A HYBRID MULTIBIOMETRIC SYSTEM FOR PERSONAL IDENTIFICATION BASED ON FACE AND IRIS TRAITS

The Development of an automated computer system for the identification of humans by integrating facial and iris features using Localization, Feature Extraction, Handcrafted and Deep learning Techniques.

Alaa Sulaiman Nassar NASSAR

Submitted for the Degree of Doctor of Philosophy from the University of Bradford

School of Electrical Engineering and Computer Science

University of Bradford

ABSTRACT

Alaa Sulaiman Nassar NASSAR

A Hybrid Multibiometric System for Personal Identification Based on Face and Iris Traits

The Development of an automated computer system for the identification of humans by integrating facial and iris features using Localization, Feature Extraction, Handcrafted and Deep learning Techniques.

Keywords: Multimodal multibiometric systems, Face recognition, Iris recognition, Iris localization, Deep learning, Feature extraction, Curvelet transform, Fractal dimension, Deep belief network, Convolutional neural network.

Multimodal biometric systems have been widely applied in many real-world applications due to its ability to deal with a number of significant limitations of unimodal biometric systems, including sensitivity to noise, population coverage, intra-class variability, non-universality, and vulnerability to spoofing. This PhD thesis is focused on the combination of both the face and the left and right irises, in a unified hybrid multimodal biometric identification system using different fusion approaches at the score and rank level.

Firstly, the facial features are extracted using a novel multimodal local feature extraction approach, termed as the *Curvelet-Fractal* approach, which based on merging the advantages of the Curvelet transform with Fractal dimension. Secondly, a novel framework based on merging the advantages of the local handcrafted feature descriptors with the deep learning approaches is proposed, Multimodal Deep Face Recognition (MDFR) framework, to address the face recognition problem in unconstrained conditions. Thirdly, an efficient deep learning system is employed, termed as *IrisConvNet*, whose architecture is based on a combination of Convolutional Neural Network (CNN) and Softmax classifier to extract discriminative features from an iris image.

Finally, The performance of the unimodal and multimodal systems has been evaluated by conducting a number of extensive experiments on large-scale unimodal databases: FERET, CAS-PEAL-R1, LFW, CASIA-Iris-V1, CASIA-Iris-V3 Interval, MMU1 and IITD and MMU1, and SDUMLA-HMT multimodal dataset. The results obtained have demonstrated the superiority of the proposed systems compared to the previous works by achieving new state-of-the-art recognition rates on all the employed datasets with less time required to recognize the person's identity.

Acknowledgements

First and foremost, I would like to express my deepest gratitude and numerous thanks to the most merciful, the most beneficent, and the most gracious Almighty Allah who gave me the courage and motivation to complete the research successfully.

This thesis is a collection of not only hard work, perseverance and continuous effort over the past four years, but also encouragement, cooperation and support from many people who deserve to be acknowledged.

I am heartily thankful to my principal supervisor Prof. Rami Qahwaji, whose constant encouragement, guidance and considerable support from the initial to the final stage enabled me to develop an understanding of the subject. Prof. Qahwaji is a supervisor, who truly made a difference in my life, I appreciate his vast knowledge and skill in many areas, whose expertise, understanding, and patience, added considerably to my graduate experience, and without him, the completion of this work could not have been achieved. Besides of being an excellent supervisor, Prof. Qahwaji encouraging me by his personal kindness, and he always was available when I needed his advises.

I also want to express deepest gratitude and thanks to my second supervisor, Dr. Stanly Ipson, whose guidance, keen interest and continuous efforts enabled me to complete this work. Dr. Ipson has offered much advice and insight throughout my research. I have been extremely lucky to work with Dr. Ipson who cared so much about my work, and who responded to my questions and queries so promptly.

I am extremely grateful to my parents for their love, prayers, caring and sacrifices for educating and preparing me for my future. I am very much thankful to my wife (Shumoos Al-Fahdawi) and my sons (Fahad and Taim) for their love, understanding, prayers and continuing support to complete this research work. Also, I express my thanks to my brothers and sisters for their continuous love, support, understanding and good wishes whenever I needed.

I should not forget to acknowledge all my lab members and colleagues for their advice and their willingness to share their bright thoughts with me, for shaping up my research.

Last but certainly not least, I would also like to thanks to people and the government of my country, the Republic of Iraq. I could not have gone through the doctoral program overseas without their financial support. I would like to express my full appreciation to the staff of the Higher Committee for Education Development in Iraq, for their generous support, both financial and moral.

Alaa S. Al-Waisy

Dedication

This work is humbly dedicated to all my valuable treasures in my life:

To my parents, brothers and sisters,

To my beautiful wife, Shumoos Al-Fahdawi

and

To my little sons, Fahad and Taim

Alaa S. Al-Waisy

List of Publications

Journal Papers:

- Alaa S. Al-Waisy, Rami Qahwaji, Stanley Ipson, Shumoos Al-Fahdawi, "A multimodal deep learning framework using local feature representations for face recognition", *Machine Vision and Applications-Springer*., vol. 29, no. 1, pp. 35–54, 2017.
- Alaa S. Al-Waisy, Rami Qahwaji, Stanley Ipson, Shumoos Al-Fahdawi, "A Multi-biometric Iris Recognition System Based on a Deep Learning Approach" , *Pattern Analysis and Applications-Springer*., pp. 1–20, 2017.
- Shumoos Al-Fahdawi, Rami Qahwaji, Alaa S. Al-Waisy, Stanley Ipson, Rayaz A. Malik, Arun Brahma, and Xin Chen, "A fully automatic nerve segmentation and morphometric parameter quantification system for early diagnosis of diabetic neuropathy in corneal images," *Comput. Methods Programs Biomed*., vol. 135, pp. 151–166, 2016.
- Tarek A M Hamad Nagem, Rami Qahwaji, Stan Ipson, Zhiguang Wang, and Alaa S. Al-Waisy, "Deep Learning Technology for Predicting Solar Flares from (Geostationary Operational Environmental Satellite) Data," *International Journal of Advanced Computer Science and Applications (IJACSA),* vol. 9, no. 1, pp. 492-498, 2018.
- Shumoos Al-Fahdawi, Rami Qahwaji, Alaa S. Al-Waisy, Stanley Ipson, Maryam Ferdousi, Rayaz A. Malik, and Arun Brahma, "A Fully Automated Cell Segmentation and Morphometric Parameter System for Quantifying Corneal Endothelial Cell Morphology," *Comput. Methods Programs Biomed.*, vol. 160, pp. 11–23, 2018.

Conference Papers:

- Alaa S. Al-Waisy, Rami Qahwaji, Stanley Ipson, Shumoos Al-Fahdawi, "A Fast and Accurate Iris Localization Technique for Healthcare Security System", *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing,* pp. 1028-1034, 2015.
- Alaa S. Al-Waisy, Rami Qahwaji, Stanley Ipson, Shumoos Al-Fahdawi, "A Robust Face Recognition System Based on Curvelet and Fractal Dimension Transforms", *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and*

Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, pp. 548-555, 2015.

- Alaa S. Al-Waisy, Rami Qahwaji, Stanley Ipson, Shumoos Al-Fahdawi, "Multimodal Biometric System for Personal Identification Based on Deep Learning Approaches", *2017 Seventh International Conference on Emerging Security Technologies (EST),* pp.163-168, 2017.
- Shumoos Al-Fahdawi, Rami Qahwaji, Alaa S. Al-Waisy, and Stanley Ipson, "An automatic corneal subbasal nerve registration system using FFT and phase correlation techniques for an accurate DPN diagnosis," *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, pp. 1035–1041, 2015.

Demos & Presentations:

- Alaa S. Al-Waisy, Rami Qahwaji, Stanley Ipson, Shumoos Al-Fahdawi, "A Multi-biometric Iris Recognition System Based on a Deep Learning Approach", *University of Bradford*, 12 Oct. 2017.
- Shumoos Al-Fahdawi, Rami Qahwaji, Alaa S. Al-Waisy, Stanley Ipson, Rayaz A. Malik, Arun Brahma, Xin Chen, "Automated Diagnosis of Diabetic Neuropathy from Corneal Images", *The Digital Health and Wellbeing Festival at the Digital Exchange, University of Bradford*, Wed. 29-Thu. 30 June, 2016.
- Shumoos Al-Fahdawi, Rami Qahwaji, Alaa S. Al-Waisy, Stanley Ipson, Rayaz A. Malik, Arun Brahma, "A Fully Automated Cell Segmentation and Morphometric Parameter System for Quantifying Corneal Endothelial Cell Morphology", *Digital Health Enterprise Zone Academic Launch Event, DHEZ Academic, Phoenix, University of Bradford*, Sep. 29, 2017.

Posters:

- Alaa S. Al-Waisy, Rami Qahwaji, Stanley Ipson, Shumoos Al-Fahdawi. "A Multimodal Deep Learning Framework Using Local Feature Representations for Face Recognition", Accepted for presentation at the 1st Annual Innovative Postgraduate Research Conference (1st AIERC), University of Bradford, July 17, 2017.
- Shumoos Al-Fahdawi, Rami Qahwaji, Alaa S. Al-Waisy, Stanley Ipson, Rayaz A. Malik, Arun Brahma, Xin Chen "A Fully Automatic Nerve Segmentation and Morphometric Parameter Quantification

System for Early Diagnosis of Diabetic Neuropathy in Corneal Images", Accepted for presentation at the 1st Annual Innovative Postgraduate Research Conference (1st AIERC), University of Bradford, July 17, 2017.

- Shumoos Al-Fahdawi, Rami Qahwaji, Alaa S. Al-Waisy, Stanley Ipson, Rayaz A. Malik, Arun Brahma, "Automated Diagnosis of Diabetic Neuropathy from Corneal Images" , *Digital Health Enterprise Zone Academic Launch Event, DHEZ Academic, Phoenix, University of Bradford*, Sep. 29, 2017.
- Rami Qahwaji, Shumoos Al-Fahdawi, Alaa Al-Waisy, Rayaz A. Malik, Maryam Ferdousi, and Arun Brahma, "Novel Imaging Technologies for the Quantification of Nerve Structures from Corneal Confocal Images ", *The transformative potential of data and image analysis for eye care, The Royal Society, London, 6-9 Carlton House Terrace,* April, 23-24, 2018.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Symbol	Definition
AAM	Active Appearance Model
AAR	Average Accuracy Rate
AER	Average Error Rate
ANN	Artificial Neural Networks
ASM	Active Shape Models
BC	Box Counting
BDBN	Boosted Deep Belief Network
BGS	Beacon Guided Search
BPSO	Binary Particle Swarm Optimization
BSA	Backtracking Search Algorithm
BSSR1	Biometric Scores Set Release-1
CC	Correlation Coefficients
CCHT	Coherent Circular Hough Transform
CD	Contrastive Divergence
CDBN	Convolutional Deep Belief Networks
CER	Classification Error Rate
CHT	Circular Hough Transform
CMC	Cumulative Match Characteristic
CNN	Convolutional Neural Network
CRR	Correct Recognition Rate
CS-LBP	Centre-Symmetric Local Binary Pattern
CSML	Cosine Similarity Metric Learning
CT-FBM	Curvelet Transform-Fractional Brownian Motion
CWT	Coiflet Wavelet Transform
DBNs	Deep Belief Networks
DCT	Discrete Cosine Transform
DCTG1	first generation Discrete Curvelet Transform
DDML	Discriminative Deep Metric Learning
DeepID2	Deep Identification-verification features
DFD	Discriminant Face Descriptor
DNNs	Deep Neural Networks
DWT	Discrete Wavelet Transform
EBGM	Elastic Bunch Graph Matching
EER	Equal Error Rate
EMD	Empirical Mode Decomposition
EMR	Electronic Medical Records
FAR	False Acceptance Rate
FBM	Fractional Brownian Motion
FDCT	Fast Discrete Curvelet Transform
FER	Failure Enrollment Rate
FERET	FacE REcognition Technology
FFNN	Feed Forward Neural Network
FFT	Fast Fourier Transform
FLD	Fisher’s Linear Discriminant
FMR	False Matching Rate
FRR	False Rejection Rate

FV	Fisher Vector
GAC	Geodesic Active Contours
GI	Gradient Isolation
GLCM	Gray Level Co-occurrence Matrices
GMMs	Gaussian Mixture Models
GPUs	Graphics Processing Units
GUIs	The Graphical User Interfaces
GVF	Gradient Vector Field
ICA	Independent Component Analysis
IDBC	Improved Differential Box Counting
iGRVM	incremental Granular Relevance Vector Machine
ITML	Information-Theoretic Metric Learning
KDDA	Kernel Direct Discriminant Analysis
KFDA	Kernel Fisher Discriminant Analysis
K-NN	K-Nearest Neighbour
LBP	Local Binary Patterns
LCRBM	Local Convolutional Restricted Boltzmann Machines
LDA	Linear Discriminant Analysis
LDML	Logistic Discriminant based Metric Learning
LFW	Labeled Faces in the Wild
LGXP	Local Gabor XOR Pattern
LPP	Locality Preserving Projections
LQP	Local Quantized Patterns
LSML	Large Scale Metric Learning
MDFR	Multimodal Deep Face Recognition
MLP	Multi-Layer Perceptron
MM-DFR	Multimodal Deep Face Representation
MMU1	Multimedia University
MNN	Modular Neural Networks
NBP	Neighborhood-Based Binary Pattern
NLP	Natural Language Processing
NNC	Nearest Neighbour Classifier
PCA	Principal Component Analysis
PINs	Personal Identification Numbers
PMML	Pairwise-constrained Multiple Metric Learning
QDC	Quadratic Discriminant Classifier
RBF	Radial Basis Function
RBM	Restricted Boltzmann Machine
RDE	Robust Direction Estimation
RDW-LBP	Regional Directional Weighted-Local Binary Pattern
ROC	Receiver Operating Characteristic
RTT	Radon Transform Thresholding
SAE	Stacked Auto-Encoder
SCER	Sparse Coding Error Ratio
SGD	Stochastic Gradient Descent
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Features
SVMs	Support Vector Machines
TCM	Texture Code Matrix
USFFT	Unequally-Spaced Fast Fourier Transforms
VAR	Validation Accuracy Rate

Chapter 1
Introduction

1.1 Overview

Over the last few decades, increasing use of many developed and sophisticated techniques of hacking and forgery have led to increasing demands for alternative methods of recognizing a person's identity [1]. Biometric systems are constantly evolving and promise technologies that can be used in automatic systems for identifying and/or verifying a person's identity uniquely and efficiently without the need for the user to carry or remember anything, unlike traditional methods. These systems have been widely employed in many governmental and civilian sensitive applications, especially those including an automatic access control to physical or virtual places, such as border checkpoints, ATM machines, security and surveillance systems, financial transactions, computer/network security, etc. [2]. Personal identification based on biometric features has many advantages over traditional knowledge-based methods (e.g., passwords or Personal Identification Numbers (PINs)) and token-based methods (e.g., driver’s license, passport, ID card, or a simple set of keys), because it is difficult to be transferred, lost, forgotten or duplicated. In addition, employing biometrics in the task of identifying a person's identity is more convenient and user-friendly than traditional methods, so that clients do not need to remember or carry anything with them [3]. Finally, the security level achieved using biometric systems can be higher than those using traditional methods.

Despite the fact that personal identification using biometric characteristics has been a matter of research for more than forty years [4], there has been a growing interest in highly secured and well-designed biometric systems in the last decade. This is evidenced by recently published books [3][5][6][7], specific international conferences on biometrics (e.g., IAPR[1], ICBTAS[2], etc.), the

[1] The 9th IAPR International Conference on Biometrics, 2016.
[2] ICBTAS 2017: 19th International Conference on Biometrics Theory, Applications and Systems.

expansion of benchmark tools and assessment campaigns for biometric systems [8][9][10]. In addition, considerable attention has been also paid by government[3,4], industry[5], and international consortia dedicated specifically to biometric systems (e.g., CITeR[6], EAB-CITeR[7], etc.) in the last few years.

Broadly, biometric systems can be divided into two main types: unimodal and multimodal biometric systems. Unimodal biometric systems are based on using a single source of information (e.g., fingerprint, iris, or face, etc.) to establish the person's identity. This type of system has a number of critical limitations and issues that can affect reliability and performance. These include the possible poor quality of the specific biometric trait of the person, non-universality, lack of uniqueness, spoof attacks, etc. All these drawbacks of unimodal systems can be efficiently addressed by systems combining evidence from multiple sources of biometric information for identifying a person's identity. Such systems are then referred to as multimodal biometric systems. For instance, face and voice, or two fingerprints (e.g., left and right index fingers), can be employed together to more accurately and robustly verify the person's identity. Hence, multimodal biometric systems can inherently improve the accuracy of unimodal systems by eliminating the dependency on one particular biometric trait as well as providing a wider population coverage [3][5].

In this introductory chapter, basic information on biometric systems is presented, including the overall structure of the biometric systems and their functionalities, some of the most common biometric modalities used in practice, and the limitations of unimodal biometric systems. In this chapter, the motivation behind multimodal biometric systems is outlined along with the challenging issues in designing and implementing such systems, from which the motivation, objectives and tasks of this PhD thesis are also derived. The chapter is finished by stating the thesis, summarizing the research contributions originated from this work and giving an outline of the dissertation.

---

[3] The UK Biometrics Working Group (BWG): http://www.idsysgroup.com/files/Biometrics%20Advice.pdf.
[4] TWS Biometrics Working Group: https://sites.google.com/site/twsbwg/.
[5] Biometrics Research Group: Mobile Biometrics Market Analysis
[6] CITeR: Center for identification technology research, US, 2011.
[7] EAB-CITeR. European Cooperative Identification Technology Research Consortium, 2015.

1.2 Biometric Systems

A biometric system is basically a classification and recognition system that captures a biometric trait from an individual, extracts a set of discriminative features from the trait captured, compares the extracted feature set against a template set (or sets) stored in the system's database. Then, the final decision is taken based on the results of this comparison. Thereby, users do not need to remember or carry anything with them, avoiding the loss, sharing or forgetting of personal information. As depicted in Fig.1.1, the overall structure of any biometric system consists of four main stages, that work in a sequential manner to obtain the result from the system [2][11].

1. **Sensor Module**: this stage represents the interface between the user and the biometric system in which an electronic sensor or reader captures the biometric trait and converts it into a digital form. For example, a digital CMOS camera may be used to capture the patterns (e.g., rings, ridges, crypts, etc.) of the iris region. This is a critical stage since the rest of the biometric system strongly depends on the quality of the scanned trait. Therefore, in some biometric systems, this stage is linked with a quality checking procedure.
2. **Pre-processing and Feature Extraction**: Typically, the captured biometric trait is subjected to pre-processing operations in order to achieve a required quality. These operations can be divided into three steps. Firstly, an assessment step to evaluate the quality of the captured biometric trait, and then based on a pre-defined threshold either use the trait for further processing or attempt to re-capture the trait from the user. Then a segmentation step, in which the region of interest of the biometric trait is separated from the background noise, for example, detecting the iris region in the eye image. Finally, a number of image quality enhancement algorithms can be applied (e.g., logarithmic transformation, histogram equalization, etc.) on the detected biometric region to improve its quality by reducing the noise level introduced by the camera and illumination variations. After all or some pre-processing operations have been applied, a set, or vector, of discriminative features are extracted from the enhanced biometric trait. Then the extracted feature set is either sent to the matching stage for the purpose of user identification, or is stored in the system’s database as a template in

an enrollment phase. Ideally, the extracted features should have minimum intra-class variations and inter-class similarities.

3. **Matching Module**: here a query trait feature vector provided by the feature extraction module is compared with all the previously stored templates in the system's database to generate matching scores. These scores are measures of either similarities, in which case a higher score indicates a closer match or dissimilarities (distance scores), in which cases a lower score indicates a closer match between the query and the templates.
4. **Decision Module**: this is the final stage in the biometric system in which the user is identified or a claimed identity is either accepted (authorizing the user) or rejected (not authorizing the user) based on the score generated by the matching module.

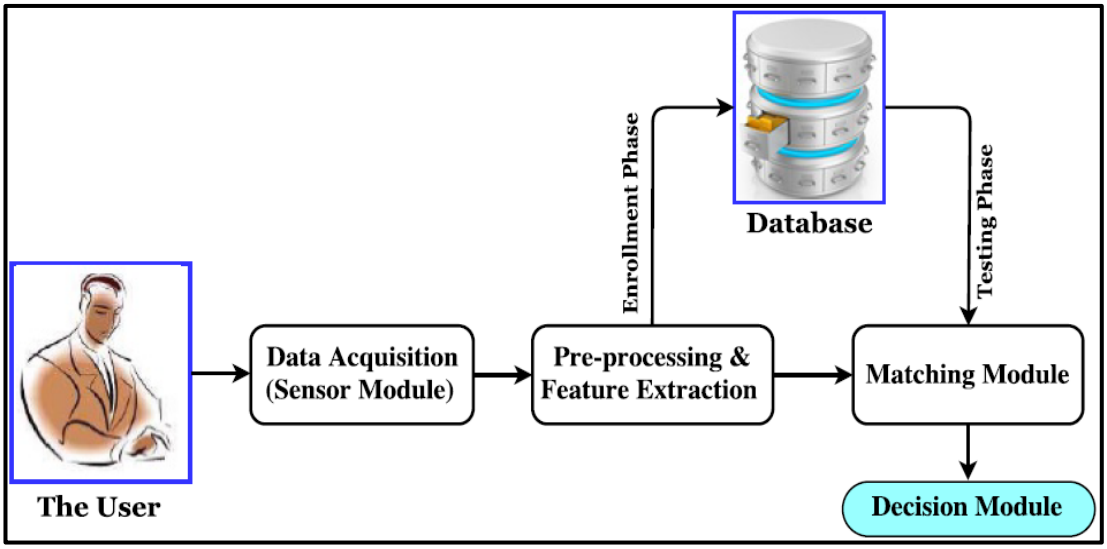


Figure 1.1: A general structure of a biometric system, adapted from Jain et al. [5].

In addition to the *Enrollment* task, the two other tasks which can be provided by a biometric system, depending on the application context, are *verification* and *identification*. In this thesis, the term *"recognition"* is used in the general case to refer to these two functionalities [5][11][12]:

- **Enrollment:** the digital representations of the features extracted from the biometric trait captured from the person using a sensor or CCD camera are known as templates. These templates are stored in digital form in the system's database through the enrolment or training process, along with

some biographic information (e.g., name, PIN, address, etc.) distinguishing the user, as shown in Fig.1.2 (a). Therefore, it can be used to register a new user or update an old user's templates, and for both of them a quality checker can be used to ensure the quality of the input trait. In some sensitive applications, these templates are kept as encrypted templates because of security and privacy concerns. Finally, this task is always coupled with the other two tasks (verification and identification).

- **Verification:** in this task a *one-to-one* comparison is conducted between the newly captured image and the claimed identity to decide whether the claim is genuine or not (e.g., "Does this biometric trait belongs to Bob?"). Firstly, the user claims an identity by providing a user-name, ID card, driver's license, or PIN, and their biometric trait to the biometric system at the same time. Secondly, the claimed identity is recalled by the system from its database corresponding to the provided user-name, ID card, driver's license, or Personal PIN, in order to compare it against the newly entered data. Finally, the system will decide if the user is accepted as an authorized user or rejected as an imposter depending on the result of this comparison, as shown in Fig. 1.2 (b). To be accepted, the similarity between the query data and stored data must be equal or higher than the pre-defined threshold set by the system's designer. Typically, the performance of the verification system is evaluated by measuring a number of criteria such as False Rejection Rate (FRR), False Acceptance Rate (FAR) and the Equal Error Rate (EER). A verification system is usually used as a positive recognition system in physical access control systems and ATM systems to prevent multiple users from using the same identity or restricting unauthorized users' access to specific services.
- **Identification:** this task is more complicated and takes longer than the verification task. In this task, users are asked to provide only their biometric trait to the system, and a one-to-many comparison is conducted by comparing the captured data with all the stored templates in the database, as shown in Fig. 1.2 (c). The user's identity is established by taking either the best match or by listing all the possible matches within the pre-defined threshold and ranking them based on the similarity score. The identification

system can be used in both positive and negative recognition applications. In the positive identification system, the question, is this person known to the system, is answered by determining the user's identity from a known set of identities. In contrast, if a user attempts to negatively identify themselves to the system by concealing their real identity, the negative identification system tries to uniquely determine the user's identity to prevent them from using multiple identities. The negative identification system answers the question, whose biometric data is this, and is usually used in forensic applications, background checks, criminal identification and law enforcement. The performance of the identification system is assessed by measuring its Correct Recognition Rate (CRR), Classification Error Rate (CER) or Rank-1 identification rate.

1.3 Biometric Modalities

Many different kinds of pattern recognition systems have been widely used in establishing the identity of a person based on different kinds of biometric characteristics. In a biometric system, these characteristics constitute a substantial and a strong link between the user and their identity [7]. The biometrics characteristics can be divided into two types, physiological characteristics such as those used in Face, Iris, Fingerprint, Hand Veins, etc., recognition or behavioral characteristics such as those used in Signature, Speech, Gait, Voice, Keystroke, etc., recognition [13]. Recently, some additional useful information, such as name, gender, height, weight, ethnicity, eye colour, age, tattoo, etc., referred to as soft biometric characteristics, can also be used to verify a person's identity. However, these biometric traits do not always provide sufficient evidence to accurately recognize the person's identity, therefore they should be incorporated into a primary biometric system along with some additional physiological or behavioral characteristics [14]. Fig.1.3 shows some examples of physiological, behavioral and soft biometric characteristics which can be employed in a biometric personal identification system. The most essential question is what biological measurements (physical/behavioral characteristics) can be interpreted as a biometric trait? Ideally, a set of criteria should be met by any human physiological or behavioral

characteristic in order for it to be used as a successful biometric characteristic, and these criteria can be summarized as follows [2][15]:

- **Uniqueness (Distinctiveness)**: is the most important requirement for choosing a good biometric trait; it means that the biometric trait should be distinguishable among all members of the community, for instance, the iris trait is unique even between identical twins.
- **Universality (Availability)**: means that the characteristic is owned by all members of the community.
- **Robustness**: means that the trait is not affected by different environmental conditions and should be constant for a long period of a person's lifetime. The face trait, for example, can be affected by ageing.
- **Accessibility**: (also referred to by Collectability and Measurability) means the trait should be easily acquired and measured quantitatively with little user interaction.

In addition to considering all these criteria when selecting the best biometric trait, there are other criteria which should be taken into account when designing and applying any person verification/identification system based on the personal biometric traits, including [5]:

- **Performance**: refers to the ability, reliability and efficiency of the biometric system to achieve a high recognition rate in different environmental conditions and at a reasonable speed and cost.
- **Acceptability**: refers to what extent the biometric system is accepted by the general public and the willingness of people to provide this biometric trait comfortably.
- **Circumvention**: refers to the ability of the biometric system to resist and discover all the fraudulent methods of an imposter, for example, using a fake fingerprint or imitating the behavior of an authorized user such as their signature.

Based on domain-knowledge from the literature, it can be observed that no single biometric trait can satisfy all the criteria mentioned above. For instance, some biometrics have easy collectability and low distinctiveness (e.g., face trait), while other biometrics (e.g., iris trait) have a very high distinctiveness

and low collectability properties where iris acquisition devices are very expensive and require more user-cooperation. As reported by Ross et al. [3], the limitations and weaknesses of using a single biometric trait can be overcome by using two or more biometric traits in multimodal biometric systems, thus enhancing the security and reliability in establishing the person's identity. In this thesis, face and iris traits are studied and, as shown in Table 1.1, almost all desired properties are well fulfilled by combining these two traits.

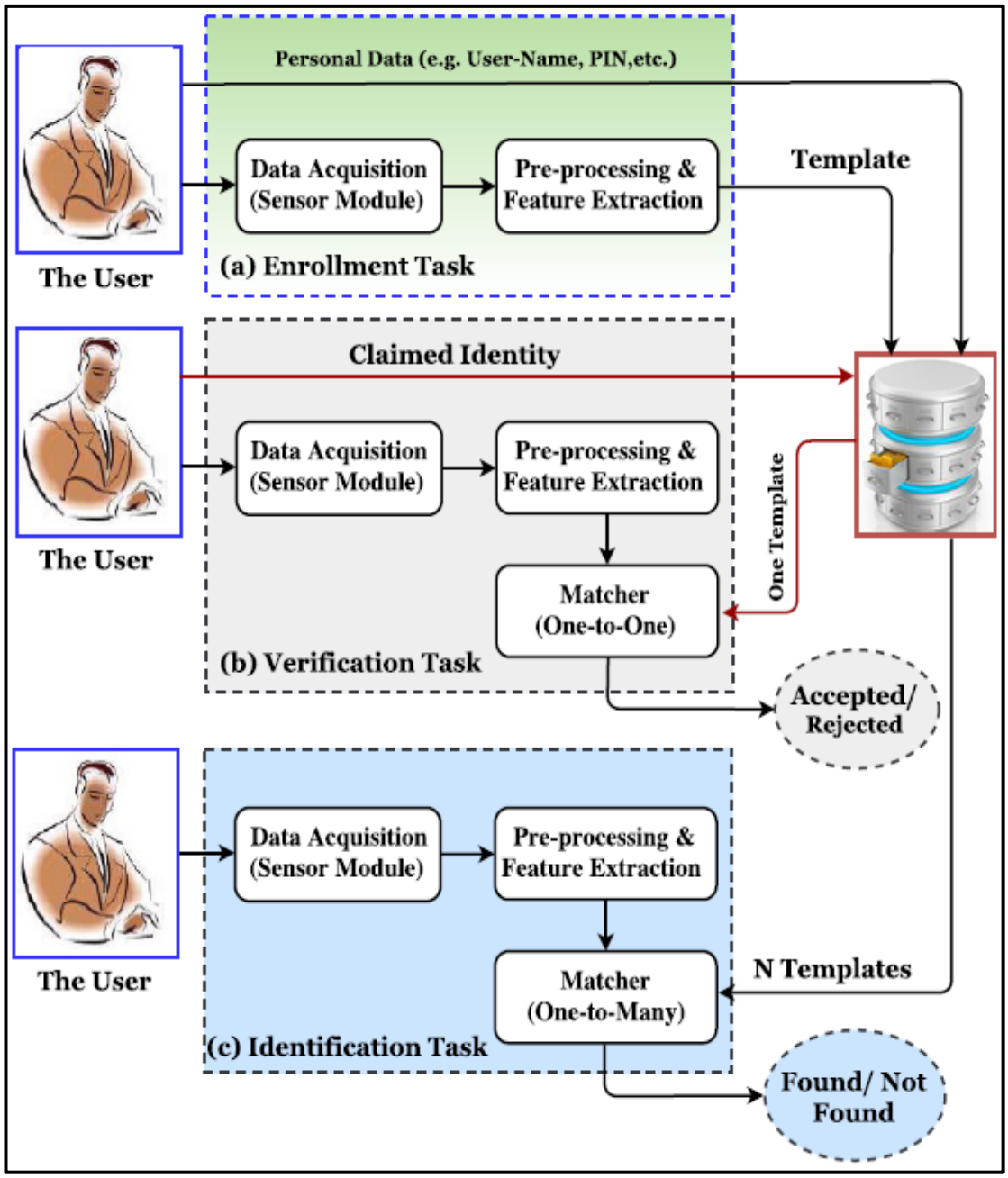


Figure 1.2: The three provided tasks by the biometric system: (a) Enrollment, (b) Authentication and (c) Identification task, adapted from Jain et al. [5].

Table 1.1: A comparison of biometric characteristics. High, Medium, and Low are denoted by H, M, and L, respectively. Adapted from Jain et al.[2].

Biometric	Universality	Distinctiveness	Robustness	Collectability	Performance	Acceptability	Circumvention
Face	H	L	M	H	L	H	H
Iris	H	H	H	M	H	L	L
Fingerprint	M	H	H	M	H	M	M
Hand Geometry	M	M	M	H	M	M	M
Gait	M	L	L	H	L	H	M
Voice	M	L	L	M	L	H	H
Signature	L	L	L	H	L	H	H

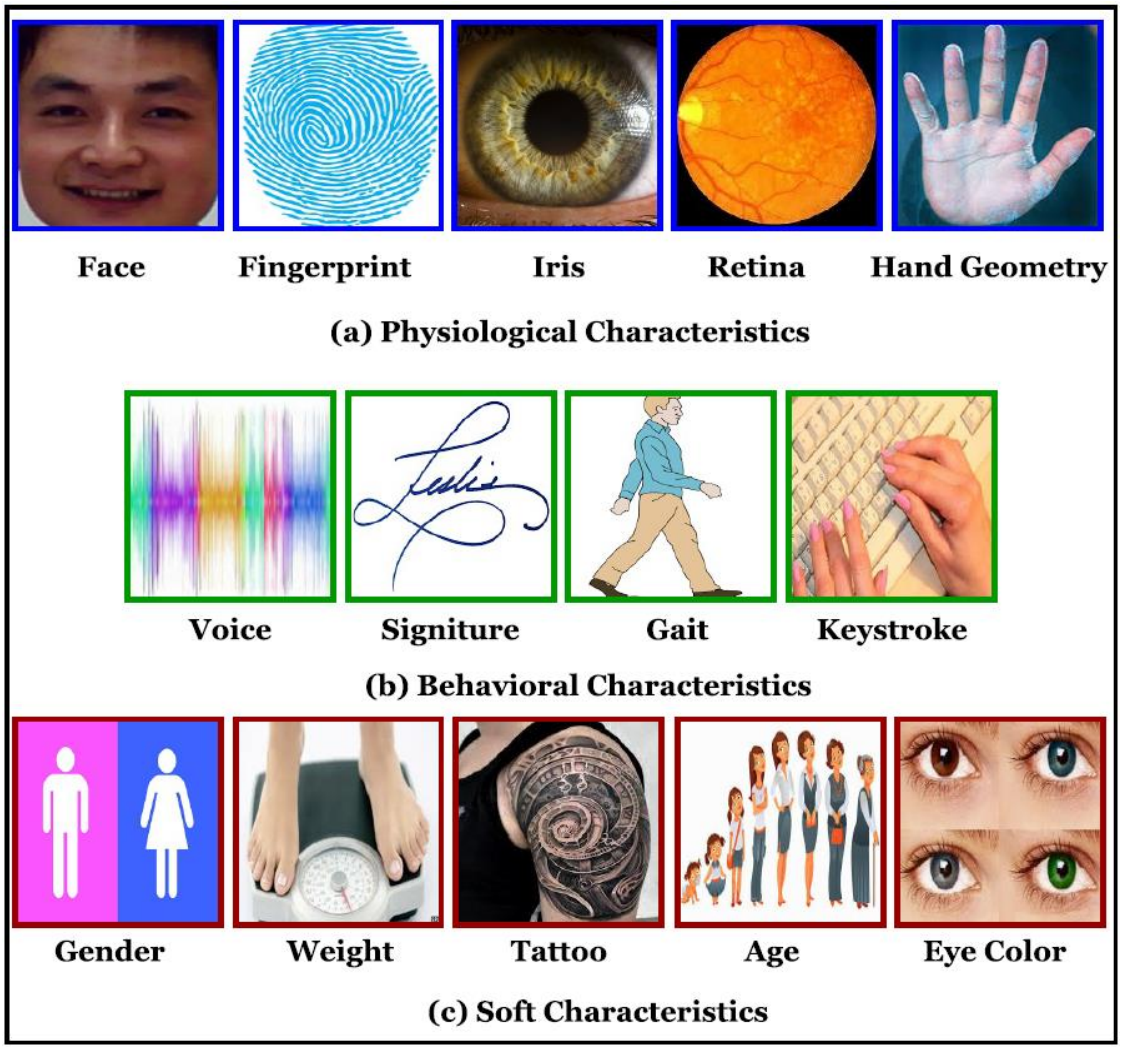


Figure 1.3: Examples of common biometrics: (a) Physiological, (b) Behavioral and (c) Soft characteristics.

1.4 The Limitations of Unimodal Biometric Systems

Over the last few years, most biometric systems that have been widely employed in sensitive governmental and civilian applications have been unimodal biometric systems. Although these can provide a high level of security in identifying a person's identity, there are a number of critical limitations and problems that can significantly affect the reliability and the performance of these systems. These limitations and problems can be divided into three main types: accuracy, scalability, security and privacy [2][3][16]:

1.4.1 Accuracy

Ideally, any biometric system should be able to offer a high level of accuracy in recognizing a user's identity when a new query sample is presented to the system. However, the accuracy of biometric systems that operate on a single trait is usually affected by a number of factors that can be summarized as follows:

- **Noise in Sensed Data:** there are a number of reasons that lead to the appearance of noise in the scanned data; some of these reasons are environmental conditions that surround the biometric acquisition process. For example, when voice data is captured in a noisy environment, or a user's face image is captured under poor illumination conditions, this affects the accuracy of the biometric system. In addition, sensors must be correctly maintained. For example, the quality of the fingerprint trait could be low as a result of the accumulation of dirt on the surface of the fingerprint reader [17]. The poor quality of the biometric data can result in a significant reduction in the overall accuracy of the system.
- **Non-Universality:** simply means that a sub-set of the population is not able to provide the required biometric trait correctly or do not possess a specific biometric trait. This can lead to increased Failure Enrollment Rate (FER). For example, persons with long eyelashes or eye abnormalities may not be able to provide their iris trait correctly to an iris recognition system. Another example is shown in Fig.1.4, where users having very dry or oily fingers and workers with bruises and cuts have fingerprints that affect and limit the performance of a fingerprint recognition system [2].

Figure 1.4: Three different examples of the non-universality of a fingerprint trait due to finger dryness [18].

- **Intra-Class Variations**: refers to the variations between user's samples acquired during the enrollment and recognition phases. These variations could be a result of incorrect operation of the reader (e.g. translation, rotation, and pressure on fingerprint sensor), inherent changes like the scars and bruises in the fingerprint trait as shown in Fig.1.5, a fraudster mimicking a particular behavioral trait, such as the voice, the effect of ageing on the face appearance, the effect of disease on the iris trait, and the use different sensors or modified sensor settings during enrollment and recognition phases. In an ideal biometric system, the features extracted from the biometric trait must be relatively invariant to these variations. However, features are very sensitive to these variations in most cases, and may need complex classification algorithms to handle these variations.

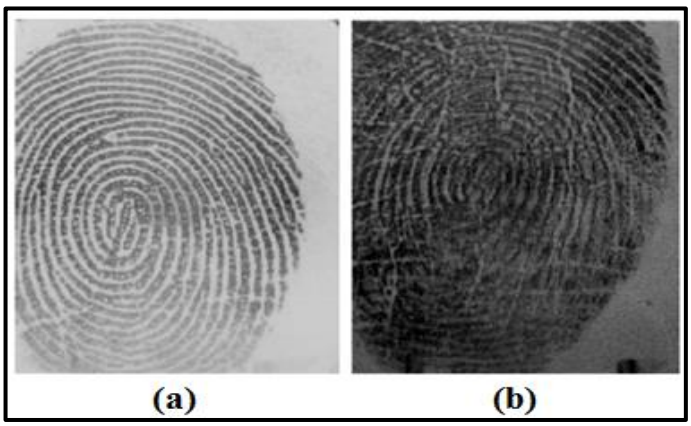


Figure 1.5: Example of an intra-class variation of a fingerprint trait obtained from the same user at two different times: (a) Enrollment phase (b) Recognition phase [2].

- **Inter-Class Similarities:** refers to the lack of individuality (uniqueness) of the biometric trait as a result of increasing number of enrolled users in the database, leading to overlapping biometric characteristics in the feature space and an increase in the FAR. Therefore, each biometric system should have an upper bound on the number of enrolled users, set by the maximum number of users who are discriminated efficiently by the system.

1.4.2 Scalability

This issue refers basically to increasing the number of enrolled users in the database of the biometric system and its effect on the speed and the performance of the system. In fact, this effect is different according to whether verification or identification is the task provided by the biometric system. In a verification system, a worsening effect does not occur with increasing number of enrolled users, because only a one-to-one comparison is needed to verify the user's identity. On the contrary, increasing the number of enrolled users in an identification system has serious negative effects, because the query template will be compared against ***N*** templates in the database (a one-to-many comparison) [19]. Thereby, increasing the value of ***N*** will decrease the throughput of the biometric system by increasing the elapsed time required for establishing the user's identity. Moreover, it can increase the False Matching Rate (FMR) of the system. To eliminate or reduce this effect on the identification system one question needs to be answered, which is how we can reduce the number of these required comparisons? In fact, there are some approaches working on indexing or filtering the enrolled users based on extrinsic factors (e.g., age, gender, etc.) or intrinsic factors (e.g., fingerprint major classes). Therefore, only a subset of the whole database will be required for comparison purpose, but these approaches still have some limitations [18].

1.4.3 Security and Privacy

This is usually related to biometric spoofing issues and the possibility that the unimodal system can be fooled, for example using a fake fingerprint or iris template. Behavioral characteristics like voice and gait are more vulnerable to such attacks than physical characteristics [3]. One of the most important

challenges in biometric systems is how to keep all the users' templates, which are stored in the database, secure from theft. There are different techniques for handling biometric spoofing issues, such as a liveness-detection techniques [20] for physical traits and a challenge-response mechanism [3] for physical traits and behavioral traits. An imposter can use a stolen template in two ways to circumvent a biometric system as follows: (i) the imposter can send the stolen template to the system to get an unauthorized access or (ii) use the stolen template by creating fake templates (physical spoof) to circumvent the system or other systems that operate on the same biometric trait [18].

1.5 Motivation for Multi-biometric Systems

Multi-biometric systems are systems that fuse evidence from multiple sources of biometric characteristics in order to enhance the security and increase the reliability in establishing the person's identity. In other words, the multi-biometric systems are mainly based on the data presented and acquired from multiple resources, as will be explained later. Recently, multi-biometric systems have been adopted for many governmental and civilian applications due to their ability to address and overcome some of the main drawbacks and limitations exhibited by unimodal biometric systems. Some of the advantages of multi-biometric systems over unimodal biometric systems are listed as follows [3][16]:

1. Increasing the efficiency and the reliability of the recognition system by considerably reducing the effect of noise or poor quality in the acquired biometric traits. The availability of other biometric sources in the multi-biometric system can enhance the performance of the system, for example, if a user cannot be identified by their voice trait due to environmental issues, then they can still be identified using another trait, such as a fingerprint.
2. Sufficient population coverage can be achieved using multi-biometric systems, due to their ability to solve problems related to the enrollment phase, such as non-universality. Thereby, a person who cannot present a particular biometric trait, can still be enrolled and identified by presenting another biometric trait, for example, a manual worker who has a poor fingerprint quality, can still be enrolled and identified using traits such as,

face, iris, voice, etc. As a result, the FER will be decreased by increasing the population coverage.

3. A multi-biometric system can greatly reduce the overlap among the feature spaces of different persons (Inter-Class similarities) by combining biometric traits and adopting a fusion approach. Combining multiple pieces of evidence from different sources can increase the dimensionality of the feature vector, but the overall accuracy of the biometric system will be increased. For example, two persons from the same family who can have the same voice trait will not have the same iris and fingerprint traits.
4. Multi-biometric systems can provide a higher accuracy and a greater resistance to unauthorized access by an intruder than unimodal biometric systems, due to the difficulty of spoofing or forging multiple biometric traits for a legitimate user at the same time. Furthermore, another mechanism can be coupled with the multi-biometric system, for example by asking users to present their biometric trait randomly at the acquisition time (e.g. fingerprint, followed by face trait and then voice) to ensure that it is a live user who is interacting with the system. This is known as a Liveness Detection or a Challenge-Response Mechanism.
5. Using a multi-biometric system can substantially improve the throughput of a biometric system, especially in the identification task where a one-to-many comparison is required. This can be carried out using the least accurate biometric trait (e.g., Fingerprint trait) to prune the database size down to an acceptable size, and then the most accurate biometric trait (e.g., iris trait) will be used on the rest of the database to make the final decision.
6. Finally, a high degree of flexibility will be offered to the user during the recognition time using a multi-biometric system. Suppose a system is implemented using three biometric traits (e.g., face, fingerprint and voice). Subsequently, at the recognition phase, a user can choose either to provide a subset or all of their biometric traits depending on the nature of the implemented application and the user's convenience.

1.6 Design Issues in Multi-biometric Systems

Based on the definition of the multi-biometric system, three main factors must be taken into account when designing and implementing a multi-biometric system. Firstly, the architecture of the multi-biometric system in which the selected biometric traits will be acquired and processed. Secondly, the source of the information used by the multi-biometric system, which could be any of six different sources, as will be explained later on. The third factor is the fusion of the obtained traits and where this takes place in the multi-biometric system. In the next subsections, each factor will be explained in detail.

1.6.1 Multi-biometric Systems Architectures

One of the most challenging issues in designing a multi-biometric system, after the proper biometric traits have been selected to establish the user's identity is choosing the architecture of the multi-biometric system. In fact, a number of factors play a significant role in deciding and selecting the best architecture type, such as the nature of the application, the required level of security, and the order in which the biometric traits from different resources are captured and processed through the system. Typically, depending on these factors, the architecture of the multi-biometric system can be categorized into three main types: serial, parallel, or hierarchical [2][21]. The serial scheme, also known as a cascade scheme in which the result of one trait can be used to prune the database size down before the next trait is used, especially when the system operates in the identification mode [16]. Therefore, the output of the next trait may be affected by the output of the previous one. For example, a multi-biometric system using iris and fingerprint can use the fingerprint trait to get the best top matches, and then the iris trait is applied to the rest of the database to make the final decision. As a result, the throughput of the system can be increased by making a decision before acquiring all the required modalities. Finally, the applied biometric traits do not have to be captured simultaneously. See Fig. 1.6 (a).

On the other hand, in the parallel scheme multiple biometric traits are acquired and processed independently at the same time, and then the results obtained from them are combined using an appropriate fusion approach to

make the final decision [3], as shown in Fig. 1.6 (b). Therefore, the parallel scheme takes a longer recognition time compared to the serial scheme, which makes the serial scheme more convenient to the user. However, the security level is higher in the parallel scheme than the serial scheme, which makes it highly recommended for sensitive applications, such as military applications. Finally, the advantages of both the parallel and serial schemes can be combined in the third scheme, which is known as a hierarchical scheme. In this scheme, a number of problems can be solved such as missing biometric data or noisy biometric traits by combining a different classifier into a tree-like structure, as shown in 1.6 (c). [18]. However, this type of multi-biometric system architecture has not received a lot of attention in the research community.

1.6.2 The Evidence Sources of Multi-biometric Systems

To satisfy the multi-biometric system concept, more than one biometric source, either physiological or behavioral characteristic is required, as illustrated in Fig.1.7. In this section, the question, “What are the sources of biometric data that can be utilized in a multi-biometric system?”, is answered. In general, five scenarios are available in the multi-biometric system, and a sixth scenario refers to the hybrid multi-biometric system, which is the combination two or more of these five scenarios [22]. In these scenarios the evidence sources of the multi-biometric system can be categorized into one of, or a combination of, the following five types:

1. **Multiple Sensors/Single Trait:** in this system one biometric trait is captured by using more than one sensor in order to get a more diverse set of features, for example, the face trait can be captured using two different cameras (e.g. infrared camera and visible light camera), or use optical and ultrasonic sensors to capture the fingerprint trait.
2. **Multiple Algorithms/Single Trait**: in this system, only one biometric trait is captured and the results from multiple feature extraction approaches and/or multiple classifiers are concatenated to improve the performance of the system. The multiple algorithms system is less expensive than previous systems due to only one sensor being required, which also reduces the user interaction with sensors.

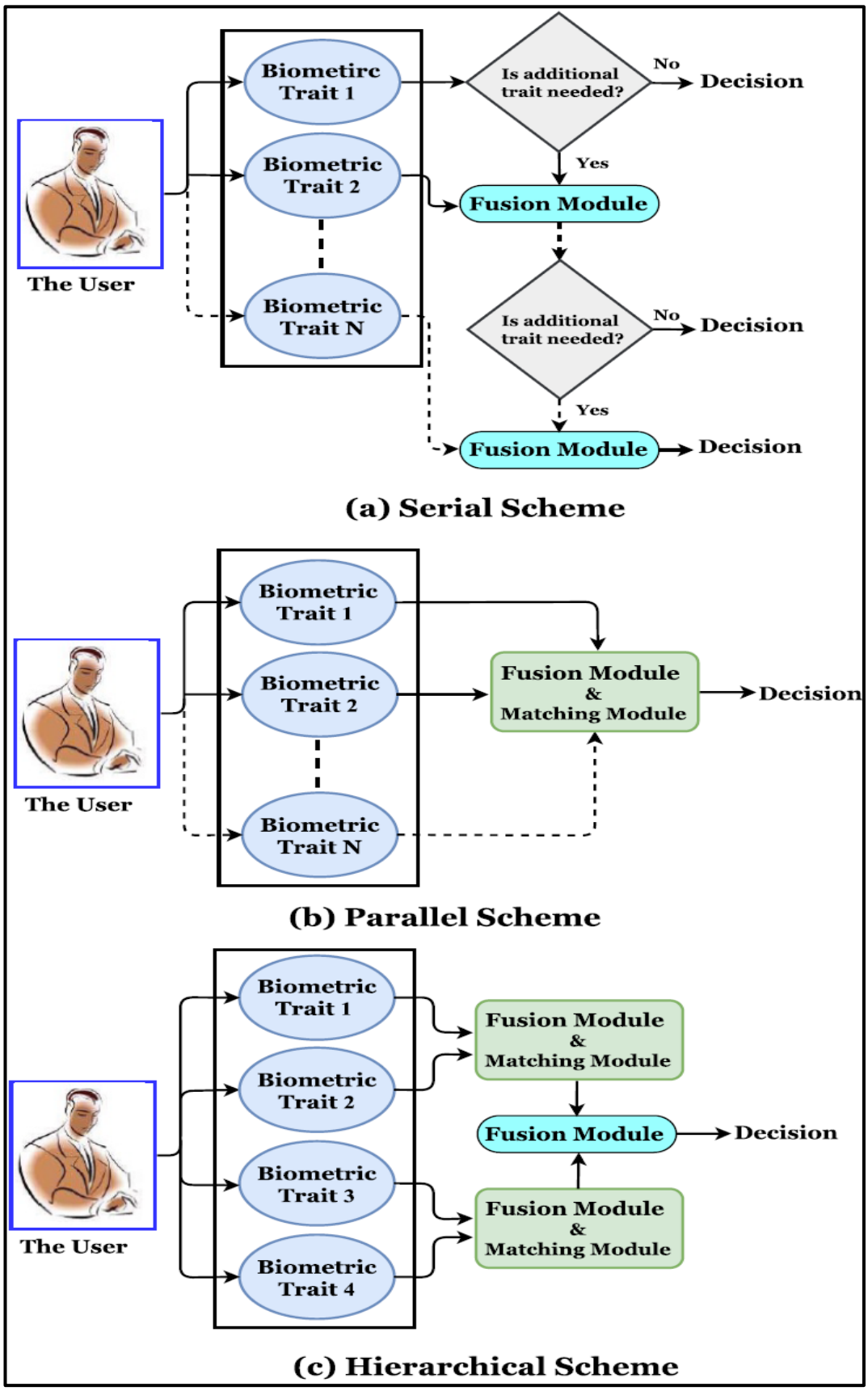


Figure 1.6: Multi-biometric system architectures: (a) Serial scheme, (b) Parallel scheme and (c) Hierarchical scheme, adapted from Ross et al. [3].

However, in the multiple algorithm systems the computational complexity and time required are increased, due to multiple feature extraction and matching modules required. For example, the same fingerprint trait can be processed by using texture-based and minutiae-based approaches [16].

3. **Multiple Samples/Single Trait:** this type of multi-biometric systems is adopted to increase the resistance of the system to variations of the biometric trait using a single sensor to capture more than one sample from the same trait. For example, in the face recognition system, the face trait can be captured from different angles (e.g. frontal, left and right profiles) using a single camera, and then the extracted information is used to address the challenges that arise from the variations of the facial pose [3].
4. **Multiple Instances/Single Trait:** again one biometric trait is used, but the information is extracted and used from multiple instances of the same biometric trait. For example, the person's identity could be established by extracting the information from his right and left iris trait or his right and left index fingers. This type of systems is cost efficient, due to no new sensors, feature extraction approaches, and/or classification approaches, being required [22].
5. **Multiple Biometric Traits:** this results in a multimodal biometric system, where multiple physiological and/or behavioral characteristics are employed to recognize the person's identity requiring the use of multiple sensors [16], for example, using the fingerprint and iris trait in establishing the person's identity. The independence of the adopted traits (e.g., fingerprint and iris) ensures that a considerable improvement in performance is obtained. However, some issues need to be addressed in this system, such as the cost of deploying multiple sensors, enrollment time, throughput time, and user’s convenience due to the number of the traits required during the enrollment and recognition phases. Finally, combining more than one trait can lead to an increase in the dimensionality of the feature vector, which results in putting an upper bound on the number of the used traits without affecting the performance [3].

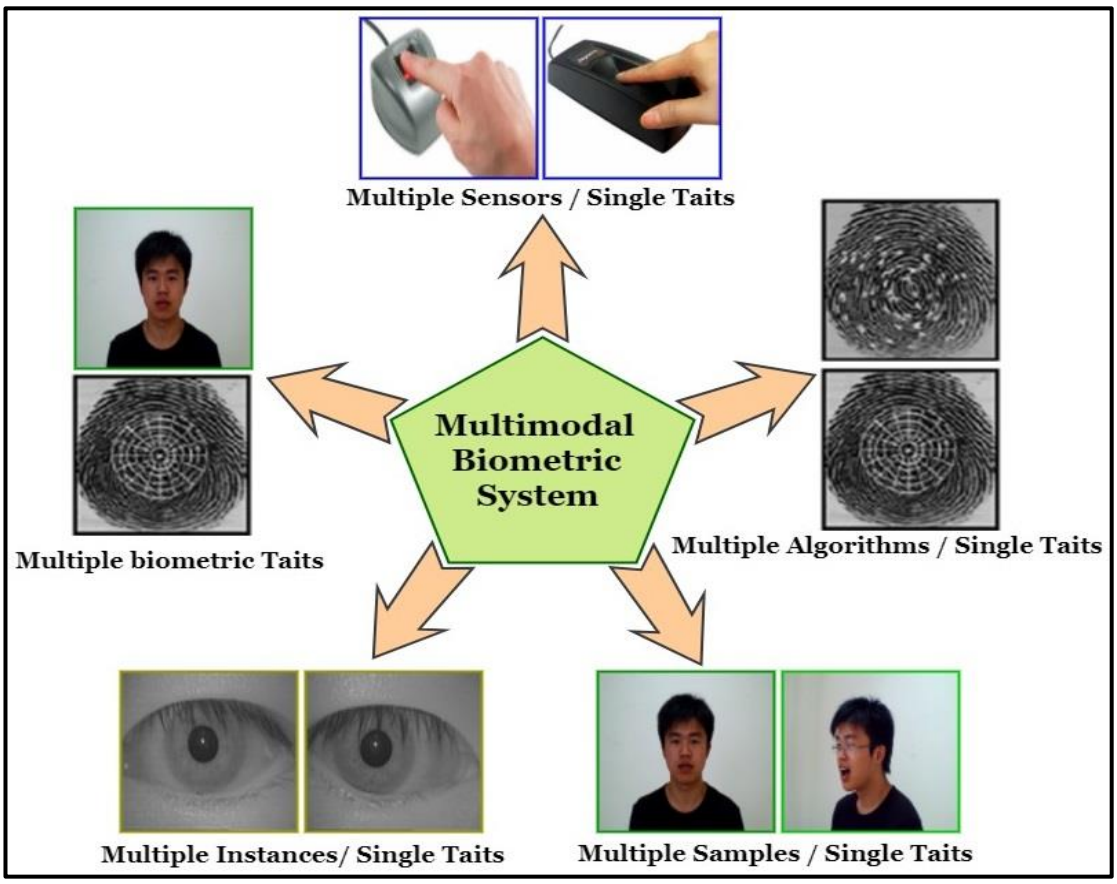


Figure 1.7: The five evidences sources of a multimodal biometric system, adapted from Ross and Jain [22].

1.6.3 Fusion Information Levels in Multi-biometric Systems

Other essential challenges for the designer of the multi-biometric system are selecting the best information sources from multiple sources in the system and finding an efficient methodology to fuse them. As it is well known, information in the biometric system can be taken from four different points: the sensor module, the feature extraction module, the matching module and the decision module. Typically, the degree of availability of the information decreases as one moves from the sensor module toward the decision module [16]. Therefore, based on these information sources and the amount of the available information, the fusion method can take place in five different levels, as shown in Fig.1.8. These levels are the sensor level, the feature level, the score level, the rank level and the decision level. The five levels of integrating the biometric information can be broadly classified into two groups. Firstly, the Prior-to-Classification fusion levels, which includes the sensor and feature levels refers to combining the information before applying any matching

algorithm. Secondly, the Post-to-Classification fusion levels, which includes the score, rank and decision levels, refers to integrating the information in the matching score and the decision spaces [13][21]. In this section, the five fusion levels are explained as follows:

1. **Sensor Level Fusion:** in this level of fusion, the sources of information either employ different compatible sensors to capture the same biometric trait or capture multiple instances for the same trait using a single sensor [23]. For example, the fingerprint trait can be obtained by using two different sensors (e.g. Optical and Ultrasonic sensors). Another example is capturing multiple snapshots of the face trait using the same camera. Although the sensor level represents the richest source of data, this data can be contaminated by noise, illumination changes and other enrollment process issues. Finally, it can quite difficult if not impossible to fuse directly two sets of raw data collected from two different modalities due to the incompatibilities between them.
2. **Feature Level Fusion:** in features level fusion, the final feature vector is constructed by integrating different feature vectors, which are acquired from different sources either by employing multiple sensors (e.g. Optical and Ultrasonic sensors are used to capture the fingerprint trait), applying different feature extraction approaches on the same biometric trait (e.g. the texture-based and the minutiae-based approaches are used to process the fingerprint trait) or using different biometric traits (e.g. face and iris trait). It is highly probable that the recognition accuracy rate can be significantly improved by applying the fusion at the feature level compared to other levels of fusion, due to the high degree of the availability of the useful information about the raw biometric data [13]. Nevertheless, integrating the data at the feature level can be coupled with some drawbacks and difficulties. Firstly, extracting the feature vectors from different biometric traits using different approaches leads to a possible increase in incompatibility between the extracted feature vectors. Secondly, it increases the computation time and storage cost due to the high dimensionality of the feature vector obtained in this level. In addition, the relationship between the feature vectors extracted from different biometric traits may be

unknown. Thereby, dimensionality reduction algorithms or feature selection approaches may be employed to address the curse of high dimensionality problem and to discard the highly correlated features, respectively. Finally, the feature vectors may not be accessible, especially within the majority of commercial biometric systems. All these difficulties have led most researchers in this field to prefer the post-to-classification fusion levels [22][23].

3. **Score Level Fusion:** also referred to as the measurement level, the expert or the confidence level. At this level, the similarity scores from one biometric trait or multiple biometric traits, by employing one or multiple classifiers, are integrated to form the final decision [13], for example, fusing the similarity scores from left and right index fingerprint trait using two different approaches, or fusing the similarity scores from the face trait and the hand trait. This level comes in the third degree in terms of the availability of information about the input biometric trait after the sensor level and feature level. In this fusion scheme, the most important thing to take into consideration is that the output from multiple classifiers might not be homogeneous [18]. For example, the matching score produced by one classifier could mean a similarity score where the highest score refers to the best match, whilst the matching score produced by another classifier could mean the dissimilarity (the distance score) where the smallest score refers to the best match. Moreover, the matching scores of multiple classifiers might not be in the same numerical range or probability distributions. This issue can be solved simply by applying a normalization technique to transform the matching scores into a common range, before this fusion takes place [16]. Depending on the treatment of the matching scores obtained from multiple classifiers, the fusion at the scores level can be classified into two different groups. Firstly, the fusion can be considered as a classification problem where the matching scores from multiple classifiers are integrated to formulate a feature vector, which is used as input to the second level classifier to make the final decision (e.g., genuine user or impostor user). Secondly, the fusion can be considered as a combination problem where the final decision is taken by normalizing into the same

range and combining the matching scores from multiple classifiers to produce a scalar fused score. The fusion at the matching score level has been extensively studied in the literature due to its simplicity, convenience of fusion and availability of information [13].

