227] Chen, Yu-Sheng, Guangjun Su, and Haihong Li. "Machine Learning for Calligraphy Styles Recognition." (2016).

[228] LeCun, Yann, et al. "Gradient-based learning applied to document recognition." Proceedings of the IEEE 86.11 (1998): 2278-2324.

[229] Arwa M Basbrain, Inas Al-Taie, Nassr Azeez, John Q Gan, and Adrian Clark. Shallow convolutional neural network for eyeglasses detection in facial images. In Computer Science and Electronic Engineering (CEEC), 2017, pages 157–161. IEEE, 2017.

[230] Xiao, Kecong, Zishuai Zhang, and Jun Wu. "Chinese text sentiment analysis based on improved Convolutional Neural Networks." 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS). IEEE, 2016.

[231] Dumoulin, Vincent, and Francesco Visin. "A guide to convolution arithmetic for deep learning." arXiv preprint arXiv:1603.07285 (2016).

[232] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.

[233] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.

[234] Harley, Adam W. "An interactive node-link visualization of convolutional neural networks." International Symposium on Visual Computing. Springer, Cham, 2015.

[235] Schmidhuber, Jürgen. "Deep learning in neural networks: An overview." Neural networks 61 (2015): 85-117.

[236] Zhao, Wenlai, et al. "F-CNN: An FPGA-based framework for training Convolutional Neural Networks." 2016 IEEE 27th International Conference on Application-specific Systems, Architectures and Processors (ASAP). IEEE, 2016.

[237] Jia, Yangqing, et al. "Caffe: Convolutional architecture for fast feature embedding." Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014.

[238] Russakovsky, Olga, et al. "Imagenet large scale visual recognition challenge." International journal of computer vision 115.3 (2015): 211-252.

[239] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[240] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[241] Al-Qizwini, Mohammed, et al. "Deep learning algorithm for autonomous driving using googlenet." 2017 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2017.

[242] Kaya, Yigitcan, Sanghyun Hong, and Tudor Dumitras. "Shallow-Deep Networks: Understanding and Mitigating Network Overthinking." International Conference on Machine Learning. 2019.