4. **Rank Level Fusion:** In multimodal biometric systems, if the system operates in the identification mode, then the output of each classifier can be viewed as a list of ranks of the enrolled candidates, which represents a set of all possible matches sorted in descending order of confidence. In this case, the fusion in the rank level can be applied using one of the ranking level fusion methods to consolidate the ranks produced by each individual classifier in order to deduce a consensus rank for each person. Then, the scores output are sorted in descending order and the identity with the lowest score is presented as the right person. As reported by Ho et al.[24], the ranks of multiple classifiers can be integrated using three different approaches: the Borda count approach, the highest rank approach, and logistic regression approach.
5. **Decision Level Fusion:** this level of fusion is also known as an abstract level fusion where the least amount of information is available. This information represents the decision outputs taken from multiple individual classifiers, which are combined to form the final decision. A number of approaches have been proposed to consolidate the final recognition decision, such as AND rule, OR rule, Bayesian decision, decision table, majority voting, weighted majority voting, and other fusion rules [13][14].

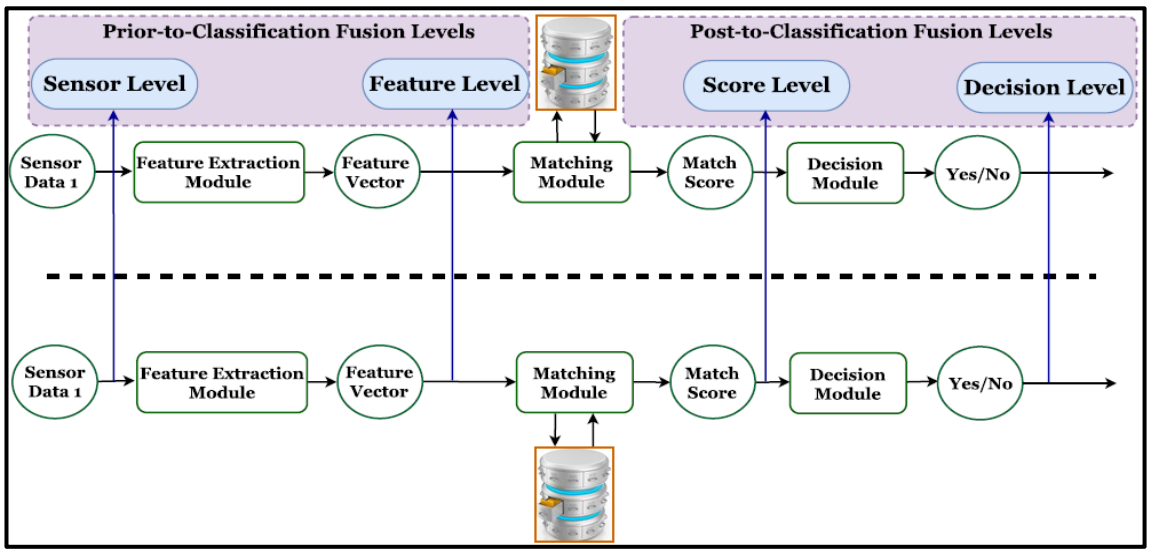


Figure 1.8: Four levels of fusion used in multi-biometric systems. Rank level fusion, not shown, is used only in the identification task, adapted from Ross [22].

1.7 Aims and Objectives

The main aim of this research is to design and implement a hybrid multi-biometric personal identification system for identifying a person's identity using a combination of face and iris biometric traits. These two biometric traits complement each other in that the face trait is the least intrusive, and the iris trait is the most accurate. Specific objectives are designing and implementing real-time, robust, and ready-to-use unimodal biometric algorithms for a multi-biometric system that integrates the strengths of these two biometric traits. Further objectives are increasing the degree of flexibility of the proposed multi-biometric system and eliminating the limitation imposed by the missing trait by selecting an appropriate fusion scheme. In order to meet and achieve the main objectives of the research the following tasks need to be investigated:

1. Reviewing and analyzing the current state-of-the-art recognition approaches for both face and iris traits, and proposing improvements to develop a powerful multi-biometric system that can meet the requirements of real-world applications. These investigations include several stages in the biometric system relating to pre-processing, feature extraction, fusion methodology, and classification. In this task, the emphasis will be on the fundamental theory, drawbacks, and hypothesis made by previous works.

2. Investigations into a proposed fully automated and robust face recognition algorithm in which a compact and discriminative facial representation is learned from face images which are taken under completely unconstrained conditions. In this thesis, a number of the most challenging face recognition problems are tackled, including a lack of training samples, changes of illumination, expressions, aging, occlusion and different poses.
3. Investigations into the design of a practical prototype and robust iris recognition system. This task involves several issues: (i) Proposal of an efficient and real-time iris localization method to separate the iris region from background features, such as eyelashes, eyelids, and specular reflection; (ii) Proposal for a novel feature extraction method to extract discriminative features from the localized iris region without any domain knowledge of the iris image data, which is differ from the previous work that depend on handcrafted features [25][26][27][28][29]. This idea is motivated by current promising findings applying the deep learning approaches to learn compact and discriminative features from other biometric trait (e.g., face trait); (iii) Proposal for a new matching approach, which could enhance the accuracy of the traditional binary iris matching approaches, such as Hamming distance.
4. Investigations into the design and implementation of a robust multi-instance system based on iris biometric trait taken from the right and the left eye of the same person.
5. Investigations into combining face and both irises traits in one multi-biometric personal identification system with multiple security levels. This task involves investigating into efficient fusion approaches for the adopted biometric traits (e.g. face and iris).
6. Conducting extensive experiments to assess the effectiveness of the proposed approaches and methodologies in different scenarios on large-scale unconstrained unimodal and multi-biometric databases. Furthermore, assessing the superiority of the proposed unimodal and multimodal systems compared to the previous state-of-the-art systems.

1.8 Thesis Contributions

In this section, the main contributions of this research to the biometrics community are highlighted. The first part of this PhD thesis addresses the problem of the unimodal biometric systems based on the face and iris trait, while the second part deals with the problem of designing and implementing a hybrid multimodal biometric identification system for identifying a person's identity using a combination of face and iris biometric traits. The main contributions of this dissertation are as follows:

1. A novel multimodal local feature extraction approach, based on merging the advantages of multidirectional and anisotropy transforms, specifically the Curvelet transform, with Fractal dimension, is proposed. Termed the *Curvelet-Fractal* approach, it is different from previously published Curvelet-based face recognition systems, which extract only the global features from the face image. The proposed method has managed to extract the local features along with the face texture roughness and fluctuations in the surface efficiently by exploiting the Fractal dimension properties, such as self-similarity. There are three main differences from the Curvelet Transform-Fractional Brownian Motion[8] (CT-FBM) approach. Firstly, unlike CT-FBM approach which used only the coarse band of the Curvelet transform as an input to the Fractal dimension stage, here other Curvelet sub-bands features are also used, which represent the most significant information in the face image (e.g. face edges and curves), which are known to be crucial in the recognition process. Secondly, a new Fractal dimension method, is proposed, based on an Improved Differential Box Counting (IDBC) method in order to calculate the Fractal dimension values from the new added Curvelet sub-bands and handle their high dimensionality. Then, the outputs of the IDBC and Fractional Brownian Motion (FBM) are combined to build an elementary feature vector. Finally, use of the Quadratic Discriminant Classifier (QDC) instead of K-Nearest

[8] The main details of the CT-FBM approach are in the conference paper, titled (A Robust Face Recognition System Based on Curvelet and Fractal Dimension Transforms), and it presented in this thesis only for comparison purposes.

Neighbour (K-NN) is proposed, because this improves the accuracy of the proposed system.

2. A novel framework is proposed, termed the Multimodal Deep Face Recognition (MDFR) framework, to learn additional and complementary features representations by training a deep neural network (e.g. Deep Belief Networks (DBNs)) on top of a *Curvelet-Fractal* approach, instead of the pixel intensity representation. This is a demonstration that the proposed framework can represent large face images, with the time required to obtain the final trained model significantly reduced compared to the direct use of the raw data. Furthermore, the proposed framework is able to efficiently handle the non-linear variations (e.g., intra-personal and inter-personal variations) of face images, and is unlikely to over fit to the training data, due to the non-linearity of a DBN.
3. An efficient and automatic iris localization model is proposed to carefully detect the iris region from the background and all extraneous features (e.g., pupil, sclera, eyelids, eyelashes, etc.), without the risk of losing important information from the iris region.
4. An efficient deep learning system is proposed called *IrisConvNet*, whose architecture is based on a combination of a CNN and Softmax classifier to extract discriminative features from the iris image without any domain knowledge, and classify it into one of ***N*** classes. To the best of the author's knowledge, this is the first work that investigates the potential use of CNNs for the iris recognition system, especially in the identification mode. It is worth mentioning that only two papers have been published recently [30][31] that investigated the performance of CNNs on the iris image. However, these two works have addressed the biometric spoofing detection problem with no more than three classes available, which is considered a simpler problem when compared to the iris recognition system where ***N*** class labels need to be correctly predicted.
5. A discriminative training scheme equipped with a number of training strategies is also presented in order to evaluate different CNN architectures, including the number of layers, the number of filters layer, input image size,

etc. To the best of the author's knowledge, this is the first work that compares the performance of these hyper-parameters in iris recognition.

6. An efficient and real-time hybrid multimodal biometric identification system is proposed, based on fusing the results obtained from the face and both the left and right irises of the same person. To the best of the author's knowledge, this is the first work that investigates the potential use of deep learning approaches (e.g. DBN and CNN) for fusing the face and both the left and right irises in a unified multimodal biometric system. In this work, a parallel architecture is considered, allowing users a high degree of flexibility to provide either a subset or all of their biometric traits, depending on the required security level and the users' convenience. In addition, a limitation due to excluding a biometric trait is eliminated.

1.9 Outline of the Thesis

This PhD thesis is organized into six chapters including this chapter. An overview of these chapters is presented below. The chapters' organization and the dependence among these chapters are illustrated in Fig. 1.9.

- **Chapter 1** introduces the basics of biometric systems, main characteristics of the biometric modalities, and the limitations of unimodal biometric systems. The motivations, objectives, contributions and outline of this PhD thesis are also presented in Chapter 1.
- **Chapter 2** presents a review of previous investigations into face recognition, iris recognition, and multimodal biometric systems. It also details the motivations for this PhD thesis based on the previous related works.
- **Chapter 3** considers the problem of the unconstrained face recognition in both identification and verification tasks, and presents two novel face recognition approaches. In particular, a novel multimodal local feature extraction approach is presented, termed *Curvelet-Fractal* approach, and a novel framework, termed the Multimodal Deep Face Recognition (MDFR) framework, to learn additional and complementary representations by training a DBN on top of existing local representations instead of the pixel intensity representations, is also presented.

- **Chapter 4** studies the problem of the iris recognition in identification task and presents an efficient and real-time iris recognition system based on building deep learning representations for images of both the right and left irises of a person. In particular, the proposed iris localization model is composed of four main stages and *IrisConvNet* system, whose architecture is based on a combination of a CNN and a Softmax classifier.
- **Chapter 5** presents an efficient and real-time hybrid multimodal biometric system for identifying a person's identity, based on fusing the matching scores generated from the face and both irises biometric traits at the score and rank level. In this chapter, different types of the multimodal biometric system are also proposed, based on the employed approach to generate the matching score and the biometric traits selected by the user at identification point.
- **Chapter 6** introduces the overall conclusions, achievements and some limitations of the thesis. Possible future directions of this research are also outlined in this chapter.

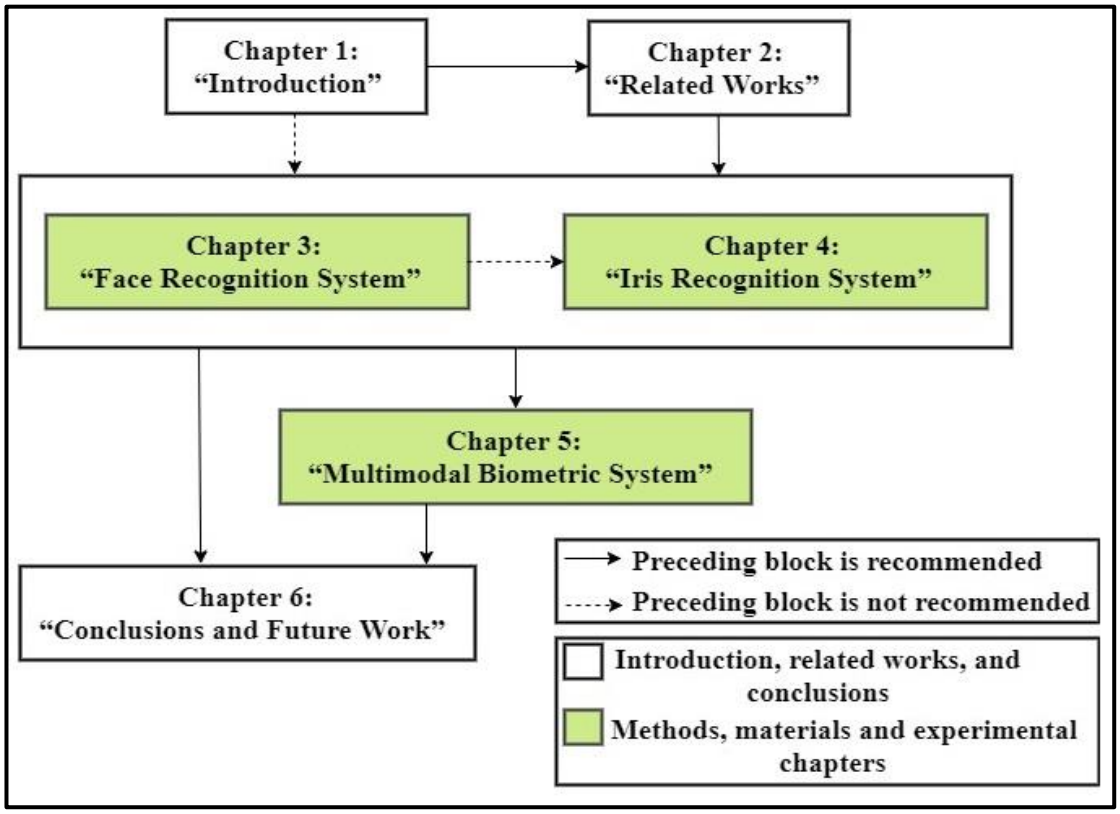


Figure 1.9: The chapters' organization and the dependence among dissertation chapters.

Chapter 2
Related Work

2.1 Introduction

This chapter briefly summarizes previous work related to this PhD thesis. The literature on unimodal and multimodal biometric systems shows a variety of approaches that have been used to establish a person's identity in a wide range of real-world applications. In this chapter, previous work on the unimodal biometric systems that use either the face or iris traits in the decision-making process is described. Also included is a discussion of some of the previous research done on multi-biometric systems that combine multiple biometric sources at different levels of fusion, with more concentration given to those works that use one or both of the adopted traits in this thesis. The main aim of this chapter is not to give a comprehensive review of all the existing work dealing with each of the face recognition, iris recognition, and multi-biometric systems, but to summarize the most recent works which are closely related to this PhD thesis and to concentrate on the most important contributions and limitations existing in the current state-of-the-art aforementioned topics.

The chapter is organized as follows: Firstly, previous research on the face recognition approaches is presented in Section 2.2. Section 2.3 outlines the most important contributions found in previous work to address the iris recognition problem. Section 2.4 discusses and analyses the main investigations carried out to date in the field of multi-biometric systems using different fusion methodologies. Section 2.5 discusses the main challenges and knowledge gaps in the previous research and finally, the summary and conclusions of this chapter being presented in Section 2.6.

2.2 Research on Face Recognition

In the last four decades, face recognition has received considerably more attention in the research community than other biometric traits, due to its wide range of commercial and governmental applications, low cost and ease of

capturing the face image in a non-intrusive way without the user's active cooperation [14]. Face recognition systems encompass two fundamental stages: Feature extraction and classification. The second stage is dependent on the first. Therefore, the task of extracting and learning useful and highly discriminating facial features in order to minimize intra-personal variations and maximize inter-personal differences is a challenging task. In this regard, a number of approaches have been proposed, implemented and refined to address some of these drawbacks and problems in the face recognition system. These approaches can be broadly divided into two categories: handcrafted-descriptor approaches, and learning-based approaches [32]. The next subsections are devoted to outlining these approaches and their usage to address face recognition problem, before discussing their respective strengths and limitations.

2.2.1 Handcrafted-Descriptor Approaches

The majority of previous face recognition systems depend on feature representations produced by local/global handcrafted-descriptor approaches, such as Scale Invariant Feature Transform (SIFT), Local Binary Patterns (LBP), 2D Gabor Wavelet, etc. Handcrafted-descriptor approaches can be further divided into three groups: Feature-Based, Holistic-Based, and Hybrid-based approaches.

In the first category, a geometric vector representing the facial features is extracted by measuring and computing the locations and geometric relationships among facial features (e.g., the mouth, eyes and nose), as shown in Fig. 2.1, and using it as an input to a structural classifier. In fact, early face recognition systems mainly relied on approaches belong to this category. One of the earliest such systems was by Wiskott et al. [33], who employed the Elastic Bunch Graph Matching (EBGM) method to reduce the input face image to a vector of geometric features. The EBGM is an example of a features-based approach, which uses the responses of Gabor filters at different orientations and frequencies at each facial feature point to extract a set of local features. This technique has the ability to provide feature representation that is invariant to affine transformations (e.g., translation, rotation, etc.) and changes in facial

expressions. A recognition rate of 98% has been achieved on a subset of 250 face images from the FERET database.

Kanade [34] proposed a simple feature extraction algorithm to extract a geometrical vector composed of 16 facial parameters, which are the ratios of distances, angles and areas between detected facial features in the input image, such as eyes, nose and mouth. Then, the Euclidean distance is employed in the matching stage to achieve a recognition rate of 75% on a small database of 20 persons (2 images per person). An extension of Kanade's approach was also proposed by Brunelli and Poggio [35], where the size of the extracted feature vector was increased to 35 geometric features. The authors were able to achieve a recognition rate of 90% when testing the system on a larger database of 47 persons (4 images per person). Biswas et al. [36] developed a face recognition system based on feature representations obtained using SIFT-based descriptors [37] at fiducial locations on the face image, and then the SIFT features of all facial landmarks were formed into a single vector representing the face. A best Rank-1 identification rate of 91% was achieved on the CMU Multi-PIE face database containing a total of 337 subjects. More recent Feature-Based approaches benefit from the great progress in facial landmark detection algorithms [38][39], which make dense facial landmarks detection more accurate, for example, the work proposed by Chen et al. [40], in which the multi-scale LBP was employed to extract features from regions around 27 facial landmarks. Similar ideas were also proposed for feature extraction in [41][42][43]. For more examples on the feature-based approaches, the reader is referred to [44][32]. The main advantage offered by these types of approach is that they are relatively robust to variations in the face image (e.g. changes of illumination, size, head position, etc.) and high-speed matching. However, the main limitation of these approaches is the difficulty of automatically detecting the facial features [45].

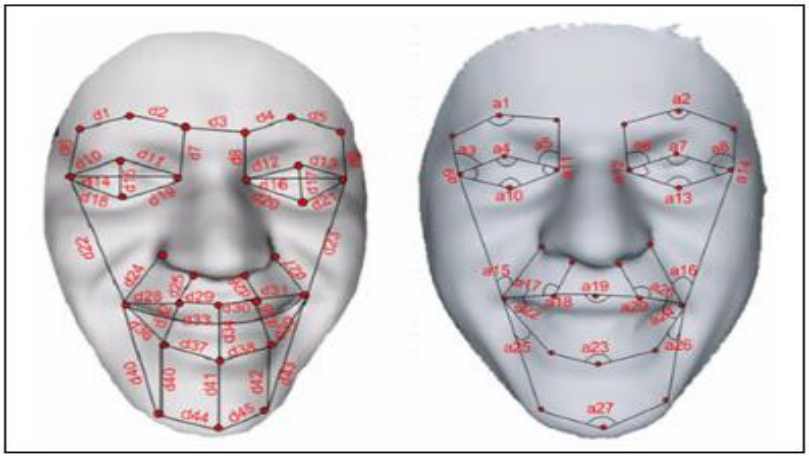


Figure 2.1: Some examples of geometrical features of the face image.

Compared with the first category, the holistic-based approaches usually extract the feature vector by operating on the whole face image instead of measuring the local geometric features. The major limitation of these approaches is performing the matching process in very high dimensionality space. Thus a number of face recognition algorithms have been developed that employ different dimensionality reduction methods to handle the issue of high dimensionality before the matching process can take place. The Eigenface methods are the best well-known examples of these approaches, which are represented by Principal Component Analysis (PCA), Independent Component Analysis (ICA), etc. The first well-known face recognition system based on the Eigenface methods was proposed by Sirovich and Kirby [46]. In this system, the PCA approach was applied on every face image in the system's database, and then projected onto a lower dimensional space, in which each face image was transformed into a vector of weights, so-called (Eigenfaces). In the testing phase, the weights vector of the query image were obtained in the same manner, and then the decision was made by measuring the similarity between each query-template pair. If the query image's weight was lower than a pre-defined threshold, it was considered to be accepted.

Based on Sirovich and Kirby’s findings, an improved system was proposed by Turk and Pentland [47], who demonstrated that these projections of the face image along Eigenfaces space could be used as input feature vectors to the classification approaches to recognize the person's identity efficiently.

Experiments on a private database consists of 2500 face images captured from 16 subjects yielded a ~100% recognition rate. Compared to the PCA approach that depends on the first and second-order statistics, the ICA approach has the ability to explore higher order statistics, which enables it to be successfully applied as a feature extraction approach to address the face recognition problem. However, the main drawback of this approach is the higher computational time required, compared to PCA approach. This problem has been partially overcome using a Fast-ICA version proposed by Bartlett et al. [48]. A comparison study presented in [49] demonstrates that a higher recognition rate can be obtained using ICA compared to PCA.

Unlike the unsupervised Eigenfaces approaches (e.g., PCA and ICA) that reconstruct a face image from a low dimensional space and discard the class information, the supervised Fisherfaces approach, based on Linear Discriminant Analysis (LDA), has also been proposed by Belhumeur et al. [50] for face recognition. Firstly, like the Eigenfaces approaches, the Fisherfaces approach aligns face images in the training set, projects them into a lower dimension space, and then instead of using maximum variation directions as in the PCA, the Fisher LDA is applied to select the projection directions. This approach aims to improve the discrimination power by maximizing the ratio of the determinant of the between-class scatter matrix to the determinant of the within-class scatter matrix of the projected samples. Fig.2.2 shows some examples of Eigenfaces and Fisherfaces approaches, and how Fisherfaces can produce more discriminative information better than Eigenfaces. In fact, several publications have demonstrated the efficiency of the Fisherfaces approaches by exploiting class variation information, as in [51][52][53][54]. Furthermore, numerous extensions and non-linear schemes have been proposed to improve the performance of the standard Eigenfaces and Fisherfaces approaches, such as Kernel PCA [55][56], Fourier-LDA [57], Pair-wise LDA [58], and incremental LDA [59]. Other well-known examples of holistic-based approaches, include the Active Appearance Model (AAM) [60], LBP [61], the 3D Morphable Model [62] and Locality Preserving Projections (LPP) [63]. The major characteristic of the holistic-based approaches is that no important information is destroyed by focusing only on restricted regions of interest. However, these types of

approach usually suffer from some drawbacks, such as performing well only under fully controlled conditions (e.g., standardized lighting, angle, scale, etc.), being computationally very expensive, and producing a feature vector with very high dimensionality. Nevertheless, as mentioned before, several of these approaches have been devised to handle these limitations.

Finally, the hybrid-based approaches use a combination of both feature and holistic approaches to address the face recognition problems. For example, one can argue that hybrid approaches that use both local and global features (e.g., obtained from both feature and holistic approaches, respectively) can be an efficient way to make use of complementary information, reduce the complexity of classifiers, and increase their generalization capability. However, the main challenging factors that can influence the performance of such approaches are in how to define which types of features should be fused, and how to fuse them [64]. For example, local features are very sensitive to the illumination changes, while expression changes have more influence on global features. Some of the recent examples belonging to this category were proposed by Fischer et al. [65], Berg and Belhumeur [66], Annan et al. [67], Zhu et al. [68], Ding et al.[69] and other examples can be found in [64].

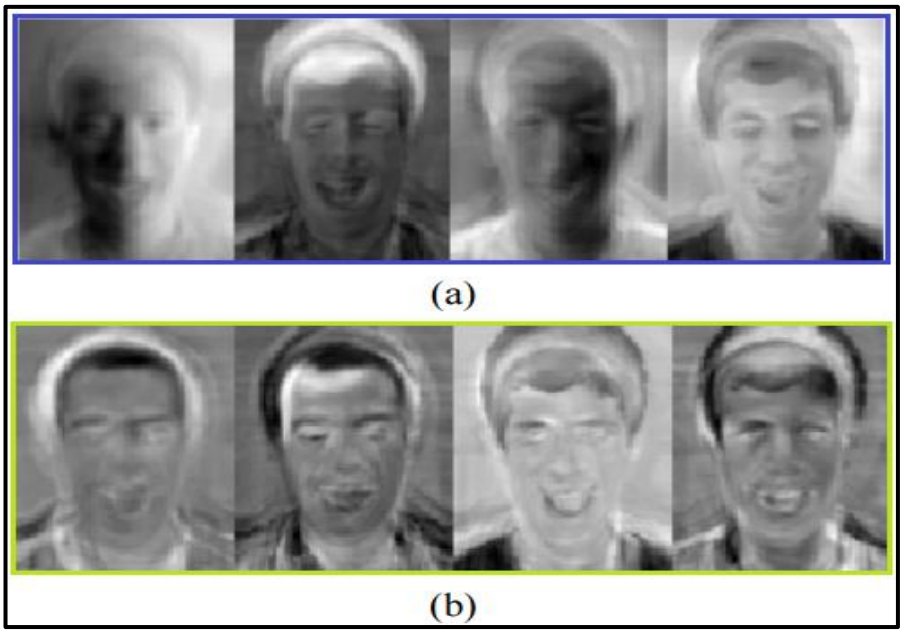


Figure 2.2: (a) Some examples of the Eigenfaces approach showing the tendency of the PCA to acquire the major variations in the training set (e.g., lighting directions), and (b) Some examples of the Fisherfaces approach show its ability to discard information unrelated to classification [70].

Multi-resolution transformation approaches (e.g., Wavelet transform, Curvelet transform, Ridgelet transform, etc.) have been widely used in conjunction with the aforementioned feature-based, holistic-based and hybrid-based approaches. For example, Mukhedkar and Powalkar [71] proposed a facial feature extraction approach based on a combination of the Discrete Wavelet Transform (DWT) and PCA. The former is used to reduce the high dimensionality of the face image with different decomposition levels, while the latter is for the feature extraction process. The Euclidean distance is employed to make the final decision. Experimental results on four face databases, namely, Face94, Face95, Face96, and SELFFACE, show that this approach produced a higher recognition rate compared to using only the PCA approach. A face recognition system based on the DWT and the Regional Directional Weighted-Local Binary Pattern (RDW-LBP) was also developed by Fengxiang [72]. Experiments on AR and ORL face databases reported recognition rates of 97.2% and 99.2%, respectively. Mandal and Wu [73] developed a face recognition system based on the Fast Discrete Curvelet Transform (FDCT) to decompose face images into Curvelet sub-bands. Then, the PCA approach was employed on selected sub-bands to produce a representative feature set. They compared the performance of the proposed system with the Wavelet transform and traditional PCA approaches, and indicated that the proposed system produced higher recognition rates of 96.6% and 100% on ORL and Essex Grimace face databases, respectively. Rziza and et al. [74] used the Curvelet transform to decompose the face image into a number of sub-bands that characterize the face texture. Then, only a selected subset of the Curvelet sub-bands was used and divided into small non-overlapping blocks, from which a set of simple statistical measures (such as mean, variance and entropy) were computed to form compact and meaningful feature vectors. These feature vectors were used as inputs to the LDA approach to make the final decision. Recognition rates of 98%, 93.33%, and 91.72% were achieved on ORL, YALE and FERET face databases, respectively.

Quite recently, Elaiwat et al. [75] presented a multimodal Curvelet-based technique for 3D face recognition. Firstly, each face image was transformed into the Curvelet domain, then a number of robust and distinctive key points (e.g.,

eyes-forehead and nose) were detected from textured 3D faces across different frequency sub-bands. Experimental results on BU-3DFE, FRGC v2, and Bosphorus face databases yielded verification rates of 95.1%, 99.2%, and 91% at 0.001 FAR. Other well-known examples of Curvelet-based face recognition systems were also proposed by Ch'ng et al. [76], Tanga and Chen [77], Zhou et al. [78], and Arivazhagan et al. [79].

2.2.2 Learning-Based Approaches

Previous research has demonstrated the efficiency of handcrafted-descriptor approaches used as robust and discriminative feature detectors to solve the face recognition problem, even when relatively few training samples per person are available, as in [80][81]. However, the performance using handcrafted-descriptors approaches declines dramatically in unconstrained conditions, due to fact that the constructed face representations are very sensitive to the highly non-linear intra-personal variations, such as expression, illumination, pose, and occlusion [82]. In addition, this type of approach usually requires significantly more effort and domain-knowledge from the researcher to find the best compact and discriminative feature set for a given problem. Thereby, they are considered to be difficult and time demanding.

To address these drawbacks, considerable attention has been paid to the use of learning-based approaches, which can learn features from labelled training samples, using machine learning techniques to recognize faces. The main advantage of the learning-based approaches over handcrafted-descriptor approaches is their ability to learn from experience and their robustness in handling non-linear variations of face images caused by pose variations, self-occlusion, etc. Some examples of traditional approaches belonging to this category are Support Vector Machines (SVMs) [83], Artificial Neural Networks (ANN) [84][85] Discriminant Face Descriptor (DFD) [86], Local Quantized Patterns (LQP) [87] and Multi-Layer Perceptron (MLP) [88], etc. For example, Ouarda et al.[89] used a set of geometric features containing a total of 33 points for each face image as input to train a Support Vector Machine (SVM). In experiments, the authors evaluated the performance of three different types of kernels (e.g., Linear, Radial Basis function and MPL kernels) with two features

selection approaches using the genetic algorithm and minimum Redundancy maximum Relevance (mRmR). The best recognition rates achieved were 88.50% (using Linear kernel and GA) and 93.07% (using Linear kernel and mRmR) on the ORL and CALTECH databases, respectively. However, these traditional approaches have a number of drawbacks that need to be overcome, such as requiring low dimensional feature vectors in order to achieve an exceptional performance, difficulty in handling deformations of the input image (e.g., translation, scaling and rotation), and the slow convergence of the shallow neural networks used (e.g., MLP networks) [64].

Inspired by their impressive ability to learn more discriminative feature representations, considerable attention has also been paid to the use of deep learning approaches, such as Deep Neural Networks (DNNs) to learn automatically a set of higher-level feature representations through hierarchical non-linear mappings, which can robustly handle the non-linear intra- and inter-personal variations of face images [90]. Moreover, in contrast to handcrafted-descriptor approaches, the applications making use of deep learning approaches can generalize well to other new fields. Some of the most relevant examples using deep learning approaches for face recognition are reviewed below.

Sun et al. [82] proposed a face recognition system, called Deep Identification-verification features (DeepID2), in which two supervisory signals (e.g., the face identification and verification signals) were combined simultaneously to boost the discriminative power of extracted features using CNNs. They showed empirically that the learned DeepID2 feature representations can generalize well to new identities in the Labeled Face in the Wild (LFW) database. A verification rate of 99.15% was achieved using the "Unrestricted, Labeled Outside Data" evaluation protocol. The same group of researchers developed the deep ConvNet-RBM model based on a combination of a set of CNNs and a Restricted Boltzmann Machine (RBM) model for face verification in the wild environment [91]. Verification rates of 91.75% and 92.52% were achieved on the LFW and CelebFaces databases, respectively. The main architecture of the ConvNet-RBM model was further improved in [92], in which very high dimensional relational features were learned for face

verification rather than the low dimensional features learned in the previous version. A verification rate of 93:83% was achieved on the LFW database using the "Unrestricted, Labeled Outside Data" evaluation protocol. Taigman et al. [93] proposed the DeepFace framework for face verification using the frontal face images generated from a 3D shape model of a large-scale face dataset. They proposed to use the softmax loss on the top of CNN as the supervisory signal to train the proposed framework. Experiments on the LFW database using the "Unrestricted, Labeled Outside Data" sitting yielded a verification rate of 97.35%. In 2015, a comprehensive deep learning framework for face recognition, called the Multimodal Deep Face Representation (MM-DFR) was proposed by Ding and Tao [94]. The main structure of the MM-DFR framework was composed of a set of CNNs to extract complementary facial representations from multimodal data (e.g., the holistic face images, the frontal face images rendered by 3D Face Model, and sampled face patches), and a three-layer Stacked Auto-Encoder (SAE) to reduce the high dimensionality of the feature vector obtained from combining the outputs of all the CNNs. Experiments on the LFW and CASIA-WebFace databases reported a verification rate of 99.02% using the "Unrestricted, Labeled Outside Data" protocol and a Rank-1 identification rate of 76.53%, respectively.

Liu et al.[95] proposed a Boosted Deep Belief Network (BDBN) framework to learn a set of features for facial expression recognition. A number of extensive experiments were conducted to assess the performance of the BDBN framework on Extended Cohn-Kanade (CK+) and JAFFE databases which yielded average classification rate of 97% and 68%, respectively. A new translation-invariant version of DBNs, called Convolutional Deep Belief Networks (CDBN) was proposed by Lee et al. [96], and used for face recognition. Experiment on the Caltech-101 face database by the authors reported a better performance compared to the state-of-the-art approaches, using SIFT operator, shape-context and geometric blur. Later, Huang et al. [97] suggested the learning of additional complementary facial representations by training a CDBN on the top of a handcrafted-descriptor approach (e.g., LBP method) instead of using the pixel intensity representations directly. In addition, a Local Convolutional Restricted Boltzmann Machines (LCRBM) was also

developed by these authors to make use of the global structure in a face image. They were able to show that complementary feature representations can be captured by combining the deep learning and handcrafted-descriptors approaches to improve the face verification results on the LFW database. A verification rate of 86.88% was achieved using the "Image-Restricted, Label-Free, Outside Data" protocol. Similar ideas were also proposed by Li et al. [98] and Yi et al. [99]. In the literature, several metric learning approaches have been proposed to address the face verification problem, including Fisher Vector (FV) [100], Information-Theoretic Metric Learning (ITML) [101], Discriminative Deep Metric Learning (DDML) [102], Pairwise-constrained Multiple Metric Learning (PMML) [103], Logistic Discriminant based Metric Learning (LDML) [104], Large Scale Metric Learning (LSML) [105], Cosine Similarity Metric Learning (CSML) [106].

The key limitations of the deep learning approaches include requiring a large amount of training data to avoid the overfitting problem and increase the generalization ability of the neural network, the number of hyper-parameters that need to be set, and all the DNNs are mainly based on a tedious iterative optimization procedure that can be computationally expensive for large-scale databases. However, these issues can be effectively alleviated using high-performance computing systems, equipped with Graphics Processing Units (GPUs) [90]. In addition to all the aforementioned examples, there have also been several comprehensive surveys on different aspects of face recognition, including 3D face recognition [107], heterogeneous face recognition [108], illumination-invariant face recognition [109], video-based face recognition [110], single image-based face recognition [111] and Unimodal and Multimodal Face Recognition [112].

2.3 Research on Iris Recognition

Iris recognition technology has been widely employed, in many mission-critical applications, for instance, the iris recognition system is being employed to check visitors to the United Arab Emirates (UAE) [113], in India’s Unique ID program [114] and some airports (e.g., The Schiphol Privium scheme at the

Amsterdam airport[9]), that uses iris-scan cards to speed up the border control procedures. Among the different biometric modalities, the iris trait is considered to be the most accurate that can play an important role in identifying a person's identity with a remarkably high recognition rate. The uniqueness of an iris pattern comes from the distinctiveness and richness of texture details within the iris region (the annular part between pupil and sclera), including rings, ridges, crypts, furrows, freckles, zigzag patterns, etc. Since 1987, when the idea of using the iris as a biometric trait was first introduced by Flom and Safir [115], many different approaches have been proposed, and great progress on iris recognition has been achieved. In 1993, the first successful and the commercially available iris recognition system was proposed by Daugman [116]. In this system, the inner and outer boundaries of the iris region were detected using an Integro-differential operator. Afterwards, the iris template was transferred into the normalized form using Daugman's Rubber Sheet method. This is followed by using a 2D Gabor filter to extract the iris features and the Hamming distance for decision making. However, as reported in [117][118][119], the key limitation of Daugman's system was that it required a high-resolution camera to capture the iris image, and its accuracy significantly decreases under non-ideal imaging conditions due to the sensitivity of the iris localization stage to noise and different lighting conditions. Another well-known iris recognition system was developed at Sarnoff Labs by Wildes [120]. In Wildes' system, the binary edge map of the iris images was computed first. This was followed by applying a Hough transform to detect the iris's region boundaries. Then, a Laplacian pyramid at multiple scales was used to produce an iris code. Finally, the normalized correlation between the query and template representations was computed, based on Fisher's Linear Discriminant (FLD). Unlike Daugman's system that uses an LED point light source coupled with a standard video camera to capture the iris image, Wildes proposed the use of a diffuse polarized illuminator with a low light level camera in the image acquisition process. It is fair to say that later all the proposed approaches for iris recognition have developed with the same concepts as these three mentioned examples. In this PhD thesis, numerous historical achievements and current

[9] https://www.dogsonacid.com/threads/iris-scanner-now-at-schiphol-airport.15949/

state-of-the-art approaches for iris biometric are reviewed and categorized into three groups, based on their primary contribution to one of the three major modules of an iris recognition systems: iris localization, iris texture analysis, and iris template matching. These three categories of approaches are reviewed in the following subsections.

2.3.1 Iris Localization Approaches

Recently, iris segmentation has been receiving extensive attention, due to its direct effects on the performance and the accuracy of the iris recognition system. In this regard, a number of iris segmentation algorithms have been proposed in the past. As mentioned earlier, Daugman [116] proposed the first implemented iris segmentation method using an integro-differential operator for detecting iris boundaries and eliminating possible extraneous features such as the eyelids and eyelashes. This operator calculates the partial derivative of the average intensity of circle points, taking into account the increasing radius (***r***), followed by convolving the operator with a Gaussian filter. Then, the centre and radius of the iris boundaries are identified by computing the maximum difference between outer and inner circle. Finally, the eyelids are detected by using a parabolic curve fitting approach. The main limitation of Daugman's method is that the Integro-differential operator can be easily affected by local gradient maxima where light spots can change the gradient greatly, which results in inaccurate iris localization. Wildes [120] proposed a Circular Hough transform-based filtering and voting procedure to localize the two circular boundaries of the iris. An edge detection filter applied to the eye image generated an edge image map which was subjected to the Hough transform voting procedure to find the desired edge map contour. The centre and radius of the circle with a maximum number of votes were detected as an iris boundary. The main limitation of this method is its computational complexity which makes it inappropriate to meet real-time requirements. Several iris localization algorithms were developed by Huang et al. [121], Liu et al. [122], Lili and Mei [123], He and Shi [124], and Feng and et al.[125], aiming to improve upon Wildes' system by reducing the computational complexity of the Hough transform.

Cui et al. [126] proposed a segmentation algorithm based on scaling the eye image using a Haar wavelet, followed by applying a Hough transform and integro-differential operator to detect the inner and outer boundaries, respectively. Then, histogram and parabolic arc based approaches were used to detect lower and upper eyelids, respectively. Ross and Shuh [127] proposed another iris segmentation algorithm using Geodesic Active Contours (GAC) to detect the iris region from input eye image. Firstly, the pupil was detected by applying a median filter on the eye image. This was followed by a simple thresholding method and circle fitting procedure. Then, an iterative process, based on the GAC was employed to detect the outer boundary as well as the eyelid contour. The performance of the segmentation algorithm was tested by applying multiple Gabor filters on a normalized iris template to observe its performance in the recognition task on the CASIA V 1.0 and WVU iris databases. A similar idea using GAC was also proposed by Pawar et al. [128]. Jan et al. [129] developed an iris localization algorithm based on a two-phase strategy. The first phase was employed to detect the pupil circle using a Hough transform, gray level statistics and adaptive thresholding method, and a new geometrical transform, the iris circle, was detected in a sub-image centred in the pupil circle in the second phase. Finally, the iris boundaries were regularized using the radial gradients and active contours. Sahmoud et al. [119] suggested enhancing the performance of the Circular Hough Transform (CHT) by applying the K-means clustering algorithm as a pre-processing step to separate the eye image into three different regions, namely the iris region, the skin region and the sclera region. Then, the non-iris regions were excluded to reduce the search space of the CHT used to estimate the radius and centre coordinates of the iris circle. Finally, a new robust algorithm was proposed to detect and isolate eyelashes, eyelids and specular reflections. The literature on iris localization shows a variety of approaches, such as Active Shape Models (ASM) [130], Genetic algorithms [131], Fuzzy logic [132][133], ANN [134][135], Adaptive thresholding and Histogram approaches [136] have also been used to detect efficiently the iris boundaries and extract the iris region from the background.

2.3.2 Iris Texture Analysis Approaches

In the last decade, working on iris texture analysis to produce discriminative representations of the iris region has attracted much attention from research teams, and various approaches have been proposed in the literature to tackle the drawbacks in this stage of the iris recognition pipeline. Some of the most relevant examples of performing texture analysis for iris recognition are reviewed below.

Sun et al. [137] proposed an iris recognition system, called Robust Direction Estimation (RDE). This system was based on estimating the local dominant direction of Gradient Vector Field (GVF), in which an iris image was convolved with a Gaussian filter to produce the local orientation at each pixel in the unwrapped iris region. The performance of the RDE system was evaluated using 2,255 iris images acquired from 213 subjects in the CASIA iris database. Harjoko et al. [138] employed four levels of Coiflet Wavelet Transform (CWT) to extract discriminatory information from the iris image. Then, the similarity between two irises representations was measured using a modified Hamming distance. A good performance was achieved on two iris databases: CASIA V 1.0 and MMU 1 (with left and right eyes of MMU 1evaluated separately), with identification rates of 84.25%, 77.78%, and 86.67%, respectively. Sun et al.[139] developed a set of multi-lobe differential filters to calculate ordinal measures for iris texture analysis with the aim to characterize the qualitative relationships between iris regions instead of precise measurements of iris pattern structures. The authors reported that producing such iris representation may lose some image specific information. However, a good trade-off between distinctiveness and robustness was achieved. Experiments on four iris databases: UBath, CASIA V1.0, CASIA-IrisV3-Interval, and ICE2005 (left & right eye) reported an EER of 4.39×10^{-3}, 3.70×10^{-3}, 3.48×10^{-3}, 6.32×10^{-3}, and 4.68×10^{-3}, respectively. Velisavljevi [140] presented a novel iris recognition algorithm which used oriented separable wavelet transforms for extracting iris features. Then, a weighted Hamming distance was employed in the decision-making process. An identification rate of 94.7% and a verification rate of 95.88 at a FAR = 0.1% were achieved on the CASIA-IrisV3-Lamp database. Gulmire and Ganorkar [141] proposed a feature extraction method, based on ICA to

generate the iris code after detecting the iris region using Daugman's integro-differential operator. A recognition rate of 89.5% was achieved on the CASIA V1.0 database.

The idea of using DWT as a feature extraction method for iris recognition was also proposed by Elgamal and Al-Biqami [142]. The high features dimensionality of the DWT's sub-bands were first reduced using PCA, and then used as input vectors for k-NN to make the final decision. Experiments on the IIT Delhi iris database yielded a recognition of 99.5%. A robust iris recognition algorithm, based on an orientation field, was developed by Patil [143]. In this algorithm, the Hough transform was employed to detect the iris boundaries, and then the detected iris region was divided into a number of non-overlapping blocks. The size of the blocks was set empirically to be (16×16) pixels. Then the block orientation was computed from the pixel gradient orientations based on an averaging and optimization procedure. This was followed by applying the Sobel gradient operator to compute the variance of orientation, which was used as a feature vector by an L2 norm classifier. Experiments with 372 iris images from 54 subjects selected from the CASIA iris database achieved genuine acceptance rates of 96.33% at 11.3% FAR and 78% at 0.17% FAR. Bharath et al. [144] proposed two novel iris recognition methods based on Gradient Isolation (GI) and Radon Transform Thresholding (RTT). GI is a pre-processing method that utilizes the edge detection characteristic of the Gradient operator to isolate the iris patterns, thus obtaining salient iris textures. Then, the RTT method was applied to extract the prominent features from the enhanced iris image. A feature selection algorithm, based on the Binary Particle Swarm Optimization (BPSO), was employed to select the optimal features for the matching stage. Experiments yielded recognition rates of 95.62%, 95.93%, and 84.17%, respectively, on three iris databases: Phoenix, IIT Delhi and CASIA-IrisV3-Interval.

More recently, Umer et al. [145] proposed a multiscale morphologic operator, based on a multiscale top-hat transformation, to extract structural/textural iris features. Good recognition performance was claimed on four iris databases, namely UPOL, MMU1, IIT Delhi, and UBIRIS with identification rates at Rank-1 of 100%, 99.55%, 98.37%, 97.51% and

verification rates of 100%, 100%, 99.55%, and 98.34%, respectively. Another work by the same group [146] introduced a novel texture feature method for iris recognition, based on a Texture Code Matrix (TCM) generated from the normalized iris image. The TCM representations were used to compute a co-occurrence matrix. Then, the prominent texture features were obtained from the co-occurrence matrix. Experiments on four iris databases, namely, UPOL, CASIA-Iris-V3, MMU1, and IIT Delhi reported identification rates at Rank-1 of 100%, 100%, 97.78%, and 99.52% and verification rate of 100%, 100%, 100%, and 99.96%, respectively. The results in [145] and [146] were obtained after combining the outputs of both left and right iris images for each person using different fusion approaches. Sangeetha [147] proposed representing the iris code by combining the local and global features of a normalized iris region using two cross over schemes in a genetic algorithm to produce a binary feature vector of size 64 bits. In this system, four iris local features, Contrast, Correlation, Homogeneity, and Energy were computed using the Gray Level Co-occurrence Matrices (GLCM), while histograms, coefficient correlations, means, and standard deviations were extracted as iris global features. A recognition rate of 98.75 was reported on the IIT Delhi database. Some examples of well-known approaches also proposed for encoding the most prominent features in the iris region include Log-Gabor Wavelet [148], a combination of DWT and Discrete Cosine Transform (DCT) [149], Fast Fourier Transform (FFT) and Moments method [150], Contourlet Transform [151], LBP [152], Neighbourhood-Based Binary Pattern (NBP) [153], PCA, Haar Transform and Block Sum algorithm [26], Speeded Up Robust Features (SURF) [154], SIFT [155], and other various approaches can be found in [156][157][158].

2.3.3 Iris Template Matching Approaches

Once the process of encoding the iris features obtained from the segmented iris region is accomplished, the iris template matching task is performed, in which those encoded feature representations of unknown individuals are compared against the reference templates previously stored in the system's database. In the literature, several theories have been proposed to develop a decision-making tool for matching iris patterns. Hao et al. [159] proposed a fast search algorithm, named the Beacon Guided Search (BGS), to

speed up the matching process for a large database of iris codes. This algorithm was based on a brute force exhaustive search over a large iris database to find the best match. Experiments with 632,500 iris codes stored in the (UAE) border control system have reported significant improvements in search speed, with a slight loss of accuracy. Mehrotra et al. [160] reduced the matching time by employing an indexing algorithm using the energy histogram. In this algorithm, a multi-resolution DCT was applied to divide the unwrapped iris image into sub-bands, and then an energy histogram was computed for each sub-band, using all the iris images stored in the database. The iris images that have similar energy values were grouped together and associated with the same key. These keys were structured into a search tree. When a new iris image was submitted to the system, the algorithm calculated the key for this image and compared it with all iris images that have the same matching keys. Low penetration rates of 0.63%, 0.06% and 0.20% were achieved with CASIA V 1.0, BATH and IITK iris databases, respectively. A similar idea was also proposed by Qiu et al. [161] by dividing iris images into different groups based on discriminative visual features, which they named Iris-Textons. In this algorithm, a K-means algorithm was applied to determine which group an iris image fell into, and a recognition rate of 95% was achieved on a five-category iris database. Chou et al. [162] proposed an edge-type matching method to perform iris recognition. They stated that this method could be easily implemented using the concept of classifier ensembles. The performance was evaluated on CASIA-V3-Interval and UBIRIS-V1 iris databases, achieving EERs of 0.031% and 0.258% respectively. Khedkar and Ladhake [163] compared the performance of three different learning approaches, including MLP, Radial Basis Function (RBF) and SVM. Experiments on the CASIA V 1.0 database reported that an MLP with one hidden layer containing 20 units outperformed the others in term of recognition rate, measured using a cross-validation procedure. Rai and Yadav [164] proposed a novel iris matching technique, based on a combination of SVM with Hamming distance to enhance the accuracy of iris recognition. In this system, the authors proposed using the former as a master classifier, and the latter as a secondary classifier. Recognition rates of 99.91% and 99.88% were achieved on CASIA and Chek

iris databases, respectively. Melin et al. [165] proposed an iris recognition system, based on a set of optimized Modular Neural Networks (MNN) using genetic algorithms. The authors evaluated different combination of their approach with other methods, including Type-1 fuzzy integration, optimized fuzzy integration using genetic algorithms, and gating network methods. A recognition rate of 99.76% was reported on the CASIA V 1.0 database using the proposed approach with the fuzzy integrator (Gaussian type MFs). In 2015, Saminathan et al. [166] introduced a novel machine learning-based multi-class SVM approach for both iris identification and verification. In their approach, three different types of kernel functions (e.g., polynomial, linear, and quadratic) were investigated and combined using three fusion methods (e.g., quadratic polynomial, sequential minimal optimization, and least squares). They also compared the results obtained from proposed approach with three other classification methods, namely Hamming distance, LBP and Feed Forward Neural Network (FFNN). Experiments on the CASIA-IrisV3-Interval database showed that the highest recognition rates were achieved using the combination of the least square method and quadratic kernel function with an identification rate of 98.5% and a verification rate of 100%. More examples belonging to this category can be found in [156][158] and also a variety of machine learning techniques for iris recognition can be found in [167].

2.4 Research on Multi-biometric Systems

In the last decade, research on multi-biometric systems has attracted much attention in the biometric community and a number of efficient fusion strategies have been proposed to fuse the biometric data at different levels of fusion. The literature shows that five possible levels of fusion are employed for fusing data in the multi-biometric systems. These levels are: the sensor level, the feature level, the score level, the rank level and the decision level. As shown in Fig. 2.3, this thesis follows the same taxonomy described by Jain et al. [13] to outline the state-of-the-art approaches of the multi-biometric systems using different levels of fusion. The following subsection provides a comprehensive review of state-of-the-art multi-biometric systems, with more emphasis on systems in which the face or iris biometric traits are used in the decision-making process.

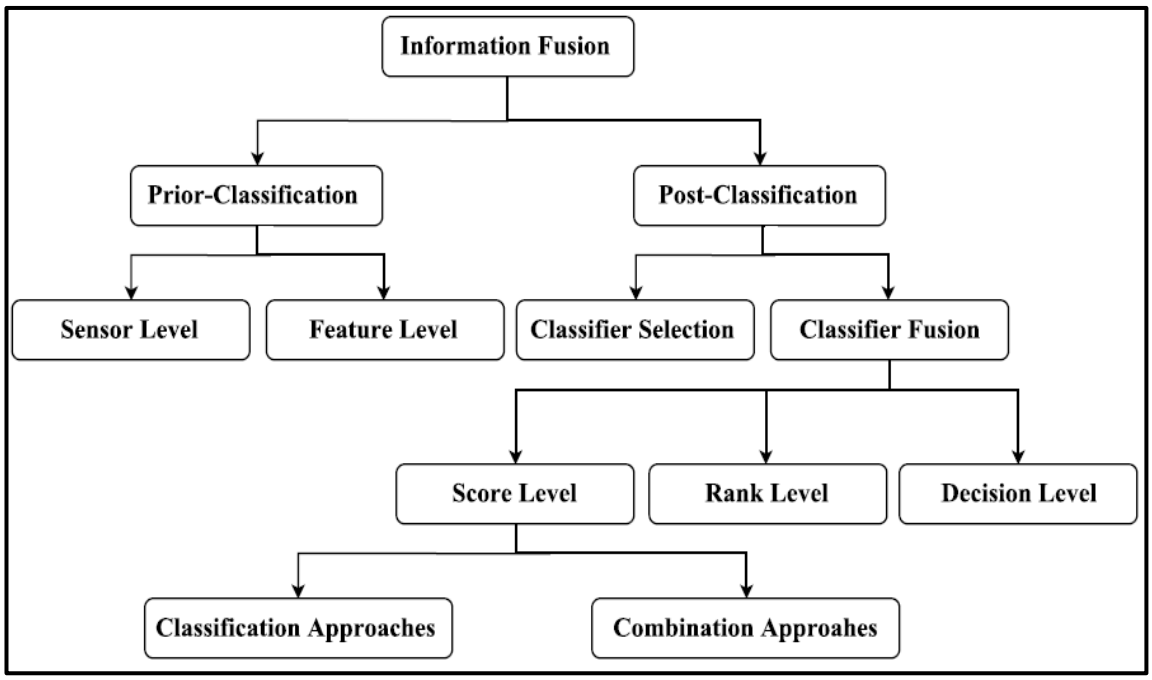


Figure 2.3: The taxonomy of information fusion approaches in multi-biometric systems, adapted from Jain et al.[13].

2.4.1 Prior-to-Classification Fusion

Prior-to-classification fusion means combining the information obtained before applying any matching algorithm at the sensor level or at the feature level. In *sensor-level* fusion, raw data of the same biometric trait acquired from multiple compatible sensors are combined. This type of fusion can perfectly serve multi-sample systems that acquire multiple samples of the same biometric trait. For example, multiple samples of person's fingerprint acquired using the same sensor can be stitched together and create more informative image that reveals more of the essential ridge structure. This process is known as a mosaicing scheme. In 2003, a similar idea was proposed for face images by Liu and Chen [168]. They introduced a novel a face mosaicing method for combining a sequence of face images captured under an orthographic camera model. In this system, each frame was unwrapped onto a certain portion of the surface of a sphere using a spherical projection. Then, a minimization process, using the Levenberg-Marquardt algorithm was applied to optimize the distance between an unwrapped frame and the sphere. Experimental results were observed on a face sequence with 58 frames in total, and a size of (44×44) pixels. Singh et al. [169] presented a multispectral image fusion technique,

based on a 2-granular SVM, for integrating visible and long wave infrared face images. In this technique, multiple SVMs were employed to learn both the local and global features of the multispectral face images at different resolution levels. Then, the 2v-GSVM was used to perform accurate classification, which was subsequently used to dynamically calculate the weight values of visible and infrared face images to generate a new fused face image. The global and local facial features are extracted from the fused image, using 2D-log Gabor transform and LBP approaches, respectively. Finally, the verification decision was made by match score level fusion, using Dezert Smarandache theory. Experiments on Notre Dame and Equinox databases reported verification rates of 95.85% and 94.71, respectively at 0.01% FAR . Raghavendra et al. [170] proposed a new biometric sensor fusion method to fuse the information obtained from face and palmprint images using Particle Swarm Optimisation (PSO). Firstly, the face and palmprint images acquired from different sensors were decomposed using a wavelet transformation. This was followed by applying the PSO method to determine the most informative wavelet sub-bands of the face and palmprint images, to produce a new combined image. Then, feature vectors extracted using Kernel Direct Discriminant Analysis (KDDA) were used as input to a Nearest Neighbour Classifier (NNC) to make the final decision. The performance was evaluated on a virtual multi-biometric database of 250 subjects (collected from the FRGC face database and the polyU palmprint database) with a verification rate of 94.26% at 0.01% FAR. A similar idea of fusing the same adopted biometric traits was also proposed by Kisku et al. [171], where the wavelet transformation was employed to decompose and fuse face and palmprint images for personal verification. Then, the SIFT method was applied to the fused image in the feature extraction stage and verification was performed using a recursive descent tree traversal approach. A verification rate of 98.19% was achieved on a private multi-biometric database consisting of 750 face and palmprint images acquired from 150 subjects.

In the literature, several approaches have been proposed to study fusion at the *feature-level*, in which the feature sets extracted from the same (or different) biometric trait are consolidated into a single feature vector. An example of a feature-level fusion scheme was proposed by Rattani and

Tistarelli [172] for fusing both face and iris traits. In this work, the SIFT method was employed to extract the most discriminative features from each trait, and then the extracted feature vectors were combined to generate a joint SIFT feature vector. The matching score between two feature vectors was computed using the Euclidean distance. Experiments on a chimeric multi-biometric database, built up by fusing an Equinox face database and a CASIA V3.0 iris database, demonstrated the efficiency of the fusion approach compared to any unimodal biometric approach.

GAN et al. [173] proposed a multi-biometric system, based on fusing face and iris traits at the feature-level using 2D-DCT and Kernel Fisher Discriminant Analysis (KFDA). A recognition rate of 100% was achieved on a chimeric database of 40 subjects. Nagar et al. [174] presented a biometric cryptosystem based on a feature-level fusion scheme to secure multiple templates of a person simultaneously. In this work, a fused multi-biometric feature vector of three biometric traits (e.g., fingerprint, iris, and face) was secured by employing two biometric cryptosystems, namely fuzzy vault and fuzzy commitment. The authors tested the trade-off between matching accuracy and security of their system on two multi-biometric databases (one real and one chimeric multimodal database), each containing the three adopted biometric traits. A verification rate of 99% was achieved on the chimeric database, while the verification rate was dramatically decreased on the real database to 68% and 75% using fuzzy vault and fuzzy commitment approaches, respectively. Another key contribution of fusing face and iris features was proposed by Lin et al. [175]. In this system, after integrating the feature set of both face and iris traits, the modified PUM was employed to measure the matching scores between the new fused feature representations. The performance was evaluated on two chimeric databases containing a total of 100 and 50 subjects, and recognition rates of 96.6% and 97.3% were achieved, respectively. Sharma and Kumar [176] studied the fusion of both face and iris traits using four different feature extraction approaches, namely, LBP, Local Gabor XOR Pattern (LGXP), Empirical Mode Decomposition (EMD) and PCA. The Euclidean distance was adopted for making the final decision. Experiments on a chimeric database of 40 subjects

reported EERs of 0.5, 0.5, 0.56, and 0.57 using LBP, LGXP, EMD and PCA, respectively.

2.4.2 Post-to-Classification Fusion

Post-to-classification fusion means combining the information after the results of the classifiers have been obtained, and it can be divided into two categories: classifier selection and classifier fusion. The first category is also known as the *winner-take-all* scheme, in which the final result depends only on the classifier which is most likely to produce the correct decision for the input pattern. Classifier fusion can be broadly divided into three different levels of fusion, based on the information to be combined, namely, *matching score level, rank level* or *decision level*.

Score level fusion refers to the integration of matching scores (e.g., similarity or distance scores) obtained from multiple classifiers applied to the same (or different) set of features to produce the final decision [13]. As shown in Fig. 2.3, depending on how the matching scores is treated by the system, the fusion approaches in this level can be further divided into two categories: combination approaches and classification approaches (See Section 1.6.3). This level of fusion is the most common level found in the multi-biometric literature and many different fusion approaches have been proposed, primarily due to their simplicity, convenience of fusion, and availability of information. In 2005, Chen and Chu [177] developed a multi-biometric system, based on combining face and iris traits at score level fusion. In this system, the face image was converted into a 1D energy profile signal and used as the feature vector for face matcher, while the iris image was divided into five circular bands, and vertical projection was applied to convert these bands into a 1-D feature vector. Neural networks were used to produce the matching scores of both traits, and then an unweighted mean rule was used to form the final decision. The results on two chimeric databases of 40 and 100 subjects reported EERs of 0.00% and 0.01%, respectively. Zhang et al. [178] proposed a multi-biometric system based on face and iris traits acquired from a single Near Infrared (NIR) image. The facial features were extracted using the PCA method, while the Gabor method was employed to encode the iris patterns. Then, the matching

scores of face and iris traits were obtained using the respective feature sets. In this work, different score fusion rules were evaluated in the decision-making process. The best verification rate of 99.75% was achieved on a database of 112 subjects, using a sum rule. Fakhar et al. [179] proposed a novel matching score fusion approach, using fuzzy set theory for a face-iris multi-biometric identification system. They claimed that their approach outperformed a number of fusion approaches, including the sum rule, product rule, max rule, and min rule. A recognition rate of 97.69% at Rank-1 was achieved on a chimeric database of 108 subjects. Another example of a multi-biometric system, based on face and iris traits, was proposed by Al-Hijaili et al. [180] to improve the security level in the hierarchical architecture of Electronic Medical Records (EMR). The matching scores obtained from each trait were firstly normalized and fused using weighted sum rule. Experiments on a chimeric database containing a total of 40 subjects improved the verification rates by 22% and 3% over face and iris unimodal systems, respectively. Kim et al. [181] proposed a multi-biometric system for personal verification based on fusing the matching scores of face and both the face and both irises. In this system, the PCA method was used to extract the facial features, while a 1D Gabor filter was applied on the segmented iris region. Then, Euclidean distance was employed to provide the matching scores of the adopted traits which are then used as an input for the SVM to make the final decision. The performance was evaluated on a real multi-biometric database of 30 subjects, and an EER of 0.131% was achieved. More recently, Mamta and Hanmandlu [182] proposed a new matching score method, named Refined Score (RS). This method was applied on three different biometric traits, including IR face, ear and iris traits under both constrained and unconstrained conditions for the verification task. A verification rate of 100% at FAR of 0.01% was achieved on a database containing a total of 100 subjects. In 2016, Aboshosha et al. [183] used the minutia-based algorithm, 1D Log-Gabor filters, and the LBP method to extract the discriminative features from three biometric traits, including fingerprints, irises and faces. The min-max normalization method was applied to convert the matching scores obtained from the adopted traits into the same numerical range. Finally, the performances of three score fusion methods are evaluated to combine the normalized scores,

such as sum, weighted sum rules, and product. A verification rate of 99.7% was achieved on a chimeric database of 100 subjects.

When the biometric system operates in the identification mode, and the output of each classifier can be viewed as a list of ranks of the enrolled candidates, the fusion can be carried out at the rank level (See Section 1.6.3). The most popular examples of rank level fusion approaches are: Borda count, Highest rank, and Logistic regression approach, as proposed by Ho et al.[24]. Several multi-biometric systems have been used these approaches with a different number of adopted biometric traits to make the final decision. Some examples of such systems can be found in[184][185]. In addition to these approaches, a number of rank level fusion approaches have been proposed in the literature to deduce a consensus rank for each person. For example, a number of modifications to enhance the performances of the Highest rank and Borda count approaches were proposed by Abaza and Ross [186]. In this work, two main issues were addressed, including the presence of weak classifiers and the low quality of input images. For the highest rank approach, a perturbation factor was proposed to handle the existing tie problem. For the Borda count approach, two modifications were introduced: (i) the Nanson function, employed to discard the worst rank(s) before the fusion takes place, and (ii) the individual classifiers were automatically weighted, based on the quality information for the input images. Experiments on the two multi-biometric databases, , WVU dataset and NIST Biometric Scores Set Release-1 (BSSR1) database, reported promising results compared to the traditional approaches. Another example of improving the performance of existing rank fusion approaches was also proposed by Sharma et al. The authors presented two ways of fusing the results obtained from various existing rank fusion approaches: serial combination and parallel combination. The performance was evaluated on three publicly-available databases: NIST BSSR1, Face Recognition Grand Challenge (FRGC V2.0), and LG4000 iris database. Rank-1 identification rates of 83.63%, 92.63%, and 93.4% were achieved using three systems, namely, multi-algorithm, multi-instance and multimodal biometric systems. Monwar and Gavrilova [187] proposed a novel rank fusion method based on a Markov chain approach for a multi-biometric system using three biometric traits (face, ear, and

iris). In this system, the facial and ear features were extracted using the fisherimage technique (a combination of PCA and LDA), while the Hough transform and Hamming distance techniques were employed for iris recognition. A Rank-1 identification rate of 98.29% was achieved using a chimeric database.

In decision level fusion, the result of the multi-biometric system was obtained by consolidating the final outputs of multiple classifiers using different decision fusion approaches, such as “AND” and “OR” rules [188], majority voting [189], weighted majority voting [190], Dempster-Shafer theory, Naive Bayesian decision fusion [191], and behavior knowledge space [192]. In 2007, Lee et al. [193] introduced a novel fusion strategy to combine the outputs decision from multiple traits using a cascading verification scheme. The features from the face, iris and voice were extracted by applying the Eigenface algorithm, multi-scale edge methods (e.g., derivative of Gaussian and Laplacian of Gaussian), and Gaussian Mixture Models (GMMs), respectively. Using the cascading scheme, if the correct match is obtained by any one of the three modules, then the subject's identity was considered to be accepted. In this system, the result of each biometric matcher was a binary output, either match (accepted) or non-match (rejected), while the final match decision was made by fusing these outputs using the “OR” rule. The performance was evaluated on very small private Multi-biometrics database containing a total of 19 subjects (14 males and 5 females).

In 2009, Yu et al. [194] proposed a multi-biometric system that consolidated three different biometric traits, namely, palmprint, fingerprint and finger geometry at the decision fusion level. Firstly, the fingertips and valley points were detected by employing an eight neighbourhood border tracing algorithm. This was followed by extracting the regions of interest for palmprint, fingerprint, and finger contour from a whole hand image. In the feature extraction stage, the LDA approach was applied to extract palmprint and fingerprint features, while the finger contour was used to obtain the finger geometry features. The final decision was formed by fusing the outputs obtained from the adopted traits using three decision fusion rules, including “AND”, “OR”, and majority voting. Experiments on a database of 86 hands image, captured using a digital camera (Canon Power Shot A75) with 10

impressions per hand, demonstrated the efficiency of the majority voting in the verification task, compared to other two decision rules.

A review of other existing multi-biometric systems using different biometric traits, different approaches, and different levels of fusion found in the literature is briefly presented in Table 2.1. Furthermore, some well-known examples of multi-biometric systems can also be found in some relevant literature surveys [3][14][16][23].

2.5 Current Challenges and Knowledge Gaps

This review of the research on unimodal biometric systems, based on the face or iris traits, and their incorporation into multimodal biometric systems, has found a number of algorithms have been proposed and developed in attempts to address various aspects of the aforementioned topics. Nevertheless, there are still some interesting and challenging problems to be addressed. The literature reveals that, although numerous face recognition approaches have been proposed, and great progress has been achieved to improve its performance over previous decades, face recognition is still a challenging open problem for the research community. This is especially when face images are taken in unconstrained conditions, due to the large intra-personal variations, such as changes in facial expression, pose, illumination, aging, and the small inter-personal differences. For instance, Zhu et al. [68] and Li et al. [195] revealed that some of the face recognition approaches, as in [40][100][196][197] that achieve high recognition rates in face images under semi-controlled environmental conditions, are unable to achieve the same recognition rates under fully unconstrained conditions. Furthermore, previous work on face recognition has only focused on extracting the facial features representation using either handcrafted approaches (e.g., SIFT, LBP, etc.), or deep learning approaches (e.g., RBM, DBN, etc.).

However, handcrafted approaches have some limitations that include: (i) they usually have difficulty with the wide variations in face images, (ii) they require domain-knowledge from the researcher to find the best feature set for a given problem, and (iii) sometimes they are extremely difficult and time demanding.

Table 2.1: A literature review of recent multimodal biometric systems. The usage of Real and Chimeric multi-biometric database is denoted by (R) and (C), respectively.

Ref.	Year	Biometric Traits	Level of Fusion	Fusion Approach
[198]	2011	Face and Iris (C)	Feature level fusion	PCA and 2D Gabor are applied to face and iris images, respectively. The normalized features are obtained using z-score method and combined using a series fusion method.
[199]	2012	Fingerprint and Iris (C)	Rank level fusion	Fusion using the Borda count, and Logistic regression approaches.
[200]	2012	Face and Palmprint (C)	Feature level fusion	Integrating the Log Gabor features using weighting scheme of PSO method
[201]	2013	Face and Ear (C)	Feature level fusion	Sparse Representation (SR) Sparse Coding Error Ratio and a novel index method called Sparse Coding Error Ratio (SCER)
[202]	2013	Face and Iris (C)	Feature level fusion	DWT and DCT are used to extract features from the face and iris traits independently and fused using PCA fusion technique.
[203]	2014	Face and Iris (C)	Score level fusion Feature level fusion	Iris-FVF and consists of PCA, LDA, LBP, sub-pattern PCA, and Modular PCA. These features are fused using PSO. LBP is used in Face-FVF. Weighted Sum rule at score level.
[204]	2014	Face and Iris (R) & (C)	Score level fusion	Fusing the output of face and iris matcher using Weighted Sum rule
[205]	2014	Iris and Fingerprints (C)	Score level fusion Decision level fusion	Fusion by Sum rule, Weighted Sum rule, and Fuzzy Logic method.
[79]	2014	Face, Fingerprint, and Iris (C)	Score level fusion	Curvelet, Ridgelet transforms and weighted sum rule.
[206]	2015	Face and Iris (C)	Feature level fusion Score level fusion	The local features are extracted from both traits using LBP Histogram, Modular PCA and sub-pattern PCA, while the global feature using PCA and LDA. Weighted Sum rule at score level.
[207]	2015	Face and Iris (C)	Feature level fusion Score level fusion	PSO and Backtracking Search Algorithm (BSA) are applied to select optimized features and weights for feature level and score level fusion, respectively.
[208]	2016	Face and Iris (C)	Feature level fusion Score level fusion Decision level fusion	Features are extracted using log Gabor and LDA. The BSA algorithm is used to reduce the number of features and select the best weights for feature level and score level fusion, respectively.
[209]	2016	Fingerprint and Iris (C)	Score level fusion	incremental Granular Relevance Vector Machine (iGRVM) classifier, which incorporates incremental and granular learning in RVM
[210]	2016	Face And Palmprint (C)	Feature level fusion	Local features are fused non-stationary feature fusion.
[211]	2017	Face and Iris (C)	Score level fusion	A framework based on bin-based classifier fusion.

On the other hand, deep learning approaches usually discard the local facial features, which are known to be important for face recognition. In addition, applying DNNs directly on the raw data can be quite computationally expensive for large-scale databases, and a very long time is required to obtain the final trained model. Therefore, there are significant potential demands for improving the performance of the face recognition system in various real-world applications. Several studies have pointed out that a better performance can be obtained by combining several feature representations from two or more handcrafted approaches, as in [64][67][68][69][212]. However, in most cases, face databases consisting of a relatively small number of subjects are employed for performance reporting, and very few of these approaches were capable of working under fully unconstrained conditions. Furthermore, very few studies have addressed and discussed the advantages of training deep learning approaches on the top of feature representations produced from handcrafted approaches instead of the raw data, to learn more informative and complementary representations.

To address all the above problems, a new fully automatic and robust face recognition algorithm, based on merging the advantages of two efficient handcrafted approaches, is proposed. Unlike previously published works that extract only the global facial features from the input image, the proposed algorithm has succeeded in extracting the local facial features, along with the face texture roughness, and fluctuations in the image's surface efficiently. In addition, a novel framework, based on merging the advantages of the handcrafted feature descriptors with the DNNs, is proposed to address the face recognition problem in unconstrained conditions. The performance of the proposed face recognition approaches is observed in both face identification and verification.

In the literature, although great progress has been achieved on iris recognition since the 1990s, some drawbacks and limitations have been observed, which indicate that identifying a person's identity using their iris trait is still an interesting area of research with many challenging and unsolved problems. Firstly, there is the lack of a real-time, accurate and fully automated iris localization system to detect the iris boundaries. The literature shows that

iris localization algorithms based on thresholding, geometric, or fuzzy logic approaches perform well when the iris images are taken under ideal imaging conditions. However, the accuracy of these approaches may decline significantly when non-ideal iris images are taken in unconstrained conditions [129]. As a result, this can significantly degrade the performance of iris recognition due to the incorrect iris localization or the poor quality of iris texture. Secondly, previous work has been limited to applying handcrafted approaches to encode the iris patterns. Despite several machine learning approaches having been employed in iris recognition systems, they mainly depend on shallow learning models. This observation is also applicable to multi-biometric systems. One of the main issues of the traditional approaches is that the input image is required to undergo several different image pre-processing stages. Furthermore, shallow learning models (e.g., an MLP) have difficulties in handling deformations of the input image, such as translation, scaling and rotation [213]. Finally, a large number of free parameters need to be tuned in order to achieve satisfactory results while avoiding the overfitting problem [214]. To overcome these drawbacks the use of deep learning approaches is proposed. Deep Learning can be viewed as an advanced subfield of machine learning techniques that depend on learning high-level representations and abstractions using a structure composed of multiple non-linear transformations. To date, as revealed in the recent survey [167], the use of deep learning approaches has not been fully explored and investigated for iris recognition. Unlike the conventional iris recognition pipeline, which consists of two separated stages (e.g., feature extraction and matching stage), one of the main contributions in this thesis is to develop an iris recognition system that can automatically learn non-linear transformation functions between the input images and their class label identities, simultaneously. Learning such general and high-level feature representations ensures that the same efficiency can be obtained on different databases and also it can be easily adapted to other new applications.

Much work on the potential of improving the accuracy of unimodal systems by using multi-biometric systems has been carried out [16], yet there are still some interesting and relevant problems that need to be addressed. Firstly, the

paucity of available multimodal databases has led to most of the previous studies basing their evaluations on “chimeric” multimodal databases obtained by arbitrarily combining two or more biometric traits from different unimodal databases. The justification for these chimeric databases is mainly based on the assumption of independence between different biometric traits. However, the study in [215] demonstrated that all the biometric measures from a single person are by necessity correlated, and hence, using a chimeric database may not reflect the real world performance of the proposed approaches. Secondly, previous work has been limited to apply either handcrafted approaches (e.g. LBP) or machine learning approaches, based on shallow learning models (e.g., MLP) to encode the biometric data. However, the main issue of the former is that the input image needs to undergo several pre-processing stages, while the latter has difficulties in handling transformations of the input image, such as translations, scaling and rotation [216]. Thirdly, most of the previous work has been shown to improve the accuracy of the biometric system in the verification task, rather than the identification task, which is more difficult than the former. Finally, there is a lack of flexibility; for instance, if one of the biometric traits is unavailable or missed, then either the whole system breaks down or the accuracy rate decreases. Thus, creating a multimodal biometric system for identifying a person's identity, with a high accuracy and acceptable system complexity, is still an issue in real-world applications.

2.6 Chapter Summary

In this chapter, a brief review of previous research related to this PhD thesis is presented. Through this review on unimodal and multimodal systems, based on face and iris traits, a number of drawbacks and limitations have been clearly highlighted. One of the main observations is the potential for improving the accuracy of face recognition, using a combination of two or more handcrafted approaches, as well as the possibility of learning additional and complementary information for face recognition by applying deep learning approaches on the top of engineered features, which will be the main focus of Chapter 3.

Other observations that will be the main focus of chapters 3 and 4, include: (i) the need for a fast and accurate iris localization algorithm to efficiently detect the iris region without losing important information; (ii) the fact that deep learning approaches have not been fully investigated in the biometrics area except for face recognition. Research on iris recognition and multi-biometric systems has shown very few papers can be found in the literature that discuss the possibility of using deep learning approaches to learn more compact and discriminative features, so the work addresses the problem of missing biometric traits by proposing an efficient and real-time multi-biometric system with a high degree of flexibility, and evaluating its performance on real multi-biometric databases using different fusion approaches.

Chapter 3

Face Recognition System

3.1 Introduction

This chapter describes the face recognition approaches proposed in this PhD thesis. Face recognition is one of the most popular biometric systems that have received a significant attention in the research community due to its distinctive accuracy and low cost. Face recognition has appeared to offer a number of characteristics over other biometric systems. Face recognition systems have the added advantage that the user can be identified without knowing he is being monitored. In addition, the biometric systems that depend on using the same sensor device to acquire the biometric trait from multiple users (e.g. Fingerprint recognition) can cause some health risks by transferring germs and/or some infectious diseases from one user to the other [217]. Moreover, there has been a growing interest in highly secured and well-designed face recognition systems in the last few years, due to their potentially wide applications in many sensitive places, such as controlling access to physical, as well as virtual places in both commercial and military associations, including check points, ATM cash dispensers, e-learning, information security, intelligent surveillance, and other daily human applications [218]. In spite of the significant improvement in the performance of face recognition over previous decades, it still a challenging task for the research community, especially when face images are taken in unconstrained conditions, due to the large intra-personal variations, such as changes in facial expression, pose, illumination, aging, the small inter- and intra-personal differences, and occlusions from wearing glasses and hats.

Generally, a number of approaches have been proposed and refined to overcome all these drawbacks and problems, but very few of them are capable of working under fully unconstrained conditions. As described in (Chapter 2, Sect.1.2), the most recent face recognition systems are mainly dependent on feature representations obtained using either local handcrafted-descriptors, such as LBP, or use a deep learning approach, such as DBN. However, the

former usually suffer from the wide variations in face images, while the latter usually discard the local facial features, which are proven to be important for face recognition. In this chapter, a novel framework based on merging the advantages of the local handcrafted feature descriptors with the DBN is proposed to address the face recognition problem in unconstrained conditions. Firstly, a novel multimodal local feature extraction approach, based on merging the advantages of the Curvelet transform with Fractal dimension is proposed and termed the *Curvelet-Fractal* approach. The main motivation of this approach is that the Curvelet transform, a new anisotropic and multidirectional transform, can efficiently represent the main structure of the face (e.g. edges and curves), while the Fractal dimension is one of the most powerful texture descriptors for face images. Secondly, a novel framework is proposed, termed the Multimodal Deep Face Recognition (MDFR) framework, to add feature representations, by training a DBN on top of the local feature representations instead of the pixel intensity representations. It is demonstrated that representations acquired by the proposed MDFR framework are complementary to those acquired by the *Curvelet-Fractal* approach. To the authors' best knowledge, very few publications can be found in the literature that discuss the potential of applying the DBN on top of pre-processed image feature representations. Huang et al. [97] have demonstrated that applying the Convolutional DBN on top of the output of LBP can increase the accuracy rate of the final system. Li et al. [98] have also reached the same conclusion by applying the DBN on top of Centre-Symmetric Local Binary Pattern (CS-LBP). However, the work in [97] was applied only to the face verification task, while the work in [98] was evaluated on a very small face dataset, where the face images were taken in controlled environments. Finally, the performance of the proposed approaches has been evaluated by conducting a number of extensive experiments on four large-scale face datasets: the SDUMLA-HMT, FERET, CAS-PEAL-R1 and LFW datasets. The results obtained from the proposed approaches outperform other state-of-the-art of approaches (e.g. LBP, DBN, WPCA and etc.) by achieving new state-of-the-art results on all the employed datasets.

This chapter focuses mainly on two different problems in the face recognition system: face identification and face verification. In this chapter, the term face recognition will be used in the general case to refer to these two problems. The remainder of the chapter is organized as follows: Section 3.2 is devoted to providing an overview of the proposed handcrafted-descriptors and deep learning approaches. Section 3.3 shows the implementation details of the proposed face recognition approaches. The experimental results are presented in Section 3.4. Finally, the summary and conclusions of this chapter are stated in the last section.

3.2 Methodology Overview

In this section, a brief description of the proposed face recognition approaches is presented, including the Curvelet transform and Fractal dimension method used in the proposed multimodal local feature extraction approach. In addition, the proposed deep learning approaches include the DBN and its building block the Restricted Boltzmann Machine (RBM) as well. The primary goal here is to review and recognize their strengths and shortcomings to empower the proposal of a novel face recognition framework that consolidates the strengths of these approaches.

3.2.1 Curvelet Transform

In recent years, many multi-resolution approaches have been proposed for facial feature extraction at different scales, aiming to improve face recognition performance. The Wavelet transform is one of the most popular multi-resolution feature extraction methods due to its ability to provide significant features in both space and transform domains. However, according to many studies in the human visual system and image analysis, the Wavelet transform is not ideal for the facial feature extraction approach [77]. A feature extraction approach cannot be optimal without satisfying conditions relating to the following: multi-resolution, localization, critical sampling, directionality and anisotropy [219]. It is believed that the Wavelet transform cannot fulfill the last two conditions, due to limitations of its basis functions in specifying the direction and the isotropic scale [220]. These restrictions lead to a weak representation of the edges and curves which are considered to be the most important facial features. Thus, a

novel transform was developed by Candes and Donoho in 1999, known as the Curvelet transform [221]. Their motivation was to overcome the drawbacks and limitations of widely used multi-resolution approaches such as the Wavelet and Ridgelet transforms. All the above five conditions can be fulfilled using the Curvelet transform. Fig.3.1 shows the edge representation capability of the Curvelet transform and Wavelet transform, where fewer Curvelet's coefficients are required to efficiently represent the edge compared to the number of required Wavelet's coefficients. Moreover, a smoother edge is produced using the Curvelet transform compared to use of the Wavelet transform.

The multi-scale transform principle is a property common to Curvelet and Wavelet transforms, where each has multiple frames indexed by location and scale parameters. However, the Curvelet transform, unlike the Wavelet transform, has a very high degree of directional flexibility. In addition, the Wavelet transform depends on the isotropic scaling principle, where the width and the length of the frame are equal, which is quite different from the anisotropic scaling principle of the Curvelet transform, where the frame size is subject to the (width ≈ length^2) law [220]. In general, two generations of the Curvelet transform are available. The first generation Discrete Curvelet Transform (DCTG1) is based on Wavelet sub-bands and the Ridgelet transform. DCTG1 is a sequence of four steps: sub-band decomposition, smooth partitioning, renormalization and Ridgelet analysis. More details on this transform, and its implementation can be found in [222]. In the first generation, the frame size obeys a parabolic or anisotropic scaling law (width ≈ length^2) due to applying different levels of the multiscale Ridgelet transform. As a result, the increased degree of information redundancy in the feature space and the lengthy execution time of this generation, make it unsuitable for facial feature extraction, especially with a large dataset. To overcome the limitations of DCTG1, a second generation Curvelet Transform (DCTG2) was introduced by Candes and Donoho in 2006. DCTG2 is less redundant, simpler and faster in execution compared to DCTG1. In DCTG2, two implementations of fast discrete Curvelet transform are available. The first depends on an Unequally-Spaced Fast Fourier Transforms (USFFT) whilst the second depends on a wrapping procedure of specially chosen Fourier samples [221].

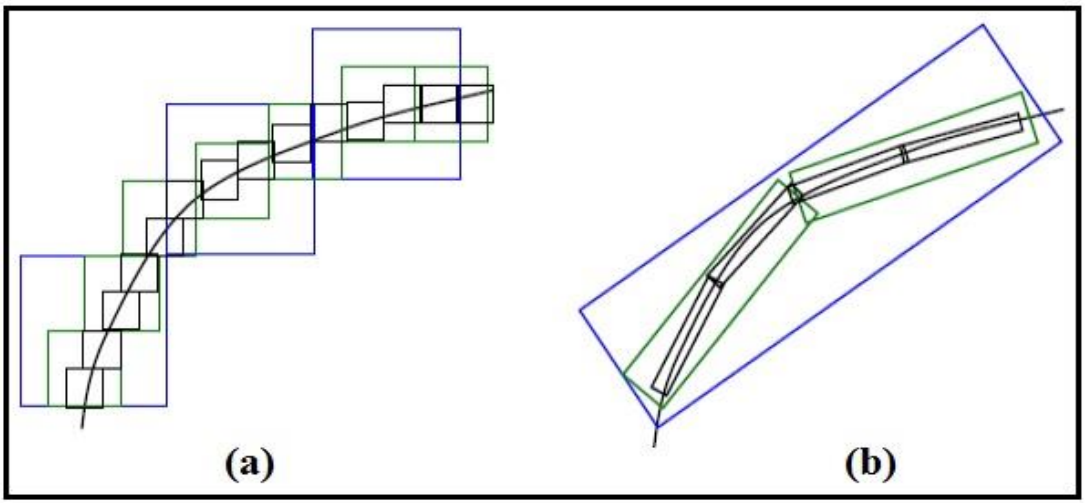


Figure 3.1: Edge representation: (a) Wavelet transform and (b) Curvelet transform [73].

The Curvelet transform has been successfully applied to solve many problems in the image processing area, such as texture classification [223], preserving edges and image enhancement [224], image compression [225], image fusion [226], and image de-noising [227]. Some work has been done to explore the potential of the Curvelet transform to help solve pattern recognition problems, for example by Lee and et al. [228], T. Mandal and et al [73] and Xie [229]. These showed that the Curvelet transform can serve as a good feature extraction method for pattern recognition problems like fingerprint and face recognition, due to its ability to represent crucial edges and curve features more efficiently than other transformation methods. However, the Curvelet transform cannot overcome the effects of large changes in illumination conditions, shadows, multiple views of face images and occlusions from wearing glasses or hats. As a result, the Curvelet transform is not able to describe the face texture roughness and fluctuations in the surface efficiently, which will have a significant effect on the recognition rate. All these factors together were behind the adoption here of the Fractal dimension to provide a better description of the face texture under unconstrained environmental conditions.

3.2.2 Fractal Dimension

The term Fractal dimension was first introduced by the mathematician Benoit Mandelbrot as a geometrical quantity to describe the complexity of objects that show self-similarity at different scales [230]. The Fractal dimension

has some important properties, such as a self-similarity, which means that an object has a similar representation to the original under different magnifications. This property can be used in reflecting the roughness and fluctuation of image's surface, where increasing the scale of magnification provides more and more details of the imaged surface. In addition, the non-integer value of the Fractal dimension gives a quantitative measure of objects that have complex geometry, and cannot be well described by an integral dimension (such as the length of a coastline) [231][232]. Many methods have been proposed to calculate Fractal dimension, such as Box Counting (BC), Differential Box Counting (DBC) and Fractional Brownian Motion (FBM), and other methods can be found here [230]. The Fractal dimension has been widely applied in many areas of image processing and computer vision, due to its simplicity and robustness in reflecting the roughness and fluctuations of the imaged surface. Hsu [233] proposed an efficient texture segmentation algorithm based on the capability of the Fractal dimension in describing the surface texture. In the medical field, Al-Kadi [234] presented an efficient algorithm for classification of histological brain tumours by using Fractal analysis combined with the Wavelet transform. Zhu et al. [235] developed a fast and accurate face detection algorithm based on the inherent advantages of Fractal dimension in reflecting the roughness and texture information of the face region to separate it from the non-face regions (e.g. background). However, not much work has been done to explore and address the potential of using the Fractal dimension to resolve pattern recognition problems. Lin and et al [236] proposed an algorithm for human eye detection by exploiting the Fractal dimension as an efficient approach for representing the texture of facial features. Farhan and et al [237] developed a personal identification system based on fingerprint images using the Fractal dimension as a feature extraction method. Therefore, it appears that the texture of the facial image can be efficiently described by using the Fractal dimension. However, Fractal estimation methods are very time consuming, and cannot meet real-time requirements. To address all the limitations and drawbacks in (Section 3.2.1and 3.2.2), a novel face recognition algorithm, based on merging the advantages of a multidirectional and anisotropy transform, specifically the Curvelet transform, with Fractal dimension, is proposed.

3.2.3 Restricted Boltzmann Machine

An RBM is an energy-based bipartite graphical model composed of two fully-connected layers via symmetric undirected edges, but there are no connections between units of the same layer. The first layer consists of m visible units $v = (v_1, v_2,..., v_m)$ that represent observed data, while the second layer consists of n hidden units $h = (h_1, h_2,..., h_n)$ that can be viewed as non-linear feature detectors to capture higher-order correlations in the observed data. In addition, $W = \{w_{11}, w_{12},...,w_{1n},...,w_{mn}\}$ is the connecting weights matrix between the visible and hidden units. A typical RBM structure is shown in Fig. 3.2(a). The standard RBM was designed to be used only with binary stochastic visible units, and is termed Bernoulli RBM (BRBM). However, using binary units is not suitable for real-valued data (e.g. pixel intensities values in images). Therefore, a new model has been developed called the Gaussian RBM (GRBM) to address this limitation of the standard RBM [238]. The energy function of the GRBM is defined as follows:

$$E(v,h) = -\sum_{i=1}^{m}\sum_{j=1}^{n} w_{i,j} h_j \frac{v_i}{\sigma_i} - \sum_{i=1}^{m} \frac{(v_i - b_i)^2}{2\sigma_i^2} - \sum_{j=1}^{n} c_i h_j \qquad (3.1)$$

Here, σ_i is the standard deviation of the Gaussian noise for the visible unit v_i, w_{ij} represents the weights for the visible unit v_i and the hidden unit h_j, and b_i and c_j are biases for the visible and hidden units, respectively. The conditional probabilities for the visible units, given hidden units, and vice versa for the hidden units are defined as follows:

$$p(v_i = 1|h) = N\left(v|b_i + \sum_j w_{i,j} h_j, \sigma_i^2\right) \qquad (3.2)$$

$$p(h_j = 1|v) = f\left(c_j + \sum_i w_{i,j} \frac{v_i}{\sigma_i^2}\right) \qquad (3.3)$$

Here, $N(\cdot|\ \mu,\sigma^2)$ refers to the Gaussian probability density function with mean μ and standard deviation σ. $f(x)$ is a sigmoid function. During the training process, the log-likelihood of the training data is maximized using stochastic

gradient descent, and the update rules for the parameters are defined as follows:

$$\Delta \boldsymbol{w}_{i,j} = \boldsymbol{\epsilon}\left(\langle \frac{\mathbf{1}}{\boldsymbol{\sigma}_i^2} \boldsymbol{v}_i \boldsymbol{h}_j \rangle_{data} - \langle \frac{\mathbf{1}}{\boldsymbol{\sigma}_i^2} \boldsymbol{v}_i \boldsymbol{h}_j \rangle_{model}\right) \quad (\mathbf{3.4})$$

$$\Delta \boldsymbol{b}_i = \boldsymbol{\epsilon}\left(\langle \frac{\mathbf{1}}{\boldsymbol{\sigma}_i^2} \boldsymbol{v}_i \rangle_{data} - \langle \frac{\mathbf{1}}{\boldsymbol{\sigma}_i^2} \boldsymbol{v}_i \rangle_{model}\right) \quad (\mathbf{3.5})$$

$$\Delta \boldsymbol{c}_i = \boldsymbol{\epsilon}\left(\langle \boldsymbol{h}_j \rangle_{data} - \langle \boldsymbol{h}_j \rangle_{model}\right) \quad (\mathbf{3.6})$$

Here, $\boldsymbol{\epsilon}$ is the learning rate and $\langle . \rangle_{data}$ and $\langle . \rangle_{model}$ represent the expectations under the distribution specified by the input data (*Positive phase*) and the internal representations of the RBM model (*Negative phase*), respectively. Finally, $\boldsymbol{b}_i$ and $\boldsymbol{c}_i$ are biases terms for visible and hidden units, respectively. As reported in the literature, it is intractable to compute the $\langle \boldsymbol{v}_i \boldsymbol{h}_j \rangle_{model}$. Therefore, the Contrastive Divergence (CD) algorithm [239] has become the standard learning method to update the RBM parameters by sampling ***k*** steps from the RBM distribution to approximate the second term in Eq. (3.4). One step of a CD algorithm for a single-sample can be implemented as follows:

1. Visible units ($\boldsymbol{v}_i$) are initialized using training data, and the probabilities of hidden units are computed with Eq. (3.2). Then a hidden activation vector ($\boldsymbol{h}_j$) is sampled from this probability distribution.
2. Compute the outer product of ($\boldsymbol{v}_i$) and ($\boldsymbol{h}_j$), which refers to the positive phase.
3. Sample a reconstruction ($\boldsymbol{v}_i'$) of the visible units from ($\boldsymbol{h}_j$) with Eq. (3.3), and then from ($\boldsymbol{v}_i'$) resample the hidden units activations ($\boldsymbol{h}_i'$). (One Gibbs sampling step).
4. Compute the outer product of ($\boldsymbol{v}_i'$)and ($\boldsymbol{h}_i'$), which refers to the negative phase.
5. Update weights matrix and biases with Eq. (3.4) - Eq. (3.6).

The computation steps of the CD-1 algorithm are graphically shown in Fig. 3.3. In the CD learning algorithm, ***k*** is usually set to **1** for many applications. More details on the GRBM model can be found in [238]. Typically, RBMs can be

used in two different ways: either as generative models or as discriminative models, as shown in Fig.3.2 (a-b). The generative models use a layer of hidden units to model a distribution over the visible units, as described above. Such models are usually trained in an unsupervised way and used as feature extractors to model only the inputs for another learning algorithm. On the other hand, discriminative models can also model the joint distribution of the input data and associated target classes. The discriminative models aim to train a joint density model using an RBM that has two layers of visible units. One represents the input data and the second is the *Softmax* label layer that represents the target classes. A discriminative RBM with ***n*** hidden units is a parametric model of the joint distribution between a layer of hidden units (referred to as features) $\boldsymbol{h} = (\boldsymbol{h_1}, \boldsymbol{h_2}, \ldots, \boldsymbol{h_n})$ and visible units made of the input data $\boldsymbol{v} = (\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_m})$ and targets $\mathbf{y} \in \{1,2,\ldots C\}$, that can be defined as follows:

$$\boldsymbol{p}(\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{h}) \propto \boldsymbol{e}^{-\boldsymbol{E}(\boldsymbol{y},\boldsymbol{v},\boldsymbol{h})} \tag{3.7}$$

where,

$$\boldsymbol{E}(\boldsymbol{y}, \boldsymbol{v}, \boldsymbol{h}) = -\boldsymbol{h}^{\boldsymbol{T}}\boldsymbol{W}\boldsymbol{v} - \boldsymbol{b}^{\boldsymbol{T}}\boldsymbol{v} - \boldsymbol{c}^{\boldsymbol{T}}\boldsymbol{h} - \boldsymbol{d}^{\boldsymbol{T}}\vec{\boldsymbol{y}} - \boldsymbol{h}^{\boldsymbol{T}}\boldsymbol{U}\vec{\boldsymbol{y}} \tag{3.8}$$

Here, $(\boldsymbol{W}, \boldsymbol{b}, \boldsymbol{c}, \boldsymbol{d}, \boldsymbol{U})$ refer to the model parameters and $\vec{\boldsymbol{y}} = (\mathbf{1}_{\boldsymbol{y}=\boldsymbol{i}})_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{C}}$ to the ***C*** classes. More details about the discriminative RBM model can be found in [240].

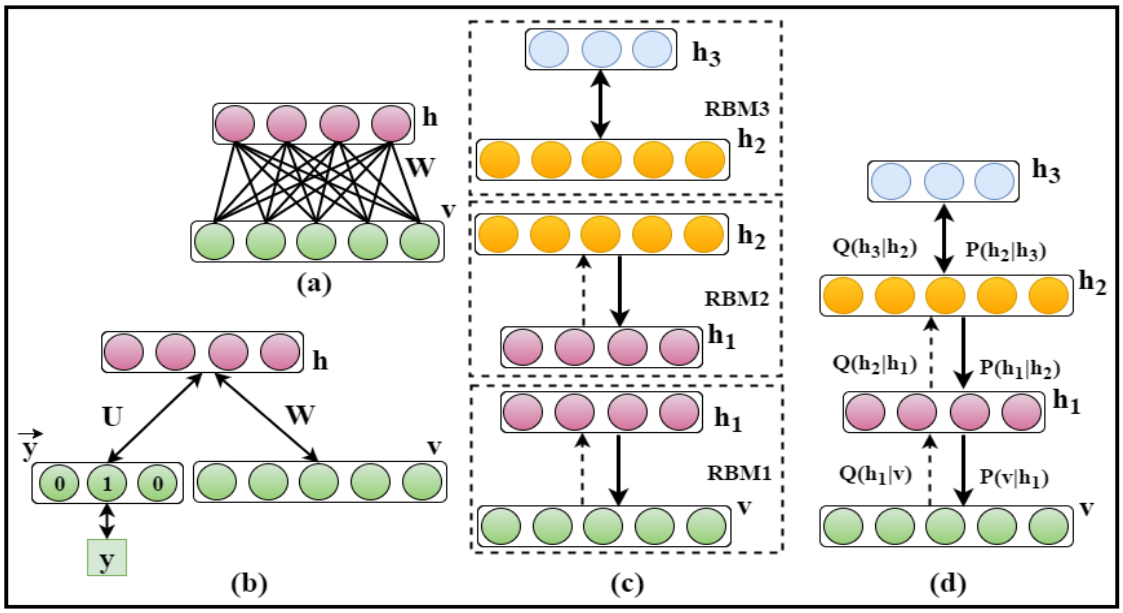


Figure 3.2: (a) A typical RBM structure, (b) A discriminate RBM modelling the joint distribution of input variables ***v*** and target classes y (represented as a one-hot vector by $\vec{y}$), (c) A greedy layer-wised training algorithm for the DBN composed of three stacked RBMs, and (d) Three layers of the DBN as a generative model, where the top-down generative path is represented by the **P** distributions (Solid arcs), and bottom-up inference and the training path are represented by the **Q** distributions (Dashed arcs).

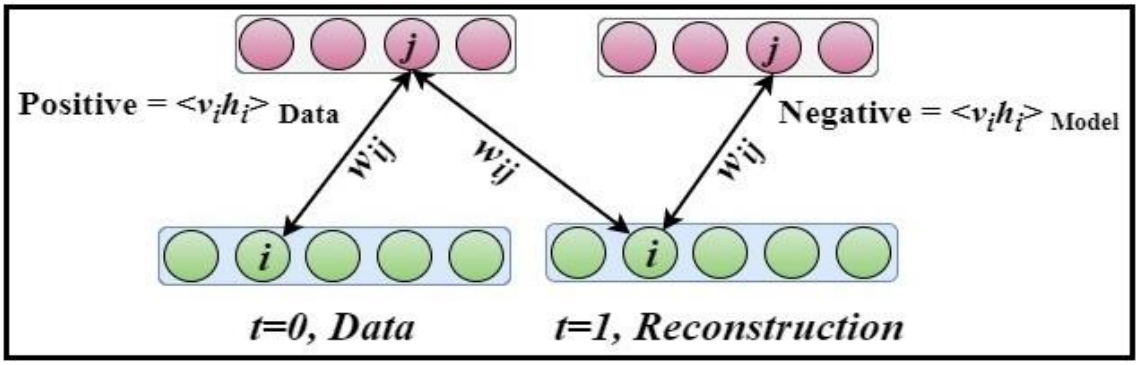


Figure 3.3: A graphical illustration of a single step of the CD algorithm.

3.2.4 Deep Belief Networks

In the last decade, DNNs have attracted much attention from research teams in the field of machine learning. Typically, DNNs comprise a set of feature detectors arranged in layers where more complex features are extracted as one moves from the lower layers towards the higher layers. In 2006, a new DNN was introduced, called the Deep Belief Network (DBNs), by Hinton et

al.[241]. DBN is a generative probabilistic model that differs from conventional discriminative neural networks. DBNs are composed of one visible layer (observed data), and many hidden layers that have the ability learn the statistical relationships between the units in the previous layer. They model the joint probability distribution over the input data (observations) and labels, which facilitates the estimation of both *P(Observations |Labels)* and *P(Labels |Observations)*, while conventional neural networks are limited only to the latter [242]. DBN has been proposed to address issues encountered when applying the Back-propagation algorithm to very deep neural networks, including:

1. The constraint of having a labelled dataset in the training phase.
2. The long time required to converge (slow learning process).
3. The increased number of free parameters that get trapped in poor local optima.

As depicted in Fig. 3.2 (c), a DBN can be viewed as a composition of bipartite undirected graphical models each of which is an RBM. Therefore, DBNs can be efficiently trained using an unsupervised greedy layer-wised algorithm, in which the stacked RBMs are trained one at a time, in a bottom to top manner. For instance, consider training a DBN composed of three hidden layers, as shown in Fig. 3.2 (c). According to the greedy layer-wised training algorithm proposed by Hinton et al. [241], the first RBM is trained using the CD algorithm to learn a layer ($\boldsymbol{h_1}$) of feature representations from the visible units, as described in (Section 3.2.3). Then, the hidden layer units ($\boldsymbol{h_1}$), of the first RBM, are used as visible units to train the second RBM. The whole DBN is trained when the learning of the final hidden layer is completed. A DBN with $\boldsymbol{l}$ layers can model the joint distribution between the observed data vector $\boldsymbol{v}$ and $\boldsymbol{l}$ hidden layers $\boldsymbol{h_k}$ as follows:

$$\boldsymbol{P}\left(\boldsymbol{v}, \boldsymbol{h}^{1}, \dots, \boldsymbol{h}^{l}\right) = \left(\prod_{k=0}^{l-2} \boldsymbol{P}\left(\boldsymbol{h}^{k} \middle| \boldsymbol{h}^{k+1}\right)\right) \boldsymbol{P}(\boldsymbol{h}^{l-1}, \boldsymbol{h}^{l}) \qquad (3.9)$$

Here, $\boldsymbol{v} = \boldsymbol{h}^{0}$, $\boldsymbol{P}\left(\boldsymbol{h}^{k} \middle| \boldsymbol{h}^{k+1}\right)$ is the conditional distribution for the visible units given hidden units of the RBM associated with level $\boldsymbol{k}$ of the DBN, and $\boldsymbol{P}(\boldsymbol{h}^{l-1}, \boldsymbol{h}^{l})$ is the visible-hidden joint distribution in the top-level RBM. An example of a three layers DBN as a generative model is shown in Fig.3.2(d),

where the symbol $\boldsymbol{Q}$ is introduced for exact or approximate posteriors of that model which are used for bottom-up inference. During the bottom-up inference, the $\boldsymbol{Q}$ posteriors are all approximate except for the top level $\boldsymbol{P}(\boldsymbol{h}^{\boldsymbol{l}}|\boldsymbol{h}^{\boldsymbol{l-1}})$, which is formed as an RBM and then the exact inference is possible.

The DBN is one of the most popular unsupervised deep learning methods, which has been successfully applied to learn a hierarchical representations from unlabelled data in a wide range of fields, including face recognition [243], speech recognition [244], audio classification [245], and natural language understanding [246]. However, a key limitation of the DBN when the pixel intensity values are assigned directly to the visible units is that the feature representations of the DBN are sensitive to the local translations of the input image. This can lead to disregarding local features of the input image known to be important for face recognition. Furthermore, scaling the DBN to work with realistic-sized images (e.g., 128×128) is computationally expensive and impractical. To improve the generalization ability and reduce the computational complexity of the DBN, a novel framework, based on merging the advantages of the local handcrafted feature descriptors with the DBN is proposed to address the face recognition problem in unconstrained conditions. It is proposed that applying the DBN on top of pre-processed image feature representations, instead of the pixel intensity representations (raw data), as a way of guiding the learning process, can greatly improve the ability of the DBN to learn more discriminating features with less training time required to obtain the final trained model.

3.3 Multimodal Deep Face Recognition Framework

As depicted in Fig.3.4, a novel face recognition framework, named the Multimodal Deep Face Recognition (MDFR) framework, is proposed to learn high-level facial feature representations by training a DBN on top of a local *Curvelet-Fractal* representation instead of the pixel intensity representation. First, the main stages of the proposed *Curvelet-Fractal* approach are described in detail. This is followed by describing how to learn additional and complementary representations by applying a DBN on top of existing local representations.

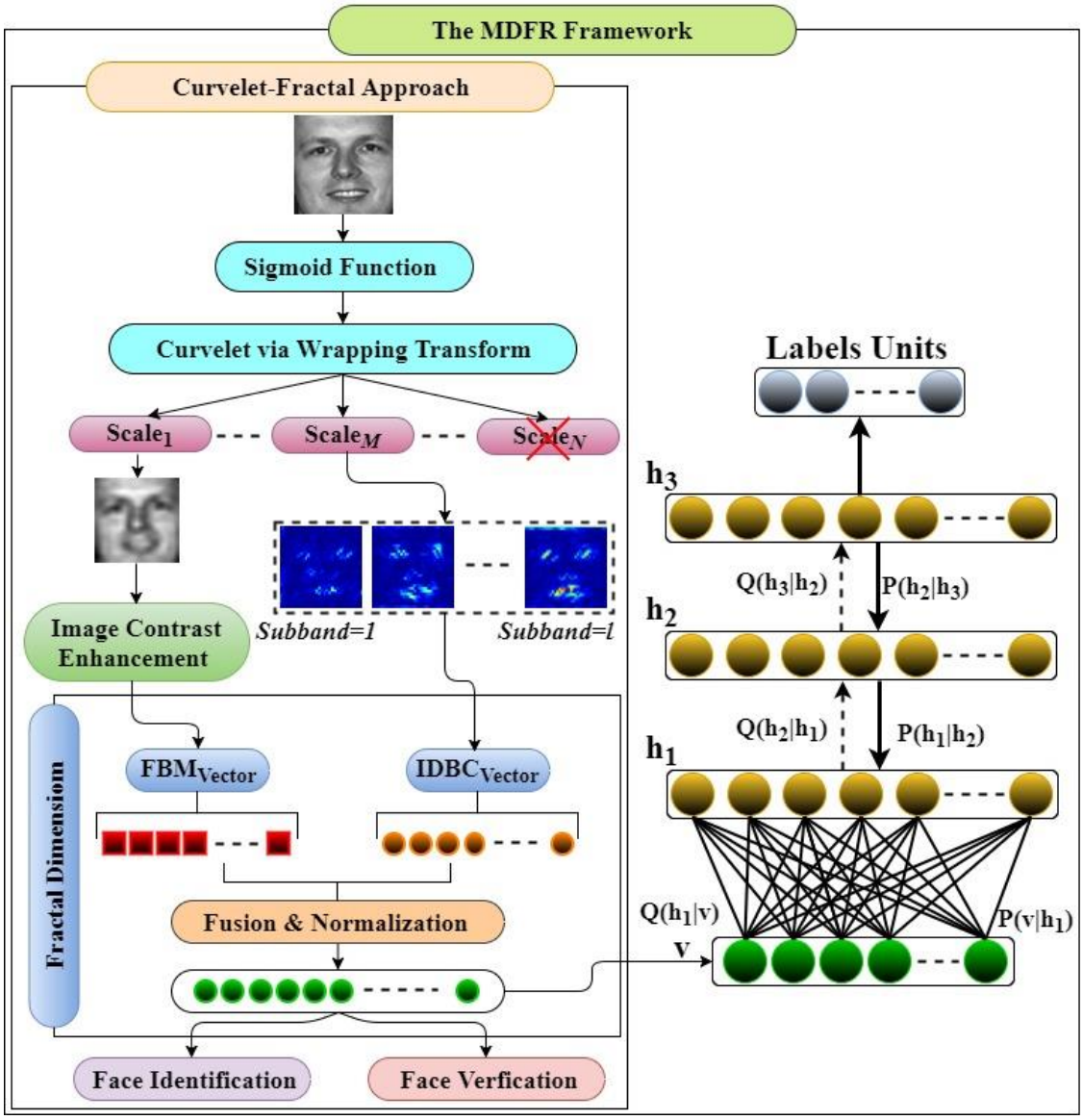


Figure 3.4: Illustration of the proposed *Curvelet-Fractal* approach with the MDFR framework.

3.3.1 The Proposed Curvelet-Fractal Approach

The proposed face recognition algorithm starts by detecting the face region using a Viola-Jones face detector [247]. Detecting the face region in a complex background is not one of the contributions in this PhD thesis. Then, after rescaling the input image $\boldsymbol{f(x,y)}$ to the range of [**0, 1**], a simple pre-processing algorithm using a sigmoid function is applied as follows:

$$\boldsymbol{g(x,y)} = \frac{\mathbf{1}}{\mathbf{1} + \boldsymbol{e}^{\left(c*\left(Th - f(x,y)\right)\right)}} \qquad (\mathbf{3.10})$$

Here, $\boldsymbol{g(x,y)}$ is the enhanced image. In this work, the contrast factor (***c***), and the Threshold value (***Th***), are empirically set to be **5** and **0.3**, respectively. The advantage of the sigmoid function is to reduce the effect of illumination changes by expanding and compressing the range of values of the dark and bright pixels in the face image, respectively. In other words, sigmoid function compresses the dynamic range of the light intensity levels and spreads the pixel values more uniformly. This operation has increased the average recognition rate by 6%. After that, the proposed *Curvelet-Fractal* approach is applied to the enhanced face image. As indicated above, the Fractal dimension has many important properties, such as its ability to reflect the roughness and fluctuations of a face image's surface, and to represent the facial features under different environmental conditions (e.g., illumination changes). However, the Fractal estimation methods can be very time consuming, and the high dimensionality of the face image makes it less suited to meet the real-time requirements. Therefore, the Fractal dimension approach is applied to the Curvelet's output to produce an illumination-insensitive representation of the face image that can meet the real-time system's demands. Hence, the Curvelet transform is used here as a powerful technique for edge and curve representation and dimensionality reduction of the face image, to increase the speed of Fractal dimension estimation. In this work, two different methods to estimate the Fractal dimension are proposed based on the FBM and IDBC methods. The FBM method is used to process only the approximation coefficients (Coarse band) of the Curvelet transform, while the IDBC method is used to process the new added Curvelet sub-bands, and handle their high dimensionality. Then, the output of the FBM and IDBC are combined to build an elementary feature vector of the input image. After the Fractal dimension feature vector $\boldsymbol{FD_{Vector}}$ is obtained, a simple normalization procedure is applied to scale the obtained features to the common range (**0, 1**), as follow:

$$\widetilde{\boldsymbol{FD}}_{\boldsymbol{Vector}} = \frac{\boldsymbol{FD_{Vector}} - \boldsymbol{min}(\boldsymbol{FD_{Vector}})}{\boldsymbol{max}(\boldsymbol{FD_{Vector}}) - \boldsymbol{min}(\boldsymbol{FD_{Vector}})} \qquad (\mathbf{3.11})$$

The main advantage of this scaling is to avoid features with greater numeric ranges dominating those with smaller numeric ranges, which can decrease the recognition accuracy. This procedure has increased the average

recognition rate by 5%. Finally, the Quadratic Discriminant Classifier (QDC) and Correlation Coefficients (CC) Classifiers are used in the recognition tasks. The main steps of the proposed *Curvelet-Fractal* approach for an input face image can be summarized as follows:

1. The sigmoid function is applied to enhance the face image illumination.
2. The Curvelet transform is applied to the image from **1**, so the input image is decomposed into **4** scales and **8** orientations. In this work, the Curvelet sub-bands are divided into three sets, as explained in (Section 3.3.1.1).
3. The FBM method is applied to a contrast enhanced version of the coarse band produced in **2,** and the result is then reshaped into a row feature vector $\boldsymbol{FBM_{Vector}}$, as explained in (Section 3.3.1.2).
4. The IDBC method is applied to the middle-frequency bands produced in **2**, and a row feature vector $\boldsymbol{IDBC_{Vector}}$ is constructed, as explained in (Section 3.3.1.3).
5. The final facial feature vector $\boldsymbol{FD_{Vector}} = \{\boldsymbol{FBM_{Vector}}, \boldsymbol{IDBC_{Vector}}\}$ is constructed. To obtain a uniform feature vector, a normalization procedure is applied to produce the normalized feature vector $\widetilde{\boldsymbol{FD}}_{\boldsymbol{Vector}}$.
6. The QDC and CC classifiers are used in the final recognition tasks. The former is used for the identification task, while the latter is used for the verification task.

The next three subsections describe in more detail the Curvelet transform and the FBM and IDBC methods mentioned above.

3.3.1.1 Curvelet via Wrapping Transform

In this work, the wrapping-based Curvelet transform described below is adopted, because it is faster to compute, more robust, and less redundant than the alternative Ridgelet and USFFT based forms of Curvelet transform. Its ability to reduce the dimensionality of the data and capture the most crucial information within face images, such as edges and curves, plays a significant role in increasing the recognition power of the proposed system. Firstly, the Curvelet transform, implemented via the wrapping function, is defined as follows:

$$(\boldsymbol{j}, \boldsymbol{l}, \boldsymbol{k}) \coloneqq \langle \boldsymbol{f}, \boldsymbol{\varphi_{j,l,k}} \rangle \qquad (\mathbf{3.12})$$

Here, ⟨ ⟩ refer to the inner product between the Curvelet function $\varphi_{j,l,k}$ and the Cartesian form of the face image ***f***, and ***j***, ***l*** and ***k*** refer to the variables' scales, orientations, and positions, respectively [221].

Secondly, the Curvelet via wrapping transform can be implemented by taking the input image as a Cartesian array $f\,[n_1, n_2]$, such that $0 \le n_1 < N_1$, $0 \le n_2 < N_2$, where N_1 and N_2 are the dimensions of the original image. Then, a number of Curvelet coefficients are generated and indexed by scale ***j*** and orientation ***l***, and with two spatial location parameters $k = (k_1, k_2)$ as outputs. The major steps implemented on a face image to obtain the Curvelet coefficients can be summarized as follows:

1. Application of the 2D-Fast Fourier Transform (2D-FFT) to the input image $f\,[n_1, n_2]$ and obtain $\hat{f}\,[n_1, n_2]$, $-n/2 \le n_1, n_2 < n/2$.
2. Division of the transformed image into a collection of Digital Corona Tiles (Wedges) so each wedge can be reached by specifying scale and angle parameters, as shown in Fig.3.5. For each scale ***j*** and angle ***l*** the product $\hat{f}\,[n_1, n_2]\,\widetilde{U}_{j,l}\,[n_1, n_2]$ is implemented, where $\widetilde{U}_{j,l}$ is a discrete localizing function, defined by a pair of windows, which are a radial window ***W (r)*** and an angular window ***V*** (**t**). These windows are calculated as follows:

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1 \qquad r \in \left(\frac{3}{4}, \frac{3}{2}\right) \qquad (3.13)$$

$$\sum_{l=-\infty}^{\infty} V^2(t - l) = 1 \qquad t \in \left(\frac{-1}{2}, \frac{1}{2}\right) \qquad (3.14)$$

3. As shown in Fig.3.6, the wrapping procedure is applied to wrap this product around the origin and obtain $\tilde{f}_{j,l}\,[n_1, n_2] = W\left(\widetilde{U}_{j,l}\,\tilde{f}\right)[n_1, n_2]$, where the range for n_1 and n_2 is now $0 \le n_1 < L_{1,j}$ and $0 \le n_2 < L_{2,j}$. $L_{1,j} \sim 2^{j}$ and $L_{2,j} \sim 2^{j/2}$ are constants and $-\pi/4 < \theta < \pi/4$. Here θ refers to the orientation.
4. The inverse 2D FFT is applied for each $\tilde{f}_{j,l}$, and then the Curvelet array is added to the collection of Curvelet coefficients.

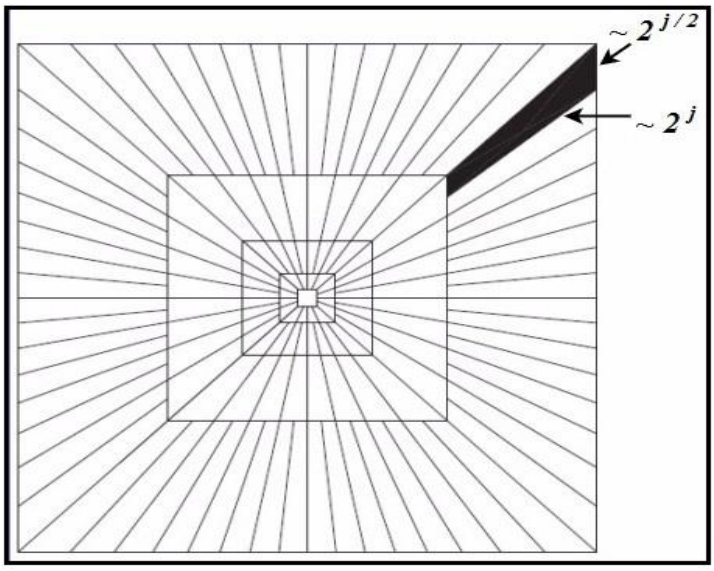


Figure 3.5: The discrete domain of the Curvelet frequency tiling [221].

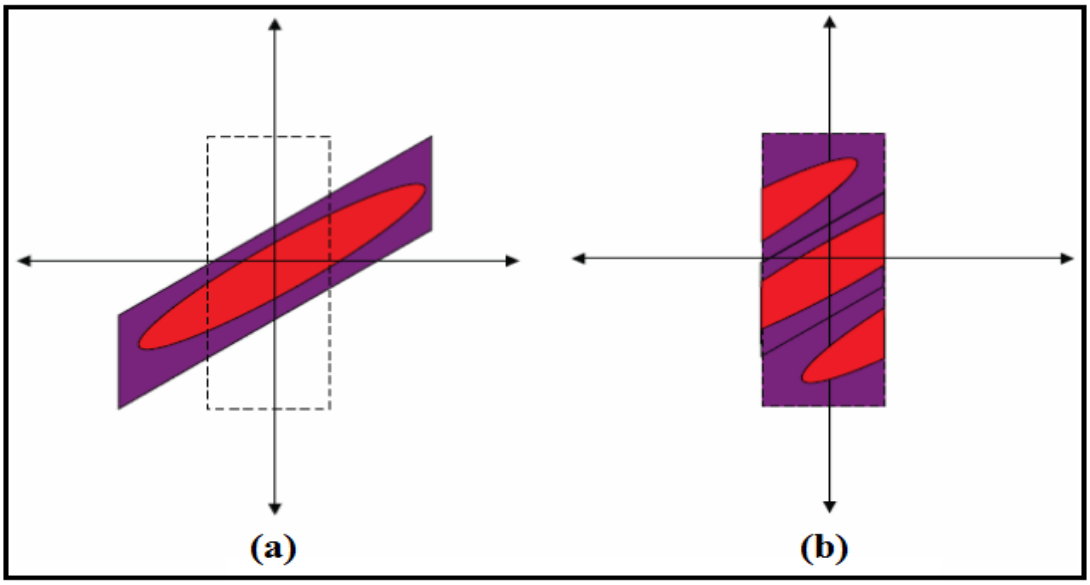


Figure 3.6: Wrapping procedure of a segment around a discrete localization window: (a) before wrapping and (b) after wrapping [248].

Based on domain-knowledge from literature, suggesting that a higher scale decomposition would only increase the number of Curvelet sub-bands (coefficients) with very marginal, or even no improvement, in recognition accuracy, the Curvelet coefficients are generated at scale **4** and orientation **8** throughout this work. This maintains an acceptable balance between the speed and performance of the proposed system. Fig.3.7 shows the Curvelet decomposition coefficients of a face image of size (128×128) pixel taken from

the FERET dataset. As indicated in Fig.3.7, the output of the Curvelet transform can be divided into three sets:

1. The coarse band, containing only the low frequency (approximation) coefficients, is stored at the centre of the display (***Scale$_1$***). These coefficients represent the main structure of the face.
2. The Cartesian concentric coronae that represents the middle-frequency bands of the Curvelet coefficients at different scales, where the outer coronae correspond to the higher frequencies **(*Scale$_2$,..., Scale$_{N-1}$*)**. Each corona is represented by four strips corresponding to the four cardinal points. These strips are further subdivided into angular panels, which represent the Curvelet coefficients at a specified scale and orientation. The coefficients in these bands represent the most significant information of the face, such as edges and curves.
3. The highest frequency band (***Scale$_N$***) of the face image, only indicated in Fig.3.7, is at scale **4**. This band has been discarded due to it being dominated by noise information.

From a practical point of view, the dimensionality of the Curvelet coefficients is extremely high due to the large amount of redundant and irrelevant information in each sub-band, especially in the middle-frequency bands. Hence, working on such a large number of Curvelet coefficients is very computationally expensive. A characteristic of the Curvelet transform is that it produces identical sub-bands coefficients at angle $\boldsymbol{\theta}$ and $(\boldsymbol{\pi}+\boldsymbol{\theta})$ for the same scale. Thus, only half of the Curvelet sub-bands need to be considered. In this work, instead of the direct use of the Curvelet coefficients, analysis and processing of these coefficients is done using other methods. For the coarse band (the lowest frequency band), an image contrast enhancement procedure is applied, as shown in Fig.3.8, to improve the illumination uniformity of the face image stored at the centre of the display by stretching the overall contrast of the image between two pre-defined lower and upper cut-offs, which are empirically set to be **0.11**, and **0.999**, respectively. This is followed by extracting the face texture roughness and fluctuations in the surface using the FBM method. For the middle-frequency bands, the IDBC method is applied to reflect the face texture information, and to reduce the high dimensionality of these bands.

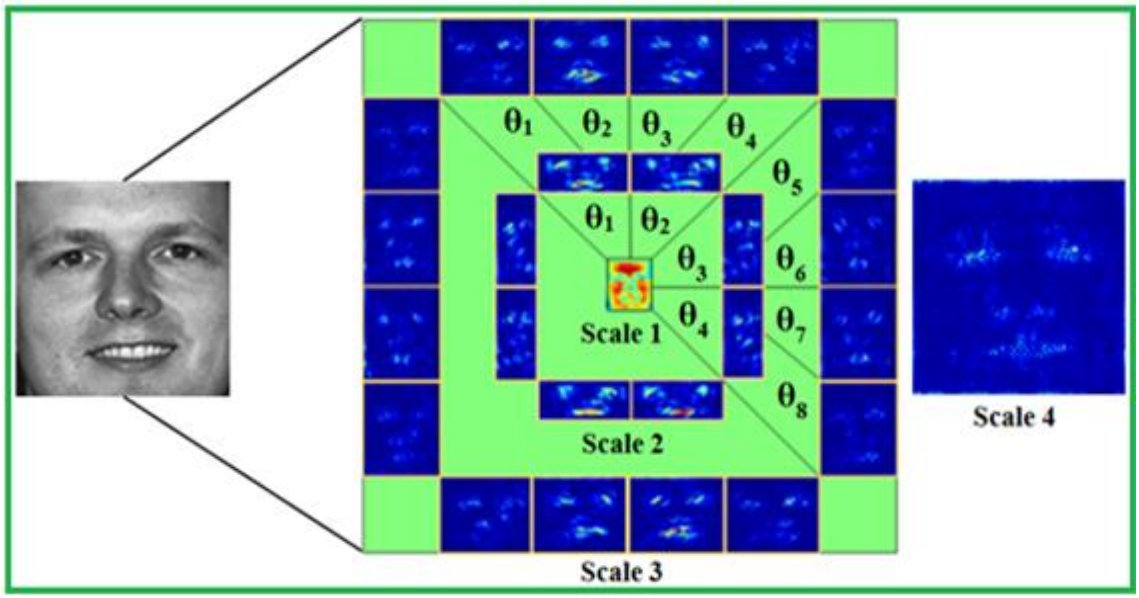


Figure 3.7: Illustration of the Curvelet decomposition coefficients obtained from a face image decomposed at scale 4 and orientations 8.

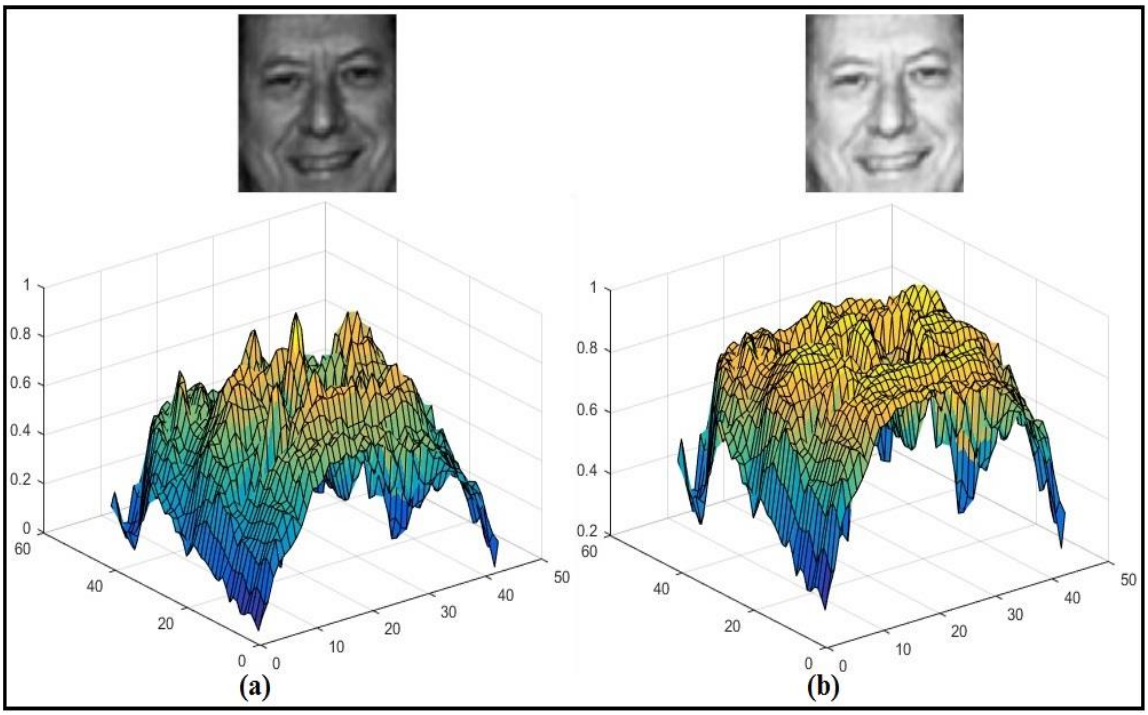


Figure 3.8: Output of the image contrast enhancement procedure: (a) The coarse band and (b) The enhanced coarse band.

3.3.1.2 Fractional Brownian Motion

As shown in Fig.3.9, the 2D face image can be considered as a 3D spatial surface that reflects the grey-level intensity value at each pixel position where the neighbourhood region around each pixel crosses the face surface, which covering a varying range of grey levels, can be processed as an FBM surface. To the best of the author's knowledge, this is the first attempt to use FBM as a facial feature extraction method. The FBM is a non-stationary model, and is widely used in medical imaging [232][249], due to its power to enhance the original image and make the statistical features more distinguishable. For example, in [250] it was found that employing the normalized FBM to extract the feature vectors from surfaces of five ultrasonic liver images improved the classification of the normal and abnormal liver tissues. Moreover, the Fractal dimension for each pixel, calculated over the whole medical image by the normalized FBM method, could be used as a powerful edge enhancement and a detection method, which can enhance the edge representation for the medical images without increasing the noise level.

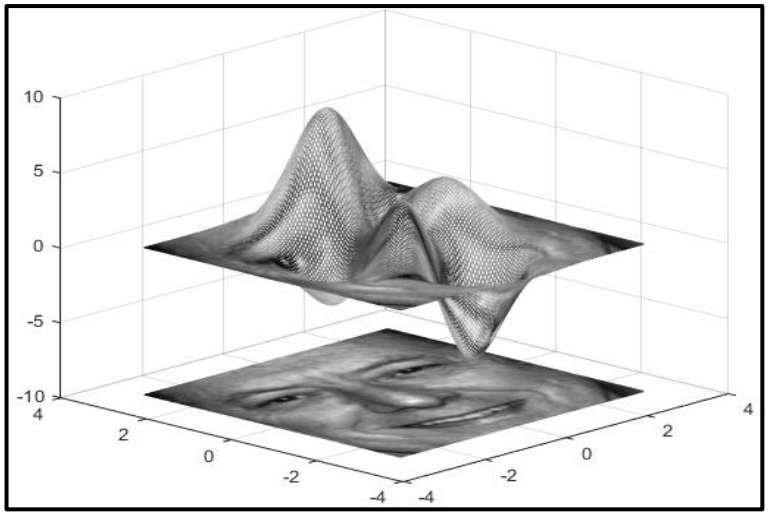


Figure 3.9: The spatial surface corresponding to a grey-scale face image.

According to Mandelbrot [231], the FBM is statistically self-affine, which means that the Fractal dimension value of the FBM is not affected by linear transformations such as scaling. Therefore, the FBM is invariant under normally observed transformations of face images. For any (***N×N***) size image region, the FBM can be defined as the mean absolute difference of pixel pairs on a surface at different scale ranges, which can be represented as follows:

$$\boldsymbol{E}(\Delta \boldsymbol{I}_{\Delta r}) = \boldsymbol{k}\Delta \boldsymbol{r}^{\boldsymbol{H}} \tag{3.15}$$

Here, $\boldsymbol{E}(\,)$ is an expectation operator, and $\Delta \boldsymbol{I}_{\Delta r} = |\boldsymbol{I}\,(\boldsymbol{x}_2, \boldsymbol{y}_2) - \boldsymbol{I}\,(\boldsymbol{x}_1, \boldsymbol{y}_1)|$ is the absolute intensity difference between pairs of pixels separated by distance $\Delta \boldsymbol{r} = [\,(\boldsymbol{x}_2, \boldsymbol{x}_1)^2 + (\boldsymbol{y}_2, \boldsymbol{y}_1)^2]^{1/2}$. ***K*** is a scaling constant (**> 0**) and ***H*** is called the Hurst exponent. Due to the discrete form of the image, the $\Delta \boldsymbol{r}$ value between all the pixels is usually an integer value ***d*** from **1** to ***N*-1**. Thus, Eq.(3.15) can be expressed as follows:

$$\frac{\mathbf{1}}{\boldsymbol{N}_{\boldsymbol{d}}} \sum_{\Delta \boldsymbol{r} = \boldsymbol{d}} \Delta \boldsymbol{I}_{\Delta r} = \boldsymbol{k}\boldsymbol{d}^{\boldsymbol{H}} \tag{3.16}$$

Here, $\boldsymbol{N}_{\boldsymbol{d}}$ is the total number of pixel pairs with a distance $\Delta \boldsymbol{r} = \boldsymbol{d}$. By taking the log of both sides of Eq.(3.16), the following form may be deduced:

$$\boldsymbol{log}\left(\frac{\mathbf{1}}{\boldsymbol{N}_{\boldsymbol{d}}} \sum_{\Delta \boldsymbol{r} = \boldsymbol{d}} \Delta \boldsymbol{I}_{\Delta r}\right) = \boldsymbol{H}.\boldsymbol{log}(\boldsymbol{d}) + \boldsymbol{log}(\boldsymbol{k}) \tag{3.17}$$

After $\Delta \boldsymbol{r}_{\boldsymbol{min}}$ and $\Delta \boldsymbol{r}_{\boldsymbol{max}}$ have been determined, a graph of Eq.(3.17) is plotted and least squares linear regression is used to estimate the slope, which represents the value of ***H***. Finally, the Fractal dimension (***FD***) of the image surface is estimated as follows:

$$\boldsymbol{FD} = \mathbf{3} - \boldsymbol{H} \tag{3.18}$$

From a theoretical viewpoint, if the surface of an image is a perfect fractal surface, then the ***H*** value is constant, and Eq. (3.15) is true over the whole domain of $\Delta \boldsymbol{r}$. However, sometimes only a specific region of the image can be described as Fractal. Therefore, a restricted range of $\Delta \boldsymbol{r}$ is required, from $\Delta \boldsymbol{r}_{\boldsymbol{min}}$ to $\Delta \boldsymbol{r}_{\boldsymbol{max}}$. This restricted range is the scale free interval of a fractal object in which the Fractal dimension value is stable, and the fitted curve is a straight line. In this work, the face image of size (***M×N***) is transformed to its Fractal dimension form by applying a kernel function ***fd(p,q)*** of size (***n×n***) on the entire face image, using the algorithm summarized in Fig.3.10. In this work, the kernel function operates by block processing on (**7×7**) neighbouring pixels of the face image, and calculating the Fractal dimension value of each pixel from its surrounding neighbours, as explained above. As a result, a fractal transformed face image is obtained.

The size of the kernel function was determined empirically, noting that increasing its size can affect the accuracy of the calculated Fractal dimension, causing the obtained image to become less distinct, while decreasing its size can result in an insufficient number of surrounding pixels to calculate accurately the Fractal dimension value. The kernel function computed and applied to the face image is defined in Eq.(3.19) and Eq.(3.20). This implementation of the FBM has the ability to enhance the edges and curves representations and create an illumination-invariant representation of the face image without increasing the noise level.

$$\boldsymbol{fd(p,q) = 3 - \left(\log\left(\frac{\Delta I}{k}\right) \Big/ \log(\Delta r) \right)} \qquad (3.19)$$

$$\boldsymbol{FDImage\,(x,y) = \sum_{p=-a}^{a} \sum_{q=-b}^{b} fd(p,q) I(x+p, y+q)} \qquad (3.20)$$

Here, ***a*** and ***b*** are non-negative integer variables, which are used to centre the kernel function on each pixel in the face image, and are defined as: ***a*** and ***b*** **= *ceil* ((*n* − 1)/ 2)**. Fig.3.11 shows examples of the approximation coefficients of the Curvelet transform and the resulting fractal transformed images. After, a fractal transformed image of size (***M×N***) has been obtained, it is reordered into a row feature vector, $\boldsymbol{FBM_{Vector}}$, for further analysis.

In this work, a mirror mapping of pixels along the edge of the input image is made to fill the part of the sliding window, when the center pixel of the sliding window is lying on or near the edge of the input image.

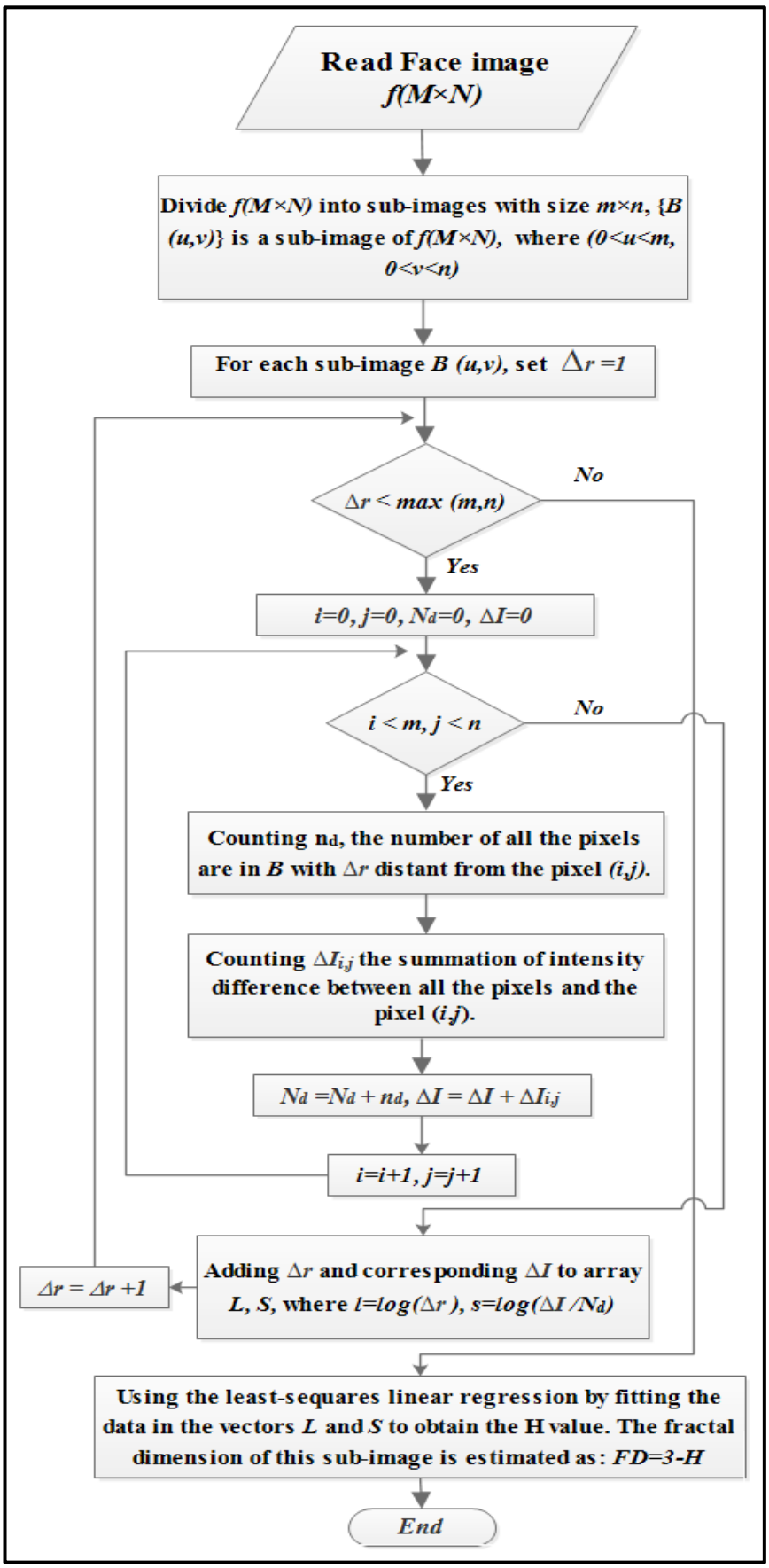


Figure 3.10: A block diagram of the implementation of the FBM method.

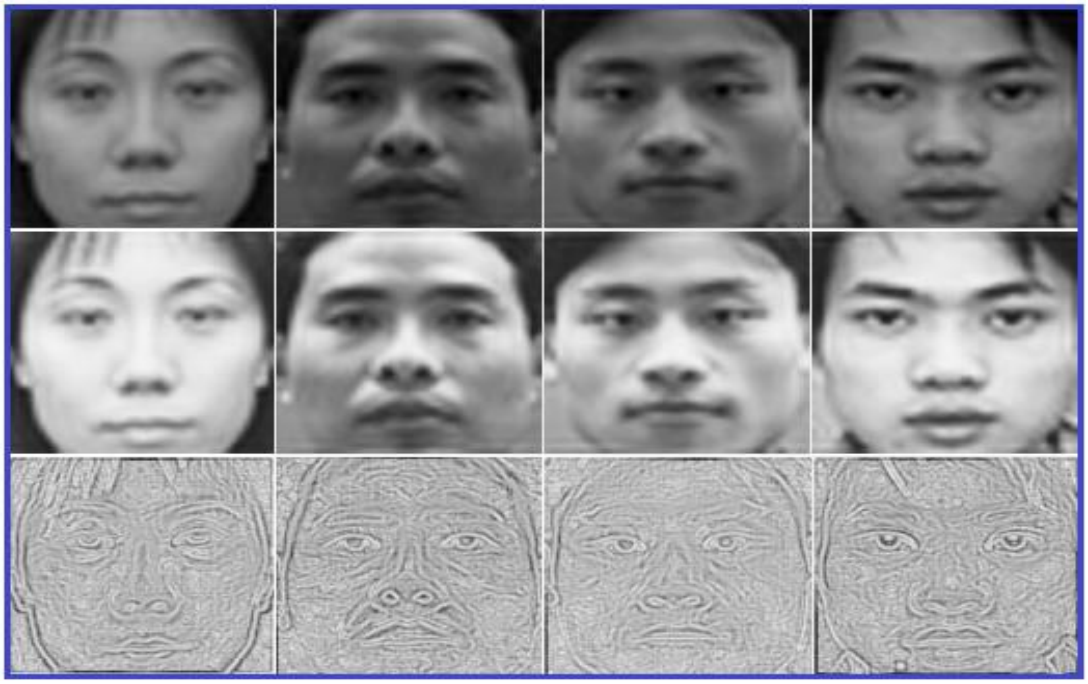


Figure 3.11: The top row shows coarse band Curvelet approximation coefficients of four images. The middle row shows the images after applying the contrast enhancement procedure, and the bottom row shows the FBM fractal transformed images.

3.3.1.3 Improved Differential Box Counting

The main purpose of the second Fractal method is to estimate the Fractal dimension features from the middle-frequency bands of the Curvelet transform, reduce the high dimensionality of these bands, and increase the speed of the proposed system. Face recognition, like other pattern recognition systems, suffers from the problem of high dimensionality. There are many possible reasons for reducing the feature vector size, such as providing a more efficient way for storing and processing the data related to the increasing number of training samples, and increasing the discriminative power of the feature vectors. The second method to compute the Fractal dimension is based on the Improved Differential Box Counting (IDBC). The basic approach of the traditional DBC is to treat any image of size ($M \times M$) as a 3D space, where (x, y) denotes the pixel position on the image surface, and the third coordinate (z) denotes the pixel intensity. The DBC starts by scaling the image down into non-overlapping blocks of size ($s \times s$), where ($M/2 > s > 1$) and s is an integer, and then the Fractal dimension is calculated as follows:

$$FD = \lim_{r \to 0} \frac{\log(N_r)}{\log(1/r)} \quad (3.21)$$

Here, $r = s$ is the scale of each block and N_r is the number of boxes required to entirely cover the object in the image, which is counted in the DBC method as follows: On each block there is a column of boxes of size ($s \times s \times s'$), where $s'=s$, and each box is assigned with a number (**1,2,…**), starting from the lowest grey level value, as shown in Fig.3.12. Let the minimum and the maximum grey level of the image in the $(i, j)^{th}$ block fall in box number k and l, respectively. The contribution of n_r in $(i, j)^{th}$ block is calculated as follows:

$$n_r(i,j) = l - k + 1 \quad (3.22)$$

The contributions from all blocks N_r is counted for different values of r as follows:

$$N_r = \sum_{i,j} n_r(i,j) \quad (3.23)$$

More information on this technique, and its implementation, can be found in [230]. The traditional DBC has many issues. The most important is how to choose the best size of the boxes that cover each block on the image surface. This can significantly affect the results of the curve fitting process, and result in inaccurate estimation of the Fractal dimension. Moreover, calculating the Fractal dimension using the traditional DBC cannot accurately reflect the local and global facial features of different and similar classes. Finally, the traditional DBC method can suffer from over or under counting of the number of boxes that cover a specific block, which leads to calculating the Fractal dimension inaccurately [251][252].

To overcome these drawbacks, an improved DBC method is proposed, which involves calculating and describing the Fractal dimension as a feature vector. This feature vector is calculated by, firstly, dividing each Curvelet sub-band into a ($k \times k$) sized sub-images, and then dividing each sub-image farther into ($n \times n$) sized blocks. In the proposed method, ($\log_2 n$) different sizes of boxes are used, which are then used to cover the same block. These boxes are represented by $(b_1, b_2, b_3, \ldots, b_{\log_2 n})$. Then, the Fractal dimension value of each block is estimated, as in Eq.(3.21). For example, a Curvelet sub-band of size

(**32×32**) will be decomposed into (**16×16**) size sub-images, and then a grid of blocks of size (**4×4**) will cover each sub-image. The Fractal dimension feature is estimated from each block using ($\boldsymbol{log_2 4}$) different sizes of boxes. Then, from each sub-image, **16** Fractal dimension features are estimated. By combining the features obtained from the four sub-images (**4×16**), a sub-row feature vector $\mathbf{V_i}$ **=** $\{Fd_1, Fd_2, \ldots, Fd_{64}\}$ is constructed for each Curvelet sub-band. As in Eq.(3.24), the final feature vector $\boldsymbol{IDBC_{Vector}}$ of the middle-frequency bands is constructed by combining the $\mathbf{V_i}$ from **4** and **8** sub-bands located at scale **2** and **3**, respectively.

$$\boldsymbol{IDBC_{Vector}} = \{\boldsymbol{V_1}, \boldsymbol{V_2}, \ldots, \boldsymbol{V_{12}}\} \qquad (\mathbf{3.24})$$

In this work, to ensure the correct division without losing any important information, the Curvelet sub-bands at scale **2** and **3** have been resized from their original sizes to (**24×24**) and (**32×32**), respectively. The experimental results have demonstrated that calculating the Fractal dimension features using different sizes of boxes covering the same block can play a significant role in increasing the discriminative power of the final feature vector, by efficiently reflecting the face texture information using the edges and curves of the face presented in the middle-frequency bands.

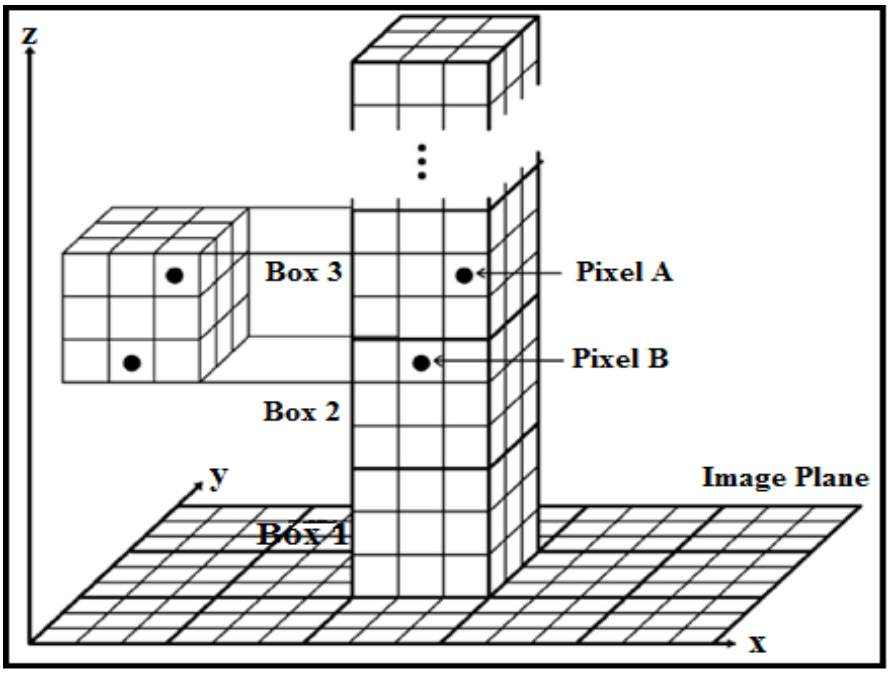


Figure 3.12: Calculating the Fractal dimension by using the traditional (DBC) [253].

3.3.1.4 Face Matching Techniques

Classification and decision making are the final steps in the proposed *Curvelet-Fractal* approach. These refer to the process of either classifying the tested samples into ***N*** classes based on the identity of the training subjects, or deciding whether two faces belong to the same subject or not. In this chapter, the QDC and CC classifiers are used in the identification and verification tasks, respectively. The QDC from PRTools[10] is a supervised learning algorithm commonly used for multi-classification tasks. It is a *Bayes-Normal-2* classifier assuming Gaussian distributions, which aims to differentiate between two or more classes using a quadric surface. Using this Bayes rule, a separate covariance matrix is estimated for each class, yielding quadratic decision boundaries. This is done by estimating the covariance matrix (***C***) for the scatter matrix (**S**), as follows:

$$\boldsymbol{C} = (\mathbf{1} - \boldsymbol{\alpha} - \boldsymbol{\beta})\boldsymbol{S} + \boldsymbol{\alpha} * \boldsymbol{diag}(\boldsymbol{S}) + \frac{\boldsymbol{\beta}}{\boldsymbol{n}} \sum \boldsymbol{diag}(\boldsymbol{S}) \qquad (\mathbf{3.25})$$

Here, ***n*** refers to the dimensionality of the feature space, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta} \in [\mathbf{0}, \mathbf{1}]$ are regularisation parameters. In this work, these parameters are determined empirically to be $\boldsymbol{\alpha} = \mathbf{0.1}$ and $\boldsymbol{\beta} = \mathbf{0.2}$, as explained in (Section 3.4.2.1). The decision making is based on calculating the similarity scores between the two face images using the CC classifier, which is defined as follows:

$$\boldsymbol{C}(\boldsymbol{A}, \boldsymbol{B}) = \frac{\sum_m \sum_n (\boldsymbol{A}_{mn} - \bar{\boldsymbol{A}})(\boldsymbol{B}_{mn} - \bar{\boldsymbol{B}})}{\sqrt{\left(\sum_m \sum_n (\boldsymbol{A}_{mn} - \bar{\boldsymbol{A}})^2\right)\left(\sum_m \sum_n (\boldsymbol{B}_{mn} - \bar{\boldsymbol{B}})^2\right)}} \qquad (\mathbf{3.26})$$

Here, ***m*** and ***n*** are the dimensions of the sample, and $\bar{\boldsymbol{A}}$ and $\bar{\boldsymbol{B}}$ are the mean values of the testing and training samples, respectively.

3.3.2 Learning Additional Features Representations

Like any deep learning approach, the DBN is usually applied directly on the pixel intensity representations. However, although DBN has been successfully applied in many different fields, scaling it to realistic-sized face images still remains a challenging task for several reasons. Firstly, the high dimensionality of the face image leads to increased computational complexity of

[10] http://www.37steps.com/prhtml/prtools.html

the training algorithm. Secondly, the feature representations of the DBN are sensitive to the local translations of the input image.

This can lead to a disregard of the local features of the input image, which are known to be important for face recognition. To address these issues of the DBN, a novel framework, based on merging the advantages of the local handcrafted image descriptors and the DBN, is proposed. It is proposed that applying the DBN on top of local features representations, instead of the pixel intensity representations (raw data), as a way of guiding the learning process, can greatly improve the ability of the DBN to learn more discriminating features with a shorter training time required to obtain the final trained model. As shown in Fig.3.4, the local facial features are first extracted using the proposed *Curvelet-Fractal* approach. Then, the extracted local features are assigned to the feature extraction units of the DBN to learn additional and complementary representations. In this work, the DBN architecture stacks **3** RBMs (**3** hidden layers). The first two RBMs are used as generative models, while the last one is used as a discriminative model associated with Softmax units for multi-class classification purposes. Finally, the hidden layers of the DBN are trained one at a time in a bottom-up manner, using a greedy layer-wised training algorithm. In this work, the training methodology to train the DBN model can be divided into three phases: pre-training, supervised, and fine-tuning.

1. In the pre-training phase, the first two RBMs are trained in a purely unsupervised way, using a greedy training algorithm, in which each added hidden layer is trained as an RBM (e.g., using the CD algorithm). The activation outputs of a trained RBM can be viewed as feature representations extracted from its input data, which will be the input data (visible units) used to train the next RBM in the stack. The unsupervised pre-training phase is finished when the learning of the second hidden layer is completed. The main advantage of the greedy unsupervised pre-training procedure is the ability to train the DBN using a massive amount of unlabelled training data, which can improve the generalization ability and prevent overfitting. In addition, the degree of complexity is reduced, and the speed of training is increased.

2. In the supervised phase, the last RBM is trained as a non-linear classifier, using the training and validation set, along with their associated labels to observe its performance in each epoch.
3. Finally, the fine-tuning phase is performed in a top-down manner using the Back-propagation algorithm to fine-tune parameters (weights) of the whole network for optimal classification.

A difference, compared with conventional neural networks, is that the DNNs require a massive amount of training data to avoid overfitting during the learning process and achieve satisfactory predictions. Hence, data augmentation is the simplest and most common method of achieving this, which artificially enlarges the training dataset using techniques such as: random crops, intensity variations, and horizontal flipping. In contrast to previous works that randomly sample a large number of face image patches [254][82], it is proposed to sample a uniformly small number of face image patches. To prevent background information from artificially boosting the results of the proposed *Curvelet-Fractal*[11] approach, and to speed up experiments when the DBN is directly applied on the pixel intensity representations, the face region is detected, and the data augmentation procedure[12] is implemented on the detected face image. In this work, for a face image of size ($\mathbf{H_{dim}} \times \mathbf{W_{dim}}$), five images patches of the same size are cropped, four starting from the corner and one centred (and their horizontally flipped counterparts), which helps maximize the complementary information contained within the cropped patches. Fig. 3.13 shows the ten image patches generated from a single input image.

[11] The data augmentation procedure is not implemented during the performance assessment of the proposed *Curvelet-Fractal* approach.

[12] In this work, the data augmentation procedure is applied only for the training set.

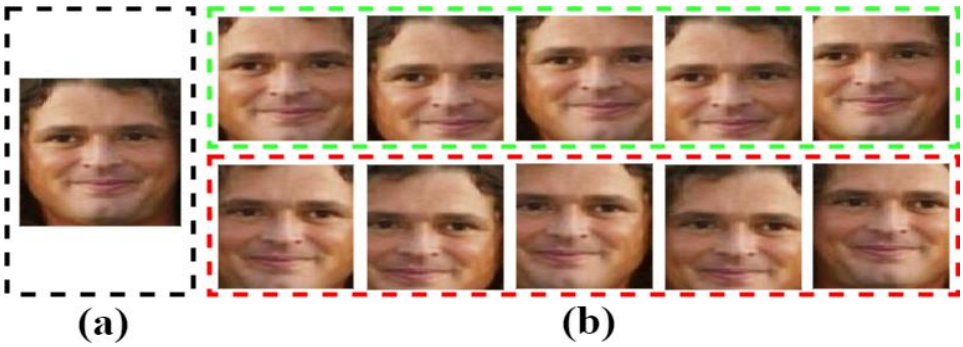


Figure 3.13: Data augmentation procedure: (a) Detected Face image, (b) The normalized face patches used as input for the MDFR where the (top) row are patches sampled from (a), and the (bottom) row their horizontal flipped versions.

3.4 Experimental Results

In this section, comprehensive experiments are described using the proposed approaches for both face identification and verification tasks, in order to demonstrate their effectiveness and compare their performance with other existing approaches. Firstly, a brief description of the face datasets used in these experiments is given. Then a detailed evaluation and comparison with the state-of-the-art approaches are presented, in addition to some insights and findings about learning additional features representations, by training a DBN on top of local feature representations.

3.4.1 Description of Face Datasets

In this work, all the experiments were conducted on four large-scale unconstrained face datasets: SDUMLA-HMT [255], FacE REcognition Technology (FERET) [256], CAS-PEAL-R1[257], and Labelled Faces in the Wild (LFW) [258]. Some examples of face images from each dataset are shown in Fig. 3.14.

- **SDUMLA-HMT face dataset** [255]**:** This includes 106 subjects, and each has 84 face images taken from **7** viewing angles, and under different experimental conditions including, facial expressions, accessories, poses, and illumination. Fig.3.15 shows the camera setting and illumination setting in the capturing process of face images in the SDUMLA-HMT dataset. The

main purpose of this dataset is to simulate real world conditions during face image acquisition. The image size is (**640×480**) pixels.

- **FERET dataset** [256]**:** This contains a total of 14,126 images taken from 1,196 subjects, with at least 365 duplicate sets of images. This is one of the largest publicly-available face datasets with a high degree of diversity of facial expression, gender, illumination conditions and age. The image size is (**256×384**) pixels.
- **CAS-PEAL-R1dataset** [257]**:** A subset of the CAS-PEAL face dataset has been released for research purposes and named CAS-PEAL-R1. This contains a total of 30,863 images taken from 1,040 Chinese subjects (595 are males and 445 are females). The image size is (**360×480**) pixels.
- **LFW dataset** [258]**:** This contains a total of 13,233 images taken from 5,749 subjects, where 1,680 subjects appear in two or more images. In the LFW dataset, all images were collected from Yahoo! News articles on the web, with a high degree of intra-personal variations in facial expression, illumination conditions, occlusion from wearing hats and glasses, etc. It has been used to address the problem of unconstrained face verification task in recent years. The image size is (**250×250**) pixels.

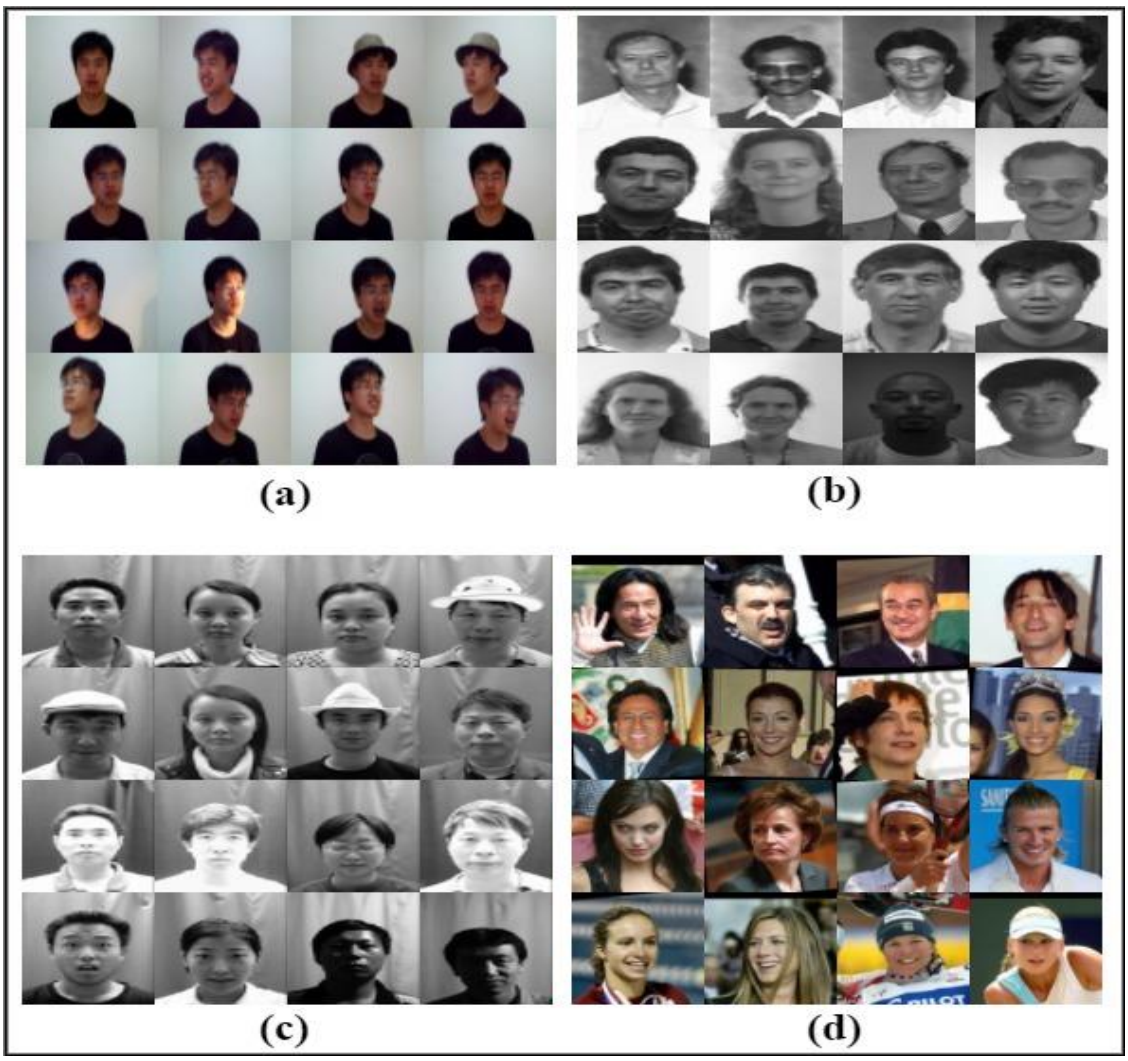


Figure 3.14: Examples of face images in four face datasets: (a) SDUMLA-HMT, (b) FERET, (c) CAS-PEAL-R1, and (d) LFW.

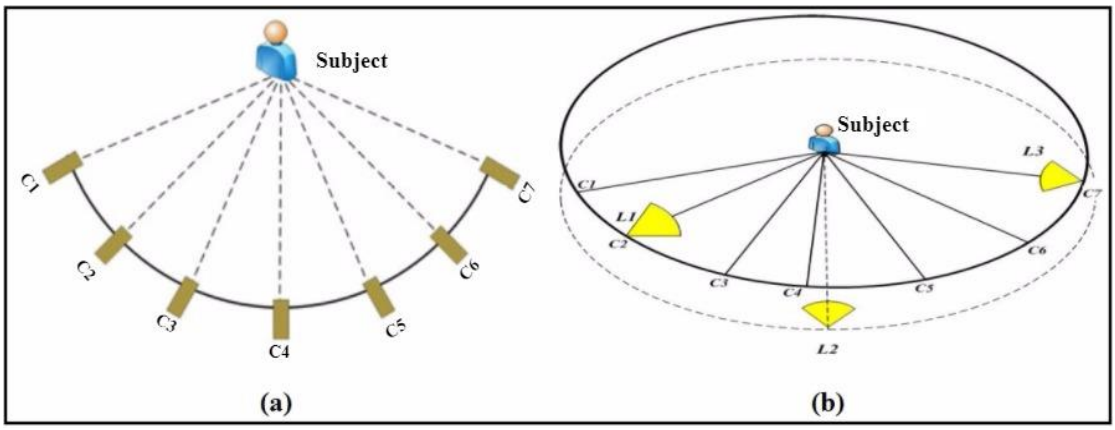


Figure 3.15: An illustration of the environmental setting during the capturing process of the face image in the SDUMLA-HMT dataset [255].

3.4.2 Face Identification Experiments

This section describes the evaluation of the proposed approach to the face identification problem on three different face datasets: SDUMLA-HMT, FERET, and CAS-PEAL-R1. To the best of the author's knowledge, this is the first work that uses the SDUMLA-HMT face dataset for evaluating a face recognition approach. In this work, the SDUMLA-HMT dataset is used as the main dataset to fine-tune the hyper-parameters of the proposed *Curvelet-Fractal* approach (e.g. regularization parameters of the QDC classifier), as well as the proposed MDFR framework (e.g. number of hidden units per layer, etc.), because it has more images per subject in its image gallery than the other datasets. This allowed more flexibility in dividing the face images into training, validation, and testing sets.

3.4.2.1 Parameter Settings of the Curvelet-Fractal Approach

In the proposed *Curvelet-Fractal* approach, the most important thing is to set the regularization parameters of the QDC classifier. In this work, these parameters are determined empirically by varying their values from 0 to 1 in steps of 0.1, starting with $\boldsymbol{\alpha} = \mathbf{0}$. and $\boldsymbol{\beta} = \mathbf{0}$. Hence, 121 experiments were conducted where each time the former was increased by 0.1 and tested it with all the possible values of the latter. Fig.3.16 shows the Validation Accuracy Rate (VAR) generated throughout these experiments. These experiments were carried out using 80% randomly-selected samples for the training set, and the remaining 20% for the testing set. In particular, the parameters optimization process is performed on the training set using the 10-fold cross-validation procedure that divides the training set into ***k*** subsets of equal size. Sequentially, one subset is used to evaluate the performance of the classifier trained on the remaining ***k*-1** subsets. Then, the Average Error Rate (AER) over 10 trials is calculated as follows:

$$\boldsymbol{AER} = \frac{\mathbf{1}}{\boldsymbol{K}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{k}} \boldsymbol{Error_i} \qquad (\mathbf{3.27})$$

Here, $\boldsymbol{Error_i}$ refers to the error rate per trial. After finding the best values of the regularization parameters, the QDC classifier is trained using the whole training set, and its performance in predicting unseen data properly is then

evaluated, using the testing set. Algorithm 1, shows pseudo-code of the procedure proposed to train the QDC classifier. Fig. 3.17, shows results comparing the present *Curvelet-Fractal* approach with the author's previous Curvelet Transform-Fractional Brownian Motion (CT-FBM) approach described in (Chapter 1, Sect. 1.8) using the Cumulative Match Characteristic (CMC) curve to visualize the performance of both approaches. It can be seen in Fig. 3.17 that the Rank-1 identification rate has dramatically increased from 0.90 to 0.95 using the CT-FBM to more than 0.95 to 1.0 using the *Curvelet-Fractal* approach.

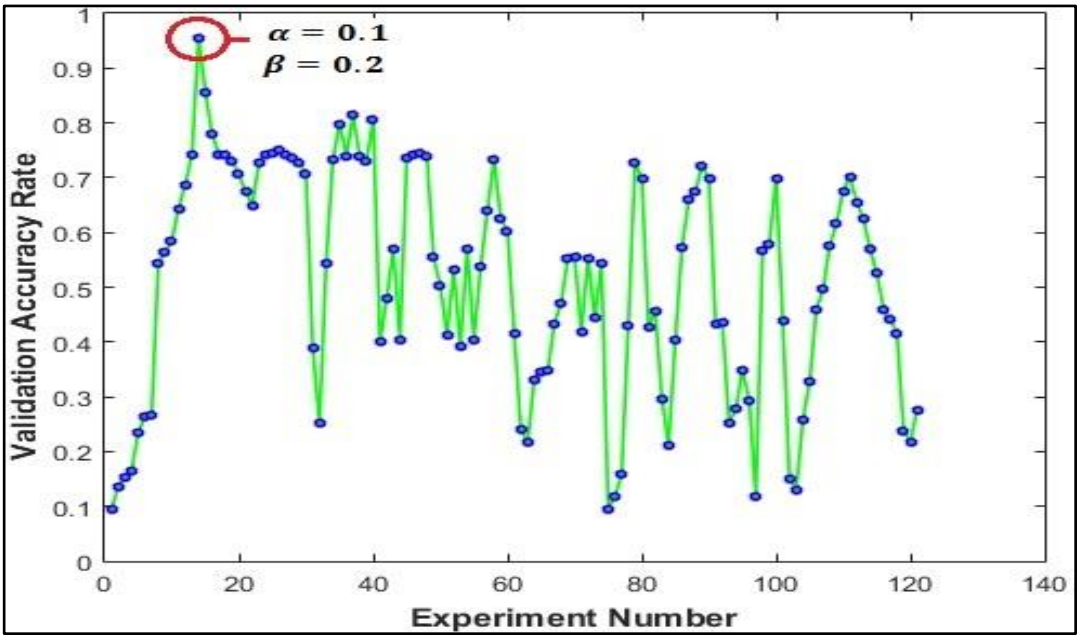


Figure 3.16: The VAR generated throughout 121 experiments of finding the best regularization parameters.

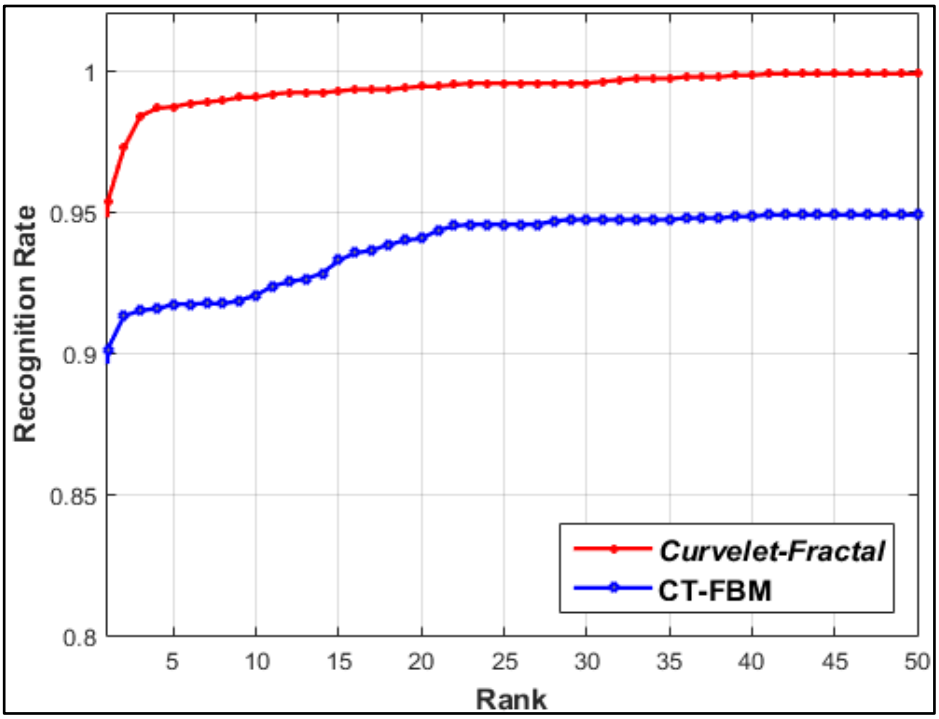


Figure 3.17: Performance comparison between the *Curvelet-Fractal* and CT-FBM approaches on the SDUMLA-HMT Dataset.

Algorithm 1. Find the optimal QDC classifier.

Input: Labeled dataset: **Training Set** $(\boldsymbol{x}_i, \mathbf{y}_i)$ and **Testing Set** $(\boldsymbol{x}_i, \mathbf{y}_i)$.
Output: Optimal QDC Classifier.
Initialization: $\boldsymbol{\alpha} = \mathbf{0}$, $\boldsymbol{\beta} = \mathbf{0}$, and **Max_Accurcy=0**;

```
for α_now = 0 : 0.1 : 1
    for β_now = 0 : 0.1 : 1
        ➢ Train the QDC classifier using the current values of α_now and β_now.
        ➢ Apply the 10-folds cross-validation evaluation procedure on the
          Training Set.
        ➢ Calculate the Average Error Rate (AER) over 10-folds using Eq.(3.27).
        ➢ Calculate the Validation_Accurcy = 1-AER;
        ➢ if  Validation_Accurcy>Max_Accurcy
              Max_Accurcy = Validation_Accurcy;
              Optimal_ α  = α_now;
              Optimal_ β_= β_now;
          end
    end
end
```

➢ Train the QDC classifier using the **Optimal_ α** and **Optimal_ β**.
➢ Assess the final trained QDC classifier using the **Testing Set**.

3.4.2.2 MDFR Architecture and Training Details

The major challenge of using DNNs is the number of the model architectures and hyper-parameters that need to be evaluated, such as the number of layers, the number of units per layer, learning rate, the number of epochs, etc. In a DBN, the value of a specific hyper-parameter may mainly depend on the values selected for other hyper-parameters. Moreover, the values of the hyper-parameters set in one hidden layer (RBM) may depend on the values of the hyper-parameters set in other hidden layers (RBMs). Therefore, hyper-parameter tuning in DBNs is very computationally expensive. Given these findings, the best hyper-parameter values are found by performing a coarse search over all the possible values. In this section, all the experiments were carried out using 60% randomly selected samples for training, and the remaining 40% samples were divided into two sets of equal size as validation and testing sets. In all experiments, the validation set is used to assess the generalization ability of the MDFR framework during the learning process before using the testing set. Following, the training methodology described in (Section 3.3.2), the MDFR framework was greedily trained using input data acquired from the *Curvelet Fractal* approach. Once the training of a given hidden layer is accomplished, its weights matrix is frozen, and its activations are served as input to train the next layer in the stack.

As shown in Table 3.1, four different 3-layer DBN models were greedily trained in a bottom-up manner using different numbers of hidden units. For the first two layers, each one was trained separately as an RBM model in an unsupervised way, using the CD learning algorithm with 1 step of Gibbs sampling (CD-1). Each individual model was trained for 300 epochs with a momentum of 0.9, a weight-decay of 0.0002, and a mini-batch size of 100. The weights of each model were initialized with small random values sampled from a zero-mean normal distribution and standard deviation of 0.02. Initially, the learning rate was to be 0.001 for each model as in [259], but it was observed this was inefficient, as each model took too long to converge, due to the learning rate being too small. Therefore, for all the remaining experiments, the learning rate was set to be 0.01. The last RBM model was trained in a supervised way as a non-linear classifier using the training and validation set,

along with their associated labels to evaluate its discriminative performance. In this phase, the same values of the hyper-parameters used to train the first two models were used, except that the last model was trained for 400 epochs. Finally, in the fine-tuning phase, the whole network was trained in a top-down manner using the Back-propagation algorithm equipped for Dropout compensation, to find optimized parameters and to avoid overfitting. The Dropout ratio was set to 0.5 and the number of epochs through the training set was determined using an early stopping procedure, in which the training process is stopped as soon as the classification error on the validation set starts to rise again. In these experiments using the validation set, it was found (see Table 3.1 and Fig. 3.18) that hidden layers with sizes **800, 800, 1000** provided considerably better results than the other hidden layer sizes that the author trained. This model trained on input data acquired from the *Curvelet-Fractal* approach is termed the MDFR framework. Table 3.1 shows the Rank-1 identification obtained from the four trained DBNs models over the validation set, while the CMC curves shown in Fig.3.18 are used to visualize their performance on the validation set.

Table 3.1: Rank-1 identification rates obtained for different DBN architectures using Validation set.

DBN Models	Accuracy Rate %
600-600-1000	92.19
700-700-1000	92.48
800-800-1000	95.38
900-900-1000	93.68

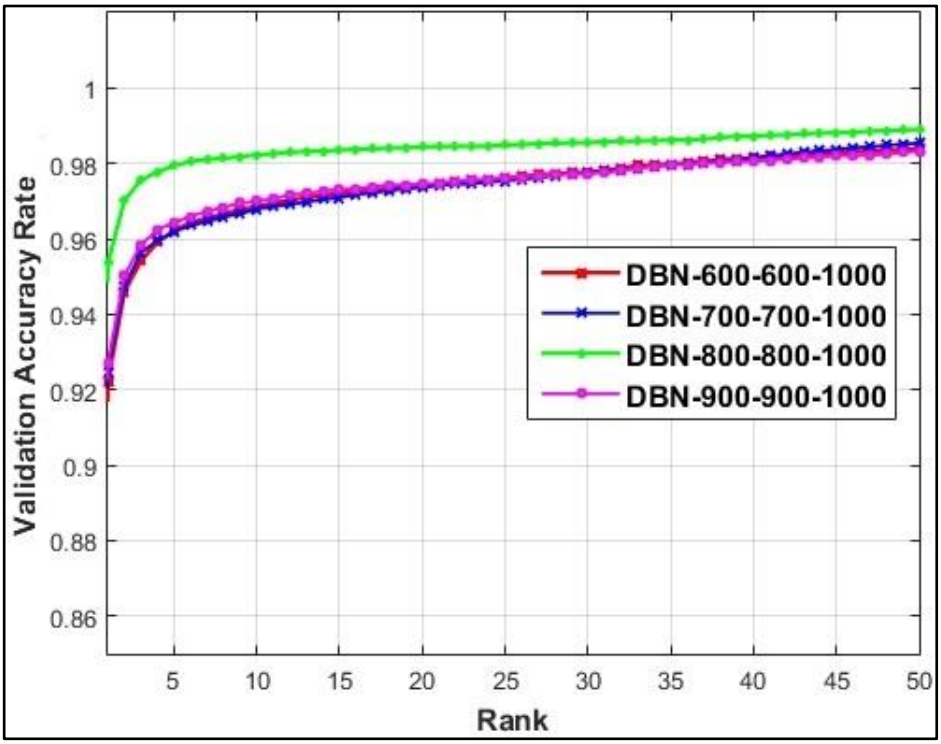


Figure 3.18: CMC curves for the four trained DBNs models over the validation set.

3.4.2.3 Comparative Study of Fractal, Curvelet-Fractal, DBN and MDFR Approaches

In this section, to evaluate the feature representations obtained from the MDFR framework, its recognition accuracy was compared with feature representations obtained by the Fractal, *Curvelet-Fractal* approach and DBN[13]. This comparison study was conducted for several reasons: Firstly, to demonstrate the efficiency of the proposed *Curvelet-Fractal* approach compared with applying the Fractal dimension individually. Secondly, to demonstrate that the feature representations acquired by the MDFR framework as a deep learning approach is complementary to the feature representations acquired by *Curvelet-Fractal* approach as a handcrafted-descriptors; Thirdly, to show that applying the DBN on top of the local feature representations, instead of the pixel intensity representations, can significantly improve the ability of the DBN to learn more discriminating features with less training time required. Finally, using these complementary feature representations, the MDFR framework was able to handle efficiently the non-linear variations of face

[13] The DBN model was trained on the top of the pixel intensity representation using the same hyper-parameters of the MDFR framework.

images due to the non-linearity of a DBN. In this work, the input image rescaled to (32×32) pixel to speed up the experiments when the Fractal dimension approaches are directly applied on the face image. Here, $\boldsymbol{Fractal_{Vector}}$ denotes applying both the FBM and IDBC approach directly on the input image. As shown in Fig.3.19, a higher identification rate was obtained using the proposed *Curvelet-Fractal* approach compared to only applying the Fractal dimension. Furthermore, It was possible to improve further the recognition rate of the *Curvelet-Fractal* approach by learning additional feature representations through the MDFR framework, as well as improve the performance of the DBN by forcing it to learn only the important facial features (e.g. edges and curves).

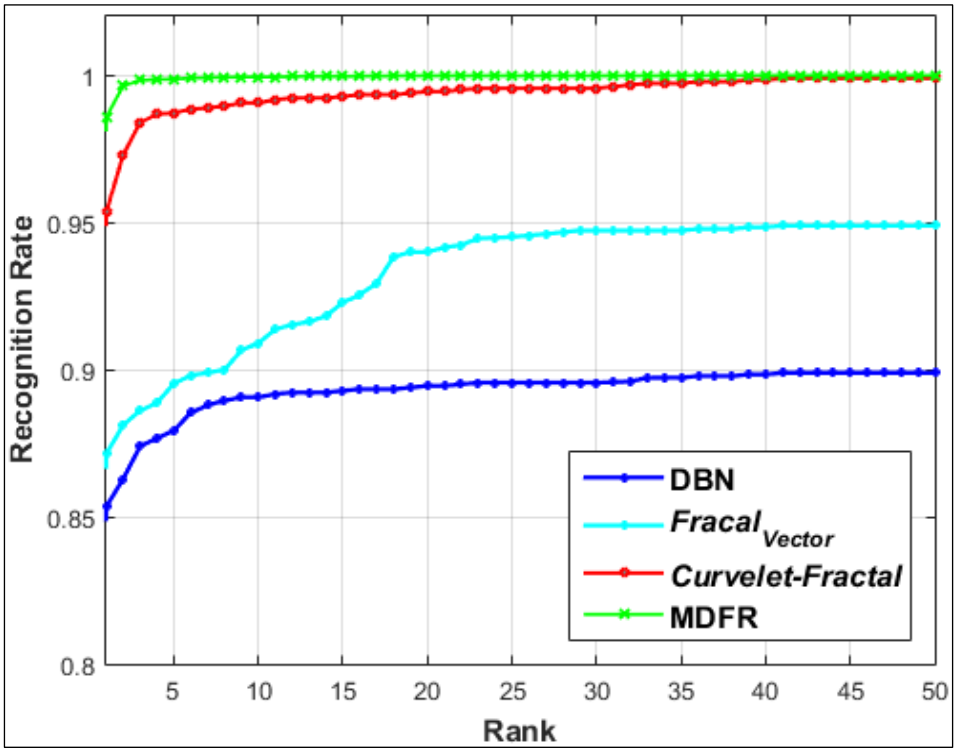


Figure 3.19: Performance comparison between the DBN, *Curvelet-Fractal* and MDFR methods on the SDUMLA-HMT Dataset.

To further examine the robustness of the proposed approaches, a number of experiments were conducted on the FERET and the CAS-PEAL-R1datasets, and the results obtained were compared with the state-of-the-art approaches. For a fair comparison, the performance of the *Curvelet-Fractal* approach was evaluated using the standard evaluation protocols of FERET and CAS-PEAL-R1dataset described in [256][257], respectively. In this work, to prevent

overfitting and increase the generalization ability of the MDFR framework, the data augmentation procedure as described in (Section 3.3.2) was applied only to the gallery set of these two datasets. Then, its performance during the learning process was observed on a separate validation set taken from the full augmented gallery set.

According to the standard evaluation protocol, the FERET dataset is divided into five distinct sets: **Fa** contains a total 1,196 subjects with one image per subject, which is used as a gallery set. The **Fb** contains 1,195 images taken on the same day and under the same lighting conditions as the **Fa** set, but with different facial expressions. The **Fc** set has 194 images taken on the same day as the **Fa** set, but under different lighting conditions. The **Dup.I** set contains 722 images acquired on different days after the **Fa** set. Finally, the **Dup.II** set contains 234 images acquired at least one year after the **Fa** set. Following the standard evaluation protocol, the last four sets are used as probe sets to address the most challenging problems in the face identification task, such as facial expression variation, illumination changes, and facial ageing. Table 3.2 lists the Rank-1 identification rates of the proposed approaches and the state-of-the-art face recognition approaches on all four probe sets of the FERET dataset.

The standard CAS-PEAL-R1evaluation protocol divides the dataset into a gallery set and six frontal probe sets without overlap between the gallery set and any of the probe sets. The gallery set consists of 1,040 images of 1,040 subjects taken under the normal conditions. The six probe sets contain face images with the following basic types of variations: Expression (PE) consists of 1570 images, Accessories (PA) consists of 2285 images, Lighting (PL) consists of 2243 images, Time (PT) consists of 66 images, Background (PB) consists of 553 images, and Distance (PS) consists of 275 images. Table 3.3 lists the Rank-1 identification rates of the proposed approaches and the state-of-the-art face recognition approaches on all six probe sets of the CAS-PEAL-R1dataset.

Table 3.2: The Rank-1 identification rates of different methods on the FERET probe sets.

Approach	Fb	Fc	Dup.I	Dup.II
DLBP-W [260]	99	99	86	85
G-LQP [261]	99.9	100	93.2	91.0
FHOGC [262]	98.3	98.3	86.3	81.2
Groupwise MRF [263]	98.5	98.8	87.7	86.2
H-Groupwise MRF [263]	99.7	99.2	94.7	93.6
LGOP+WPCA [264]	99.2	99.5	89.5	88.5
DFD(S=3)+WPCA [86]	99.3	99	88.8	87.6
DFD(S=5)+WPCA[86]	99.4	100	91.8	92.3
AMF [81]	99.9	100	96.4	93.6
GOM [80]	99.9	100	95.7	93.1
DBN	99.95	100	95.15	93.35
Fractal$_{Vector}$	97.5	96.65	92	90.34
Curvelet-Fractal	**100**	**98.97**	**97.92**	**95.72**
MDFR Framework	**100**	**100**	**98.40**	**97.86**

It can be seen from the results listed in Table 3.2 and 3.3, it was possible to achieve competitive results with the state-of-the-art face identification results on the FERET and the CAS-PEAL-R1datasets, using only the *Curvelet-Fractal* approach. Its performance was compared with popular and recent feature descriptors, such as G-LQP, LBP, WPCA, etc. Although some approaches, such as DFD(S=5)+WPCA [86], GOM [80], AMF [81], and DBN achieved a slightly higher identification rate on the **Fc** probe set, they obtained inferior results on the other probe sets of the FERET dataset. In addition, the *Curvelet-Fractal* approach achieved a higher identification rate on all the probe sets of the CAS-PEAL-R1dataset. Some of the existing approaches, such as H-Groupwise MRF [263] and FHOGC [262] also achieved a 100% identification rate on the **PB** and **PT** probe set, respectively, but they obtained inferior results on the other probe sets of the CAS-PEAL-R1 dataset. Finally, a further improvement and a new state-of-the-art recognition accuracy were achieved using the MDFR framework on the FERET and the CAS-PEAL-R1 datasets. In particular, this was the case when the most challenging probe sets are under

consideration, such as **Dup.I** and **Dup.II** in FERET dataset and **PE**, **PA**, **PL**, and **PS** in the CAS-PEAL-R1dataset.

Table 3.3: The Rank-1 identification rates of different methods on the CAS-PEAL_R1 probe sets.

Approach	PE	PA	PL	PT	PB	PS
RBFNN [265]	84.8	93.4	63.4	96.9	-	-
DT-LBP [266]	98	92	41	-	-	-
DLBP-W [260]	99	92	41	-	-	-
1D-CFA [267]	83.12	74.84	31.43	71.21	98.19	98.55
Groupwise MRF [263]	94.8	90.3	66.9	99.2	98.8	99.5
H-Groupwise MRF [263]	96.4	90.3	66.9	99.8	100	99.6
LGOP+WPCA [264]	99.6	96.8	69.9	-	-	-
DFD(S=3)+WPCA [86]	99	96.9	63.9	-	-	-
DFD(S=5)+WPCA [86]	99.6	96.9	58.9	-	-	-
FHOGC [262]	94.9	90.3	68.7	100	-	-
LBP [268]	92.93	82.58	32.46	-	-	-
DBN	98.93	75.36	80.60	95.45	96.01	97.09
Fractal$_{Vector}$	95.12	92.55	78.01	92.33	95.23	96.03
Curvelet-Fractal	**99.87**	**98.07**	**89.48**	**100**	**100**	**99.64**
MDFR Framework	**100**	**99.43**	**89.92**	**100**	**100**	**100**

3.4.3 Face Verification Experiments

In this section, the robustness and the effectiveness of the proposed approaches were examined to address the unconstrained face verification problem using the LFW dataset. The face images in the LFW dataset were divided into two distinct Views. "View 1" is used for selecting and tuning the parameters of the recognition model, while "View 2" is used to report the final performance of the selected model. In "View 2", the face images are paired into 6,000 pairs, with 3,000 pairs labelled as positive pairs, and the rest as negative pairs. The final performance is reported as described in [258] by calculating the mean accuracy rate ($\hat{\mu}$) and the standard error of the mean accuracy (S_E) over 10-folds cross-validation, with 300 positive and 300 negative image pairs per each fold. For a fair comparison between all face recognition algorithms, the

creators of LFW dataset have pre-defined six evaluation protocols, as described in [269]. In this work, the "*Image-Restricted, Label-Free Outside Data*" protocol is followed, where only the outside data is used to train the MDFR framework. Furthermore, the aligned LFW-a[14] dataset is used, and the face images were resized to (128×128) pixel after the face region has been detected using pre-trained Viola-Jones[15] face detector.

For the proposed *Curvelet-Fractal* approach, the feature representation of each test sample is obtained first, and then the similarity score between each pair of face images is calculated using the CC classifier. In the training phase, the *Curvelet-Fractal* approach does not use any data augmentation or outside data (e.g. creating additional positive/negative pairs from any other source). It just uses the pre-trained Viola-Jones face detector, which has been trained using outside data. The final results over 10-folds are reported, where each of the 10 experiments is completely independent of the others, and the decision threshold of the CC classifier is learnt from the training set according to the standard evaluation protocol. Then, the accuracy rate in each round of 10-folds cross-validation is calculated as the number of correctly classified pairs of samples divided by the total number of test sample pairs. For further evaluation, the results obtained from the *Curvelet-Fractal* approach were compared to state-of-the-art approaches on LFW dataset, such as DDML [102], LBP, Gabor [43], and MSBSIF-SIEDA [270] using the same evaluation protocol (Restricted), as shown in Table 3.4. It can be seen that the accuracy rate, 0.9622 ± 0.0272, of the *Curvelet-Fractal* approach is higher than the best results reported on the LFW dataset, which is 0.9463 ± 0.0095. In this work, further improvements and a new state-of-the-art result were achieved by applying the MDFR framework on the LFW dataset. This experiment can be considered as an examination of the MDFR's generalization ability to address the unconstrained face verification problem on the LFW dataset. In this work, the final performance of two pre-trained DBNs models was evaluated, while the first model was applied directly on top of pixel intensity representations the second was applied on top of local features representations, and referred to the MDFR framework. Following the

[14] http://www.openu.ac.il/home/hassner/data/lfwa/

[15] The incorrect face detection results have been detected manually to ensure that all the subjects are contributed in the subsequent evaluation of the proposed approaches.

same evaluation protocol mentioned above, the hyper-parameters of the MDFR framework were find-tuned using data from the SDUMLA-HMT dataset, as described in (Section 3.4.2.2).

In the MDFR framework, the feature representations, $\mathbf{f_x}$ and $\mathbf{f_y}$, of a pair of two images, $\mathbf{I_x}$ and $\mathbf{I_y}$, are obtained firstly by applying *Curvelet-Fractal* approach, and then a feature vector **F** for this pair is formed using element-wise multiplication ($\mathbf{F} = \mathbf{f_x} \odot \mathbf{f_x}$). Finally, these feature vectors **F** (extracted from pairs of images) are used as input data to the DBN, to learn additional features representations and perform face verification in the last layer. The performance of the MDFR framework is reported over 10-folds, each time one fold was used for testing and the other nine folds for training. For each round of the 10 experiments, the data augmentation procedure was applied for the training set, to avoid overfitting and increase the generalization ability of the network. Table 3.4 lists the mean accuracy of the recent state-of-the-art methods on the LFW dataset, and the corresponding Receiver Operating Characteristic (ROC) curves are shown in Fig. 3.20. Considering the results of the MDFR framework, it is significantly improved over the mean accuracy rate of the *Curvelet-Fractal* approach and the DBN model applied directly on top of pixel intensity representations, by 2.6% and 5.3% respectively. In this work, the performance of the proposed MDFR framework is also compared with several state-of-the-art deep learning approaches, including, DeepFace [93], DeepID [254], ConvNet-RBM [92], ConvolutionalDBN [97] and DDML [102]. The first three approaches were mainly trained using the "*Unrestricted, Labeled Outside Data*" protocol, in which a private dataset consisting of a large number of training images (> 100K), is employed. The accuracy rate has been improved by 1.38%, compared to the next highest results reported by DeepID [254]. These promising results demonstrate the good generalization ability of the MDFR framework, and its feasibility for deployment in real applications.

Table 3.4: Performance comparison between the proposed approaches and the State-of-the-Art approaches on LFW dataset under different evaluation protocols.

Approach	Acc. ($\hat{\mu} \pm S_E$)	Protocol
DeepFace [93]	0.9735 ±0.0025	Unrestricted
DeepID [254]	0.9745 ± 0.0026	Unrestricted
ConvNet-RBM[92]	0.9252 ± 0.0038	Unrestricted
ConvolutionalDBN [97]	0.8777 ± 0.0062	Restricted
DDML [102]	0.9068 ± 0.0141	Restricted
VMRS [271]	0.9110 ± 0.0059	Restricted
HPEN+HD-LBP+ DDML [43]	0.9257 ± 0.0036	Restricted
HPEN+HD-Gabor+ DDML [43]	0.9280 ± 0.0047	Restricted
Sub-SML+Hybrid+LFW3D [272]	0.9165 ± 0.0104	Restricted
MSBSIF-SIEDA [270]	0.9463 ± 0.0095	Restricted
DBN	0.9353 ± 0.0165	Restricted
Curvelet-Fractal	**0.9622 ± 0.0272**	Restricted
MDFR Framework	**0.9883 ± 0.0121**	Restricted

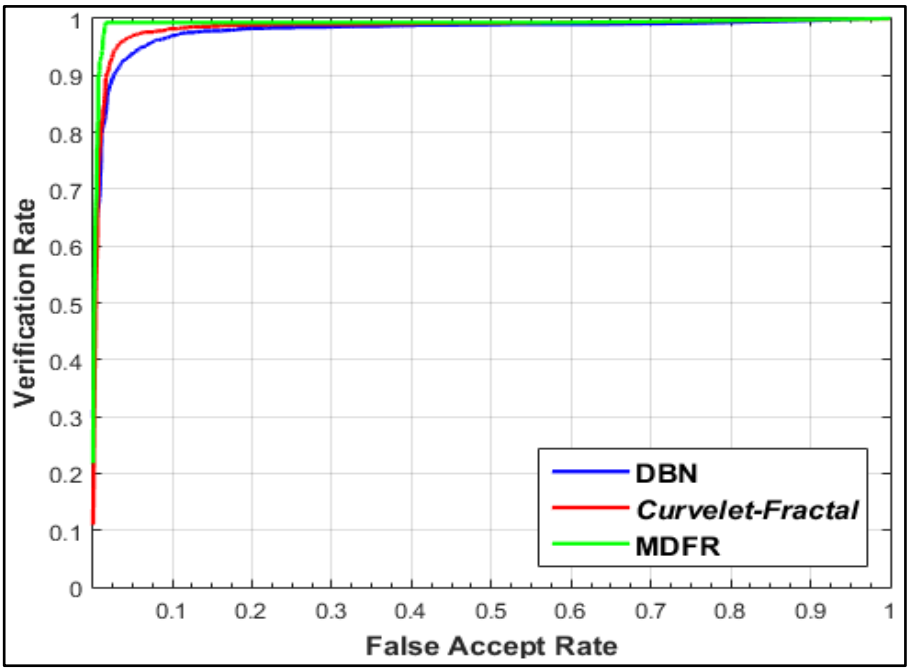


Figure 3.20: ROC curves averaged over 10-folds of "View 2" of the LFW-a dataset. Performance comparison between the DBN, Curvelet-Fractal, and MDFR framework on the face verification task.

3.4.4 Running Time

In this section, the running time of the proposed approaches, including the *Curvelet-Fractal*, DBN, and MDFR framework, was measured by implementing them on a personal computer with the Windows 8 operating system, a 3.60 GHz Core i7-4790 CPU and 24 GB of RAM. The system code was written in MATLAB R2015a and later versions. It should be noted that the running time of the proposed approaches is proportional to the number of subjects and their images in the dataset. The training time using the different datasets is given in Table 3.5. It is clear from the table that the training time of the proposed MDFR framework has significantly reduced the training time of the DBN when it is applied directly on top of the pixel intensity representations. Moreover, the computational efficiency of the proposed MDFR framework can be further improved using GPUs and code optimization. The test time per image, from image input until the recognition decision, for both the *Curvelet-Fractal* approach and MDFR framework, is about 1.3ms and 1.80ms, respectively, which is fast enough to be used for real-time applications.

Table 3.5: The average training time of the proposed approaches using different datasets.

Datasets	DBN	Curvelet-Fractal	MDFR Framework
SDUMLA-HMT	18 Hours & 35 Minutes	35 Minutes	4 Hours & 15 Minutes
FERET	16 Hours & 45 Minutes	17 Minutes	3 Hours & 33 Minutes
CAS-PEAL-R1	15 Hours & 27 Minutes	14 Minutes	3 Hours & 32 Minutes
LFW-a	13 Hours & 41 Minutes	7 Minutes	2 Hours & 56 Minutes

3.5 Chapter Summary

In this chapter, a novel multimodal local feature extraction approach is proposed, based on merging the advantages of multidirectional and anisotropy transforms, such as the Curvelet transform with Fractal dimension, termed the *Curvelet-Fractal* approach. Using this approach, effective local facial features (e.g., the edges and curves) were efficiently captured by the Curvelet transform, while the Fractal dimension was employed to produce an illumination invariant representation of the face image, and to describe the face texture roughness and fluctuations in the surface efficiently, which play a significant role in increasing the recognition power of the proposed system. In this work, Curvelet coefficients in different frequency bands have been processed separately using two different Fractal dimension methods. The FBM method was used to process only a Coarse band of the Curvelet transform, while the IDBC method was used to handle the high dimensionality of other adopted Curvelet sub-bands. Furthermore, a novel framework is proposed, termed the Multimodal Deep Face Recognition (MDFR) framework, to learn additional and complementary representations by training a DBN on top of existing local representations (*Curvelet-Fractal*'s representations), instead of the pixel intensity representations. It has been observed that more discriminating features can be learned with less training time required to obtain the final trained model, by feeding the DBN approach with only the usefulness features, instead of with raw data. Finally, comprehensive experiments were conducted using the proposed approaches for both face identification and verification tasks, on four large-scale unconstrained face datasets (e.g. SDUMLA-HMT, FERET, CAS-PEAL-R1 and LFW dataset) with high variation in facial expressions, lighting conditions, noise, etc. It was possible to achieve competitive results with the state-of-the-art approaches (e.g. G-LQP, LBP, WPCA, etc.) using only the *Curvelet-Fractal* approach. Furthermore, new state-of-the-art accuracy rates were achieved by applying the proposed MDFR framework on all the employed datasets.

Chapter 4
Iris Recognition System

4.1 Introduction

This chapter describes the iris recognition approaches proposed in this PhD thesis. Iris recognition is considered as the most accurate and highly reliable biometric system for personal verification and/or identification [273]. Several studies have demonstrated that the iris trait has a number of advantages over other biometric traits (e.g. face, fingerprint, etc.), which make it commonly accepted for application in high reliability and accurate biometric systems. Firstly, the iris trait represents a highly protected part of the human eye; it is located between the black pupil and the white sclera. This position makes the iris trait completely isolated and protected from external environmental conditions [274][275]. Secondly, it is believed that the iris texture provides a very high degree of uniqueness and randomness, so it very unlikely for any two iris patterns to be the same, even irises from identical twins, or from the right and left eyes of an individual person. This complexity in iris patterns is due to the distinctiveness and richness of the texture details within the iris region, including rings, ridges, crypts, furrows, freckles, zigzag patterns, etc. [276]. Thirdly, the iris trait provides a high degree of stability during a person's lifetime from one year of age until death. Finally, it is considered the most secure biometric trait against fraudulent methods and spoofing attacks by an imposter, where any attempt to change its patterns, even with a surgery, is a high risk, unlike the fingerprint trait which is relatively easier to tamper with [277]. Despite these advantages, implementing an iris recognition system is considered a challenging problem due to the iris acquisition process possibly acquiring irrelevant parts, such as eyelids, eyelashes, eyelashes, pupil and specular reflections, which may greatly influence the iris segmentation and recognition outcomes.

Typically, the iris recognition system consists of four main stages: image acquisition, pre-processing, feature extraction and classification. The image

acquisition stage employs an imaging device (e.g., LG2200 camera) to capture the iris image of high-resolution and sharpness [156]. The image pre-processing stage can be divided further into three sub-steps: iris localization, iris normalization and iris enhancement.

Iris localization refers also to the iris segmentation, which is a crucial step to build an efficient, robust and real-time iris recognition system. The main aim of the iris localization is to detect the iris region from the background and all surrounding features, such as sclera, pupil, eyelashes, eyelids, eyebrows and specular reflections. Accurate iris localization can be achieved by delimiting the inner border between the pupil and iris and the outer border between the iris and sclera. However, the task becomes more challenging, due to the portions of the iris region occluded by eyelids and eyelashes. In addition, the small difference of the intensity levels between the iris and the sclera regions makes detecting the outer boundary more difficult compared to the detection of the inner boundary. Finally, varying illumination conditions that can affect the appearance of the iris patterns could greatly influence the iris segmentation outcome [278][279]. In this chapter, an efficient and automatic method is proposed for the inner and outer iris boundary localization, which addresses the main concerns in many previous methods, for instance, the computational cost and high accuracy. Firstly, the pupil boundary is detected after eliminating the specular reflections using a simple thresholding technique with the morphological operations. Then, the outer iris boundary is detected using the generated edge map along with the Coherent Circular Hough Transform (CCHT). At this stage, an efficient enhancement procedure is proposed to enhance the iris boundary by applying 2D Gaussian filter and Histogram equalization processes. Finally, robust and fast eyelids detection algorithm is developed by employing an anisotropic diffusion filter with Radon transform to fit the upper and lower eyelids boundaries.

After the pre-processing stage, the most discriminative feature representations of the iris region are extracted and used for establishing the person's identity in the classification stage. In this chapter, two discriminative learning techniques are proposed, based on the combination of a Convolutional Neural Networks (CNN) and the Softmax classifier as a multinomial logistic

regression classifier. The trained deep learning system proposed is called the *IrisConvNet* system. CNNs are efficient and powerful DNNs which are widely applied in image processing and pattern recognition with the ability to automatically extract distinctive features from input images, even without a pre-processing step. Moreover, CNNs have a number of advantages compared to other DNNs, such as fast convergence, simpler architecture, adaptability and fewer free parameters. In addition, CNNs are invariant to image deformations, such as translation, rotation, and scaling [280]. The Softmax classifier is a discriminative classifier widely used for multi-class classification purposes. It was chosen for use on top of the CNN because it has produced outstanding results compared to other popular classifiers, such as Support Vector Machines (SVMs), in terms of accuracy and speed [281]. In this work, the efficiency and learning capability of the proposed techniques are investigated by employing a training methodology based on the Back-propagation algorithm with the mini-batch Adagrad optimization method. In addition, other training strategies are also used, including dropout and data augmentation to prevent the overfitting problem and increase the generalization ability of the neural network [282][283], as will be explained later on. The performance of the proposed iris recognition system has been tested on five public datasets collected under different conditions: SDUMLA-HMT, CASIA-Iris-V1, CASIA-Iris-V3 Interval, MMU1 and IITD iris datasets. The results obtained from the proposed system outperform other state-of-the-art of approaches (e.g. Wavelet Transform, Scattering Transform, LBP and PCA) by achieving new state-of-the-art Rank-1 identification rates on all the employed datasets and a recognition time less than one second per person.

The remainder of the chapter is organized as follows: Section 4.2 provides an overview of the proposed approaches for iris localization, feature extraction and pattern matching processes. The implementation details of the proposed iris recognition system including the proposed iris localization and *IrisConvNet* system are presented in Section 4.3. The experimental results are presented in Section 4.4. Finally, the summary and conclusions of this chapter are stated in the last section.

4.2 Methodology Overview

In this section, since the proposed iris localization system is based on the CCHT and Radon transform, they are briefly reviewed here. Moreover, a brief description of the proposed deep learning approach is given, which incorporates two discriminative learning techniques: a CNN and a Softmax classifier. The main aim here is to inspect their internal structures, and identify their strengths and weaknesses to enable the proposal of an iris recognition system that integrates the strengths of these proposed techniques.

4.2.1 Coherent Circular Hough Transform

The standard Circular Hough transform (CHT) works on detecting the circular shapes of a given radius within the image. The edge map of the eye image is generated by computing the first derivatives of intensity values. Each point in the edge map donates a circle of radius ***r*** and centre **(x_c, y_c)** to an output accumulator array. Then, a voting procedure is used to find the largest peak in the resulting accumulator array in the parameter space, which corresponds to the circle best defined by the edge points [120]. Considering the obtained set of edge points as **(x_i, y_i), i = (1,2, . . .,n)**, the CHT for fitting a circle with radius ***r*** and centre coordinates **(x_c, y_c)** is defined as follows:

$$H(x_c, y_c, r) = \sum_{i=1}^{n} h(x_i, y_i, x_c, y_c, r) \tag{4.1}$$

where

$$h(x_i, y_i, x_c, y_c, r) = \begin{cases} 1, & if \quad g(x_i, y_i, x_c, y_c, r) = 0 \\ 0, & otherwise. \end{cases} \tag{4.2}$$

And the parametric function **g** is defined as follows:

$$g(x_i, y_i, x_c, y_c, r) = (x_i - x_c)^2 + (y_i - y_c)^2 - r^2 \tag{4.3}$$

The value of the function **(g)** of each point is equal to 1, if it is located over the circle with parameters **(x_c, y_c, r)** by Eq.(4.2). Finally, the voting procedure is implemented in the Hough space to detect the correct circle. A number of modifications have been applied to the CHT to either improve the localization

rate or to decrease its computational complexity. In this work, the CCHT developed by Atherton and Kerbyson [284] is employed to improve the inner and outer boundary localization rate and to reduce the execution time, using a complex phase coding to produce a complex accumulator array. The word complex here refers to the real and imaginary numbers in the generated accumulator array. In the CCHT, the phase is used to code for the radius of the circle, where the phase is relative to the distance travelled far from the edge points. After projecting the edge points along a line in the direction of the edge orientation, their complex values are added to the accumulator array. The projections of the edge points lying on the same circle will be intersected at a common point and within the same phase, which means they have the same distance from this intersection point. Edge projections are usually not lying on a circle border (e.g. noise edge points), instead, they will be associated with a random phase and tend to be discarded. In other words, the contribution of each edge point to the accumulator array is within the phase, if this point is lying on the circle and out of the phase elsewhere, as shown in Fig. 4.1. Finally, the circle parameters are obtained where the centre coordinates are represented by the positions of the peaks in the magnitude of the accumulator array, while the radius is represented by the phase at the peak position. More details on CCHT can be found in [284].

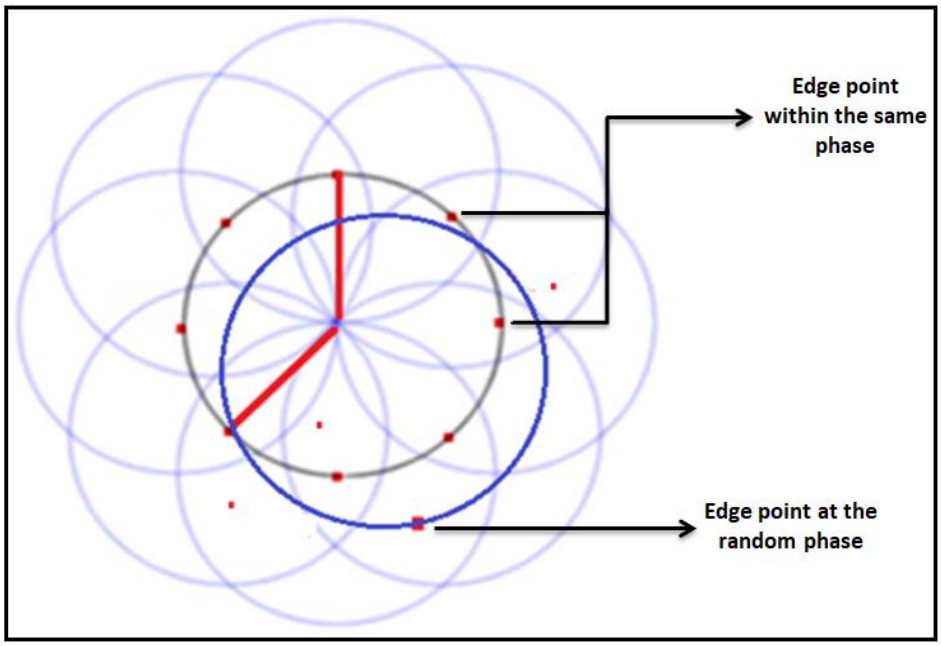


Figure 4.1: The edge points and accumulator space of the CCHT approach.

4.2.2 Radon Transform

In 1917, an integral transform was introduced by the Austrian mathematician named Johann Radon. The Radon transform, as it became known, is now an important mathematical tool that is widely used in many medical imaging, geophysical imaging and radar imaging applications [285]. The Radon transform is a set of 1D projections of a given 2D function ***f*(*x, y*)**. In this case, the given 2D function is an image, and the projection is a set of line integrals. These line integrals are computed from multiple resources by the Radon transform in a certain angle. Generally, the Radon transform of the 2D function ***f*(*x, y*)** is the line integral of ***f*** at an angle **θ**, that is parallel to the ***y´***-axis defined as follows:

$$\boldsymbol{R_\theta}\ (\boldsymbol{x'}) = \int_{-\infty}^{\infty} \boldsymbol{f}(\boldsymbol{x'}\ \boldsymbol{cos}\ \boldsymbol{\theta} - \ \boldsymbol{y'}\ \boldsymbol{sin}\ \boldsymbol{\theta},\ \boldsymbol{x'}\ \boldsymbol{sin}\ \boldsymbol{\theta} + \boldsymbol{y'}\ \boldsymbol{cos}\ \boldsymbol{\theta})\boldsymbol{dy'} \qquad (\mathbf{4.4})$$

Here $\boldsymbol{x'}$ and $\boldsymbol{y'}$ are computed as follows:

$$\begin{bmatrix} \boldsymbol{x'} \\ \boldsymbol{y'} \end{bmatrix} = \begin{bmatrix} \boldsymbol{cos}\ \boldsymbol{\theta} & \boldsymbol{sin}\ \boldsymbol{\theta} \\ -\boldsymbol{sin}\ \boldsymbol{\theta} & \boldsymbol{cos}\ \boldsymbol{\theta} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix} \qquad (\mathbf{4.5})$$

The Radon transform and the Hough transform are similar in the sense they map the data points from the image space to a parameter space. However, they differ in their mapping form. The Radon transform derives a point in parameter space from image space, and this mapping is known as the reading paradigm. The Hough transform, on the other hand, depends on the writing paradigm in mapping the data points from image space to the parameter space. Moreover, unlike the Hough transforms, the Radon transform has a well-formed mathematical construct [286]. In this work, the Radon transform is employed to detect the parts of the upper and lower eyelids' boundaries that can be modelled as a line without losing important information from the iris region.

4.2.3 Convolutional Neural Network

As is well known, the success of any biometric system defined as a classification and recognition system mainly depends on the efficiency and robustness of the feature extraction and classification stages. In the literature, several publications have documented the high accuracy and reliability of neural networks, such as the MLP, in many real-world pattern recognition and

classification applications [287][288]. Inspired by a number of characteristics of such systems (e.g. a powerful mathematical model, the ability to learn from experience and robustness in handling noisy images), neural networks are considered as one of the simplest and powerful of classifiers [289]. However, traditional neural networks have a number of drawbacks and obstacles that need to be overcome. Firstly, the input image is required to undergo several different image processing stages, such as image enhancement, image segmentation and feature extraction, to reduce the size of the input data and achieve a satisfactory performance. Secondly, designing a handcrafted feature extractor needs a good domain-knowledge and a significant amount of time. Thirdly, an MLP has difficulty in handling deformations of the input image, such as translations, scaling and rotation [216]. Finally, a large number of free parameters need to be tuned in order to achieve satisfactory results while avoiding the overfitting problem. The large number of these free parameters is due to the use of full connections between the neurons in a specific layer, and all activations in the previous layer [214]. To overcome these limitations and drawbacks the use of deep learning techniques has been proposed. Deep learning can be viewed as an advanced subfield of machine learning techniques that depend on learning high-level representations and abstractions, using a structure composed of multiple non-linear transformations. In deep learning, the hierarchy of automatically learning features at multiple levels of representations can provide a good understanding of data, such as image, text and audio, without depending completely on any domain-knowledge and handcrafted features [280]. In the last decade, deep learning has attracted much attention from research teams with promising and outstanding results in several areas, such as Natural Language Processing (NLP) [290], texture classification [291], object recognition [283], face recognition [292], speech recognition [293], information retrieval [294], traffic sign classification [295], etc.

A CNN is a feed-forward multi-layer neural network, which differs from traditional fully-connected neural networks by combining a number of locally-connected layers aimed at automated feature recognition, followed by a number of fully-connected layers aimed at classification [242]. The CNN architecture, as illustrated in Fig. 4.2, comprises several distinct layers including sets of locally-

connected convolutional layers (with a specific number of different learnable kernels in each layer), sub-sampling layers named pooling layers, and one or more fully-connected layers. The internal structure of the CNN combines three architectural concepts, which make the CNN successful in different fields, such as image processing, pattern recognition, speech recognition, and NLP. The first concept is applied in both convolutional and pooling layers, in which each neuron receives input from a small region of the previous layer called the local receptive field equals in size to a convolution kernel [289]. This local connectivity scheme ensures that the trained CNN produces strong responses to capture local dependencies and extracts elementary features in the input image (e.g., edges, ridges, curves, etc.) which can play a significant role in maximizing the inter-class variations and minimizing the intra-class variations, and hence increasing the CRR of the iris recognition system. Secondly, the convolutional layer applies the sharing parameters (weights) scheme in order to control the model capacity and reduce its complexity. At this point, a form of translational invariance is obtained using the same convolution kernel to detect a specific feature at different locations in the iris image [296]. Finally, the non-linear down sampling applied in the pooling layers reduces the spatial size of the convolutional layer's output and reduces the number of the free parameters of the model. Together, these characteristics make the CNN very robust and efficient at handling image deformations and other geometric transformations, such as translation, rotation and scaling [242]. In more detail these layers are:

- **Convolutional layer:** In this layer, the parameters (weights) consist of a set of learnable kernels that are randomly generated and learned by the Back-propagation algorithm. These kernels have a few local connections, but connect through the full depth of the previous layer. The result of each kernel convolved across the whole input image is called the activation (or feature) map, and the number of the activation maps is equal to the number of applied kernels in that layer. Fig. 4.2. shows a first convolution layer consisting of 6 activation maps stacked together and produced from 6 kernels independently convolved across the whole input image. Hence, each activation map is a grid of neurons that share the same parameters. The activation map of the convolutional layer is defined as:

$$y^{j(r)} = max\left(0, b^{j(r)} + \sum_{i} k^{ij(r)} * x^{i(r)}\right) \quad (4.6)$$

Here, $x^{i(r)}$ and $y^{j(r)}$ are the ***i-th*** input and the ***j-th*** output activation map, respectively. $b^{j(r)}$ is the bias of the ***j-th*** output map and $(*)$ denotes convolution. $k^{ij(r)}$ is the convolution kernel between the ***i-th*** input map and the ***j-th*** output map. The ReLU activation function ***(y =max (0, x))*** is used here to add non-linearity to the network, as will be explained later on.

- **Max-Pooling layer:** Its main function is to reduce the spatial size of the convolutional layers' output representations, and it produces a limited form of the translational invariance. Once a specific feature has been detected by the convolutional layer, only its approximate location relative to other features is kept. As shown in Fig. 4.2, each depth slice of the input volume (convolutional layer's output) is divided into non-overlapping regions, and for each sub-region, the maximum value is taken. A commonly used form is max-pooling with regions of size (**2×2**) pixels and a stride of 2. The depth dimension of the input volume is kept unchanged. The max-pooling layer can be formulated as follows:

$$y^{i}_{j,k} = \max_{0 \le m,n < s}\left(x^{i}_{j.s+m,k.s+n}\right) \quad (4.7)$$

Here, $y^{i}_{j,k}$ represents a neuron in the ***i-th*** output activation map, which is computed over an (***s×s***) non-overlapping local region in the ***i-th*** input map $x^{i}_{j,k}$.

- **Fully-connected layers:** the output of the last convolutional or max-pooling layer is fed to one or more fully-connected layers as in a traditional neural network. In those layers, the outputs of all neurons in layer **(*l-1*)** are fully-connected to every neuron in layer ***l***. The output $y^{(l)}(j)$ of neuron ***j*** in a fully-connected layer ***l*** is defined as follows:

$$y^{(l)}(j) = f^{(l)}\left(\sum_{i=1}^{N^{(l-1)}} y^{(l-1)}(i).\ w^{(l)}(i,j) + b^{(l)}(j)\right) \quad (4.8)$$

where $N^{(l-1)}$ is the number of neurons in the previous layer **(*l-1*)**, $w^{(l)}(i,j)$ is the weight for the connection from neuron ***j*** in layer **(*l-1*)** to neuron ***j*** in layer

l, and $b^{(l)}(j)$ is the bias of neuron j in layer l. As for the other two layers, $f^{(l)}$ represents the activation function of layer l.

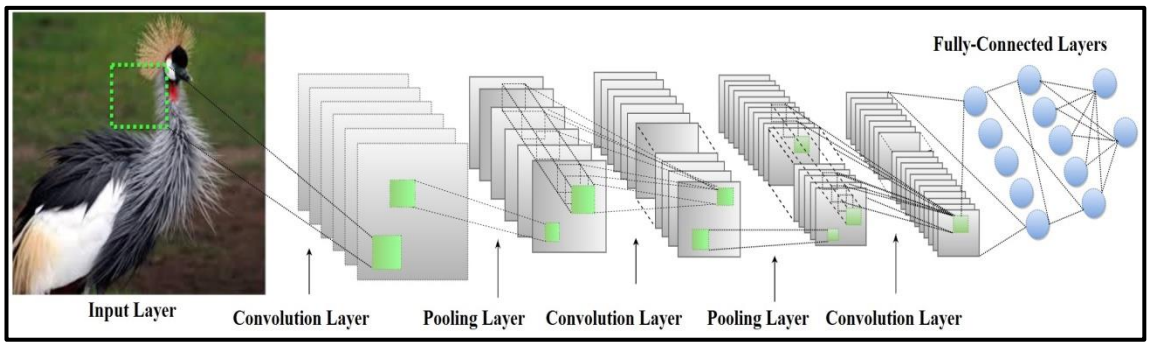


Figure 4.2: An illustration of the CNN architecture, where the grey and green squares refer to the activation maps and the learnable convolution kernels, respectively. The cross lines between the last two layers refer to the fully-connected neurons.

4.2.4 Softmax Regression Classifier

The classifier implemented in the fully-connected part of the system, shown in Fig.4.2, is the Softmax regression classifier, which is a generalized form of binary Logistic Regression classifier intended to handle multi-class classification tasks. Suppose that there are K classes and n labelled training samples $\{(x_1, y_1),\cdots, (x_n, y_k)\}$, where $x_i \in R^m$ is the *i-th* training example and $y_i \in \{1,\cdots,K\}$ is the class label of x_i. Then, for a given test input x_i, the Softmax classifier will produce a K-dimensional vector (whose elements sum to **1**), where each element in the output vector refers to the estimated probability of each class label conditioned on this input feature. The hypothesis, $h_\theta(x_i)$, to estimate the probability vector of each label, can be defined as follows:

$$h_\theta(x_i) = \begin{bmatrix} p(y_i = 1|\, x_i;\ \theta) \\ p(y_i = 2|\, x_i;\ \theta) \\ \vdots \\ p(y_i = K|\, x_i;\ \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{K} e^{\theta_j^T x_i}} \begin{bmatrix} e^{\theta_1^T x_i} \\ e^{\theta_2^T x_i} \\ \vdots \\ e^{\theta_K^T x_i} \end{bmatrix} \qquad (4.9)$$

Here, $(\theta_1, \theta_2, \dots, \theta_K)$ are the parameters to be randomly-generated and learned by the Back-propagation algorithm. The cost function used for the Softmax classifier is called as the cross-entropy loss function and can be defined as follows:

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{K} 1\{y_i = j\} log \frac{e_{j^{x_i}}^{T}}{\sum_{l=1}^{K} e_{l^{x_i}}^{T}}\right] + \frac{\lambda}{2}\sum_{i=1}^{K}\sum_{j=0}^{n}\theta_{ij}^{2} \qquad (4.10)$$

Here, **1{·}** is a logical function, that is, when a true statement is given, **1{·}= 1**, otherwise, **1{·}= 0**. The second term is a weight decay term that tends to reduce the magnitude of the weights, and prevents the overfitting problem. Finally, the gradient descent method is used to solve the minimum of the $J(\theta)$, as follows:

$$\nabla_{\theta_j} J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[x_i\left(1\{y_i = j\} - p(y_i = j|\, x_i;\, \theta)\right)\right] + \lambda\theta_j \qquad (4.11)$$

In Eq. (4.10), the gradients are computed for a single class j, and for each iteration, the parameters will be updated for any given training pair **(x_i, y_i)**, as follows: $\theta^{new} = \theta^{old} - \alpha\nabla_{\theta}J(\theta)$, where the symbol α refers to the learning rate [297].

4.3 The Proposed Iris Recognition System

An overview of the proposed iris recognition system is shown in Fig. 4.3. Firstly, a pre-processing procedure is implemented, based on employing an efficient and automatic iris localization to carefully detect the iris region from the background and all extraneous features, such as pupil, sclera, eyelids, eyelashes, and specular reflections. In this work, the main reason for defining the iris area as the input to CNN instead of the whole eye image is to reduce the computational complexity of the CNN. Another reason is to avoid the performance degradation of the feature extraction and matching processes resulting from the appearance of eyelids and eyelashes. After detection, the iris region is transformed into a normalized form with fixed dimensions, in order to allow direct comparison between two iris images with initially different sizes.

The normalized iris image is further used to provide robust and distinctive iris features by employing the CNN as an automatic feature extractor. Then, the matching score is obtained using the generated feature vectors from the last fully-connected layer as the input to the Softmax classifier. Finally, the matching scores from either the right or left iris images are used to establish the identity

of the person whose iris images are under investigation. During the training phase, different CNN configurations are trained on the training set and tested on the validation set, to obtain the best one with the smallest error that is called the *IrisConvNet* system. Its performance on test data is then assessed in the testing phase.

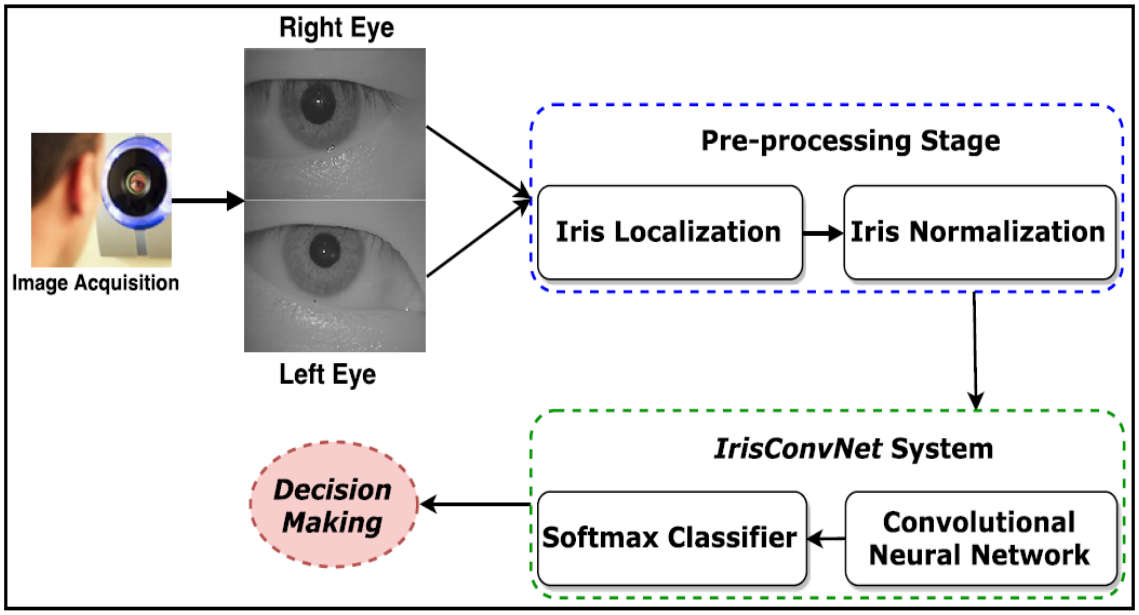


Figure 4.3: Overall stages of the proposed iris recognition system.

4.3.1 The Proposed Iris Localization System

As mentioned previously, accurate iris localization plays a significant role in improving the accuracy and reliability of an iris recognition system, as the performance of the following stages of the system directly depends on the quality of the detected iris region. The iris localization procedure aims to detect the two iris region boundaries: the inner (pupil-iris) boundary, and the outer (iris-sclera) boundary. However, the task becomes more difficult, when parts of the iris are covered by eyelids and eyelashes. In addition, changes in the lighting conditions during the acquisition process can affect the quality of the extracted iris region, and then affect the iris localization and the recognition outcome. As depicted in Fig. 4.4, the proposed iris localization method can be divided broadly into four stages: specular reflection removal, pupil localization, iris localization, and eyelid and eyelash detection.

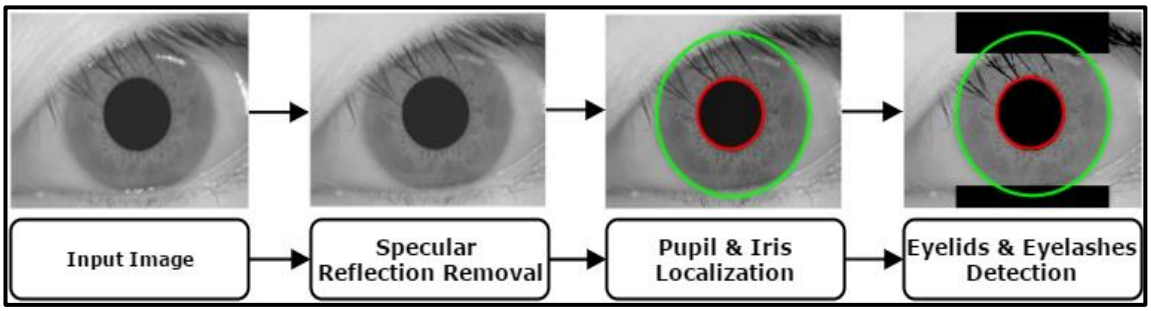


Figure 4.4: Overall stages of the proposed iris localization procedure.

4.3.1.1 Specular Reflection Removal

One of the major issues in the iris segmentation and recognition system is the presence of reflection spots, which can result in inaccurate iris localization and thereby lead to poor iris recognition system performance. In this case, these spots need to be detected and isolated to get a clean iris pattern that can contribute to increasing the recognition rate. In this work, a reflection mask (R_m) is calculated to detect and remove the reflection spots along the eye image in two steps. Firstly, an adaptive threshold is applied, and, if the intensity of the pixel $I(x, y)$ is greater than the determined threshold (T), then it will be considered as a reflection noise.

$$R_m(x,y) = \begin{cases} 1, & I(x,y) < T \\ 0, & otherwise \end{cases} \qquad (4.12)$$

Secondly, a morphological dilation operation is applied using a square shaped structure element whose width is **5** pixels for perfect detection and elimination of the reflection spots. Finally, the specular reflection spots in the eye image are painted using a defined reflection mask and a *roifill* Matlab function. As shown in Fig. 4.5, the specular reflections spots in the eye image have been removed efficiently; thereby this could contribute to improving the quality of iris patterns and the performance of the iris localization method.

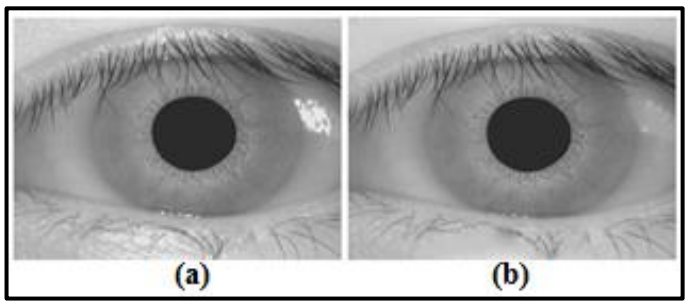


Figure 4.5: Specular reflection removal stage: (a) The original eye image and (b) The output image.

4.3.1.2 Pupil Localization

The inner boundary of the iris region is detected before the outer boundary, due to the fact that the pupil region is the darkest region in the eye image, and can be detected easily. Moreover, this can contribute to improving the accuracy and the speed of detecting the outer boundary, as will be explained later on. The pupil localization is carried out by transforming the grey-scale eye image into a binary image using the Hysteresis thresholding method. In this method, all pixels that have values above upper threshold, T_{up}, are marked as edge points. In addition, all the adjacent pixels to these edge points with values greater than lower threshold, T_{low}, are marked as edge points as well. In this method, eight connectivity is used to detect connected regions to each edge point. As shown in Fig. 4.6 (b), there could exist some noise present in the binary image, due to other dark regions, such as eyelashes and eyelids. A morphological erosion operation using a disk shaped structure element of 1-pixel radius is applied to eliminate such noise. This is followed by discarding all the connected components smaller than 80 pixels in order to produce the final binary image. As shown in Fig. 4.6 (c), the pupil region is almost completely detected in the binary image. Since it can be modelled as a circle, the pupil is detected correctly by employing the CCHT to obtain the centre coordinates and radius of the pupil circle. Fig. 4.6 (d) shows the detected inner boundary of the iris region.

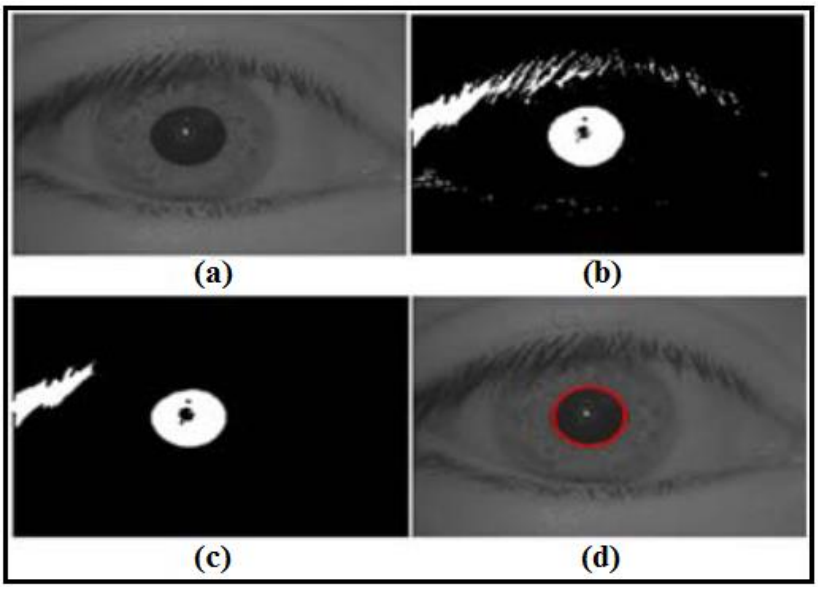


Figure 4.6: Pupil localization stage: (a) The input image, (b) Applying the Hysteresis thresholding method, (c) The output of the morphological operation, and (d) The localized pupil boundary.

4.3.1.3 Iris Localization

The most challenging stage of the iris region localization model is detecting the outer boundary between the iris and sclera, for several reasons. Firstly, no clearly defined border exists between the iris and sclera because of the low variation in intensity between them. Secondly, the upper part of the iris is occluded by the upper eyelid and eyelashes in most cases. In this work, the edge map is generated using a Canny edge detector, followed by the CCHT, to localize the iris outer boundary. Each edge point in the edge map casts a vote in the Hough transform space and produces a circle of radius ***r*** to the output accumulator array. The circle with the highest number of votes is then chosen. However, the main concern here is that the Hough transform is computationally expensive, and increasing the number of unnecessary edge points can result in inaccurate iris localization. Therefore, to reduce the search space, the eye image needs to be enhanced before and after generating the edge map. In this work, a 2D Gaussian filter is employed, in order to smooth the eye image and reduce noise, as shown in Fig. 4.7 (b). The Gaussian filter is a low-pass filter, whose 2D filter coefficients are computed as follows:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (4.13)$$

Here, x and y are the distances from the origin along the horizontal and vertical axes, and $\sigma = 0.8$ is the standard deviation of the Gaussian distribution. After that, the Histogram equalization is used to enhance the contrast between the iris and sclera region, and to make detecting the iris outer boundary easier, as shown in Fig. 4.7 (c). In the next step, an edge map of the eye image is generated by applying the Canny edge detector. As can be seen in Fig. 4.7 (d), there are a lot of unnecessary edge points that need to be identified and eliminated to improve the performance of the CCHT. In this work, based on prior knowledge from the literature that the centre of the iris and pupil are relatively close, and to the established fact in [298][299] that the ratio of iris radius/pupil falls between 4 and 1.75, a circle using the pupil centre is drawn and the pixels within the circle are set to zero. The radius of the drawn circle was selected empirically to be $(r_{pupil}+6)$ for all the employed datasets. This procedure eliminates edge points within the iris region that represent the pupil boundary and iris tissue without the risk of affecting the iris outer boundary. Then, all the connected components smaller than 99 pixels are excluded, as shown in Fig. 4.7 (e). These parameters are adjusted only once for all the datasets considered. Finally, the centre coordinates and radius of the iris circle are obtained by applying the CCHT, as shown in Fig. 4.7 (f).

4.3.1.4 Eyelids and Eyelashes Detection

This stage of detecting the eyelids boundaries and the eyelashes has a significant role in either improving or degrading the performance of the iris localization model and the iris recognition system. Therefore, these artefacts need to be identified and eliminated to obtain a clean iris template. However, detecting the eyelids boundaries is a challenging problem as they can be occluded by the eyelashes in many cases. A number of eyelids detection algorithms have been proposed in the literature, as described in [300], where the boundary between the eyelids and iris is modeled as a parabolic arc or line. In this work, an efficient and robust algorithm is proposed by fitting these boundaries as straight lines, where the line form has fewer parameters than the parabolic arc, and thereby requires less processing time.

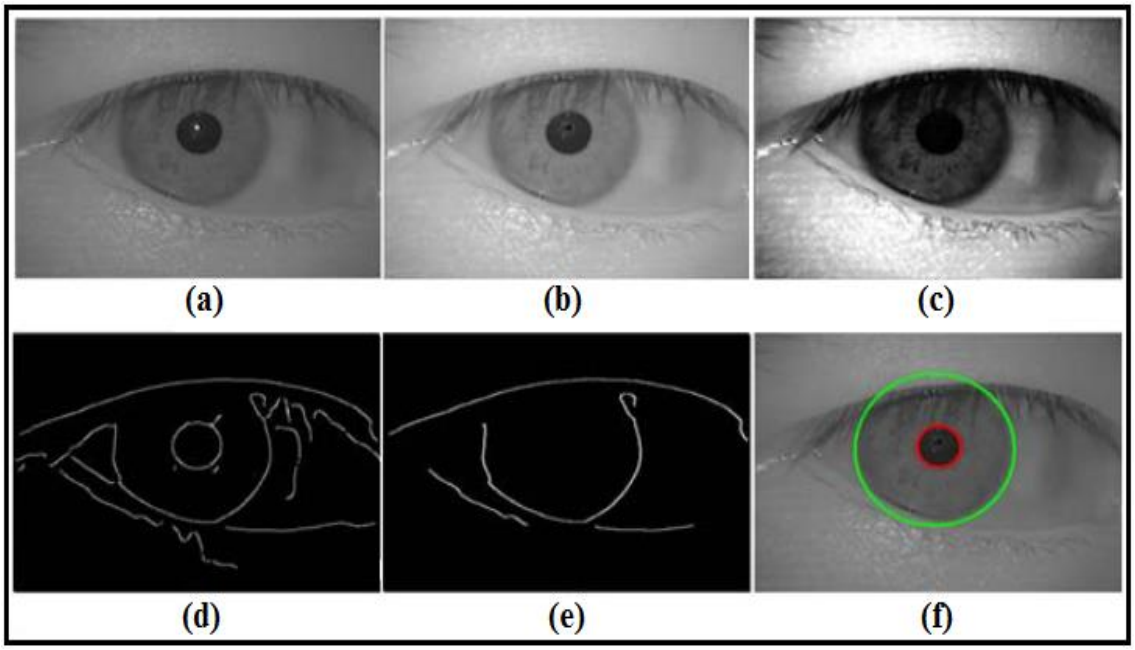


Figure 4.7: Iris localization stage: (a) The input image, (b) Applying the 2D Gaussian filter, (c) Histogram equalization is employed, (d) Generated edge map, (e) Further processing for removing noise, and (f) The final detected iris boundaries.

Firstly, an anisotropic diffusion filter [301], is applied on the eye image in order to enhance the eyelids boundaries and reduce the eyelashes effect, as shown in Fig. 4.8 (b). The anisotropic diffusion filter can be defined as follows:

$$I_{i,j}^{t+1} = I_{i,j}^{t} + \lambda[c_N \,.\, \nabla_N I + c_S \,.\, \nabla_S I + c_E \,.\, \nabla_E I + c_W \,.\, \nabla_W I]_{i,j}^{t} \qquad (4.14)$$

Here, I is the eye image, $0 \leq \lambda \leq 1/4$, c refers to the conduction coefficients that updated each iteration as a function of the brightness gradient, t refers to the iteration index, and ∇ points to the nearest neighbour differences in all the directions N,S,E and W, as follows:

$$\nabla_N I_{i,j} = I_{i-1,j} - I_{i,j}$$

$$\nabla_S I_{i,j} = I_{i+1,j} - I_{i,j}$$

$$\nabla_E I_{i,j} = I_{i,j+1} - I_{i,j}$$

$$\nabla_W I_{i,j} = I_{i,j-1} - I_{i,j} \qquad (4.15)$$

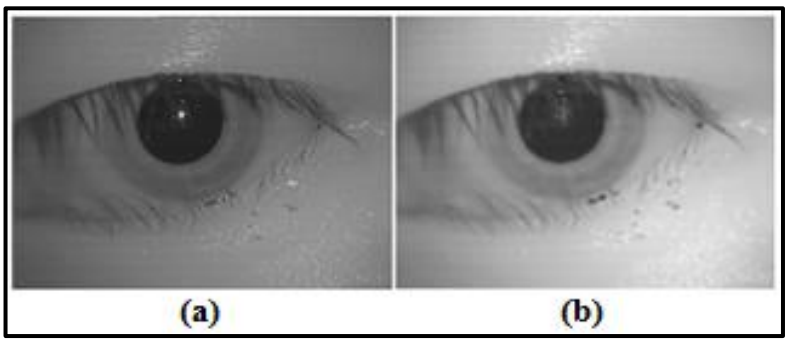


Figure 4.8: (a) An occluded eye image and (b) the output of the anisotropic diffusion filter.

In this work, $\lambda = \mathbf{0.21}$ and ***t*=6**. Secondly, the search time for the Radon transform is reduced by extracting the iris and dividing it into upper and lower parts, as shown in Fig. 4.9 (a) and (b). This is followed by generating the edge map of each part using a modified Canny edge detector where only the horizontal edges are considered, as shown in Fig. 4.9 (c). Finally, the Radon transform is implemented to fit the upper and lower eyelids boundaries. In the proposed method, there is no line to fit, if the maximum of the Radon transform space is less than the pre-defined threshold, which means that the iris region is not occluded by the eyelids. Generally, the eyelashes can be divided into types: separable and multiple eyelashes that grow along the eyelid boundaries. In addition, it is observed that the eyelashes have a lower intensity value than the iris region. In this work, a simple thresholding technique is employed to detect eyelashes carefully to avoid remove important iris information, as shown in Fig. 4.9 (d).

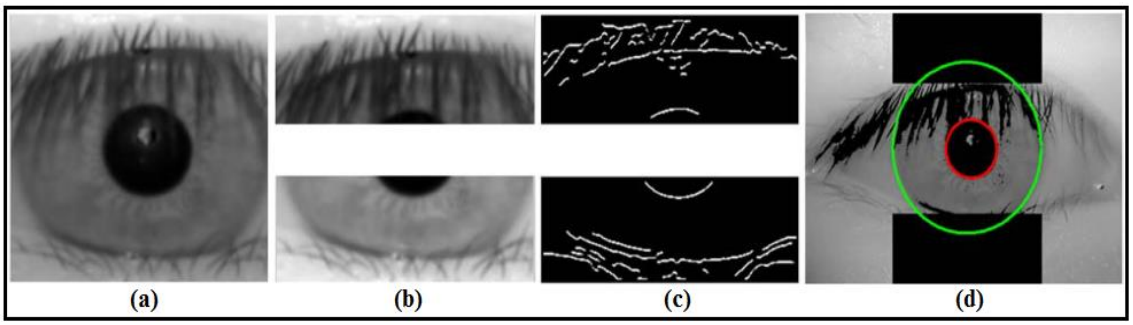


Figure 4.9: Eyelids and Eyelashes Detection: (a) The iris boundary region, (b) The top and bottom parts of iris region, (c) The edge map the top and bottom parts, and (d) The final detected eyelids and eyelashes.

4.3.1.5 Iris Normalization

Once, the iris boundaries have been detected, iris normalization is implemented to produce a fixed dimension feature vector that allows comparison between two different iris images. The main advantage of the iris normalization process is to remove the dimensional inconsistencies that can occur due to stretching of the iris region caused by pupil dilation with varying levels of illumination. Other causes of dimensional inconsistencies include, changing imaging distance, elastic distortion in the iris texture that can affect the iris matching outcome, rotation of the camera or eye and so forth. To address all these mentioned issues the iris normalization process is applied using Daugman's Rubber Sheet mapping to transform the iris image from Cartesian coordinates to polar coordinates, as shown in Fig. 4.10. Daugman's mapping takes each point **(*x, y*)** within the iris region to a pair of normalized non-concentric polar coordinates **(*r, θ*)**, where ***r*** is on the interval **[0, 1]** and ***θ*** is the angle on the interval **[0, 2π]**. This mapping of the iris region can be defined mathematically as follows:

$$\boldsymbol{I}\big(\boldsymbol{x}(\boldsymbol{r},\boldsymbol{\theta}),\boldsymbol{y}(\boldsymbol{r},\boldsymbol{\theta})\big) \rightarrow \boldsymbol{I}(\boldsymbol{r},\boldsymbol{\theta})$$

$$\boldsymbol{x}(\boldsymbol{r},\boldsymbol{\theta}) = (\mathbf{1}-\boldsymbol{r})\boldsymbol{x_p}(\boldsymbol{\theta})\boldsymbol{r}\boldsymbol{x_l}(\boldsymbol{\theta})$$

$$\boldsymbol{y}(\boldsymbol{r},\boldsymbol{\theta}) = (\mathbf{1}-\boldsymbol{r})\boldsymbol{y_p}(\boldsymbol{\theta})\boldsymbol{r}\boldsymbol{y_l}(\boldsymbol{\theta}) \qquad (\mathbf{4.16})$$

Here $\boldsymbol{I}(\boldsymbol{x},\boldsymbol{y})$ is the intensity value at $(\boldsymbol{x},\boldsymbol{y})$ in the iris region image. The parameters $\boldsymbol{x_p}$, $\boldsymbol{x_l}$, $\boldsymbol{y_p}$ and $\boldsymbol{y_l}$ are the coordinates of the pupil and iris boundaries along the $\boldsymbol{\theta}$ direction.

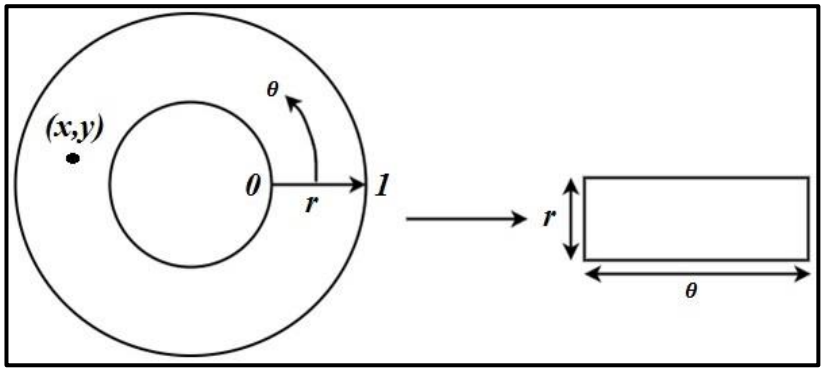


Figure 4.10: Daugman's Rubber Sheet model to transfer the iris region from the Cartesian coordinates to the polar coordinates.

4.3.2 The Proposed IrisConvNet System

Once a normalized iris image is obtained, feature extraction and classification is performed using a deep learning approach that combines a CNN and a Softmax classifier. In this work, the structure of the proposed CNN involves a combination of convolutional layers and sub-sampling max-pooling. The top layers in the proposed CNN are two fully-connected layers for the classification task. Then, the output of the last fully-connected layer is fed into the Softmax classifier, which produces a probability distribution over the ***N*** class labels. Finally, a cross-entropy loss function, a suitable loss function for the classification task, is used to quantify the agreement between the predicted class scores and the target labels, and calculate the cost value for different configurations of CNN. In this section, the proposed methodology for finding the best CNN configuration to be used for the iris recognition task is explained. Based on domain-knowledge from the literature [31][291], there are three main aspects that have a great influence on the performance of a CNN, which need to be investigated. These include: (i) Training methodology, (ii) Network configuration or architecture, and (iii) Input image size. The performance of some carefully proposed training strategies, including the dropout method, AdaGrad method and data augmentation is investigated as part of this work. These training strategies have a significant role in preventing the overfitting problem during the learning process, and increasing the generalization ability of the neural network for new unseen data. These three aspects are described in more details in the next section.

4.3.2.1 Training Methodology

In this work, all of the experiments were carried out, given a particular set of sample data, using 80% randomly selected samples for training, and the remaining 20% for testing. The training methodology is similar to [254][302]. It starts with training a particular CNN configuration by dividing the training set into four sets: three sets are used to train the CNN, and the last one is used as for validation. Validation is used for testing the generalization ability of the network during the learning process and storing the weights configuration that performs best on it with a minimum validation error, as shown in Fig. 4.11. In this work, the training procedure is performed using the Back-propagation

algorithm, with the mini-batch AdaGrad optimization method introduced in [303], where each set of the three training data is divided into mini-batches, and the training errors are calculated upon each mini-batch in the Softmax layer and get back-propagated to the lower layers.

After each epoch (passing through the entire training samples), the validation set is used to measure the accuracy of the current configuration by calculating the cost value and the Top-1 validation error rate. Then, according to the Adagrad optimization method, the learning rate is scaled by a factor equal to the square root of the sum of squares of the previous gradients, as shown in Eq.(4.17). An initial learning rate must be selected, hence two of the most common used learning rate values are analysed herein, as shown in (Section 4.4.3.1). To avoid the overfitting problem, the training procedure is stopped as soon as the cost value and the error on the validation set starts to rise again, which means that the network starts to overfit the training set. This process is one of the regularization methods called the early stopping procedure. In this work, different numbers of epochs are investigated as explained in (Section 4.4.3.1). Finally, after the training procedure is finished, the testing set is used to measure the efficiency of the final configuration obtained in predicting the unseen samples by calculating the identification rate at Rank-1 as an optimization objective, which is maximized during the learning process. Then, the CMC curve is used to visualize the performance of the best configuration obtained as the iris identification system. The main steps of the proposed training methodology are summarized as follows:

1. Split the dataset into three sets: Training, Validation and Test set.
2. Select a CNN architecture and a set of training parameters.
3. Train the each CNN configuration using the training set.
4. Evaluate each CNN configuration using the validation set.
5. Repeat steps **3** through **4** using ***N*** epochs.
6. Select the best CNN configuration with a minimal error on the validation set.
7. Evaluate the best CNN configuration using the test set.

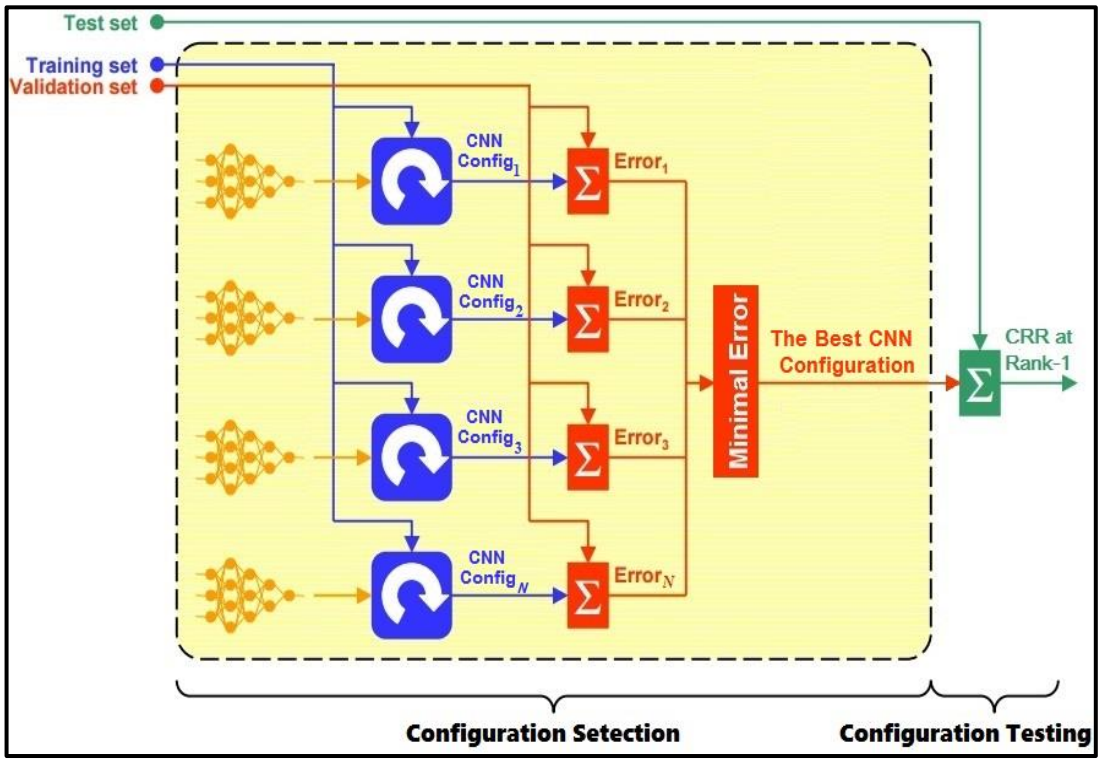


Figure 4.11: An overview of the proposed training methodology to find the best CNN architecture, where CRR refers to the Correction Recognition Rate at Rank-1.

4.3.2.2 Network Architecture

Once the parameters of the training methodology are determined (e.g. learning rate, number of epochs, etc.), it is used to identify the best network architecture. From the literature, it appears that choosing the network architecture is still an open problem, and is application dependent. The main concern in finding the best CNN architecture is the number of the layers to employ transforming from the input image to high-level feature representations, along with the number of convolution filters in each layer. Therefore, some CNN configurations using the proposed training methodology are evaluated by varying the number of convolutional and pooling layers, and the number of filters in each layer, as explained in (Section 4.4.3.2). To reduce the number of configurations to be evaluated, the number of the fully-connected layers is fixed at two, as in [304][305], and the size of filters for both the convolutional and pooling layers are kept as the same as in [30], except in the first convolutional layer where it is set to (3×3) pixels, to avoid a rapid decline in the amount of input data.

4.3.2.3 Input Image Size

The input image size is one of the hyper-parameters in the CNN that has a significant influence on the speed and the accuracy of the neural network. In this work, the influence of input image size is investigated using the sizes (64×64) pixels and (128×128) pixels (generated from original images of larger size as described in the Data Augmentation section below), given that, for lower values than the former, the iris patterns become invisible, while for higher values than the latter, the larger memory requirements and higher computational costs are potential problems. In order to control the spatial size of the input and output volumes, a zero-padding (of **1**-pixel) is applied only to the input layer.

4.3.2.4 Training Strategies

In this section, a number of carefully-designed training techniques and strategies are used to prevent overfitting during the learning process and increase the generalization ability of the neural network. These techniques are:

1. **Dropout Method:** this is a regularization method recently introduced by Srivastava et al. [282] that can be used to prevent neural networks from overfitting the training set. The dropout technique is implemented in each training iteration by completely ignoring individual nodes with a probability of **0.5**, along with their connections. This method decreases the complex co-adaptations of nodes by preventing the interdependencies from emerging between them. The nodes which are dropped do not participate in both forward and backward passing. Therefore, as shown in Fig. 4.12 (b), only a reduced network is left and is trained on the input data in that training iteration. As a result, the process of training a neural network with $\boldsymbol{n}$ nodes will end up with a collection of $\boldsymbol{(2^n)}$ possible "thinned "neural networks that share weights. This allows the neural network to avoid overfitting, learn more robust features that generalize well to new unseen data, and speeds up the training process. Furthermore, it provides an efficient way of combining many neural networks with different architectures, which make the combination more beneficial. In the testing phase, it is not practical to average the predictions from $\boldsymbol{(2^n)}$ "thinned" neural networks, especially for a

large value of ***n***. However, this can be easily addressed by using a single network without dropout, and with the outgoing weights of each node multiplied by a factor of **0.5** to ensure that the output of any hidden node is the same as in the training phase. In this work, the dropout method is applied only to the two fully-connected layers, as they include most of the parameters in the proposed CNN, and are more vulnerable to overfitting. More information on the dropout method can be found in [282].

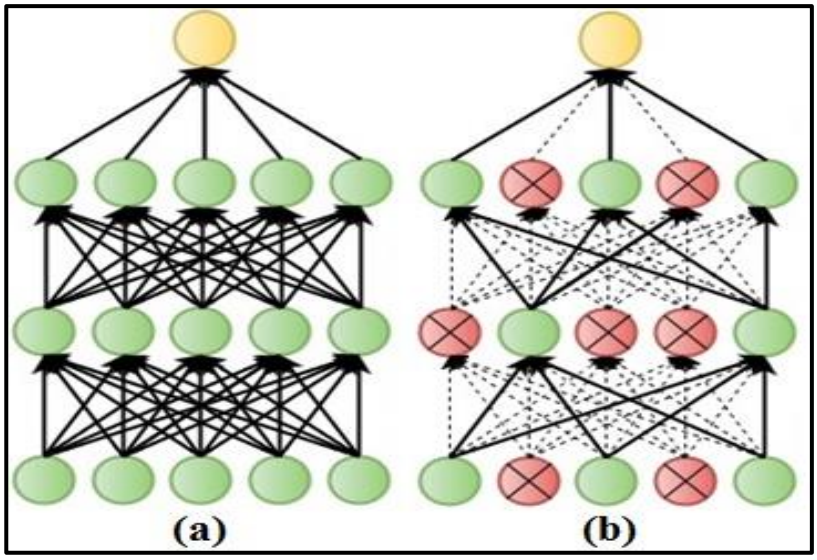


Figure 4.12: An illustration of applying the dropout method to a standard neural network: (a) A standard neural network with 2 hidden layers before applying dropout method. (b) An example of a reduced neural network after applying dropout method. The crossed units and the dashed connections have been dropped.

2. **AdaGrad Algorithm:** in the iris recognition system, infrequent features can significantly contribute to improving the accuracy of the system through minimizing intra-class variations and inter-class similarities, which is caused by several factors, including pupil dilation/constriction, eyelid/eyelash occlusion, and specular reflections spots. However, in the standard Stochastic Gradient Descent (SGD) algorithm for learning rate adaptation, both infrequent and frequent features are weighted equally in terms of learning rate, which means that the influence of the infrequent features is practically discounted. To counter this, the AdaGrad algorithm is implemented to increase the learning rate for more sparse data, which is

translated into a larger update for infrequent features, and decreased learning rate for less sparse data, which is translated into a smaller update for the frequent features. The AdaGrad algorithm also has the advantage of being simpler to implement than the SGD algorithm [303]. The AdaGrad technique has been shown to improve the convergence performance stability of neural networks over the SGD in many different applications (e.g. NLP, document classification, etc.) in which the infrequent features are more useful than the more frequent features. The AdaGrad algorithm computes the learning rate ($\boldsymbol{\eta}$) for every parameter ($\boldsymbol{\theta_i}$) at each time step $\boldsymbol{t}$, based on the previous gradients of the same parameter as follows:

$$\boldsymbol{\theta_i^{(t+1)}} = \boldsymbol{\theta_i^{(t)}} - \frac{\boldsymbol{\eta}}{\sqrt{\boldsymbol{G_{t,ii}} + \boldsymbol{e}}} \cdot \boldsymbol{g_{t,i}} \qquad (\mathbf{4.17})$$

Here, $\boldsymbol{g_{t,i}} = \boldsymbol{\nabla_\theta J(\theta_i)}$ is the gradient of the objective function at time step $\boldsymbol{t}$, and $\boldsymbol{G_{t,ii}} = \sum_{\boldsymbol{r=1}}^{\boldsymbol{t}} \boldsymbol{g_{t,i}^2}$ is the diagonal matrix, where each diagonal element (***i, i***) is the sum of the squares of the gradients for the parameter ($\boldsymbol{\theta_i}$) at time step $\boldsymbol{t}$. Finally, $\boldsymbol{e}$ is a small constant to avoid division by zero. More details on the AdaGrad algorithm can be found in [303].

3. **Data Augmentation:** it is well-known that DNNs need to be trained on a large number of training samples to achieve satisfactory prediction and prevent overfitting. Data augmentation is a simple and commonly-used method to artificially enlarge the dataset by methods such as: random crops, intensity variations, horizontal flipping, etc. In this work, data augmentation is implemented similarly to [283]. Initially, a given rectangular image is rescaled so that the longest side is reduced to the length of the shortest side instead of cropping out a square central patch from the rectangle image, as in [283], which can lose crucial features from the iris image. Then, five image regions are cropped from the rescaled image corresponding to the four corners and central region. In addition, their horizontally flipped versions are also acquired. As a result, ten image patches are generated from each input image. During prediction time, the same ten image patches are extracted from each input image, and the mean of the predictions on the ten patches is taken at the Softmax layer. In this work, the performance of the CNN is evaluated using two different input

image sizes, so the data augmentation procedure is implemented twice, once for each size. Image patches of size (**64×64**) pixels are extracted from original input images of size (**256×70**) pixels, and image patches of size (**128×128**) pixels are extracted from original input images of size (**256×135**) pixels.

4. **The ReLU Activation Function:** is applied on the top of the convolutional and fully-connected layers, in order to add non-linearity to the network. As reported by Krizhevsky [283], the ReLU $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{max}\,(\mathbf{0}, \boldsymbol{x})$ has been found to be crucial to learning when using DNNs, especially for CNNs, compared to other activation functions, such as the sigmoid and tangent. In addition, it results in neural network training several times faster than with other activation functions, without making a significant difference to generalization accuracy.
5. **Weight Decay:** is used in the learning process as an additional term in calculating the cost function and updating the weights. Here, the weight decay parameter is set to **0.0005** as in [306].

4.4 Experimental Results

In this section, a number of extensive experiments are described to assess the effectiveness of the proposed iris localization, the deep learning approach for iris recognition, and compare their performance with other existing approaches. Firstly, a brief description of the iris datasets used in these experiments is given. Secondly, the performance of the proposed iris localization system was evaluated and compared with current state-of-the-art approaches in term of Average Accuracy Rate (AAR) and average localization time. Finally, extensive experiments performed to find the best CNN (called *IrisConvNet*) for the iris recognition system, are described and compared with current state-of-the-art approaches, in term Rank-1 identification rate and recognition time.

4.4.1 Description of Iris Datasets

In this work, the performances of the proposed iris localization and recognition systems were evaluated on the most challenging iris datasets currently available in the public domain. Five iris datasets, namely, SDUMLA-

HMT [255], CASIA-Iris-V1 [307], CASIA-Iris-V3 Interval [308], Multimedia University (MMU1) [309], and IITD [309], were employed as testing benchmarks and for comparing the results obtained with current state-of-the-art approaches. In most cases, the iris images in these datasets were captured under different conditions of pupil dilation, eyelids/eyelashes occlusion, head-tilt, a slight shadow of eyelids, specular reflection spots, etc. Some examples of iris images from each dataset are shown in Fig. 4.13, while the basic characteristics of these five datasets are summarized in Table 4.1.

- **SDUMLA-HMT iris dataset** [255]**:** This comprises of 106 subjects with each one providing 10 images (e.g. 5 images for each eye). Therefore, the total is 1,060 images taken using an intelligent iris capture device with a distance between the device and the eye of between 6cm and 32cm. The images are stored in grey-scale level and “BMP” format, with an image size of (768×576) pixels. To the best of the author’s knowledge, this is the first work that uses the SDUMLA-HMT iris dataset for evaluating an iris recognition system operating in the identification mode.
- **CASIA-Iris-V1 dataset** [307]**:** This contains a total of 756 images captured from 108 subjects with 7 images for each one. The images in this dataset were collected in two sessions with some irises occluded by the upper and/or lower eyelids. All images were stored in “BMP” format, with an image size of (320×280) pixels.
- **CASIA-Iris-V3 Interval dataset** [308]**:** The CASIA-Iris-V3 Interval dataset consists of 2,566 images from 249 subjects, which were captured from both left and right eye using a self-developed close-up iris camera. In this dataset, the number of images of each subject differs, and 129 subjects have less than 14 iris images. These were not used in the experiments. The images were stored in “JPEG” format with an image size of (320×280) pixels.
- **MMU1 dataset** [309]**:** This contains a total of 450 images captured from 45 subjects with each one providing 10 images (e.g. 5 images for each eye). These images were captured using an LG IrisAccess semi-automated camera with a distance between the eye and the camera of between 7cm and 25cm. The images were stored in “BMP” format with an image size of

(280×320) pixels. In this work, the left and right iris images of the MMU1 dataset were treated separately for a comparison purpose, and hence the MMU1 dataset has 90 subjects during the experimental setup.

- **IITD dataset** [309]**:** This contains a total of 1,120 iris images captured from 224 subjects (176 males and 48 females) in the age group 14–55 years, who are students and staff at IIT Delhi, New Delhi, India. Each subject has 5 images for each eye, which were captured using three different cameras: JIRIS, JPC1000, digital CMOS cameras. The size of each image is (320×240) pixels, and they were stored in the “BMP” format.

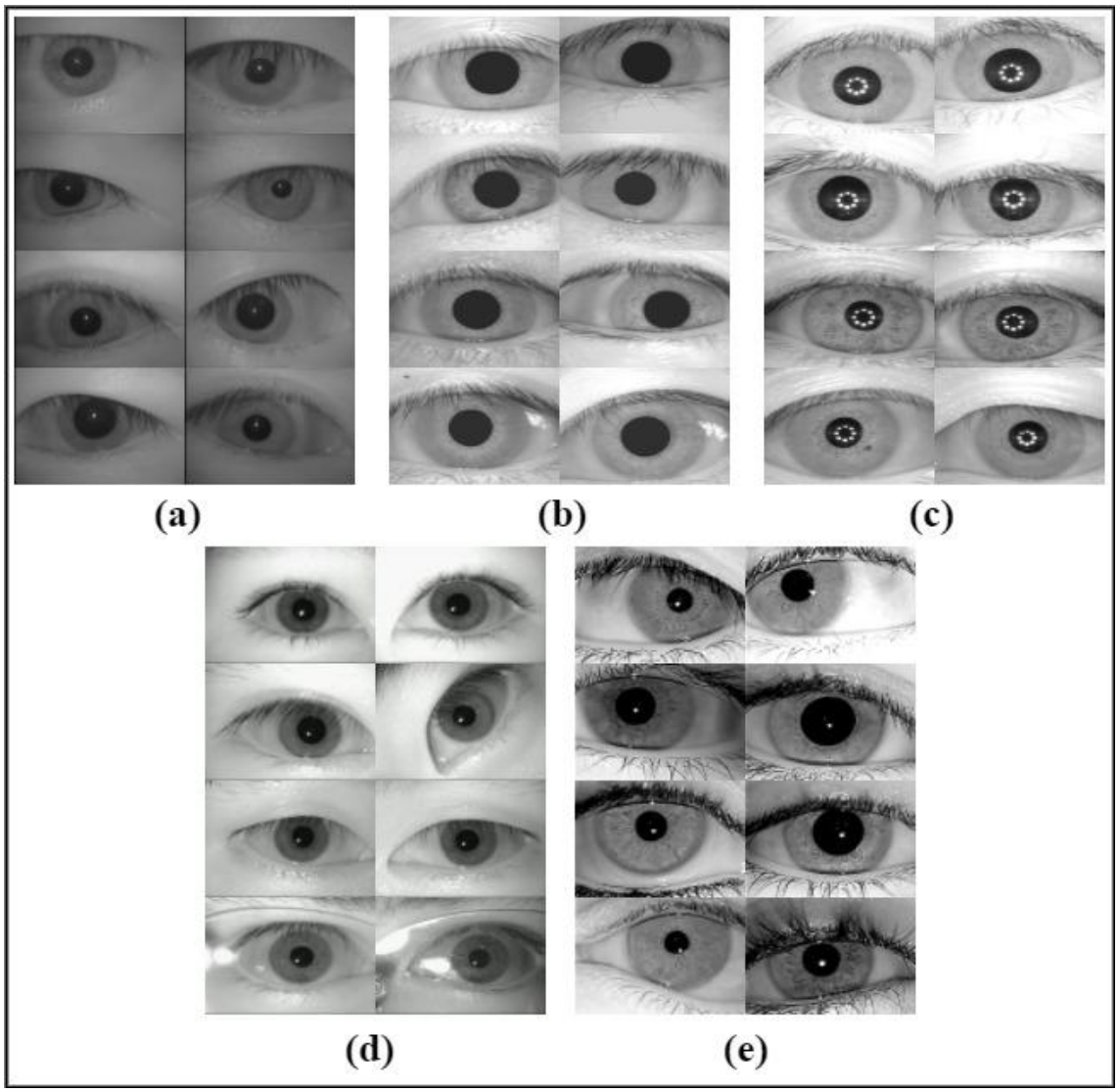


Figure 4.13: Examples of iris images in five iris datasets: (a) SDUMLA-HMT, (b) CASIA-Iris-V1, (c) CASIA-Iris-V3 Interval (d) MMU1, and (e) IITD.

Table 4.1: The characteristics of the adopted iris image datasets.

Iris Datasets	The Main Characteristics				
	Number of Classes	Samples Per Subject	Number of Images	Image Size	Image Format
SDUMLA-HMT	106	5 Right & 5 Left	1,060	(768×576)	BMP
CASIA-Iris-V1	108	7 Per Subject	756	(320×280)	BMP
CASIA-Iris-V3 Interval	120	7 Right & 7 Left	1,680	(320×280)	JPEG
MMU1	90	5 Per Subject	450	(320×240)	BMP
IITD	224	5 Right & 5 Left	2,240	(320×240)	BMP

4.4.2 Iris Localization Experiments

In this section, extensive experiments were conducted on five challenging and well-known iris datasets to evaluate the performance of the proposed iris localization system. In this work, the iris localization is considered as accurate only if two conditions are satisfied. Firstly, the inner and outer boundaries are correctly localized. Secondly, the upper and the lower eyelids are correctly detected, as shown in Fig. 4.14. If one of the above conditions is not provided, then the iris localization is not precise, as illustrated in Fig. 4.15. The AAR of the proposed iris localization method was computed by dividing the number of correctly localized iris images with the number of all the iris images in the dataset, defined as follows:

$$\boldsymbol{AAR} = \frac{\boldsymbol{Correctly\ Localized\ Iris\ Images}}{\boldsymbol{Total\ Number\ of\ Images}} \times \mathbf{100} \qquad (\mathbf{4.18})$$

Finally, the average localization time was computed by calculating the localization time for all the correctly localized iris images, divided by the number of the correctly localized iris images in the dataset.

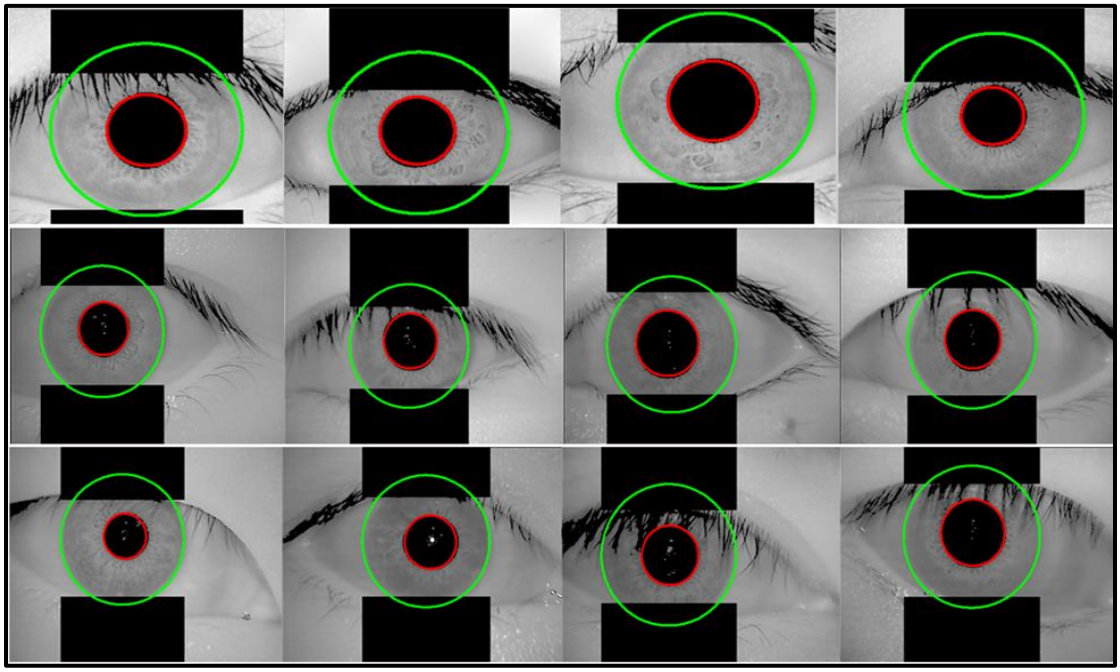


Figure 4.14: Examples of correct localized iris images: The top row represents examples from the CASIA-Iris-V1 dataset, while the second and bottom rows represent examples from the SDUMLA-HMT dataset, the right and left eye image, respectively.

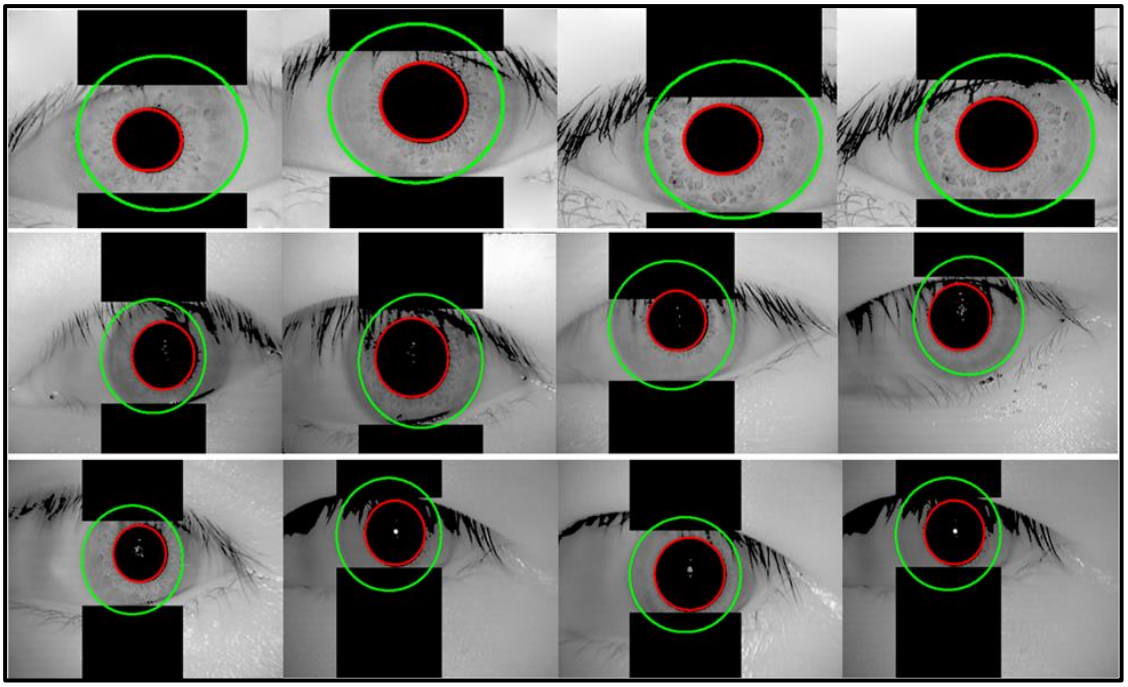


Figure 4.15: Examples of failed iris localization: The top row represents examples from the CASIA-Iris-V1 dataset, while the second and bottom rows represent examples from the SDUMLA-HMT dataset, the right and left eye image, respectively.

As can be seen from Table 4.2, excellent results with an overall accuracy of 96.99%, 99.07%, 99.82%, 99.11%, and 99.87%, obtained with times of 0.72s, 0.63s, 0.62s, 0.32s, and 0.51s, were achieved by applying the proposed iris localization model on the SDUMLA-HMT, CASIA-Iris-V1, CASIA-Iris-V3 Interval, MMU1, and IITD dataset, respectively. The proposed iris localization model managed to correctly localize the iris region from 1,028 out of 1,060 eye images, 749 out of 756 eye images, 1,677 out of 1,680 eye images,446 out of 450 eye images, and 2,237 out of 2,240 eye images in the SDUMLA-HMT, CASIA-Iris-V1, CASIA-Iris-V3 Interval, MMU1, and IITD datasets, respectively. Moreover, it has been found that, using the Radon transform in the eyelids boundary detection increases the speed of the proposed system, which has the advantage over other fitting algorithms, such as the parabolic Hough transform. In this work, the main causes of the error iris localization are due to the weak contrast between the iris and sclera regions, the presence of eyelashes that cover the iris and/or the eyelids, and the low illumination. However, in most of the failing situations, the proposed method is still able to detect correctly at least one of the iris boundaries. The incorrect iris localization results have been taken into account manually, to ensure that all the subjects have the same number of images for the subsequent evaluation of the overall proposed iris recognition system.

Table 4.2: The accuracy rate and running time of the proposed iris localization system.

Iris Datasets	AAR (%)	Running Time (s)
SDUMLA-HMT	96.99	0.72
CASIA-Iris-V1	99.07	0.63
CASIA-Iris-V3 Interval	99.82	0.62
MMU1	99.11	0.32
IITD	99.87	0.51

For further evaluation, the performance of the proposed iris localization system was compared with other established iris localization approaches, as listed in Table 4.3. The results obtained demonstrate that the proposed system

outperforms the indicated state-of-the-art of approaches in terms of accuracy in 33 out of 35 cases, and, in terms of running time, in 15 out of 19 cases, where this information is available. Although some approaches have achieved slightly a higher AAR (e.g., Jan et al. [129]) or a lower running time (e.g., Hentati et al. [310], Uhl et al. [311], Mehrotra et al. [154], etc.) compared with the proposed system, they obtained inferior results either in another term or in other datasets.

4.4.3 Iris Identification Experiments

In this section, extensive experiments performed to find the best CNN model (called *IrisConvNet*) for the iris recognition system are described. Based on the domain-knowledge from the literature, sets of training parameters and CNN configurations, as illustrated in Fig. 4.16, were evaluated to study their behavior and to obtain the best CNN. Then, the performance of this best system was used later on, to make comparisons with current state-of-the-art iris recognition systems.

In this work, due to the sufficient number of iris images, the datasets namely: SDUMLA-HMT, CASIA-Iris-V3 Interval, and IITD were used as the main datasets to fine-tune the hyper-parameters of the CNN to find the best recognition model. Then, the best obtained CNN model was used for subsequent evaluations on CASIA-Iris-V1 and MMU1 dataset.

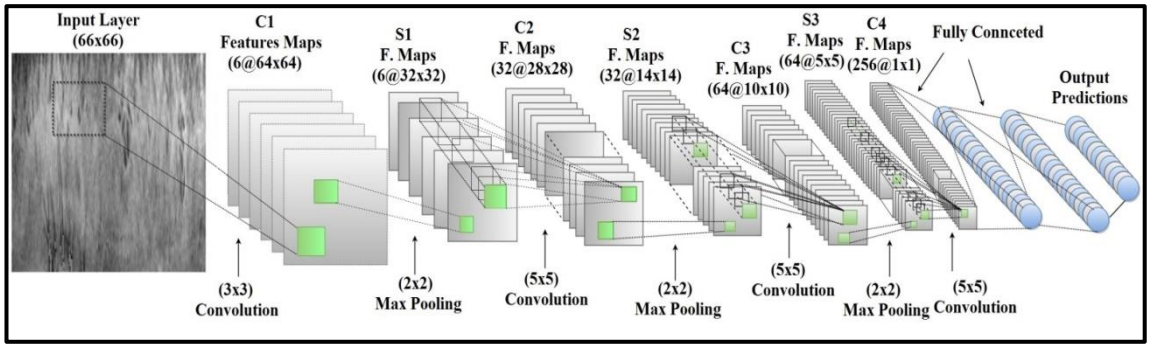


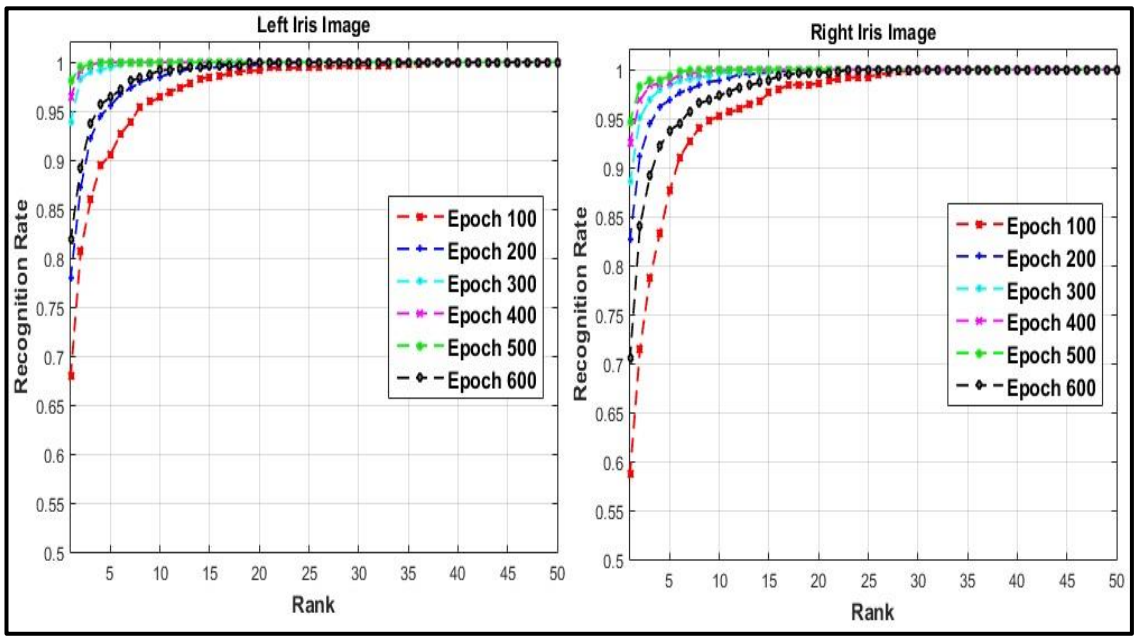
Figure 4.16: An illustration of the *IrisConvNet* model for iris recognition.

Table 4.3: Comparison of the proposed iris localization model with previous approaches using four different iris datasets.

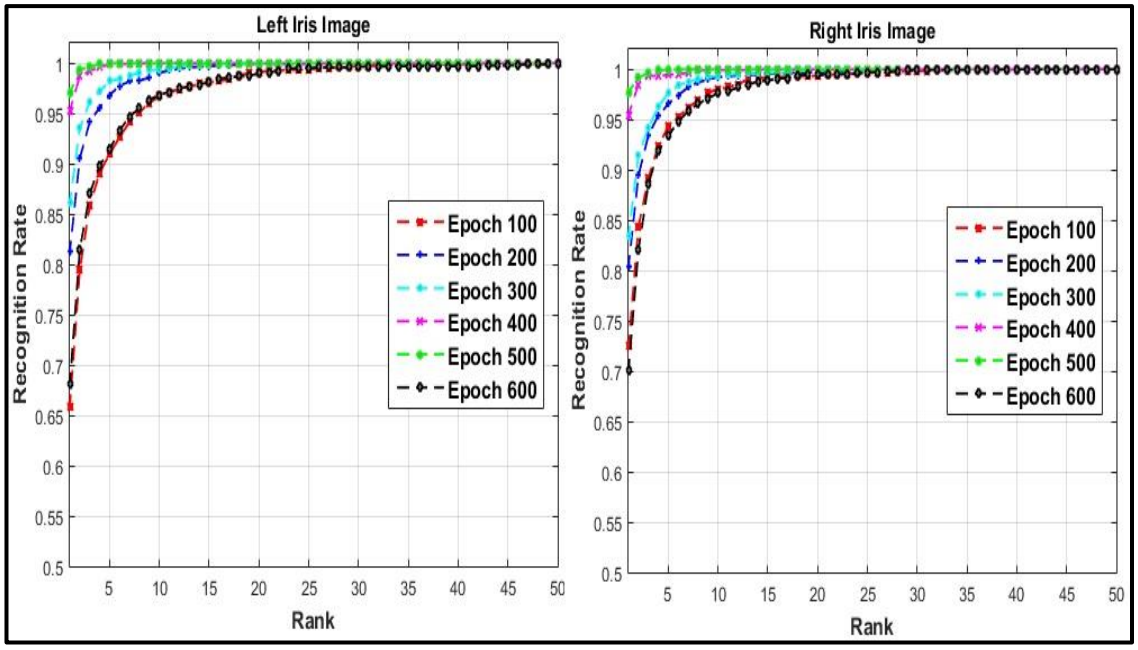
Datasets	Approach	AAR (%)	Time (s)
CASIA-Iris-V1	**Faundra et al.** [312]	98.88	-
	Shamsi et al. [313]	98.00	-
	Masek [314]	83.00	9.37
	Mahmoud and Ali [315]	98.37	-
	Hentati et al. [310]	92.6	**0.290**
	GuangZhu et al. [316]	98.42	-
	Ng et al. [317]	97.22	-
	Jan et al. [129]	**100**	7.20
	Ibrahim et al. [318]	92.00	-
	Proposed Iris Localization	99.07	0.63
CASIA-Iris-V3 Interval	**Rao et al.** [319]	91.00	0.26
	Mahmoud and Ali [315]	99.18	-
	Jan et al. [129]	99.50	7.75
	Wang et al. [320]	96.95	165.4
	Uhl et al. [311]	74.00	**0.21**
	Ugbaga et al. [321]	98.90	-
	Umer et al. [146]	95.87	0.89
	Wild et al. [322]	98.13	-
	Aydi et al. [274]	96.51	9.049
	Pawar et al. [128]	96.88	-
	Mehrotra et al. [154]	99.55	0.396
	Wan et al. [323]	97.29	1.67
	Wild et al. [324]	94.03	-
	Proposed Iris Localization	**99.82**	0.62
MMU1	**Shamsi et al.** [313]	99.00	-
	Jan et al. [129]	**100**	2.67
	Umer et al. [146]	98.22	0.58
	P´erez et al. [325]	84.00	0. 379
	Ibrahim et al. [318]	93.00	-
	Wan et al. [323]	97.83	1.52
	Valentina et al. [326]	92.89	-
	Fernandez et al. [327]	97.60	0.557
	Proposed Iris Localization	99.11	**0.32**
IITD	**Jan et al.** [129]	99.40	8.52
	Wang et al. [320]	96.07	145.4
	Umer et al. [146]	98.48	0.77
	Wild et al. [322]	97.60	-
	Wild et al. [324]	96.77	-
	Proposed Iris Localization	**99.87**	**0.51**

4.4.3.1 Training Parameters Evaluation

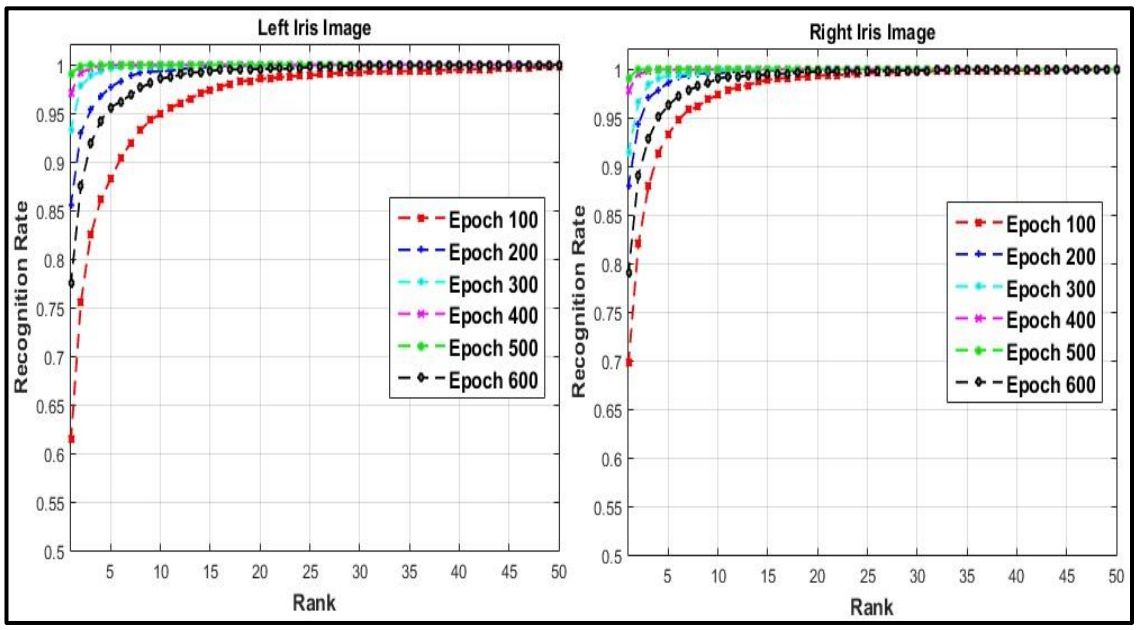
As mentioned previously, a set of training parameters is needed in order to study and analyse their influence on the performance of the proposed deep learning approach and to design a powerful network architecture. All these experiments were conducted on the three different iris datasets, and the parameters with the best performance (e.g. lowest validation error rate and best generalization ability) were kept to be used later in finding the best network architecture. For an initial network architecture, the *Spoofnet* architecture, as described in [30], was used with only a few changes. The receptive field in the first convolutional layer was set to be (3×3) pixels rather than (5×5) pixels to avoid a rapid decline in the amount of input data, and the output of the Softmax layer was set to ***N*** units (the number of classes) instead of 3 units, as in the *Spoofnet*. Finally, the (64×64) pixels input image size, rather than (128×128) pixels, was used in these experiments with a zero-padding of 1-pixel value applied only to the input layer. The first evaluation was to analyse the influence of the learning rate parameter using the AdaGrad optimization method. Based on the proposed training methodology described in (Section 4.3.2.1), an initial learning rate of 10^{-3} was employed as in [328]. However, it was observed that the model takes too long to converge because the learning rate was too small and it reduced continuously after each epoch according to the AdaGrad method. Therefore, for all the remaining experiments, an initial learning rate of 10^{-2} was used. For the first time, the initial number of epochs was set to 100 epochs as in [283]. After that, larger numbers of epochs were also investigated using the same training methodology, including 200, 300, 400, 500 and 600 epochs. The CMC curves shown in Fig. 4.17 are used to visualize the performance of the last obtained model on the validation set. It can be seen that, as long as the number of epochs is increased, the performance of the last model gets better. However, when 600 epochs were evaluated, it was observed that the obtained model started overfitting the training data, and poor results were obtained on the validation set. Therefore, 500 epochs were taken as the initial number of epochs in the assessment procedure for all remaining experiments, since the learning process still achieved good generalization without overfitting.



(a) SDUMLA-HMT Dataset.



(b) CASIA-Iris-V3 Interval Dataset.



(c) IITD Dataset.

Figure 4.17: CMC curves for epoch number parameter evaluation using three different iris datasets: (a) SDUMLA-HMT, (b) CASIA-Iris-V3, and (c) IITD.

4.4.3.2 Network Architecture and Input Image Size Evaluation

The literature on designing powerful CNN architectures shows that this is an open problem, and usually approached using previous knowledge of related applications. Generally, the CNN architecture is related to the size of the input image. A smaller network architecture (a smaller number of layers) is required for a small image size to avoid degrading the quality of the last generated feature vectors, by increasing the number of layers, while a deeper network architecture can be employed for input images with a larger size, along with a large number of training samples to increase the generalization ability of the network by learning more distinctive features from the input samples. In this study, when the training parameters have been determined, the network architecture and input image size were evaluated simultaneously by performing extensive experiments using different network configurations. Based on the proposed training methodology, the evaluation strategy starts from a relatively small network (three layers), and then the performance of the network was observed by adding more layers and filters within each layer. In this work, the influence of input image size was investigated using image sizes of (64×64) pixels and (128×128) pixels, each with two different network configurations. For

example, the (64×64) size was assessed using network topologies with **3** and **4** convolutional layers, while the (128×128) size was assessed using network topologies with **4** and **5** convolutional layers.

The results obtained by applying the proposed system on the three different iris datasets with image sizes of (64×64) pixels and (128×128) pixels are presented in Table 4.4 and 4.5, respectively. As can be seen in these tables, the number of the filters in each layer is tending to increase as one moves from the input layer toward the higher layers, as has been done in previous works in the literature, to avoid memory issues and control the model capacity. In general, it has been observed that the performance of a CNN improves as the number of the employed layers is increased, along with the number of the filters per each layer. For instance, in Table 4.4 the recognition rate dramatically increased for all employed datasets by adding a new layer on the top of the network. However, adding a new layer on the top of the network, and/or altering the number of the filters within each layer, should be carefully controlled. For instance, in Table 4.5, it can be seen that adding a new layer led to a decrease in the recognition rate from 93.02% to 80.09% for the left iris image in the SDUMLA-HMT dataset, and from 99.17% to 95.23% for the right iris image in the CASIA-Iris-V3 Interval dataset. In addition, changing the number of filters within each layer has a significant influence on the performance of the CNN. There are examples of this shown in Table 4.4 (e.g. Configuration number **10** and **11**), and Table 4.5 (e.g. Configuration number **18** and **19**) where altering the number of filters in some layers has led to either an increase or a decrease in the recognition rate.

As indicated in Fig. 4.16, the last CNN configuration in Table 4.4 is preferred as the adopted CNN architecture for identifying a person's identity, for several reasons. Firstly, it provides the highest identification rate at Rank-1 for both the left and right iris images for all the employed datasets with less complexity (fewer parameters). Secondly, although this model has given promising results using an input image of size (128×128) pixels, the input image size might be a major constraint in some applications; hence the smaller one is used as the input image size for *IrisConvNet*. In addition, the training time required to train such a configuration is less than one day, as shown in (Section

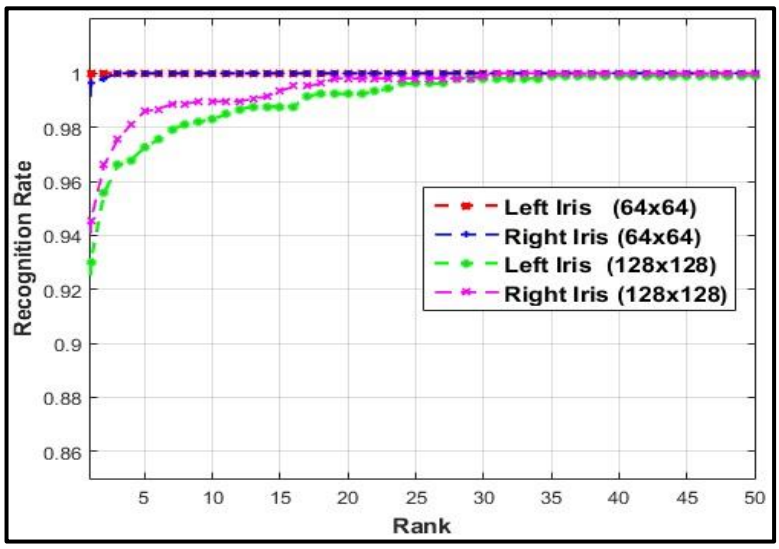
4.4.4). Finally, a larger CNN configuration, along with a larger image size, drives significant increases in memory requirements and computational complexity. The performance of *IrisConvNet* for iris identification for both employed input images sizes is expressed through the CMC curve, as shown in Fig. 4.18. Fig. 4.19 shows the feature maps for the first two layers of the *IrisConvNet* system, with corresponding input images from SDUMLA-HMT dataset. One can see that the filters in the first two layers of the *IrisConvNet* system can efficiently capture the edges and corners within the input image, which are proven to be important for iris recognition.

Table 4.4: Rank-1identification rates obtained for different CNN architectures using the input image size of (64×64) pixels. Each configuration has either 3 or 4 layers and indicates the number of filters in each layer.

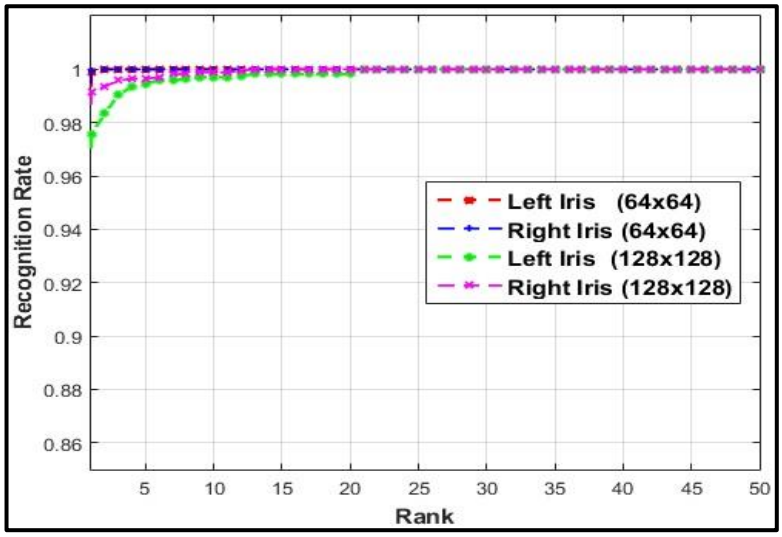
Configuration	SDUMLA-HMT		CASIA-Iris-V3		IITD	
	R. Iris	L. Iris	R. Iris	L. Iris	R. Iris	L. Iris
[6 6 6]^{C1}	46.30	44.71	7.79	0.85	0.44	0.44
[6 6 20]^{C2}	48.77	44.33	0.83	0.84	0.45	0.46
[6 20 6]^{C3}	48.96	40.94	76.60	69.46	0.47	0.44
[6 20 36]^{C4}	46.22	46.41	62.69	60.89	47.76	0.46
[6 20 36 36]^{C5}	86.50	92.73	87.68	96.79	88.04	86.47
[6 20 36 64]^{C6}	93.30	96.22	94.64	97.62	84.46	82.45
[6 20 36 96]^{C7}	97.54	95.94	96.84	98.21	94.82	94.15
[6 20 36 128]^{C8}	95.66	98.68	96.85	98.57	95.54	96.56
[6 20 36 150]^{C9}	98.88	97.64	98.04	98.27	95.94	96.74
[6 20 36 256]^{C10}	98.77	98.08	98.87	99.10	97.00	97.77
[6 32 36 64]^{C11}	94.15	98.67	98.33	97.02	99.10	99.12
[6 32 36 96]^{C12}	99.25	99.43	99.52	97.86	99.02	99.50
[6 32 36 128]^{C13}	99.15	99.71	99.29	99.64	99.33	99.64
[6 32 36 150]^{C14}	98.68	98.08	99.16	99.11	99.28	98.88
[6 32 36 256]^{C15}	99.05	98.96	99.70	99.64	99.46	99.50
[6 32 64 256]^{C16}	**99.62**	**100**	**99.94**	**99.88**	**99.82**	**99.92**

Table 4.5: Rank-1identification rates obtained for different CNN architectures using the input image size of (128×128) pixels. Each configuration has either 4 or 5 layers and indicates the number of filters in each layer.

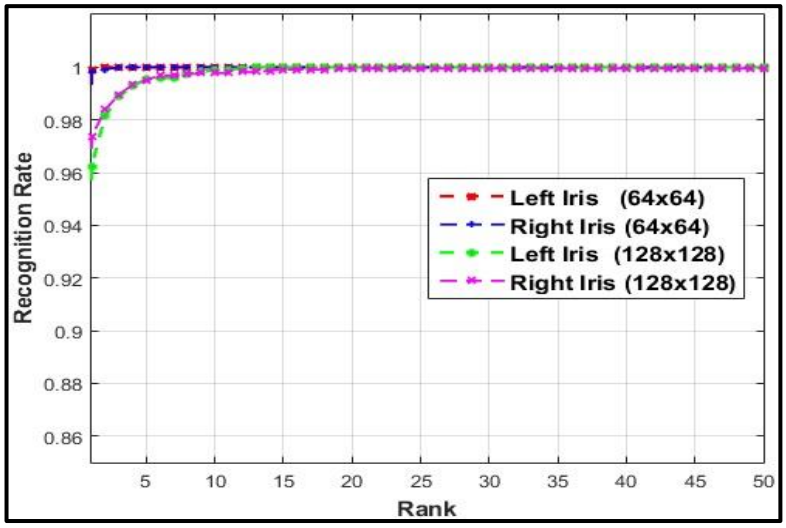
Configuration	SDUMLA-HMT		CASIA-Iris-V3		IITD	
	R. Iris	L. Iris	*R. Iris*	L. Iris	R. Iris	L. Iris
[6 6 16 16]^{C1}	0.97	0.94	45.35	11.78	34.50	15.89
[6 16 16 16]^{C2}	56.79	56.45	59.46	66.13	40.80	37.86
[6 16 16 32]^{C3}	57.55	71.51	72.38	72.20	46.38	34.06
[6 16 32 32]^{C4}	78.77	80.28	55.54	57.97	94.41	94.73
[6 16 32 64]^{C5}	85.94	64.76	96.13	94.70	97.67	95.93
[6 16 32 96]^{C6}	92.26	95.18	96.66	97.14	98.48	98.30
[6 16 32 128]^{C7}	93.58	94.52	98.51	98.21	96.07	98.12
[6 16 32 256]^{C8}	95.75	95.66	98.15	98.92	98.48	97.36
[6 32 32 32]^{C9}	32.54	66.13	82.38	94.70	85.17	84.11
[6 32 32 64]^{C10}	92.07	81.41	92.55	92.73	89.19	93.83
[6 32 32 96]^{C11}	93.77	92.16	97.32	98.09	96.25	85.71
[6 32 32 128]^{C12}	94.52	92.35	97.02	98.09	96.25	96.60
[6 32 32 256]^{C13}	93.49	92.92	96.90	96.93	94.91	93.48
[6 32 64 256]^{C14}	94.53	93.02	99.17	97.56	97.37	96.25
[6 16 32 32 64]^{C15}	96.42	80.09	95.23	99.04	98.43	98.17
[6 16 32 32 96] ^{C16}	97.45	93.27	99.28	99.34	98.34	98.83
[6 16 32 32 128] ^{C17}	98.87	96.98	99.34	99.40	99.73	96.92
[6 16 32 32 256] ^{C18}	98.49	97.83	99.22	99.64	97.09	99.28
[6 16 32 64 64] ^{C19}	98.49	91.04	92.92	96.90	99.78	**99.64**
[6 16 32 64 96] ^{C20}	98.58	98.39	99.64	99.82	99.11	98.75
[6 16 32 64 128] ^{C21}	99.43	**99.71**	99.16	99.82	99.50	95.76
[6 16 32 64 256] ^{C22}	**99.43**	99.62	**99.88**	**100**	99.41	98.75
[6 16 64 64 256] ^{C23}	97.07	99.39	99.40	99.64	**99.91**	99.15



(a) SDUMLA-HMT Dataset.



(b) CASIA-Iris-V3 Interval Dataset.



(c) IITD Dataset.

Figure 4.18: CMC curves for *IrisConvNet* for iris identification: (a) SDUMLA-HMT, (b) CASIA-Iris-V3, and (c) IITD.

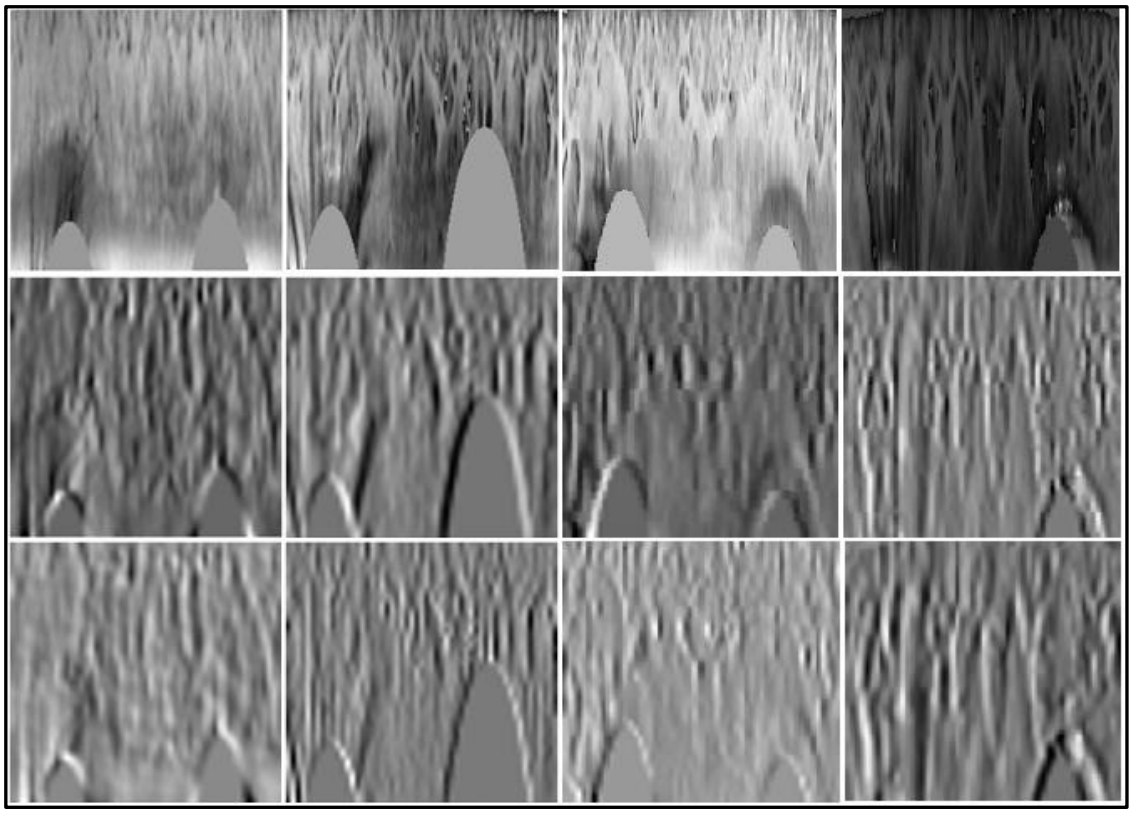


Figure 4.19: The feature maps for the first two layers of the *IrisConvNet* system with corresponding input images from SDUMLA-HMT dataset: The top row shows the input image, while the middle and bottom rows represent the feature maps of the first and second layer, respectively.

The comparison of the proposed *IrisConvNet* system with the current state-of-the-art approaches using CASIA-Iris-V1, CASIA-Iris-V3 Interval, MMU1 and IITD dataset is demonstrated in Table 4.6. In this table, the average Rank-1 identification rates obtained from both the right and left iris images are reported for both CASIA-Iris-V3 Interval and IITD dataset, while the test time per image is calculated from image input until the recognition decision. The feature extraction and classification techniques used in these approaches, along with their evaluation protocols, are shown in Table 4.7. We have assumed that these existing approaches shown in Table 4.6 are customized for these four iris datasets and the best results they obtained are quoted herein.

As can be seen from inspection of Table 4.6, that the performance of the proposed *IrisConvNet* system on CASIA-Iris-V1 and MMU1 dataset is better than all the state-of-art iris recognition systems, in terms of Rank-1 identification rate and running time. The results obtained demonstrate that the proposed

system outperforms the indicated state-of-the-art of approaches in terms of Rank-1 identification rate in 14 out of 14 cases, and, in terms of running time, in 5 out of 5 cases, where this information is available using both datasets. The performance of the proposed *IrisConvNet* system, with respect to CASIA-Iris-V3 Interval dataset, is comparable with the results of Chen et al. [329], but the work in [329] has been evaluated using a small number of subjects (100 classes) compared to the proposed *IrisConvNet* system. Moreover, the proposed system achieved a better running time to establish the person's identity from 120 subjects using the same dataset instead of 100 subjects as in [329]. Finally, the proposed system outperforms the indicated state-of-the-art of approaches in terms of Rank-1 identification rate in 9 out of 10 cases and in terms of running time in 6 out of 7 cases, where this information is available using CASIA-Iris-V3 Interval dataset. Although some approaches have achieved a lower recognition time compared with the proposed system on the CASIA-Iris-V3 Interval and the IITD dataset, such as Bharath et al. [144], they obtained inferior results in term of Rank-1 identification rates on both datasets. For the IITD iris dataset, the proposed system managed to outperform the previous approaches in terms of Rank-1 identification rate in 10 out of 10 cases and in terms of running time in 5 out of 6 cases, where this information is available.

From Table 4.7, it has been seen that the proposed deep learning approaches have overall, outperformed all the state-of-art feature extraction methods, which include the Discrete Wavelet Transform (DWT), the Discrete Cosine Transform (DCT), Principal Component Analysis (PCA), Average Local Binary Pattern (ALBP), Texture Code Matrix (TCM) Empirical Mode Decomposition (EMD), Multi-Perturbation Shapley Analysis (MSA), etc. In term of the Rank-1identification rate, the highest results were obtained by the proposed system using these four iris datasets.

Table 4.6: Comparison of the proposed *IrisConvNet* system with other existing approaches using four different iris datasets. Data labelled by (*) come from Chowhan et al. [330].

Dataset	Approach	Average	RT (s)
CASIA-Iris-V1	**Daugman** [330]	99.25*	-
	Wildes [330]	97.43*	-
	Chowhan et al. [330]	94.00	-
	Kumar et al. [331]	99.07	-
	Liang [25]	97.55	30.2
	Gale and Salankar [26]	98.00	-
	Khan et al. [332]	90.25	-
	Nalla and Chalavadi [148]	75.00	-
	***IrisConvNet* System**	**100**	**0.82**
CASIA-Iris-V3 Interval	**Kerim and Mohammed** [333]	99.40	2
	Umer et al. [146]	96.82	0.98
	De Costa and Gonzaga [334]	99.10	-
	Zhang and Guan [335]	99.60	-
	Roy et al. [336]	97.21	0.995
	Li et al. [27]	99.91	-
	Bharath et al. [144]	84.17	**0.44**
	Chen et al. [329]	**99.82**	138.3
	Umer et al. [29]	99.57	1.05
	Umer et al. [28]	99.38	0.93
	***IrisConvNet* System**	**99.82**	0.89
MMU1	**Umer et al.** [146]	94.22	0.92
	Chen et al. [329]	99.75	-
	Umer et al. [29]	99.56	0.89
	Umer et al. [28]	97.78	0.76
	Umer et al. [145]	98.89	0.94
	Rahulkar and Holambe [337]	98.16	-
	***IrisConvNet* System**	**99.89**	**0.45**
IITD	**Nalla and Chalavadi** [148]	86.00	-
	Umer et al. [146]	98.80	1.11
	Bharath et al [144]	95.93	**0.10**
	Umer et al. [29]	99.38	1.13
	Umer et al. [28]	98.03	1.00
	Umer et al. [145]	97.40	1.18
	Elgamal and Al-Biqami [142]	99.50	-
	Minaee et al. [338]	99.20	-
	Dhage et al. [149]	97.81	93.24
	Abhiram et al. [339]	97.12	-
	***IrisConvNet* System**	**99.87**	0.81

Table 4.7: Summary of the compared iris recognition approaches and their evaluation protocols.

Approach	Feature Extraction	Classification	Evaluation Protocol
Chowhan et al. [330]	MFHSNN	Bhattacharyya distance	4:3 (Training: Testing)
Kumar et al. [331]	PCA+DWT	KNN	6:1 (Training: Testing)
Liang [25]	Gabor filter	Self-Adaptive	5:2 (Training: Testing)
Gale and Salankar [26]	Haar Transform + PCA + Block Sum	Artificial Neural Network	___
Khan et al. [332]	2D Gabor filter	SVM	5:2 (Training: Testing)
Nalla and Chalavadi [148]	Log-Gabor Wavelet	On-Line Dictionary Learning	5:2 (Training: Testing) 3:2 (Training: Testing)
Kerim et al. [333]	Co-occurrence Matrix	Euclidean distance	___
Umer et al. [146]	TCM with ordered PB	SVM + Fusion	Leave-one-out
Abhiram et al. [339]	Circular Sector DCT	Euclidean distance	3:2 (Training: Testing)
De Costa and Gonzaga [334]	Dynamic features	Euclidean distance	Cross-Validation
Zhang and Guan [335]	EMD	KNN	___
Roy et al. [336]	MSA	SVM	Cross-Validation
Li et al. [27]	Average LBP	KNN + SVM	4:1 (Training: Testing)
Bharath et al. [144]	Radon Transform + Gradient-Based Isolation	Euclidean distance	4:1 (Training: Testing)
Chen et al. [329]	SIFT + OPDF	Weighted Sub-Region Matching Fusion	Cross-Validation
Umer et al. [29]	Feature Learning Techniques	Linear SVM	10-Fold Cross-Validation
Umer et al. [28]	Textural Edgeness Descriptors	Linear SVM	10-fold cross-validation
Umer et al. [145]	Morphologic features	SVM	Leave-one-out
Rahulkar and Holambe [337]	Wavelet Filter Bank	Post-Classifier	Leave-one-out
Elgamal and Al-Biqami [142]	DWT+PCA	KNN	___
Minaee et al. [338]	Scattering Transform	Minimum distance	Cross-Validation
Dhage et al. [149]	DWT+DCT	Euclidean distance	9:1 (Training: Testing)
IrisConvNet System	CNN	Softmax Classifier	Cross-Validation 3:2 (Training: Testing) 4:3 (Training: Testing)

4.4.4 Running Time

In this work, the running time was measured by implementing the proposed approaches using a laboratory in Bradford University consisting of 25 PCs with the Windows 8.1 operating system, Intel Xeon E5-1620 CPUs and 16 GB of RAM. The system code was written to run in MATLAB R2015a and later versions. Table 4.8 shows the overall average of the training time of the proposed system, which mainly depends on the input image size, the number of subjects in each dataset, and the CNN architecture. Although the training time is not as important as the testing time in many applications, just for as the iris recognition system, using high-performance GPUs and code optimization can significantly reduce the time required to train proposed *IrisConvNet* system. The test time per image from image input until the recognition decision is less than one second per subject, which is fast enough to be utilized for a real-time iris recognition system.

Table 4.8: The average training time of the proposed deep learning approach.

Iris Datasets	Input Image Size	
	(64×64)	(128×128)
SDUMLA-HMT	6 Hours & 30 minutes	20 Hours & 33 minutes
CASIA-Iris-V1	2 Hours & 12 minutes	4 Hours & 13 minutes
CASIA-Iris-V3 Interval	9 Hours & 18 minutes	53 Hours & 14 minutes
MMU1	1 Hours & 38 minutes	2 Hours & 10 minutes
IITD	17 Hours & 33 minutes	60 Hours & 46 minutes

4.5 Chapter Summary

In this chapter, a robust and fast iris recognition system has been proposed called the *IrisConvNet* system, which has an architecture based on the combination of two discriminative learning techniques, namely a CNN and the Softmax classifier to identify the person's identity, using either the right or left irises of the same person. The architecture of the proposed iris recognition system starts by firstly employing an efficient and automatic iris localization

model to carefully detect the iris region from the background and all extraneous features, such as the pupil, sclera, eyelids, eyelashes, and specular reflections. The proposed iris localization model is composed of four main stages: specular reflection removal, pupil localization, iris localization, and eyelid and eyelash detection. It was noticed that using the CCHT in the first two stages has significantly increased the overall accuracy and reduced the processing time of the subsequent stages in the proposed iris localization model and the iris recognition system. It was also observed that the proposed eyelids and eyelashes algorithm, which employs an anisotropic diffusion filter to enhance the eyelids boundaries and reduce the eyelashes effect, along with the Radon transform for fitting these boundaries as straight lines, has significantly increased the iris recognition performance. Once a normalized iris image is obtained, the *IrisConvNet* system is applied to extract the most distinctive and robust feature representations of iris region and to establish the person's identity in the classification stage. Furthermore, a powerful training methodology equipped with a number of training strategies (e.g. Back-propagation, Dropout method, AdaGrad algorithm, etc.) is proposed, in order to control overfitting during the learning process and increase the generalization ability of the neural network to new unseen data. In this chapter, the *IrisConvNet* system has been applied separately for the left and right iris of the same person, and the average recognition rate was used for comparison purpose with the state-of-the-art iris recognition approaches.

In the experimental part, comprehensive have been conducted to evaluate the performance of the proposed iris recognition approaches using five challenging datasets: SDUMLA-HMT, CASIA-Iris-V1, CASIA-Iris-V3 Interval, MMU1 and IITD iris dataset. The iris images in these datasets were captured under different conditions of pupil dilation, eyelids/eyelashes occlusion, specular reflection, etc. The results obtained demonstrated the superiority of the proposed approaches over many existing feature extraction approaches, such as the DWT, DCT, TCM, EMD, etc. In addition, a new state-of-the-art accuracy, in term of rate Rank-1 identification rate, has been achieved on all the five datasets, with less than one second required to establish the person's identity.

Chapter 5
Multimodal Biometric System

5.1 Introduction

In the previous two chapters, the effectiveness and robustness of the proposed face recognition and iris recognition systems were discussed as unimodal biometric systems. In this chapter, a novel multimodal biometric system is proposed, based on fusing the results obtained from both the face and the left and right irises using different approaches and evaluated under different scenarios based on the biometric traits selected by the user at identification point. As mentioned earlier in Chapter 1, most biometric systems that have been widely employed in governmental and civilian sensitive applications have been unimodal biometric systems. Although these unimodal biometric systems have been used efficiently in identifying a person's identity, there are a number of critical limitations and issues (e.g. noisy data, non-universality, spoof attacks, etc.) that can affect the reliability and the performance of these systems [16]. As reported by Ross et al. [3], these limitations of unimodal biometric systems can be overcome using multimodal biometric systems, by extracting the features from two or more biometric traits in order to enhance the security and increase the reliability in establishing the person's identity. Recently, considerable attention has been paid to employing multimodal biometric systems in many governmental and private sectors, due to their ability to improve significantly the recognition performance of biometric systems, besides adding a number of other advantages compared to unimodal biometric systems, including: (i) improving population coverage; (ii) improving the biometric system's throughput; (iii) deterring spoofing attacks; (iv) maximizing the inter-personal similarities and minimizing the intra-personal variations; and (v) providing a high degree of flexibility allowing people to choose either to provide a subset or all of their biometric traits depending on the nature of the implemented application and the user's convenience [3][22][23]. Hence, research on multimodal biometric systems has attracted much attention in the biometric community, and a number of efficient fusion strategies have

been proposed to fuse the biometric data at different levels of fusion in the last decade. Although existing multimodal biometric systems have been shown to improve effectively the accuracy of biometric system, there are still some interesting and relevant problems that need to be addressed, including: (i) employing a real multimodal biometric dataset to evaluate the performance of the proposed multimodal biometric system instead of using a chimeric dataset, in order to reflect the real performance of the proposed system, (ii) investigations into combining the advantages of both the local handcrafted feature descriptors and deep learning approaches for encoding the biometric traits, (iii) improving the accuracy of the multimodal biometric system in the identification task, rather than just in the verification task, and (iv) providing a high degree of flexibility and handling the problem of missing biometric traits while maintaining a high recognition rate.

In this chapter, an efficient and real-time hybrid multimodal biometric system for identifying a person's identity, using a combination of the face and both irises biometric traits, is proposed to addresses the problems mentioned above. These two types of biometric traits complement each other, as the face trait is unobtrusive, while the iris trait is more accurate. For face recognition, the matching scores are obtained from two distinctive face recognition algorithms, described in Chapter 3. These are the *Curvelet-Fractal* approach and MDFR framework, in which additional and complementary features representations are learnt by training a DBN on top of *Curvelet-Fractal* approach, instead of the pixel intensity representations. An efficient deep learning system described in Chapter 4, and referred to as *IrisConvNet*, is used for iris recognition. In this work, a parallel architecture is considered allowing, users a high degree of flexibility to provide either a subset or all of their biometric traits, depending on the required security level and the users' convenience. In addition, a limitation due to excluding a biometric trait is eliminated. Furthermore, three different types of the multimodal biometric system are proposed, namely, multiple algorithms face recognition, multiple instances iris recognition, and hybrid multimodal biometric system. The Graphical User Interface (GUI) of these multimodal biometric systems are shown in the Appendix (Fig. A.1-A-3). The performance of the proposed multiple algorithms face recognition system was

tested in both identification and verification task, using large-scale unconstrained face datasets: FERET, CAS-PEAL-R1 and LFW datasets. On the other hand, the performances of the last two types were evaluated in the identification task using real multimodal datasets. The CASIA-Iris-V3 Interval and IITD iris datasets were employed to assess the efficiency of the proposed multiple instances iris recognition, while the SDUMLA-HMT multimodal dataset, in which the face and both irises traits were acquired from the same person, was used to validate the accuracy of the proposed hybrid multimodal biometric system.

The remainder of the chapter is organized as follows: The implementation details of the proposed hybrid multimodal biometric system are presented in Section 5.2. Section 5.3 provides an overview of the employed fusion approaches for combining multiple biometric modalities. The experimental results of using the different types of the multimodal biometric system are presented in Section 5.4. Finally, the summary and conclusions of this chapter are stated in the last section.

5.2 The Proposed Hybrid Multimodal Biometric System

The proposed hybrid multimodal biometric identification system for face and both irises traits is shown in the Fig. 5.1. Firstly, the data obtained from each biometric trait is pre-processed to detect the region of interest from the input image (e.g. face and iris regions). Secondly, discriminative features and matching scores are obtained by applying different approaches for each individual biometric trait. Finally, the person's identity is assigned to one of ***N*** classes stored in the dataset, after calculating the similarity scores between the query trait and training traits. In this work, no normalization of outputs is needed, due to all the employed classifiers producing outputs within the same numeric range [**0, 1**]. The next two subsections describe the main implementation steps of the unimodal biometric systems for both face and iris traits.

5.2.1 Face Recognition Matcher

In this work, the matching scores of the face biometric trait are obtained using two different face recognition algorithms, as described in Chapter 3. In the first algorithm, a novel multimodal local feature extraction approach, based on merging the advantages of multidirectional and anisotropy transforms, specifically the Curvelet transform, with Fractal dimension, is employed, referred to as the *Curvelet-Fractal* approach. The key contribution of this approach is that the Fractal dimension approach is applied on the Curvelet's output to produce an illumination invariant representation of the face image that can meet the real-time system's demands. Hence, the Curvelet transform is used here as a powerful technique for edge and curve representations and dimensionality reduction of the face image, to increase the speed of fractal dimension estimation. After the face region has been detected using the Viola-Jones face detector [247], the main steps of the proposed *Curvelet-Fractal* approach for an input face image are implemented as explained in (Chapter 3, Sect. 3.3.1).

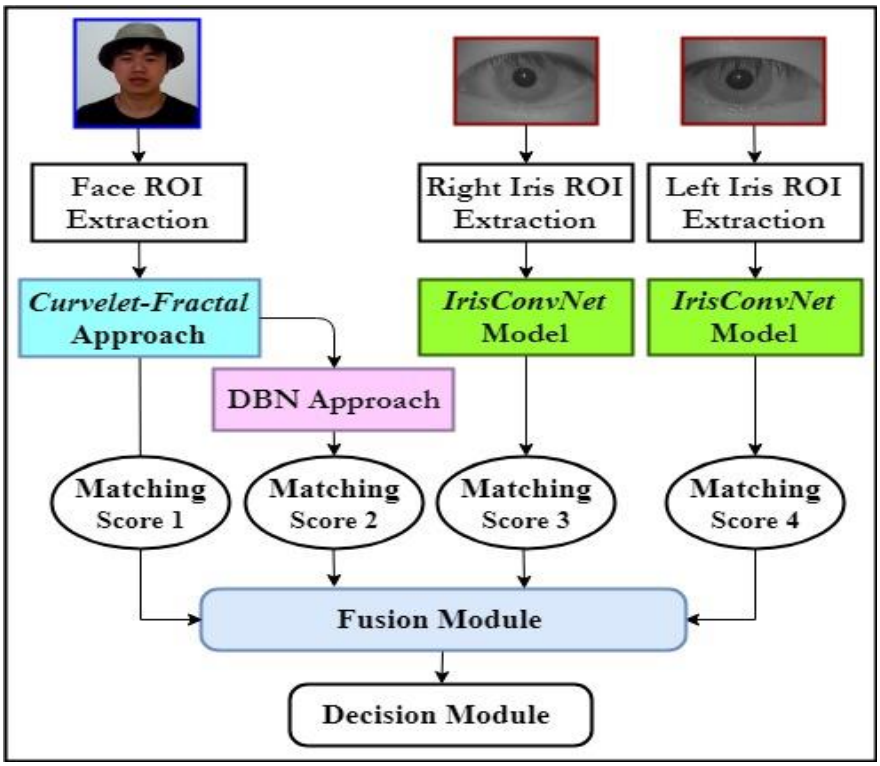


Figure 5.1: Block diagram of the proposed hybrid multimodal biometric system for face and both iris biometric traits.

The second face recognition algorithm employs a novel face recognition framework, referred to as the MDFR framework, in which additional and complementary features representations are learnt by training a DBN on top of *Curvelet-Fractal* representations instead of the pixel intensity representations. Like any deep learning approach, the DBN is usually applied directly on the pixel intensity representations. However, although the DBN has been successfully applied in many different fields, scaling it to realistic-sized face images still remains a challenging task, for several reasons. Firstly, the high dimensionality of the face image leads to increased computational complexity of the training algorithm. Secondly, the feature representations of the DBN are sensitive to the local translations of the input image. This can lead to a disregard of the local features of the input image, which are known to be important for face recognition. To address these issues of the DBN, a novel framework, based on merging the advantages of the local handcrafted image descriptors and the DBN is proposed. It is argued that applying the DBN on top of local features representations instead of the pixel intensity representations (raw data), as a way of guiding the learning process, can greatly improve the ability of the DBN to learn more discriminating features with a shorter training time required to obtain the final trained model. Initially, the local facial features are first extracted using the proposed *Curvelet-Fractal* approach. Then, the extracted local features are assigned to the feature extraction units of the DBN to learn additional and complementary representations. In this work, the DBN architecture stacks **3** RBMs (**3** hidden layers). The first two RBMs are used as generative models, while the last one is used as a discriminative model, associated with Softmax units for multi-class classification purpose. Finally, the hidden layers of the DBN are trained, one at a time, in a bottom-up manner, using a greedy layer-wise training algorithm. In this chapter, a less complex architecture of a hybrid multimodal biometric identification system was also investigated, which completely depends on deep learning approaches by employing only a DBN approach to obtain the matching scores of the face biometric trait, as shown in Fig. 5.2. To the best of the author's knowledge, this is the first work that investigates the potential use of deep learning approaches (e.g. DBN and CNN) for fusing the face and both the left and right irises in a unified multimodal biometric system. Fig. 5.3 shows the main structure of the

hybrid multimodal biometric identification system, in which the face and both iris biometric traits are encoded using only deep learning approaches.

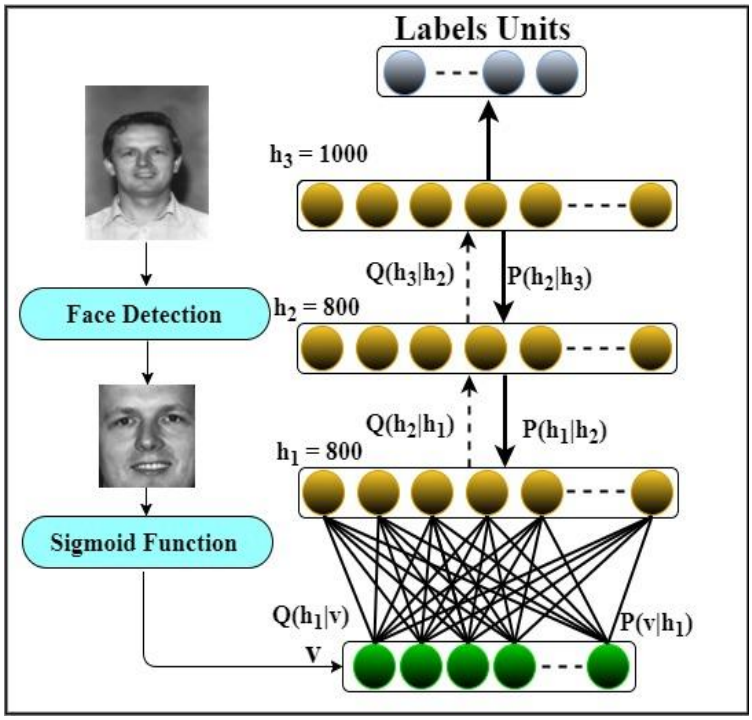


Figure 5.2: Block diagram of the proposed face recognition system using the DBN approach.

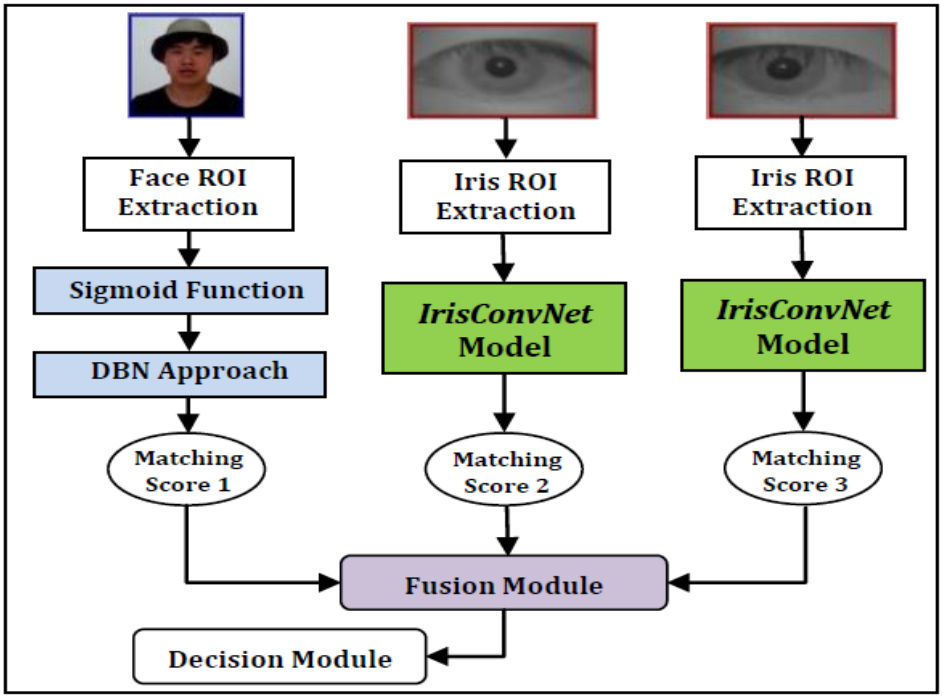


Figure 5.3: Block diagram of a less complex face and iris multimodal biometric system using only deep learning approaches.

5.2.2 Iris Recognition Matcher

An efficient deep learning system described in Chapter 4 and referred to as *IrisConvNet*, is used for iris recognition. The architecture of the *IrisConvNet* system is based on a combination of a CNN and a Softmax classifier to extract automatically discriminative features from the iris image without any domain knowledge, and then classify it into one of ***N*** classes. Firstly, a pre-processing procedure is implemented, which employs an efficient and automatic iris localization algorithm to detects carefully the iris region from the background and all extraneous features, such as the pupil, sclera, eyelids, eyelashes, and specular reflections. Daugman's Rubber Sheet Model is then applied to transfer the detected iris region from Cartesian to polar coordinates, and allow direct comparison between two iris images with initially different sizes. Once a normalized iris image is obtained, feature extraction and classification is performed using a deep learning approach that combines a CNN and a Softmax classifier. In this work, the structure of the proposed CNN involves a combination of convolutional layers and sub-sampling max-pooling. The top layers in the proposed CNN are two fully-connected layers for the classification task. Then, the output of the last fully-connected layer is fed into the Softmax classifier, which produces a probability distribution over the ***N*** class labels. Finally, a cross-entropy loss function, a suitable loss function for the classification task, is used to quantify the agreement between the predicted class scores and the target labels, and calculate the cost value for different configurations of CNN. As described in (Chapter 4, Sect. 4.3.2.1), the *IrisConvNet* system is equipped with a number of carefully-designed training techniques and strategies (e.g. Back-propagation, Dropout method, AdaGrad algorithm, etc.) to prevent overfitting during the learning process, and to increase the generalization ability of the neural network for unseen testing data.

5.3 Fusion Techniques

One of the main factors that must be taken into account when designing and implementing multimodal biometric systems is the fusion of the obtained biometric traits, and where the fusion takes place in the multimodal biometric system. Generally, the application context of the biometric system can also play

a significant role in determining where the fusion process can take place in the multimodal biometric system. For instance, in verification mode, only a single match score for each matcher is obtained, by comparing the query trait with only the template trait of the claimed identity. In contrast, using the identification mode, the query trait is compared with all the templates enrolled in the dataset, resulting in ***N*** matching scores for each individual matcher. Hence, the rank level fusion can only be applied when the biometric system operates in the identification mode. In this chapter, the fusion of the three scores obtained from three biometric traits (the face and both irises) is evaluated and tested at two different fusion levels: Score level fusion and Rank level fusion. As all the employed classifiers produce the same type of output (Similarity score), and all of them have the same numeric range [**0, 1**], no normalization procedure is needed before applying the score fusion methods.

5.3.1 Score Level Fusion

For score level fusion, the problem of classifying a given input pattern **F** into one of ***N*** possible classes is considered, based on the scores provided by ***M*** different classifiers, where ***N*** represents the number of persons enrolled in the dataset. Let $(\boldsymbol{f_i})$ be the feature vector extracted from the input pattern **F**, and presented to the $\boldsymbol{i^{th}}$ classifier with output $\boldsymbol{P(r_j|f_i)}$, referring to the posterior probability of the input pattern **F** belonging to class $\boldsymbol{r_j}$ given the feature vector $(\boldsymbol{f_i})$. Suppose that the input pattern **F** is assigned to class $\boldsymbol{c} = \{\mathbf{1}, \mathbf{2}, \dots \mathbf{N}\}$, then there are five matching scores fusion rules to determine $(\boldsymbol{c})$. An example which represents the general flow of information using the match score level fusion scheme is shown in Fig. 5.4. In this example, matching scores generated by face and iris matchers are similarity measures within the same numeric range [**0, 1**].

1. **Product Rule (PR):** this rule mainly depends on the assumption of statistical independence between the ***M*** feature representations $\{\boldsymbol{f_1}, \boldsymbol{f_2}, \dots, \boldsymbol{f_M}\}$. The product rule assigns the input pattern **F** to class $(\boldsymbol{c})$, as follows:

$$\boldsymbol{c} = \underset{j}{\mathbf{argmax}} \prod_{i=1}^{M} \boldsymbol{P(r_j|f_i)} \qquad (\mathbf{5.1})$$

2. **Sum Rule (SR):** this rule is more practical than the product rule, especially when the input pattern **F** tends to be more noisy, leading to ambiguity in the estimation of posteriori probabilities. In addition, it also assumes that the posteriori probabilities calculated by the different classifiers do not deviate dramatically from the prior probabilities for each class. The input pattern **F** is assigned to the class $(\boldsymbol{c})$, as follows:

$$\boldsymbol{c} = \underset{j}{\mathbf{argmax}} \sum_{i=1}^{M} \boldsymbol{P}(\boldsymbol{r_j}|\boldsymbol{f_i}) \qquad (\mathbf{5.2})$$

3. **Weighted Sum Rule (WSR):** this is an extended version of sum rule that assigns the input pattern **F** to class $\boldsymbol{c}$, as follows:

$$\boldsymbol{c} = \underset{j}{\mathbf{argmax}} \sum_{i=1}^{M} \boldsymbol{P}(\boldsymbol{r_j}|\boldsymbol{f_i}) * \boldsymbol{w_i} \qquad (\mathbf{5.3})$$

Here, $\boldsymbol{w_i}$ is the weights assigned to the $\boldsymbol{i^{th}}$ biometric trait, with condition $\sum_{i=1}^{M} \boldsymbol{w_i} = \mathbf{1}$. In this work, the weights are calculated as in [13][204] based on the biometric traits selected by the user. In general, a higher weight value is assigned to the iris trait compared to the face trait when both of them are under consideration.

4. **Max Rule:** this assigns the input pattern **F** to class $(\boldsymbol{c})$ using the maximum score from the ***M*** different Classifiers, as follows:

$$\boldsymbol{c} = \underset{j}{\mathbf{argmax}} \ \underset{i}{\mathbf{max}} \ \boldsymbol{P}(\boldsymbol{r_j}|\boldsymbol{f_i}) \qquad (\mathbf{5.4})$$

5. **Min Rule:** this assigns the input pattern **F** to class $(\boldsymbol{c})$ using the minimum score from the ***M*** different Classifiers, as follows:

$$\boldsymbol{c} = \underset{j}{\mathbf{argmax}} \ \underset{i}{\mathbf{min}} \ \boldsymbol{P}(\boldsymbol{r_j}|\boldsymbol{f_i}) \qquad (\mathbf{5.5})$$

5.3.2 Rank Level Fusion

In this work, rank level fusion is employed, where each individual classifier produces a ranked list of possible matching scores for each user (a higher rank indicates a better match). Then, these ranks are integrated to create a new ranking list that is used to make the final decision on user identity. A simple example of rank level fusion using three different methods is shown in Fig. 5.5.

In this example, the iris matchers are more accurate than the face matcher. Hence, a weight of 0.4 is assigned for both of them, and due to this significant difference in their weights, the reordered ranks using logistic regression method are very similar to the ranks assigned by the iris matchers. Suppose that there are **P** users registered in the dataset, and the number of employed classifiers is **C**. Let $\boldsymbol{r_{i,j}}$ is the rank assigned to $\boldsymbol{j^{th}}$ user in the dataset by the $\boldsymbol{i^{th}}$ classifier, $\boldsymbol{i} = \{\mathbf{1}, \mathbf{2}, \ldots \mathbf{C}\}$ and $\boldsymbol{j} = \{\mathbf{1}, \mathbf{2}, \ldots \mathbf{P}\}$. Then, the consensus ranks $\mathbf{R_c}$ for a particular class are obtained using the following fusion methods:

1. **Highest Rank (HR):** this is a useful method for fusing the ranks only when the number of registered users is large compared to the number of classifiers, which is the usual scenario in the identification system. The consensus rank of a particular class is computed as the lowest rank generated by different classifiers (minimum $\boldsymbol{r_{i,j}}$ value) and the ties are broken by incorporating a small factor epsilon (***e***), as follows:

$$\boldsymbol{R_c} = \min_{\mathbf{1} \leq \boldsymbol{i} \leq \boldsymbol{C}} \boldsymbol{r_{i,j}} + \boldsymbol{e_i} \qquad (\mathbf{5.6})$$

   Here,

$$\boldsymbol{e_i} = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{C}} \boldsymbol{r_{i,j}} / \boldsymbol{K} \qquad (\mathbf{5.7})$$

   Here, the value of $(\boldsymbol{e_i})$ is ensured to be small by assigning a large value to parameter ***K***.

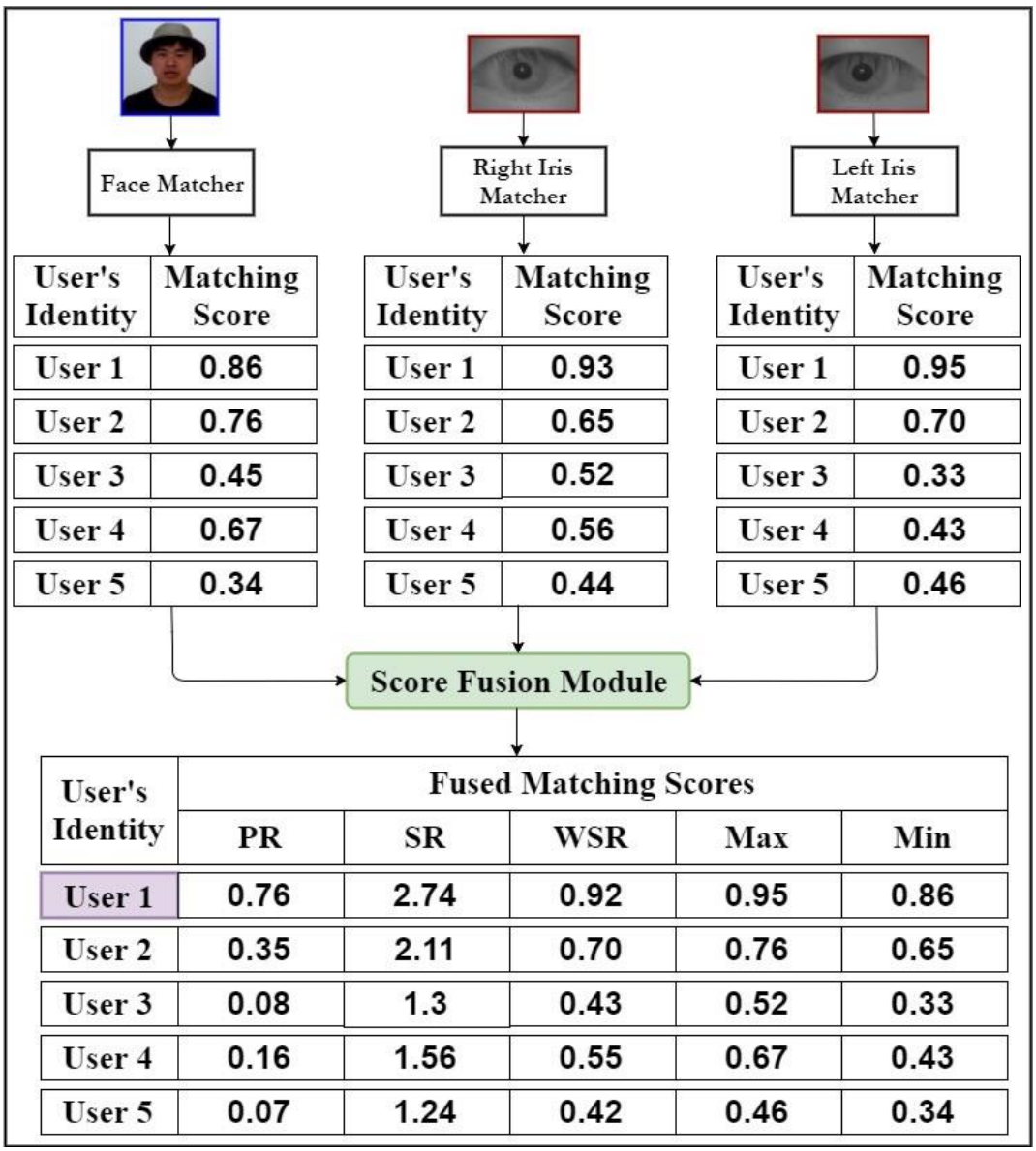
2. **Borda Count (BC):** using this fusion method, the consensus rank of a query identity is computed as the sum of ranks assigned by individual classifiers independently, as follows:

$$\boldsymbol{R_c} = \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{C}} \boldsymbol{r_{i,j}} \qquad (\mathbf{5.8})$$

3. **Logistic Regression (LR):** this is a generalized form of the Borda count method to solve the problem of the uniform performance of the individual classifiers. The consensus rank is calculated by sorting the users according to the summation of their ranks obtained from individual classifiers, as follows:

$$R_c = \sum_{i=1}^{C} w_{i*} r_{i,j} \tag{5.9}$$

Here, w_i is the weight to be assigned to the ith classifier, which is determined by logistic regression according to the biometric traits selected by the user, such that the maximum recognition rate is achieved, and the possibility of spoofing attacks is reduced.



User's Identity	Matching Score
User 1	0.86
User 2	0.76
User 3	0.45
User 4	0.67
User 5	0.34

User's Identity	Matching Score
User 1	0.93
User 2	0.65
User 3	0.52
User 4	0.56
User 5	0.44

User's Identity	Matching Score
User 1	0.95
User 2	0.70
User 3	0.33
User 4	0.43
User 5	0.46

User's Identity	Fused Matching Scores				
	PR	SR	WSR	Max	Min
User 1	0.76	2.74	0.92	0.95	0.86
User 2	0.35	2.11	0.70	0.76	0.65
User 3	0.08	1.3	0.43	0.52	0.33
User 4	0.16	1.56	0.55	0.67	0.43
User 5	0.07	1.24	0.42	0.46	0.34

Figure 5.4: Example of score level fusion using five score fusion methods: Product, Sum, Weighted Sum, Max, and Min Rule. Adapted from [3].

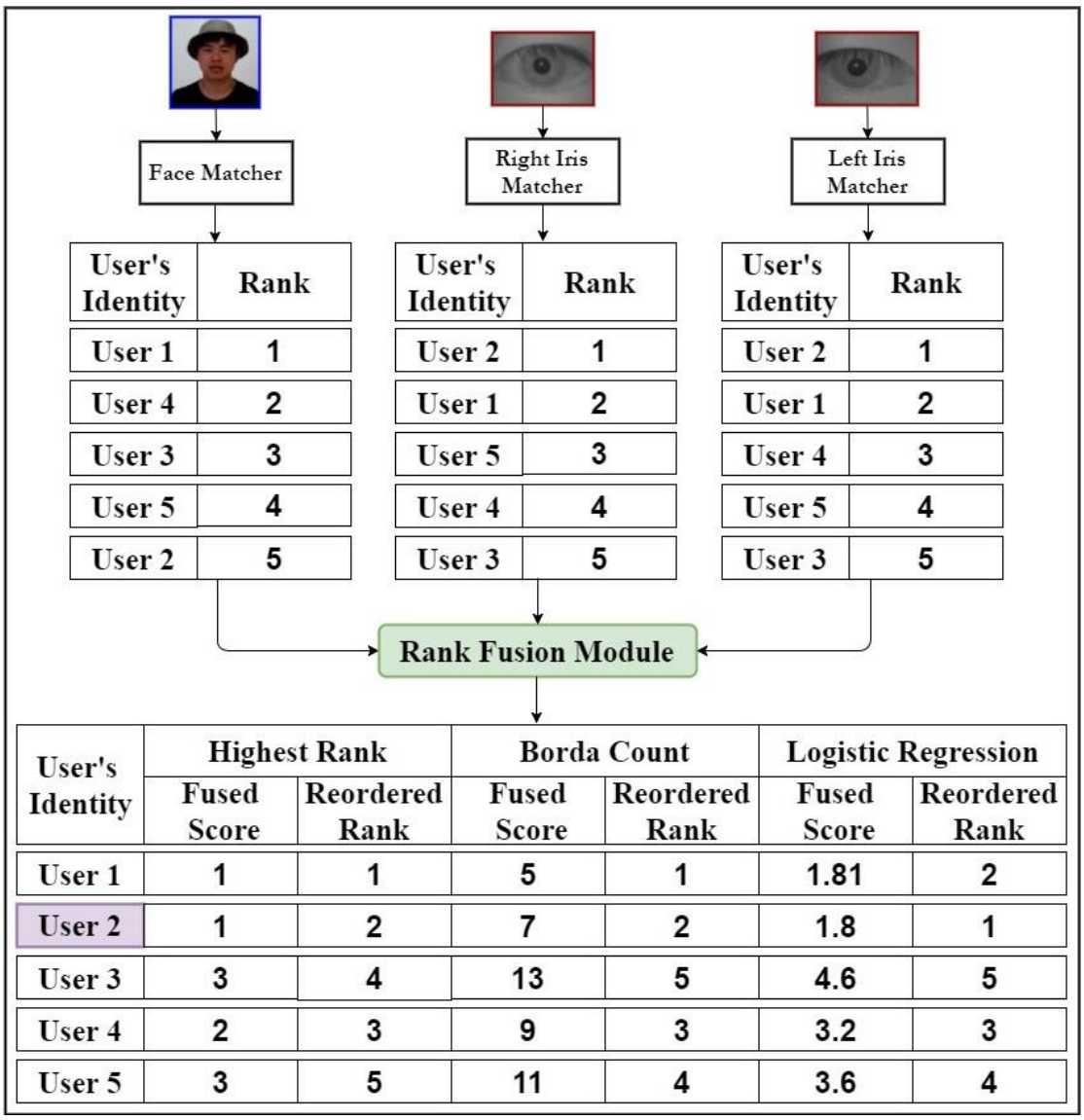


Figure 5.5: Example of rank level fusion using the Highest rank, Borda count and Logistic Regression method. Adapted from [3].

5.4 Experimental Results

In this section, the performances of the proposed multimodal systems are tested on the most challenging face and iris datasets currently available in the public domain, in order to demonstrate their effectiveness and compare their performances with other existing approaches. Firstly, a number of experiments are carried out using the proposed multiple algorithms face recognition system for both face identification and verification tasks, in order to demonstrate its effectiveness and compare its performance with the proposed unimodal approaches, described in Chapter 3. The performances of the proposed

multiple algorithms face recognition system was tested using large-scale unconstrained face datasets: FERET, CAS-PEAL-R1 and LFW datasets. Secondly, the performance of the last two types of the proposed multimodal biometric systems was assessed in the identification task using real multimodal datasets, and compared with the state-of-the-art approaches. The CASIA-Iris-V3 Interval and IITD iris datasets were employed to assess the efficiency of the proposed multiple instances iris recognition, while the SDUMLA-HMT multimodal dataset, in which the face and both irises traits were acquired from the same person, was used to validate the accuracy of the proposed hybrid multimodal biometric system, using two different architectures. Finally, different fusion scenarios of the face and both irises traits are evaluated and discussed, using different fusion methods at the score and rank level. In the implementation of weighted fusion methods, the assigned weights are based on the biometric traits selected by the user, such that the maximum recognition rate is achieved, and the possibility of spoofing attacks is reduced. Hence, in the proposed hybrid multimodal biometric system, the iris trait is assigned with higher weight compared to the face trait. The recognition time of the proposed approaches was measured by implementing them on a personal computer with the Windows 8 operating system, a 3.60 GHz Core i7-4790 CPU and 24 GB of RAM. The system code was written to run in MATLAB R2015a and later versions. It should be noted the execution time of the proposed multimodal biometric system is proportional to the number of biometric traits selected by the user, the number of subjects, and their images in the dataset. However, the test time is less than 4 seconds on average, which is fast enough to be utilized for real-time applications.

5.4.1 Multiple Algorithms Face Recognition

The performance of the proposed multiple algorithms face recognition system was evaluated by conducting a number of extensive experiments on FERET and CAS-PEAL-R1 for the identification task, while the LFW dataset was used for the verification task. In these experiments, as described in (Chapter 3, Sect. 3.4.2), regularization parameters ($\boldsymbol{\alpha}$, $\boldsymbol{\beta}$) of the QDC classifier used in the *Curvelet-Fractal* approach were set to 0.1 and 0.2, respectively. On the other hand, the MDFR framework, trained on input data acquired from the

Curvelet-Fractal approach using a **3**-layer DBN model with hidden layers of sizes (**800, 800, 1000)**, was greedily trained in a bottom-up manner. The first two layers were each trained separately as non-linear feature detectors (RBM models) in an unsupervised way, using the CD learning algorithm, while the last layer was trained in a supervised way as a non-linear classifier (a discriminative RBM). Each individual RBM was trained for 300 epochs, with a momentum of 0.9, a weight-decay of 0.0002, a mini-batch size of 100, and the learning rate was set to be 0.01. Finally, in the fine-tuning phase, the whole network was trained in a top-down manner, using the back-propagation algorithm equipped for Dropout compensation to find optimized parameters and to avoid overfitting. The Dropout ratio was set to 0.5, and the number of epochs through the training set was determined using an early stopping procedure to be around 1000 epochs.

In the identification task, the robustness of the proposed multiple algorithms face recognition system was tested on the FERET and CAS-PEAL-R1 dataset. In these experiments, the matching scores obtained from the present *Curvelet-Fractal* approach and the MDFR framework were fused using several fusion methods at the score and rank level fusion. The Rank-1 identification rates (%), using different fusion methods at the score and rank level fusion, are listed in Table 5.1 and 5.2, respectively. In general, the results obtained show a noticeable improvement in the performance of the proposed multiple algorithms face recognition system, compared to only applying the *Curvelet-Fractal* approach or the MDFR framework, by achieving higher Rank-1 identification rates at all the implemented fusion methods. From Table 5.1 and 5.2, one can see that the highest Rank-1 identification rates on all the probe sets of both face datasets were achieved by employing the weighted sum rule and Highest rank method at the score and rank level fusion, respectively. Finally, additional improvements and new state-of-the-art recognition rates were reached, using the proposed multiple algorithms face recognition system on the FERET and the CAS-PEAL-R1 dataset; in particular, when the most challenging probe sets are under consideration, such as **Dup.I** and **Dup.II** in the FERET dataset and **PE**, **PA**, **PL**, and **PS** in the CAS-PEAL-R1dataset.

Table 5.1: Rank-1 identification rate (%) of the proposed multiple algorithms face recognition system on FERET and CAS-PEAL-R1 dataset using score level fusion.

Dataset		*Curvelet-Fractal*	MDFR	Score Fusion Methods				
				SR	WSR	PR	Max	Min
FERET	Fc	100	100	100	100	100	100	100
	Fc	98.97	100	98.97	100	98.99	98.96	98.98
	Dup.I	97.92	98.40	97.94	98.92	98.22	97.97	98.20
	Dup.II	95.72	97.86	95.75	98.72	96.42	95.77	96.75
CAS-PEAL_R1	PE	99.87	100	99.89	100	99.93	99.90	99.91
	PA	98.07	99.43	99.63	99.93	99.49	98.49	98.52
	PL	89.48	89.92	89.97	91.94	89.99	89.52	89.49
	PT	100	100	100	100	100	100	100
	PB	100	100	100	100	100	100	100
	PS	99.64	100	99.94	100	99.71	99.68	99.66

Table 5.2: Rank-1 identification rate (%) of the proposed multiple algorithms face recognition system on FERET and CAS-PEAL-R1 dataset using rank level fusion.

Dataset		*Curvelet-Fractal*	MDFR	Rank Fusion Methods		
				HR	BC	LR
FERET	Fc	100	100	100	100	100
	Fc	98.97	100	100	99.95	100
	Dup.I	97.92	98.40	98.86	98.12	98.62
	Dup.II	95.72	97.86	97.94	96.85	97.85
CAS-PEAL_R1	PE	99.87	100	100	99.97	99.95
	PA	98.07	99.43	99.87	98.63	99.32
	PL	89.48	89.92	90.62	89.56	89.85
	PT	100	100	100	100	100
	PB	100	100	100	100	100
	PS	99.64	100	100	99.87	100

Furthermore, the LFW dataset was used to evaluate the robustness and the effectiveness of the proposed multiple algorithms face recognition system to address the unconstrained face verification problem. As described in (Chapter 3, Sect. 3.4.3), the "*Image-Restricted, Label-Free Outside Data*" protocol is followed, where only the outside data is used to train MDFR framework, and the aligned LFW-a version is employed as a testing benchmark. The general flow of information, when the proposed multiple algorithms face recognition system operates in the verification mode, is shown in Fig. 5.6. The final performance is reported in Table 5.3 by calculating the mean accuracy rate ($\hat{\mu}$) and the standard error of the mean accuracy (S_E) over 10-folds cross-validation using different score fusion methods, and the corresponding ROC curves are shown in Fig. 5.7. From Table 5.3 and Fig, 5.7, the highest accuracy rate was obtained using the weighted sum rule as a fusion mothed in the score level, where a higher weight was assigned to the MDFR framework compared to the *Curvelet-Fractal* approach. The accuracy rate has been improved by 3.58% and 0.97% compared to the *Curvelet-Fractal* approach and the MDFR framework, respectively.

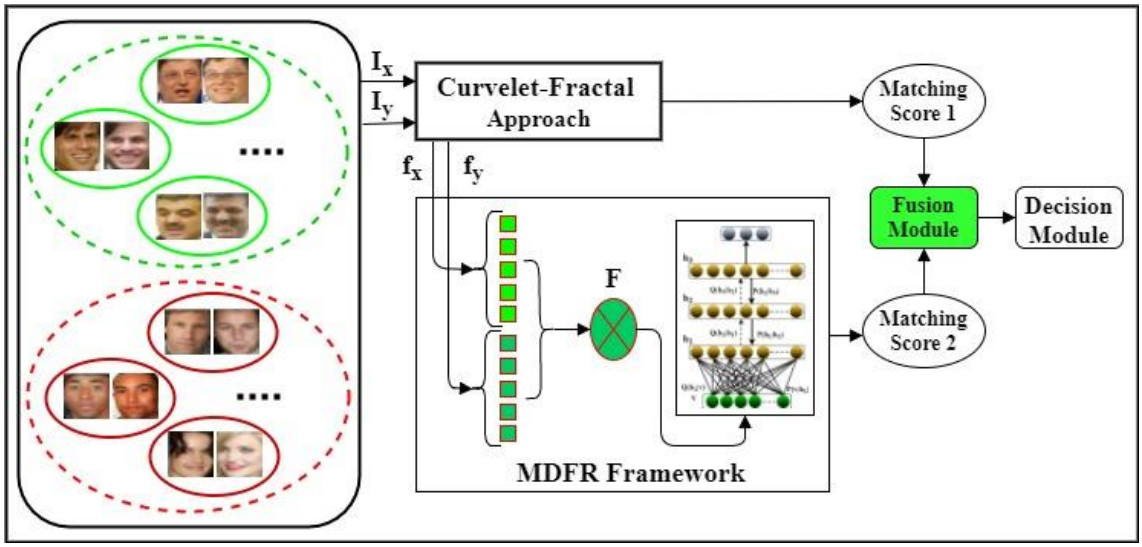


Figure 5.6: Flow of information when the proposed multiple algorithms face recognition system operates in the verification mode.

Table 5.3: Performance comparison between the proposed multiple algorithms face recognition system on LFW dataset using different score fusion methods.

Curvelet-Fractal	MDFR	Acc. ($\hat{\mu} \pm S_E$)				
		SR	WSR	PR	Max	Min
0.9622 ± 0.0272	0.9883 ± 0.0121	0.9832 ± 0.0186	0.9980 ± 0.0198	0.9837± 0.0194	0.9748 ± 0.0187	0.9840 ± 0.0201

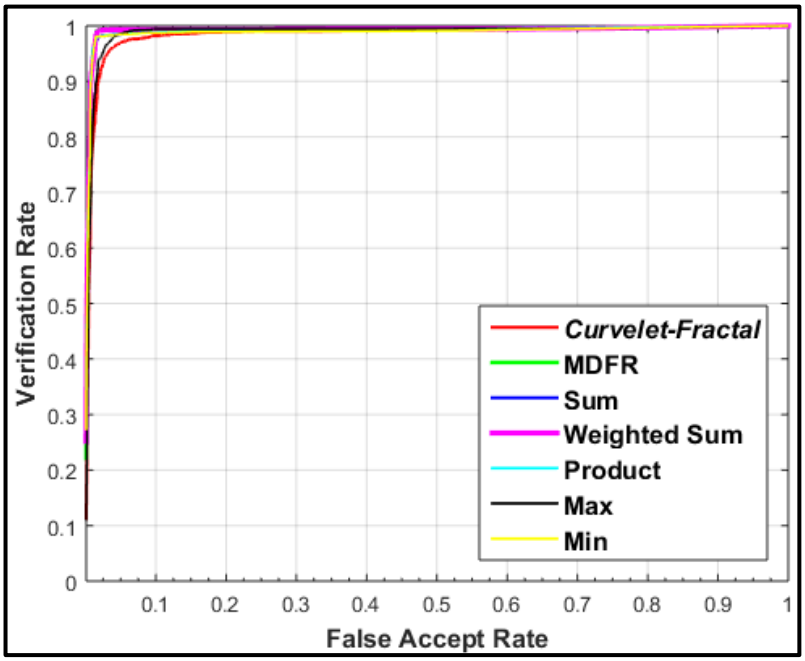


Figure 5.7: ROC curves using different score fusion methods.

5.4.2 Multiple Instances Iris Recognition

In this section, a number of extensive experiments to evaluate the robustness of the proposed multiple instance iris recognition system on the most challenging iris datasets currently available in the public domain is described. Two iris datasets, namely, CASIA-Iris-V3 Interval, and IITD were employed as testing benchmarks, and for comparing the results obtained with current state-of-the-art approaches. Used as an iris identification system, each time a query sample is presented, the similarity score is computed by comparing it against the templates of ***N*** different subjects registered in the dataset and a vector of ***N*** matching scores is produced by the classifier. These matching scores are obtained from both the left and right iris images of the same person and fused using different fusion methods at the score and rank

level. In the rank level fusion, the matching scores from each iris are arranged in descending order to form the ranking list of matching identities, where a higher rank number indicates a better match.

Table 5.4 and 5.5 show the Rank-1 identification rate (%) for both left and right iris images in the CASIA-Iris-V3 Interval and IITD dataset, and their fusion rates at the score and rank level, respectively. In these experiments, the same weights were assigned for both the left and right iris images in the application of the weighted fusion methods. Although all the fusion methods produced the same level of accuracy, the Highest ranking method was adopted for comparing the performance of the proposed multiple instance iris recognition system with that of other existing systems, due to its efficiency compared to the other fusion methods in exploiting the strength of each classifier effectively, and breaking the ties between the subjects in the final ranking list. In addition, it is simpler than the weighted fusion methods, such as weighted sum rule and logistic regression method, which need a training phase to find the weight for each individual classifier.

Table 5.4: Rank-1identification rate (%) of the proposed multiple instance iris recognition system using score level fusion.

Dataset	R. Iris	L. Iris	Score Fusion Methods				
			SR	WSR	PR	Max	Min
CASIA-Iris-V3 Interval	99.94	99.88	100	100	100	100	100
IITD	99.82	99.92	100	100	100	100	100

Table 5.5: Rank-1identification rate (%) of the proposed multiple instance iris recognition system using rank level fusion.

Dataset	R. Iris	L. Iris	Rank Fusion Methods		
			HR	BC	LR
CASIA-Iris-V3 Interval	99.94	99.88	100	100	100
IITD	99.82	99.92	100	100	100

The comparison of performance of the proposed multiple instance iris recognition system with the other existing methods on the CASIA-Iris-V3 Interval and ITD dataset is demonstrated in Table 5.6. In this table, the percentage accuracy as CRR for identification, is presented along with the recognition time in seconds per iris image. The feature extraction and classification techniques used in most of these methods, along with their evaluation protocols are shown in (Chapter 4, Table 4.7). It is assumed that these existing methods, as shown in Table 5.6, are customized for these two iris datasets and the best results they obtained are quoted herein. From Table 5.6, the proposed deep learning approach has overall, outperformed all the state-of-art feature extraction methods, which include DWT, DCT, PCA, ALBP, etc. In term of the Rank-1 identification rate, the highest results were obtained by the proposed system using these two datasets. Although the works described in [146][29] also achieved a 100% recognition rate for the CASIA-Iris-V3 Interval dataset, the proposed system achieved a better running time to establish the person's identity from 120 persons from the same dataset instead of 99 persons, as in [146], and 100 persons as in [29]. In addition, they obtained inferior results on the IITD dataset in terms of both Rank-1 identification rate and running time. For the IITD dataset, even though the methods of Elgamal and Al-Biqami [142] show a good performance, they evaluated their system on a small subset of 80 persons, compared to 244 persons in the author’s experiments. Finally, the results obtained demonstrate that the proposed multiple instance iris recognition system outperforms the indicated state-of-the-art of approaches in terms of Rank-1 identification rate, in 19 out of 21 cases, and, in terms of running time, in 12 out of 12 cases, where this information is available using both datasets.

Table 5.6: Comparison of the proposed system with other existing approaches using two different iris datasets.

Dataset	Approach	CRR (%)	Time (s)	Fusion Method
CASIA-Iris-V3 Interval	**Vatsa et al.** [340]	97.21	1.82	2ν-SVM
	Kerim and Mohammed [333]	99.40	2	-
	Umer et al. [146]	**100**	0.98	BC
	Umer et al. [29]	**100**	1.05	BC
	Umer et al. [28]	99.38	0.93	-
	De Costa and Gonzaga [334]	99.10	-	-
	Ng et al. [341]	98.45	-	-
	Zhang and Guan [335]	99.60	-	-
	Roy et al. [336]	97.21	0.995	-
	Li et al. [27]	99.91	-	-
	Tsai et al. [342]	99.97	-	-
	***The Proposed* System**	**100**	**0.89**	HR
IITD	**Umer et al.** [146]	99.52	1.11	BC
	Umer et al. [29]	99.46	1.13	BC
	Umer et al. [28]	98.03	1.00	-
	Umer et al. [145]	98.37	1.18	BC
	Elgamal and Al-Biqami [142]	99.50	3.00	-
	Nalla and Chalavadi [148]	86.00	-	-
	Elgamal and Al-Biqami [142]	99.50	-	-
	Minaee et al. [338]	99.20	-	-
	Dhage et al. [149]	97.81	93.24	-
	Abhiram et al. [339]	97.12	-	-
	***The Proposed* System**	**100**	**0.81**	HR

5.4.3 Hybrid Multimodal Biometric System

The performance of the proposed hybrid multimodal biometric system was tested on the real multi-biometric dataset, SDUMLA-HMT, where the face and both irises traits were acquired from the same person, is described. In this work, two different architectures of a hybrid multimodal biometric identification system were investigated, termed as *Architecture_1* and *Architecture_2,* as shown in Fig. 5.1 and 5.3, respectively. Starting with a less complex architecture that may be called *Architecture_2*, the matching scores of the face and both irises biometric traits were obtained using deep learning approaches by employing a DBN approach and *IrisConvNet* system for encoding face and iris trait, respectively. On the other hand, the feature representations of the face and both irises traits were extracted using different approaches, the facial features

were extracted using *Curvelet-Fractal*/MDFR approach, while the iris features were extracted using the *IrisConvNet* system in the *Architecture_1*. Then, the fusion of the face and both irises matching scores for both architectures was evaluated using different fusion methods at the score and rank level. In this work, the effectiveness and robustness of the proposed hybrid multimodal biometric systems have been evaluated with different scenarios, based on the biometric traits selected by the user at the identification point. The Rank-1 identification rates (%) for *Architecture_1* and *Architecture_2*, using different fusion methods are listed in Table 5.7 and 5.8, respectively. In general, the results obtained show a very clear improvement in the performance of the proposed multimodal biometric system for both architectures compared to unimodal systems, by achieving higher Rank-1 identification rates at all the implemented scenarios. In addition, a better performance is obtained by fusing the face along with one or both irises compared to the performance of fusion solely the face trait (e.g. Using Multi-Algorithms), due to the high recognition accuracy obtained using iris biometric trait compared with the face biometric trait. From Table 5.7, the highest identification rate was obtained using the weighted sum rule to fuse the matching scores generated from the *Curvelet-Fractal* approach and the MDFR framework, when only the face trait is selected to establish the person's identity.

Although high identification rates obtained using both architectures, the proposed multimodal biometric system using *Architecture_1* provides a higher degree of flexibility than *Architecture_2* allowing the user to provide a subset or all of his/her biometric traits depending on the nature of the implemented application, required security level and the user's convenience. Finally, the authors in [343] have proposed a multimodal biometric system using only the face trait in SDUMLA-HMT dataset. Therefore, for purpose of comparison, the best result obtained from fusion only of face recognition approaches using the weighted sum rule is employed, and the results are listed in Table 5.9. It can be seen that the accuracy rate of 100 % of the proposed system is higher than the best results reported in [343], which is 96.54 % using the CLVQ approach.

Table 5.7: Rank-1identification rate (%) of the proposed hybrid multimodal biometric system (*Architecture_1*) using different scenarios of fusion at score and rank level fusion.

Scenarios	Score Fusion Methods					Rank Fusion Methods		
	PR	SR	WSR	Max	Min	HR	BC	LR
Face + Face (Using Multi-Algorithms)	99.95	99.34	100	99.55	99.68	99.97	99.21	99.95
R. Iris + L. Iris (Using Multi-Instances)	100	100	100	100	100	100	100	100
R. Iris + L. Iris + Face (Using Curvelet-Fractal)	100	100	100	100	100	100	100	100
R. Iris + L. Iris + Face (Using MDFR Framework)	100	100	100	100	100	100	100	100
R. Iris + Face (Using Curvelet-Fractal)	100	100	100	100	100	100	100	100
L. Iris+ Face (Using MDFR Framework)	100	100	100	100	100	100	100	100
Hybrid Multimodal biometrics System	100	100	100	100	100	100	100	100

Table 5.8: Rank-1identification rate (%) of the proposed hybrid multimodal biometric system (*Architecture_2*) using different scenarios of fusion at score and rank level fusion.

Scenarios	Score Fusion Methods					Rank Fusion Methods		
	PR	SR	WSR	Max	Min	HR	BC	LR
R. Iris +L. Iris	100	100	100	100	100	100	100	100
L. Iris + Face	99.91	99.95	100	99.88	99.98	100	99.83	99.92
R. Iris + Face	100	100	100	100	100	100	100	100
R. Iris + L. Iris + Face	100	100	100	100	100	100	100	100

Table 5.9: Comparison of the proposed multimodal system using face trait with the state-of-the-art approaches on SDUMLA-HMT dataset.

Approach	CRR (%)
MLP [343]	93.35
CLVQ [343]	96.54
CRBF[343]	92.25
Face + Face (Using Multi-Algorithms)	**100**

5.5 Chapter Summary

In this Chapter, an efficient and real-time hybrid multimodal biometric identification system has been proposed using a parallel architecture to fuse the results obtained from the face and both irises of a subject. This type of architecture can offer people a high degree of flexibility to provide either a subset or all of their biometric traits, depending on the required security level and their convenience. In addition, it is efficient in handling the problem of missing biometric traits, whilst maintaining a high recognition rate. In this work, each individual biometric trait has been processed using different approaches. The facial features are extracted using a *Curvelet-Fractal*/MDFR approach, while the iris features are extracted using the *IrisConvNet* system. Three different types of the multimodal biometric system are proposed, namely, multiple algorithms face recognition, multiple instances iris recognition, and hybrid multimodal biometric system. The performances of the proposed multiple algorithms face recognition system was evaluated in the identification task using the FERET and CAS-PEAL-R1dataset, in which the face images were captured under different conditions of facial expressions, lighting conditions, etc. The results obtained have indicated the reliability and efficiency of the proposed face recognition approaches when there is only one image in the gallery set. The same system was also tested to address the unconstrained face verification problem, and a new state-of-the-art accuracy rate has been achieved on the LFW dataset. Further, new state-of-the-art results were produced, by applying the proposed multiple instances iris recognition system on real multimodal datasets, namely, CASIA-Iris-V3 Interval and IITD iris datasets, in which the left and right iris images were captured from the same person. Finally, the SDUMLA-HMT as a real multibiometric dataset has been used to assess the performance of the proposed hybrid multimodal biometric system for different scenarios based on the biometric traits selected by the user at the identification point. The results obtained have clearly shown the efficiency of the proposed multimodal biometric system using two different architectures with these scenarios compared to unimodal systems in term of Rank-1 identification rates using different fusion methods at the score and rank level.

Chapter 6

Conclusions and Future Work

This thesis has addressed the problem of designing and implementing a novel multimodal biometric identification system, based on fusing the matching scores obtained from the face and both irises using different fusion approaches. After a summary of the state-of-the-art works in both the unimodal and multimodal biometric systems that employ either the face or iris traits in the decision-making process, a number of efficient and robust algorithms have been proposed for encoding the most prominent features in the face and iris traits, based on combining the advantages of the local handcrafted-descriptors/deep learning approaches towards building a novel hybrid multimodal biometric system for personal identification. This chapter briefly summarizes the work presented in this PhD thesis. Firstly, the main contributions achieved in each chapter are briefly discussed and summarized. This is followed by pointing out some of the suggestions and recommendations to be addressed in future work in order to further enhance the security and increase the reliability of both the unimodal and multimodal biometric systems.

6.1 Conclusions

Nowadays, using traditional methods such as username-passwords, ID cards and PINs does not achieve the required level of security, because the security of the system can be easily broken when the ID card is stolen or the password is guessed by an imposter. This problem can be solved by verifying persons' identity based on their biometric traits. Multimodal biometric systems seek to alleviate some of these limitations of unimodal biometric systems by combining multiple pieces of evidence of the same person in the decision-making process. In this work, the author believes that a considerable progress has been made with regards to integrating both the face and the left and right irises in a unified hybrid multimodal biometric identification system.

As demonstrated in Chapter 2, the main limitation of the experimental results in most of the previous works is using a "chimeric" multimodal database

instead of a real multimodal biometric database, along with the lack of flexibility when one of the biometric traits is unavailable or missed.

In Chapter 3, a novel multimodal local feature extraction approach was proposed, based on merging the advantages of multidirectional and anisotropy transforms like the Curvelet transform with Fractal dimension. The main contribution of this approach is to apply the Curvelet transform as a fast and powerful technique for representing edges and curves of the face structure, and then to process the Curvelet coefficients in different frequency bands using two different Fractal dimension approaches to efficiently reflect the face texture under unconstrained environmental conditions. The proposed approach has been tested on four large-scale unconstrained face datasets (e.g. SDUMLA-HMT, FERET, CAS-PEAL-R1 and LFW dataset), with high diversity in facial expressions, lighting conditions, noise, etc. The results obtained demonstrated the reliability and efficiency of the *Curvelet-Fractal* approach by achieving competitive results with the state-of-the-art approaches (e.g. G-LQP, LBP, WPCA, etc.), especially when there is only one image in the gallery set. The author is also aware that a possible source of error using the *Curvelet-Fractal* approach could result from the full-profile face images, which suffer from severe self-occlusion. However, this drawback has been significantly alleviated using a novel MDFR framework, in which additional and complementary information were automatically and conjointly learned by applying the DBN on top of the local feature representations obtained from the *Curvelet-Fractal* approach. Extensive experiments were conducted, and a new state-of-the-art accuracy rate was achieved by applying the proposed MDFR framework on all the employed datasets. Based on the results obtained, it can be concluded that the proposed *Curvelet-Fractal* approach and MDFR framework can be readily used in real face recognition system for both the identification and the verification task with different face variations.

In the iris recognition module, described in Chapter 4, a robust and fast iris recognition system called *IrisConvNet* system, to identify the person's identity by constructing a deep learning based system for either the right or left irises of the same person, was developed. The proposed system starts by applying an automatic and real-time iris localization model to detect the iris region using

CCHT, which has significantly increased the overall accuracy and reduced the processing time of the subsequent stages in the proposed system. Despite the fact that the proposed iris localization model was designed to detect the iris region within grayscale images, the main contribution of the proposed model lies in its ability to detect accurately the iris boundaries accurately, which can greatly improve the accuracy of the proposed iris recognition system. Furthermore, its speed in detecting the iris boundaries is less than one second, which is comparable to the face detection speed.

For iris texture encoding features, an efficient deep learning system based on a combination of the CNN and Softmax classifier, was proposed. A powerful training methodology equipped with a number of training strategies has also been proposed, in order to control overfitting during the learning process and increase the generalization ability of the neural network. The effectiveness and robustness of the proposed iris localization and recognition approaches have been tested on five challenging datasets: SDUMLA-HMT, CASIA-Iris-V1, CASIA-Iris-V3 Interval, MMU1 and IITD iris datasets. Extensive experiments have been conducted on these datasets to evaluate different numbers of training parameters (e.g. learning rate, number of layers, number of filters per each layer, etc.), in order to build an optimum CNN as the framework for the proposed iris identification system. The experimental results demonstrated the superiority of the proposed system over recently reported iris recognition systems, by achieving a new state-of-the-art Rank-1 identification rate on all the five datasets. The system is fast, as it requires less than one second to establish the person's identity.

In the fusion module, an efficient and real-time hybrid multimodal biometric identification system was proposed, using a parallel architecture to fuse the results obtained from the face and both irises of a subject, as described in Chapter 5. Three different types of the multimodal biometric system were proposed, namely, multiple algorithms face recognition, multiple instances iris recognition, and hybrid multimodal biometric system. The performances of the proposed multiple algorithms face recognition system was evaluated in the identification task using the FERET and CAS-PEAL-R1dataset, in which the face images were captured under different conditions of facial expressions,

lighting conditions, etc. The experimental results indicate the reliability and efficiency of the proposed face recognition approaches, especially when there is only one image in the gallery set. New state-of-the-art recognition rates are achieved, using all the probe sets of these two face databases. The same system was also tested to address the unconstrained face verification problem, and a new state-of-the-art accuracy rate is achieved on the LFW dataset. Furthermore, a new state-of-the-art Rank-1 identification rate of 100% produced by applying the proposed multiple instances iris recognition system on real multimodal datasets, namely, CASIA-Iris-V3 Interval and IITD iris datasets, in which the left and right iris images were captured from the same person. Finally, the SDUMLA-HMT as a real multibiometric dataset, has been used to assess the performance of the proposed hybrid multimodal biometric system for different scenarios, based on the biometric traits selected by the user at the identification point. The results obtained have clearly shown the efficiency of the proposed multimodal biometric system using two different architectures with these scenarios, compared to unimodal systems in term of Rank-1 identification rates using different fusion methods at the score and rank level. All of the proposed multimodal biometric systems can work in a real-time to meet the requirements of real-world applications.

To conclude, this thesis has accomplished and fulfilled all objectives and aims, as stated in Chapter 1. Despite the long training time required, the hybrid multimodal biometric identification system presented in this PhD thesis is highly customizable, in terms of selecting different recognition scenarios and different approaches to encode the selected biometric traits at the decision-making process. Furthermore, it has been designed and implemented in a way that allows adding more approaches and biometric traits (e.g., fingerprint trait) at a later time. Hence, the proposed hybrid multimodal biometric system can be improved significantly by selecting robust recognition approaches that offer a better performance in different recognition scenarios, based on the selected biometric traits. All experimental results were carefully documented and presented in this thesis, and, unlike previous multimodal biometric systems, real multi-biometric databases were employed to assess the performance of the proposed approaches.

6.2 Future Work

Although a significant progress has been achieved in this PhD thesis in recognising the identities of people based on a combination of the face and both irises biometric traits using the proposed unified multimodal biometric system, there are some limitations that need to be addressed in the near future, such as:

1. Firstly, the performance of the proposed hybrid multimodal biometric system presented in Chapter 5 should be tested using other real multimodal datasets. In particular, it would be necessary to validate further the efficiency and reliability of the proposed hybrid multimodal biometric system using a larger real multimodal dataset, containing more individuals with images captured under more challenging conditions. Unfortunately, to the best of our knowledge, there is no other free variable, real dataset, which contains the face and both irises biometric traits of the same person that can be used to evaluate the performance of the proposed system.
2. The performance of the proposed hybrid multimodal biometric system presented in Chapter 5 can be enhanced further by incorporating the quality measures of the biometric traits and feeding them to the adopted classifiers, either as separate features, weights, or combined with the obtained scores from each individual classifier. Several image quality assessment algorithms, for both the face and iris biometric traits, have been proposed in the literature, as in [344][345] [346] [347] [348]. For face recognition, the face image quality measures are divided into two categories: texture quality measures (e.g., sharpness, resolution, contrast and light intensity, compression ratio, etc.) and other characteristics directly associated with the facial features (e.g., position, symmetry, rotation, eyes visibility, the appearance of shadows or glare on the face). On the other hand, some of quality metrics used for iris recognition include occlusion (e.g., eyelids eyelashes, hair, and glasses) focus and motion blur (e.g., due to hand-held sensors), off-angle (e.g., angular deformation), and image resolution [349].
3. As the performance of the proposed approaches in chapter 4 and 5 is limited to the identification mode, it could be interesting to evaluate the performance in different and new scenarios, such as verification

applications. In particular, the performance of the proposed *IrisConvNet* system could be further assessed in addressing the problem of the heterogeneous iris verification, which can be further divided into cross-resolution iris verification and cross-sensor iris verification. In the cross-resolution iris verification, the aim is to match low-resolution testing iris images with high-resolution training images captured at the enrolment phase [350]. On the other hand, the aim of the cross-sensor iris verification is to measure the similarity between iris images captured using different sensors along with different sensors settings [351][352].

4. The proposed iris localization model is designed to detect the iris region within grayscale images. This model could be further extended in the future to deal with coloured iris images, which might contain additional information and can help to enhance the accuracy of detecting iris boundaries and minimize the identification error rates.

5. A well-known limitation of using deep learning approaches is the long time required to train the neural network. This is the main limitation of the proposed *IrisConvNet* system. Although the training time is not as important as the testing time in many applications, like the iris recognition system, using high-performance GPUs and code optimization can significantly reduce the time required to train the proposed *IrisConvNet* system. Furthermore, a number of powerful textural edgeness descriptors can be employed to process the detected iris region (e.g., LBP, ALBP, FBM, etc.), and then to apply the proposed *IrisConvNet* system on the feature representations generated from one or more of these descriptors. It can be argued that the overall performance can be improved, and the training time can be reduced, by feeding the CNN with only useful iris feature representations.

6. The proposed approaches in this dissertation can also be applied to other biometric traits. Firstly, the performance of the hybrid multimodal biometric system presented in Chapter 5 has been tested using only two biometric traits (face and iris trait). It would be interesting to integrate other biometric traits in conjunction with the employed traits to enhance the recognition performance. For instance, physiological traits (e.g., fingerprint, palm vein, etc.) or behavioural traits (e.g., signature, voice, etc.) can be potential

candidates for the proposed approaches. Secondly, studying the application of the proposed face recognition approaches (*Curvelet-Fractal* approach/MDFR framework) to other biometric traits, in which the edges and curves representations can play a significant role in improving their recognition performance, such as fingerprint, finger/palm vein, etc. Finally, studying the potential use of the DBN approach to address the problem of the iris recognition system in both identification and verification tasks. This has led Baqar et al. [353] to propose an iris recognition system that operates in the identification mode. However, the main weakness in their study is that they evaluated their system using very small iris dataset of 10 classes. Therefore, there are still some interesting and relevant problems to be addressed.

References

[1] H. Imtiaz and S. A. Fattah, “A Curvelet Domain Face Recognition Scheme Based on Local Dominant Feature Extraction,” *ISRN Signal Process.*, vol. 2012, pp. 1–13, 2012.

[2] A. K. Jain, A. Ross, and S. Prabhakar, “An Introduction to Biometric Recognition,” *IEEE Trans. CIRCUITS Syst. VIDEO Technol.*, vol. 14, no. 1, pp. 1–29, 2004.

[3] A. Ross, K. Nandakumar, and J. Anil K, “Handbook of Multibiometrics,” *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2006.

[4] B. S. Atal, “Automatic Recognition of Speakers from Their Voices,” *Proc. IEEE*, vol. 64, no. 4, pp. 460–475, 1976.

[5] A. K. Jain, A. A. Ross, and K. Nandakumar, “Introduction to Biometrics,” in *Springer Publishing Company*, 2011.

[6] D. Zhang and G. Lu, “3D Biometrics Systems and Applications,” in *Springer Science+Business Media New York*, pp. 1–290, 2013.

[7] S. Marcel, M. Nixon, Li, S. Z., “Handbook of Biometric Anti-Spoofing-Trusted Biometrics under Spoofing Attacks,” *Advances in Computer Vision and Pattern Recognition-Springer,* pp.1-279, 2014.

[8] P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. O’Toole, D. S. Bolme, J. Dunlop, Y. M. Lui, H. Sahibzada, and S. Weimer, “An introduction to the good, the bad, & the ugly face recognition challenge problem,” *IEEE Int. Conf. Autom. Face Gesture Recognit. Work. FG 2011*, pp. 346–353, 2011.

[9] P. J. Phillips, P. J. Flynn, J. R. Beveridge, W. T. Scruggs, A. J. O’Toole, D. Bolme, K. W. Bowyer, B. A. Draper, G. H. Givens, Y. M. Lui, H. Sahibzada, J. A. Scallan, and S. Weimer, “Overview of the multiple biometrics grand challenge,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5558 LNCS, pp. 705–714, 2009.

[10] N. Houmani, A. Mayoue, S. Garcia-Salicetti, B. Dorizzi, M. I. Khalil, M. N. Moustafa, H. Abbas, D. Muramatsu, B. Yanikoglu, A. Kholmatov, M. Martinez-Diaz, J. Fierrez, J. Ortega-Garcia, J. Roure Alcobé, J. Fabregas, M. Faundez-Zanuy, J. M. Pascual-Gaspar, V. Cardeñoso-Payo, and C. Vivaracho-Pascual, “BioSecure signature evaluation campaign (BSEC2009): Evaluating online signature algorithms depending on the quality of signatures,” *Pattern Recognit.*, vol. 45, no. 3, pp. 993–1003, 2012.

[11] A. Jain, P. Flynn, and A. A. Ross, “Handbook of Biometrics,” in *Springer-Verlag US*, pp. 1–556, 2008.

[12] M. Tistarelli, S. Z. Li, and C. Rama, “Handbook of Remote Biometrics- for Surveillance and Security,” in *Springer-Verlag London*, pp. 1–281, 2009.

[13] A. Jain, K. Nandakumar, and A. Ross, “Score normalization in multimodal biometric systems,” *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, 2005.

[14] B. Bhanu and V. Govindaraju, “Multibiometrics for Human Identification,” in *Cambridge University*, pp. 1–408, 2011.

[15] S. Prabhakar, S. Pankanti, and a K. Jain, “Biometric recognition: security and privacy concerns,” *IEEE Secur. Priv. Mag.*, vol. 1, pp. 33–42, 2003.

[16] R. Gad, A. EL-SAYED, M. Zorkany, and N. El-fishawy, “Multi-Biometric Systems : A State of the Art Survey and Research Directions,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 6, no. 6, pp. 128–138, 2015.

[17] Y. Chen, S. Dass, and A. Jain, “Fingerprint quality indices for predicting authentication performance,” *Lect. Notes Comput. Sci.*, pp. 160–170, 2005.

[18] K. Nandakumar, “Multibiometric Systems : Fusion Strategies and Template Security,” *PhD thesis, Dep. Comput. Sci. Eng.*, pp. 1–209, 2008.

[19] F. Matta, “Video person recognition strategies using head motion and facial appearance,” *PhD thesis, Univ. Nice Sophia-antip.*, pp. 1–148, 2008.

[20] Z. Wei, X. Qiu, Z. Sun, T. Tan, Zhuoshi Wei, Xianchao Qiu, Zhenan Sun, and Tieniu Tan, “Counterfeit iris detection based on texture analysis,” in *In Pattern Recognition, 2008. ICPR 19th IEEE International Conference on*, pp. 1–4, 2008.

[21] S. Sahoo, T. Choubisa, and S. Mahadeva Prasanna, “Multimodal Biometric Person Authentication : A Review,” *IETE Tech. Rev.*, vol. 29, no. 1, p. 54, 2012.

[22] A. Ross and A. K. Jain, “Multimodal Biometrics : an Overview,” in *Signal Processing Conference, 12th European. IEEE*, pp. 1221–1224, 2004.

[23] A. Lumini and L. Nanni, “Overview of the combination of biometric matchers,” *Inf. Fusion*, vol. 33, pp. 71–85, 2017.

[24] T. K. Ho, J. J. Hull, and S. N. Srihari, “Decision Combination in Multiple Classifier Systems,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. I, pp. 66–75, 1994.

[25] J. Z. Liang and E. Power, “Iris Recognition based on Block Theory and Self-adaptive Feature Selection,” *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 8, no. 2, pp. 115–126, 2015.

[26] A. G. Gale and S. S. Salankar, “Evolution of Performance Analysis of Iris Recognition System By using Hybrid Methods of feature Extraction and Matching by Hybrid Classifier for Iris Recognition System,” in *Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on. IEEE*, pp. 1–5, 2016.

[27] C. Li, W. Zhou, and S. Yuan, “Iris recognition based on a novel variation of local binary pattern,” *Vis. Comput.*, vol. 31, no. 10, pp. 1419–1429, 2015.

[28] S. Umer, B. C. Dhara, and B. Chanda, “An Iris Recognition System Based on Analysis of Textural Edgeness Descriptors,” *IETE Tech. Rev.*, vol. 4602, no. September, pp. 1–12, 2017.

[29] S. Umer, B. C. Dhara, and B. Chanda, “A novel cancelable iris recognition system based on feature learning techniques,” *Inf. Sci. (Ny).*, vol. 406–407, pp. 102–118, 2017.

[30] D. Menotti, G. Chiachia, A. Pinto, W. R. Schwartz, H. Pedrini, A. X. Falcão, and A. Rocha, “Deep Representations for Iris, Face, and Fingerprint Spoofing Detection,” *IEEE Trans. Inf. Forensics Secur.*, vol. 10, no. 4, pp. 864–879, 2015.

[31] P. Silva, E. Luz, R. Baeta, H. Pedrini, A. X. Falcao, and D. Menotti, “An Approach to Iris Contact Lens Detection Based on Deep Image Representations,” *28th SIBGRAPI Conf. Graph. Patterns Images*, pp.

157–164, 2015.

[32] C. Ding and D. Tao, “A Comprehensive Survey on Pose-Invariant Face Recognition,” *ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 3, pp. 1–40, 2016.

[33] L. Wiskott, J. M. Fellous, N. Krüger, and C. von der Malsburg, “Face recognition using by Elastic Bunch Graph Matching,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, 1997.

[34] T. Kanade, “Picture Processing System by Computer Complex and Recognition of Human Faces,” *Dr. Diss. Kyoto Univ.*, p. 178, 1973.

[35] R. Brunelli and T. Poggio, “Face Recognition Features versus Templates,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 10, pp. 1042–1052, 1993.

[36] S. Biswas, G. Aggarwal, P. J. Flynn, and K. W. Bowyer, “Pose-robust recognition of low-resolution face images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 3037–3049, 2013.

[37] S. Keypoints and D. G. Lowe, “Distinctive Image Features from,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[38] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N. Kumar, “Localizing parts of faces using a consensus of exemplars,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 12, pp. 2930–2940, 2011.

[39] X. Cao, Y. Wei, F. Wen, and J. Sun, “Face alignment by explicit shape regression,” *Int. J. Comput. Vis.*, vol. 107, no. 2, pp. 177–190, 2012.

[40] D. Chen, X. Cao, F. Wen, and J. Sun, “Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3025–3032, 2013.

[41] S. J. Prince and J. Elder, “Tied factor analysis for face recognition across large pose changes,” *Procedings Br. Mach. Vis. Conf. 2006*, p. 91.1-91.10, 2008.

[42] Y. Zhang, M. Shao, E. K. Wong, and Y. Fu, “Random faces guided sparse many-to-one encoder for pose-invariant face recognition,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 2416–2423, 2013.

[43] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li, “High-Fidelity Pose and Expression Normalization for Face Recognition in the Wild,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 787–796, 2015.

[44] R. Jafri and H. R. Arabnia, “A Survey of Face Recognition Techniques,” *J. Inf. Process. Syst.*, vol. 5, no. 2, pp. 41–68, 2009.

[45] I. J. Cox, G. Joumana, and P. N. Yianilos, “Feature-Based Face Recognition Using Mixture-Distance,” in *In Computer Vision and Pattern Recognition, 1996. Proceedings CVPR’96, 1996 IEEE Computer Society Conference on*, pp. 209–216, 1996.

[46] L. Sirovich and M. Kirby, “Low-dimensional procedure for the characterization of human faces.,” *Journal of the Optical Society of America. A, Optics and image science*, vol. 4, no. 3. pp. 519–524, 1987.

[47] M. a. M. Turk and A. A. P. Pentland, “Face Recognition Using Eigenfaces,” *Journal of Cognitive Neuroscience*, vol. 3, no. 1. pp. 72–86, 1991.

[48] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, “Face recognition by independent component analysis.,” *IEEE Trans. Neural Netw.*, vol. 13, no. 6, pp. 1450–64, 2002.

[49] B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge, “Recognizing faces with PCA and ICA,” *Comput. Vis. Image Underst.*, vol. 91, no. 1–2, pp. 115–137, 2003.

[50] P. Belhumeur, J. Hespanha, and D. Kriegman, “Eigenfaces Vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, 1997.

[51] J. Lu, S. Member, K. N. Plataniotis, and A. N. Venetsanopoulos, “Face Recognition Using Kernel Direct Discriminant Analysis Algorithms,” *IEEE Trans. NEURAL NETWORKS*, vol. 14, no. 1, pp. 117–126, 2003.

[52] C. Kambhamettu, X. X. W. and X. Tang, J. Y. and C. K. Qi Li, and X. Wang, “Linear Projection Methods in Face Recognition under Unconstrained Illuminations: A Comparative Study,” *Proc. 2004 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition, 2004. CVPR 2004.*, vol. 0, no. C, pp. 259–265, 2004.

[53] S. Prince and J. Elder, “Probabilistic Linear Discriminant Analysis for Inferences About Identity,” in *In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference o*, pp. 1–8, 2007.

[54] H. Zhao and P. C. Yuen, “Incremental linear discriminant analysis for face recognition.,” *IEEE Trans. Syst. man Cybern. Part B Cybern. a Publ. IEEE Syst. Man Cybern. Soc.*, vol. 38, no. 1, pp. 210–221, 2008.

[55] Vo Dinh Minh Nhat and S. Lee, “An improvement on PCA algorithm for Face Recognition,” *Second Int. Symp. Neural Networks, Chongqing, China, May 30 - June 1, 2005, Proceedings, Part I*, vol. 3496, pp. 1016–1021, 2005.

[56] N. Sun, H. X. Wang, Z. H. Ji, C. R. Zou, and L. Zhao, “An efficient algorithm for Kernel two-dimensional principal component analysis,” *Neural Comput. Appl.*, vol. 17, no. 1, pp. 59–64, 2008.

[57] X.-Y. Jing, Y.-Y. Tang, and D. Zhang, “A Fourier–LDA approach for image recognition,” *Pattern Recognit.*, vol. 38, no. 3, pp. 453–457, 2005.

[58] M. Loog, D. R.P.W, and R. Haeb-Umbach, “Multiclass Linear Dimension Reduction by Weighted Pairwise Fisher Criteria,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 7, pp. 762–766, 2001.

[59] H. Zhao and P. C. Yuen, “Analysis for Face Recognition,” *IEEE Trans. Syst. Man Cybern. Part B Cybern.*, vol. 38, no. 1, pp. 210–221, 2008.

[60] T. F. Cootes, G. J. Edwards, and C. J. Taylor., “Active Appearance Models,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 681–685, 2001.

[61] B. Yang and S. Chen, “A comparative study on local binary pattern (LBP) based face recognition: LBP histogram versus LBP image,” *Neurocomputing*, vol. 120, pp. 365–379, 2013.

[62] V. Blanz and T. Vetter, “Face recognition based on fitting a 3D morphable model,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, 2003.

[63] C. Lu, X. Liu, and W. Liu, “Face recognition based on two dimensional locality preserving projections in frequency domain,” *Neurocomputing*, vol. 98, pp. 135–142, 2012.

[64] K. Meethongjan and D. Mohamad, “A Summary of literature review : Face Recognition,” *Rev. Lit. Arts Am.*, vol. 2007, no. July, pp. 1–12, 2007.

[65] M. Fischer, H. K. Ekenel, and R. Stiefelhagen, “Analysis of partial least squares for pose-invariant face recognition,” *2012 IEEE Fifth Int. Conf. Biometrics Theory, Appl. Syst.*, pp. 331–338, 2012.

[66] T. Berg and P. Belhumeur, “Tom-vs-Pete Classifiers and Identity-Preserving Alignment for Face Verification,” *Procedings Br. Mach. Vis. Conf. 2012*, p. 129.1-129.11, 2012.

[67] A. Li, S. Shan, and W. Gao, “Coupled Bias – Variance Tradeoff for Cross-Pose Face Recognition,” *IEEE Trans. IMAGE Process.*, vol. 21, no. 1, pp. 305–315, 2012.

[68] Z. Zhu, P. Luo, X. Wang, and X. Tang, “Multi-View Perceptron : a Deep Model for Learning Face Identity and View Representations,” *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2014.

[69] C. Ding, J. Choi, D. Tao, and L. S. Davis, “Multi-Directional Multi-Level Dual-Cross Patterns for Robust Face Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 518–531, 2016.

[70] D. L. Swets, “Using discriminant eigenfeatures for image retrieval,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 831–836, 1996.

[71] M. M. Mukhedkar, “Fast Face Recognition Based on Wavelet Transform on PCA,” in *International Conference on Energy Systems and Applications (ICESA 2015)*, pp. 761–764, 2015.

[72] W. Fengxiang, “Face Recognition Based on Wavelet Transform and Adaptive Local Binary Pattern,” *J. Multimed.*, vol. 9, no. 8, pp. 158–166, 2014.

[73] T. Mandal and Q. M. J. Wu, “Face Recognition using Curvelet Based PCA,” in *In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*, pp. 1–4, 2008.

[74] M. Rziza, M. El Aroussi, M. El Hassouni, S. Ghouzali, and D. Aboutajdine, “Local Curvelet Based Classification Using Linear Discriminant Analysis for Face Recognition,” *J. Comput. Sci.*, pp. 72–77, 2009.

[75] S. Elaiwat, M. Bennamoun, F. Boussaid, and A. El-Sallam, “A Curvelet-based approach for textured 3D face recognition,” *Pattern Recognit.*, vol. 48, no. 4, pp. 1231–1242, 2015.

[76] S. I. Ch’ng, K. P. Seng, and L.-M. Ang, “Curvelet-based illumination invariant feature extraction for face recognition,” in *2010 International Conference on Computer Applications and Industrial Electronics*, pp. 458–462, 2010.

[77] M. Tang and F. Chen, “Facial expression recognition and its application based on curvelet transform and PSO-SVM,” *Opt. - Int. J. Light Electron Opt.*, vol. 124, no. 22, pp. 5401–5406, Nov. 2013.

[78] L. Zhou, W. Liu, Z. Lu, and T. Nie, “Face recognition based on curvelets and local binary pattern features via using local property preservation,” *J. Syst. Softw.*, vol. 95, pp. 209–216, Sep. 2014.

[79] S. Arivazhagan, J. R. Sekar, and S. S. Priyadharshini, “Curvelet and Ridgelet-based Multimodal Biometric Recognition System using Weighted Similarity Approach,” *Def. Sci. J.*, vol. 64, no. 2, pp. 106–114, 2014.

[80] Z. Chai, Z. Sun, H. Mendez-Vazquez, R. He, and T. Tan, “Gabor ordinal measures for face recognition,” *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 1, pp. 14–26, 2014.

[81] Z. Li, D. Gong, X. Li, and D. Tao, “Learning compact feature descriptor and adaptive matching framework for face recognition,” *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2736–2745, 2015.

[82] Y. Sun, X. Wang, and X. Tang, “Deep Learning Face Representation by Joint Identification-Verification,” in *In Advances in neural information processing systems*, pp. 1988–1996, 2014.

[83] H. H. -, H. L. -, and G. L. -, “Face Recognition Using Pyramid Histogram of Oriented Gradients and SVM,” *Int. J. Adv. Inf. Sci. Serv. Sci.*, vol. 4, no. 18, pp. 1–8, 2012.

[84] Z. M. Z. Mu-chun, “Face Recognition Based on FastICA and RBF Neural Networks,” *2008 Int. Symp. Inf. Sci. Eng.*, vol. 1 , pp. 2005–2009, 2008.

[85] T. H. Le, “Applying Artificial Neural Networks for Face Recognition,” *Adv. Artif. Neural Syst.*, vol. 2011, pp. 1–16, 2011.

[86] Z. Lei, M. Pietikainen, and S. Z. Li, “Learning discriminant face descriptor,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 289–302, 2014.

[87] S. U. Hussain, B. Triggs, S. U. Hussain, B. Triggs, V. Recognition, Q. Patterns, S. Lazebnik, P. Perona, Y. Sato, and C. S. Eccv, “Visual Recognition using Local Quantized Patterns To cite this version : Visual Recognition using Local Quantized Patterns,” in *ECCV 2012 - 12th European Conference on Computer Vision*, pp. 716–729, 2012.

[88] F. Z. Chelali and A. Djeradi, “Face Recognition Using MLP and RBF Neural Network with Gabor and Discrete Wavelet Transform Characterization: A Comparative Study,” *Math. Probl. Eng.*, vol. 2015, 2015.

[89] W. Ouarda, H. Trichili, A. M. Alimi, and B. Solaiman, “Face recognition based on geometric features using Support Vector Machines,” *2014 6th Int. Conf. Soft Comput. Pattern Recognit.*, pp. 89–95, 2014.

[90] L. Deng, “Three Classes of Deep Learning Architectures and Their Applications: A Tutorial Survey,” *Research.Microsoft.Com*, 2013.

[91] Y. Sun, X. Wang, and X. Tang, “Hybrid deep learning for face verification,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1489–1496, 2013.

[92] Y. Sun, X. Wang, and X. Tang, “Hybrid Deep Learning for Face Verification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 10, pp. 1997–2009, 2016.

[93] Y. Taigman, M. A. Ranzato, T. Aviv, and M. Park, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708, 2014.

[94] C. Ding and D. Tao, “Robust Face Recognition via Multimodal Deep Face Representation,” *IEEE Trans. Multimed.*, vol. 17, no. 11, pp. 2049–2058, 2015.

[95] P. Liu, S. Han, Z. Meng, and Y. Tong, “Facial Expression Recognition via a Boosted Deep Belief Network,” in *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1805–1812, 2014.

[96] H. Lee, R. Grosse, R. Ranganath, and Y. N. Andrew, “Unsupervised Learning of Hierarchical Representations with Convolutional Deep Belief Networks,” *Commun. ACM*, vol. 54, no. 10, pp. 95–103, 2011.

[97] G. B. Huang, H. Lee, and E. Learned-Miller, “Learning hierarchical representations for face verification with convolutional deep belief networks,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2518–2525, 2012.

[98] C. Li, Wei Wei, J. Wang, W. Tang, and S. Zhao, “Face Recognition Based on Deep Belief Network Combined with Center-Symmetric Local Binary Pattern,” *Adv. Multimed. Ubiquitous Eng. Springer Singapore*, vol. 354, pp. 277–283, 2016.

[99] D. Yi, Z. Lei, S. Liao, and S. Z. Li, “Shared Representation Learning for Heterogeneous Face Recognition,” in *In Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, pp. 1–7, 2015.

[100] K. Simonyan, O. Parkhi, A. Vedaldi, and A. Zisserman, “Fisher Vector Faces in the Wild,” *Procedings Br. Mach. Vis. Conf. 2013*, p. 8.1-8.11, 2013.

[101] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, “Information-Theoretic Metric Learning,” in *in Proceedings of the 24th International Conference on Machine Learning*, pp. 209–216, 2007.

[102] J. Hu, J. Lu, and Y. P. Tan, “Discriminative deep metric learning for face verification in the wild,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1875–1882, 2014.

[103] Z. Cui, W. Li, D. Xu, S. Shan, and X. Chen, “Fusing robust face region descriptors via multiple metric learning for face recognition in the wild,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3554–3561, 2013.

[104] M. Guillaumin, J. Verbeek, C. Schmid, M. Guillaumin, J. Verbeek, C. Schmid, M. Guillaumin, J. Verbeek, C. Schmid, and L. J. Kuntzmann, “Is that you ? Metric learning approaches for face identification To cite this version,” in *Proceedings of ICCV*, pp. 498--505, 2009.

[105] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, “Large scale metric learning from equivalence constraints,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2288–2295, 2012.

[106] H. V Nguyen and L. Bai, “Cosine Similarity Metric Learning for Face Verification,” in *In: Asian Conference on Computer Vision. Springer Berlin Heidelberg*, pp. 1–12, 2011.

[107] K. W. Bowyer, K. Chang, and P. Flynn, “A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition,” *Comput. Vis. Image Underst.*, vol. 101, no. 1, pp. 1–15, 2006.

[108] S. Ouyang, T. Hospedales, Y.-Z. Song, and X. Li, “A survey on heterogeneous face recognition:sketch, infra-red,3D and low-resolution,” *Image Vis. Comput.*, vol. 56, no. 2016, pp. 28–48, 2016.

[109] K. Hardeep and K. Amandeep, “Illumination Invariant Face Recognition,” *Int. J. Comput. Appl.*, vol. 64, no. 21, pp. 23–27, 2013.

[110] J. Barr and K. Bowyer, “Face recognition from video: A review,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 16, no. 2, pp. 1–56, 2012.

[111] X. Tan, S. Chen, Z. H. Zhou, and F. Zhang, “Face recognition from a single image per person: A survey,” *Pattern Recognit.*, vol. 39, no. 9, pp. 1725–1745, 2006.

[112] H. Zhou, A. Mian, L. Wei, D. Creighton, M. Hossny, and S. Nahavandi, “Recent advances on singlemodal and multimodal face recognition: A survey,” *IEEE Trans. Human-Machine Syst.*, vol. 44, no. 6, pp. 701–716, 2014.

[113] A. N. Al-Raisi and A. M. Al-Khouri, “Iris recognition and the challenge of homeland and border control security in UAE,” *Telemat. Informatics*, vol. 25, no. 2, pp. 117–132, 2008.

[114] K. Ricanek, “Dissecting the human identity,” *IEEE Comput. Soc.*, vol. 44, no. 1, pp. 96–97, 2011.

[115] L. Flom and A. Safir, “Iris recognition system,” *U.S. Patent No. 4,641,349 A. Washington, DC: U.S. Patent and Trademark Office*, 1987.

[116] J. G. Daugman, “High confidence visual recognition of persons by a test of statistical independence,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1148–1161, 1993.

[117] X. Ren, Z. Peng, Q. Zeng, C. Peng, J. Zhang, S. Wu, and Y. Zeng, “An improved method for Daugman’s iris localization algorithm,” *Comput. Biol. Med.*, vol. 38, no. 1, pp. 111–115, 2008.

[118] H. Proenc and L. A. Alexandre, “Iris segmentation methodology for non-cooperative recognition,” *IEE Proceedings-Vision, Image Signal Process.*, vol. 153, no. 2, pp. 199–205, 2006.

[119] S. A. Sahmoud and I. S. Abuhaiba, “Efficient iris segmentation method in unconstrained environments,” *Pattern Recognit.*, vol. 46, no. 12, pp. 3174–3185, 2013.

[120] R. P. Wildes, “Iris recognition: an emerging biometric technology,” *Proc. IEEE*, vol. 85, no. 9, pp. 1348–1363, 1997.

[121] Y.-P. H. Y.-P. Huang, S.-W. L. S.-W. Luo, and E.-Y. C. E.-Y. Chen, “An efficient iris recognition system,” *Proceedings. Int. Conf. Mach. Learn. Cybern.*, vol. 1, pp. 4–5, 2002.

[122] Y. Liu, S. Yuan, X. Zhu, and Q. Cui, “A practical iris acquisition system and a fast edges locating algorithm in iris recognition,” *Instrumentation and Measurement Technology Conference, 2003. IMTC ’03. Proceedings of the 20th IEEE*, vol. 1. pp. 166–168, 2003.

[123] P. Lili and X. Mei, “The algorithm of iris image preprocessing,” *Proc. - Fourth IEEE Work. Autom. Identif. Adv. Technol. AUTO ID 2005*, vol. 2005, pp. 134–138, 2005.

[124] X. He and P. Shi, “A new segmentation approach for iris recognition based on hand-held capture device,” *Pattern Recognit.*, vol. 40, no. 4, pp. 1326–1333, 2006.

[125] X. Feng, C. Fang, X. Ding, and Y. Wu, “Iris localization with dual coarse-to-fine strategy,” *Proc. - Int. Conf. Pattern Recognit.*, vol. 4, no. c, pp. 553–556, 2006.

[126] J. Cui, W. Yunhong, T. Tan, L. Ma, and Z. Sun, “A fast and robust iris localization method based on texture segmentation,” in *Defense and Security. International Society for Optics and Photonics*, pp. 401–408, 2004.

[127] A. Ross and S. Shah, “Segmenting non-ideal irises using geodesic active contours,” in *In Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the IEEE*, pp. 1–6, 2006.

[128] M. K. Pawar, P. G. Student, and G. Jhajjav, “Iris Segmentation using Geodesic Active Contour for Improved Texture Extraction in Recognition,” *Int. J. Comput. Appl.*, vol. 47, no. 16, pp. 40–47, 2012.

[129] F. Jan, I. Usman, and S. Agha, “Iris localization in frontal eye images for less constrained iris recognition systems,” *Digit. Signal Process. A Rev. J.*, vol. 22, no. 6, pp. 971–986, 2012.

[130] A. Abhyankar and S. Schuckers, “Active Shape Models for Effective Iris Segmentation,” *Def. Secur. Symp.*, pp. 1–10, 2006.

[131] L. Sun, T. C. Lin, H. C. Huang, B. Y. Liao, and J. S. Pan, “An optimized approach on applying genetic algorithm to adaptive cluster validity index,” *Proc. - 3rd Int. Conf. Intell. Inf. Hiding Multimed. Signal Process. IIHMSP 2007.*, vol. 2, no. 4, pp. 582–585, 2007.

[132] A. D. S. Sierra, J. G. Casanova, C. S. Avila, and V. J. Vera, “Iris segmentation based on fuzzy mathematical morphology, neural networks and ontologies,” in *In: Security Technology, 2009. 43rd Annual 2009 International Carnahan Conference on. IEEE*, pp. 355–360, 2009.

[133] P. Puranik, P. Bajaj, A. Abraham, P. Palsodkar, and A. Deshmukh, “Using Comprehensive Learning Particle Swarm Optimization,” *J. Inf. Hiding Multimed. Signal Process.*, vol. 2, no. 3, pp. 227–235, 2011.

[134] R. P. Broussard, L. R. Kennell, D. L. Soldan, and R. W. Ives, “Using artificial neural networks and feature saliency techniques for improved iris segmentation,” *IEEE Int. Conf. Neural Networks - Conf. Proc.*, pp. 1283–1288, 2007.

[135] R. P. Broussard and R. W. Ryes, “Using artificial neural networks and feature saliency to identify iris measurements that contain the most discriminatory information for iris segmentation,” *2009 IEEE Work. Comput. Intell. Biometrics Theory, Algorithms, Appl. CIB 2009 - Proc.*, pp. 46–51, 2009.

[136] T. M. Khan, M. Aurangzeb Khan, S. A. Malik, S. A. Khan, T. Bashir, and A. H. Dar, “Automatic localization of pupil using eccentricity and iris using gradient based method,” *Opt. Lasers Eng.*, vol. 49, no. 2, pp. 177–187, 2011.

[137] Z. Sun, T. Tan, and Y. Wang, “Robust Encoding of Local Ordinal Measures: A General Framework of Iris Recognition.,” in *ECCV Workshop BioAW*, pp. 270–282, 2004.

[138] A. Harjoko, S. Hartati, and H. Dwiyasa, “A Method for Iris Recognition Based on 1D Coiflet Wavelet,” *World Acad. Sci. Eng. Technol.*, vol. 56, no. 8, pp. 126–129, 2009.

[139] T. Tan and Z. Sun, “Ordinal Measures for Iris Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2211–2226, 2009.

[140] V. Velisavljevic, “Low-complexity iris coding and recognition based on directionlets,” *IEEE Trans. Inf. Forensics Secur.*, vol. 4, no. 3, pp. 410–417, 2009.

[141] K. Gulmire and S. Ganorkar, “Iris Recognition Using Independent Component Analysis,” *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 7, pp. 433–437, 2012.

[142] M. Elgamal and N. Al-Biqami, “An Efficient Feature Extraction Method for Iris Recognition Based on Wavelet Transformation,” *Int. J. Comput. Inf. Technol.*, vol. 2, no. 3, pp. 521–527, 2013.

[143] P. S. Patil, “Robust Iris Recognition based on Orientation Field,” *Int. J. Comput. Appl.*, vol. 95, no. 19, pp. 19–24, 2014.
[144] B. V Bharath, A. S. Vilas, K. Manikantan, and S. Ramachandran, “Iris Recognition using Radon Transform Thresholding based Feature Extraction with Gradient-based Isolation as a Pre-processing Technique,” in *9th International Conference on Industrial and Information Systems (ICIIS). IEEE*, pp. 1–8, 2014.
[145] S. Umer, B. C. Dhara, and B. Chanda, “Iris recognition using multiscale morphologic features,” *Pattern Recognit. Lett.*, vol. 65, pp. 67–74, 2015.
[146] S. Umer, B. C. Dhara, and B. Chanda, “Texture code matrix-based multi-instance iris recognition,” *Pattern Anal. Appl.*, vol. 19, no. 1, pp. 283–295, 2016.
[147] S. P. Sangeetha, “Fusion of Local and Global Iris Features to Construct Feature Vector Using Genetic Algorithm,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 2, pp. 1486–1489, 2015.
[148] P. R. Nalla and K. M. Chalavadi, “Iris classification based on sparse representations using on-line dictionary learning for large-scale de-duplication applications,” *Springerplus*, vol. 4, no. 1, pp. 1–10, 2015.
[149] S. S. Dhage, S. S. Hegde, K. Manikantan, and S. Ramachandran, “DWT-based Feature Extraction and Radon Transform Based Contrast Enhancement for Improved Iris Recognition,” *Procedia Comput. Sci.*, vol. 45, no. 2015, pp. 256–265, 2015.
[150] B. Jain, “Efficient Iris Recognition Algorithm Using Method of Moments,” *Int. J. Artif. Intell. Appl.*, vol. 3, no. 5, pp. 93–105, 2012.
[151] A. Azizi and H. R. Pourreza, “A New Method for Iris Recognition Based on Contourlet Transform and Non Linear Approximation,” in *International Conference on Intelligent Computing. Springer Berlin Heidelberg,* pp. 307–316, 2009.
[152] M. Z. Rashad, M. Y. Shams, O. Nomir, and R. M. El-Awady, “Iris Recognition Based on LBP And Combined LVQ Classifier,” *Int. J. Comput. Sci. Inf. Technol.*, vol. 3, no. 5, pp. 405–419, 2011.
[153] I. Hamouchene and S. Aouat, “Efficient approach for iris recognition,” *Signal, Image Video Process.*, vol. 10, no. 7, pp. 1361–1367, 2016.
[154] H. Mehrotra, P. K. Sa, and B. Majhi, “Fast segmentation and adaptive SURF descriptor for iris recognition,” *Math. Comput. Model.*, vol. 58, no. 1–2, pp. 132–146, 2013.
[155] C. Belcher and Y. Du, “Region-based SIFT approach to iris recognition,” *Opt. Lasers Eng.*, vol. 47, no. 1, pp. 139–147, 2009.
[156] K. W. Bowyer, K. Hollingsworth, and P. J. Flynn, “Image understanding for iris biometrics: A survey,” *Comput. Vis. Image Underst.*, vol. 110, no. 2, pp. 281–307, 2008.
[157] M. J. Burge and Ke. W. Bowyer, *Handbook of Iris Recognition*, vol. 1542. 2013.
[158] A. Fitrie, H. Sallehuddin, M. I. Ahmad, R. Ngadiran, and M. N. Isa, “A Survey of Iris Recognition System,” *J. Telecommun. Electron. Comput. Eng.*, vol. 8, no. 4, pp. 133–138, 2016.
[159] F. Hao, J. Daugman, and P. Zieli??ski, “A fast search algorithm for a large fuzzy database,” *IEEE Trans. Inf. Forensics Secur.*, vol. 3, no. 2, pp. 203–212, 2008.

[160] H. Mehrotra, B. G. Srinivas, B. Majhi, and P. Gupta, “Indexing iris biometric database using energy histogram of dct subbands,” *Commun. Comput. Inf. Sci.*, vol. 40, pp. 194–204, 2009.

[161] X. Qiu, Z. Sun, and T. Tan, “Coarse Iris Classification by Learned Visual Dictionary,” in *International Conference on Biometrics - Advances in Biometrics*, pp. 770–779, 2007.

[162] C. Te Chou, S. W. Shih, W. S. Chen, V. W. Cheng, and D. Y. Chen, “Non-orthogonal view iris recognition system,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 3, pp. 417–430, 2010.

[163] M. M. Khedkar and S. a. Ladhake, “Neural Network Based Iris Pattern Recognition System Using Discrete Walsh Hadamard Transform Features,” in *IEEE International Conference on Advances in Computing, Communications and Informatics*, pp. 388–393, 2013.

[164] H. Rai and A. Yadav, “Iris recognition using combined support vector machine and Hamming distance approach,” *Expert Syst. Appl.*, vol. 41, no. 2, pp. 588–593, 2014.

[165] P. Melin, V. Herrera, D. Romero, F. Valdez, and O. Castillo, “Genetic Optimization of Neural Networks for Person Recognition Based on the Iris.,” *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 10, no. 2, pp. 309–320, 2012.

[166] K. Saminathan, T. Chakravarthy, and M. C. Devi, “Iris Recognition Based on Kernels of Support Vector Machine,” *ICTACT J. SOFT Comput. Spec. ISSUE SOFT – Comput. THEORY, Appl. Implic. Eng. Technol.*, vol. 5, no. 2, pp. 889–895, 2015.

[167] M. De Marsico, A. Petrosino, and S. Ricciardi, “Iris recognition through machine learning techniques: A survey,” *Pattern Recognit. Lett.*, vol. 82, pp. 106–115, 2016.

[168] X. Liu and T. Chen, “Geometry-assisted statistical modeling for face mosaicing,” *Proc. 2003 Int. Conf. Image Process. (Cat. No.03CH37429)*, vol. 2, pp. 883–886, 2003.

[169] R. Singh, M. Vatsa, and A. Noore, “Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition,” *Pattern Recognit.*, vol. 41, no. 3, pp. 880–893, 2008.

[170] R. Raghavendra, A. Rao, and G. Hemantha Kumar, “Multisensor biometric evidence fusion of face and palmprint for person authentication using Particle Swarm Optimisation (PSO),” *Int. J. Biom.*, vol. 2, no. 1, pp. 19–33, 2010.

[171] D. R. Kisku, J. K. Sing, M. Tistarelli, and P. Gupta, “Multisensor biometric evidence fusion for person authentication using wavelet decomposition and monotonic-decreasing graph,” *Proc. 7th Int. Conf. Adv. Pattern Recognition, ICAPR 2009*, pp. 205–208, 2009.

[172] A. Rattani and M. Tistarelli, “Robust Multi-Modal and Multi-Unit Feature Level Fusion af Face and Iris Biometrics,” in *nternational Conference on biometrics. Springer Berlin Heidelberg,* pp. 960–969, 2009.

[173] J. Gan, J. Gao, and J. Liu, “Research on face and iris feature recognition based on 2DDCT and Kernel Fisher Discriminant Analysis,” *2008 Int. Conf. Wavelet Anal. Pattern Recognit.*, pp. 401–405, Aug. 2008.

[174] A. Nagar, K. Nandakumar, and A. K. Jain, “Multibiometric cryptosystems based on feature-level fusion,” *IEEE Trans. Inf. Forensics Secur.*, vol. 7, no. 1, pp. 255–268, 2012.

[175] J. I. E. Lin, J. Li, H. U. I. Lin, and J. I. Ming, “Robust person identification with face and iris by modified PUM method,” in *2009 International Conference on Apperceiving Computing and Intelligence Analysis*, pp. 321–324, 2009.

[176] D. Sharma and A. Kumar, “An Empirical Analysis Over the Four Different Feature-Based Face and Iris Biometric Recognition Techniques,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 3, no. 10, pp. 76–83, 2012.

[177] C. Chen and C. Te Chu, “Fusion of Face and Iris Features for Multimodal Biometrics,” in *In International Conference on Biometrics*, pp. 571–580, 2005.

[178] Z. Zhang, R. Wang, K. Pan, S. Li, and P. Zhang, “Fusion of Near Infrared Face and Iris Biometrics,” *Adv. Biometrics. Springer Berlin Heidelb.*, pp. 172–180, 2007.

[179] K. Fakhar, M. El Aroussi, and M. N. Saidi, “Score Fusion in Multibiometric Identification Based on Fuzzy Set Theory,” *Int. Conf. Image Signal Process. Springer Berlin Heidelb.*, pp. 1–8, 2012.

[180] S. J. Al-Hijaili, “Biometrics in Health Care Security System, Iris-Face Fusion System. ,” *Int. J. Acad. Res.*, vol. 3, no. 1, pp. 11–19, 2011.

[181] Y. G. Kim, K. Y. Shin, E. C. and K. R. P. Lee, and 1, “Multimodal Biometric System Based on the Recognition of Face and Both Irises,” *Int. J. Adv. Robot. Syst.*, vol. 9, no. 3, 2012.

[182] Mamta and M. Hanmandlu, “Multimodal biometric system built on the new entropy function for feature extraction and the Refined Scores as a classifier,” *Expert Syst. Appl.*, vol. 42, no. 7, pp. 3702–3723, 2015.

[183] A. Aboshosha and K. A. El Dahshan, “Score Level Fusion for Fingerprint, Iris and Face Biometrics,” *Int. J. Comput. Appl.*, vol. 111, no. 4, pp. 47–55, 2015.

[184] M. M. Monwar and M. L. Gavrilova, “Multimodal Biometric System Using Rank-Level Fusion Approach,” *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 39, no. 4, pp. 867–878, 2009.

[185] D. T. Meva and K. C. K., “Comparative Study of Different Fusion Techniques in Multimodal Biometric Authentication,” *Int. J. Comput. Appl.*, vol. 66, no. 19, pp. 16–19, 2013.

[186] A. Abaza and A. Ross, “Quality based rank-level fusion in multibiometric systems,” *Proc. 3rd IEEE Int. Conf. Biometrics Theory, Appl. Syst.*, pp. 1–6, 2009.

[187] M. M. Monwar and M. Gavrilova, “Markov chain model for multimodal biometric rank fusion,” *Signal, Image Video Process.*, vol. 7, no. 1, pp. 137–149, 2013.

[188] G. Amirthalingam and G. Radhamani, “A Multimodal Approach for Face and Ear Biometric System,” *Int. J. Comput. Sci. Issues*, vol. 10, no. 5, pp. 234–241, 2013.

[189] L. Lam and C. Y. Suen, “Optimal combinations of pattern classifiers,” *Pattern Recognit. Lett.*, vol. 16, no. 9, pp. 945–954, 1995.

[190] M. Razzak and R. Yusof, “Multimodal face and finger veins biometric authentication,” *Sci. Res. Essays*, vol. 5, no. 17, pp. 2529–2534, 2010.

[191] L. Xu, A. Krzyzak, and C. Y. Suen, “Methods of Combing Multiple Classifiers and Their Application to Handwriting Recognition,” *IEEE Trans. Syst. Man. Cybern.*, vol. 22, no. 3, pp. 418–435, 1992.

[192] Y. S. Huang and C. Y. Suen, “A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 1, pp. 90–94, 1995.

[193] P. H. Lee, L. J. Chu, Y. P. Hung, S. W. Shih, C. S. Chen, and H. M. Wang, “Cascading Multimodal Verification Using Face, Voice and Iris Information,” in *Multimedia and Expo, 2007 IEEE International Conference on*, pp. 847–850, 2007.

[194] P. Yu, D. Xu, H. Zhou, and H. Li, “Decision fusion for hand biometric authentication,” *Proc. - 2009 IEEE Int. Conf. Intell. Comput. Intell. Syst. ICIS 2009*, vol. 4, no. 8, pp. 486–490, 2009.

[195] S. Li, X. Liu, X. Chai, H. Zhang, S. Lao, and S. Shan, “Maximal likelihood correspondence estimation for face recognition across pose,” *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4587–4600, 2014.

[196] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang, “Probabilistic elastic matching for pose variant face verification,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 3499–3506, 2013.

[197] Y. Taigman, L. Wolf, and T. Hassner, “Multiple One-Shots for Utilizing Class Label Information,” *BMVC*, vol. 2, no. September, pp. 1–12, 2009.

[198] Z. Wang, E. Wang, S. Wang, and Q. Ding, “Multimodal Biometric System Using Face-Iris Fusion Feature,” *J. Comput.*, vol. 6, no. 5, pp. 931–938, May 2011.

[199] N. Radha, “Rank Level Fusion Using Fingerprint and Iris Biometrics,” *Indian J. Comput. Sci. Eng.*, vol. 2, no. 6, pp. 917–923, 2012.

[200] R. Raghavendra, “PSO based framework for weighted feature level fusion of face and palmprint,” *Proc. 2012 8th Int. Conf. Intell. Inf. Hiding Multimed. Signal Process. IIH-MSP 2012*, pp. 506–509, 2012.

[201] Z. Huang, Y. Liu, C. Li, M. Yang, and L. Chen, “A robust face and ear based multimodal biometric system using sparse representation,” *Pattern Recognit.*, vol. 46, no. 8, pp. 2156–2168, 2013.

[202] S. A. H. Nair, P. Aruna, and M. Vadivukarassi, “PCA Based Image Fusion of Face and Iris Biometric Features,” *Int. J. Adv. Comput. Theory Eng. information.*, vol. 1, no. 2, pp. 106–112, 2013.

[203] M. Eskandari, Ö. Toygar, and H. Demirel, “Feature extractor selection for face–iris multimodal recognition,” *Signal, Image Video Process.*, vol. 8, no. 6, pp. 1189–1198, Jun. 2014.

[204] H. M. Sim, H. Asmuni, R. Hassan, and R. M. Othman, “Multimodal biometrics: Weighted score level fusion based on non-ideal iris and face images,” *Expert Syst. Appl.*, vol. 41, no. 11, pp. 5390–5404, Sep. 2014.

[205] H. Benaliouche and M. Touahria, “Comparative study of multimodal biometric recognition by fusion of iris and fingerprint,” *Sci. World J.*, vol. 2014, 2014.

[206] V. Azom, A. Adewumi, and J. R. Tapamo, “Face and Iris biometrics person identification using hybrid fusion at feature and score-level,” *Proc. 2015 Pattern Recognit. Assoc. South Africa Robot. Mechatronics Int. Conf. PRASA-RobMech 2015*, pp. 207–212, 2015.

[207] M. Eskandari and Ö. Toygar, “Selection of optimized features and weights on face-iris fusion using distance images,” *Comput. Vis. Image Underst.*, vol. 137, pp. 63–75, 2015.

[208] O. Sharifi and M. Eskandari, “Optimal Face-Iris Multimodal Fusion Scheme,” *Symmetry (Basel).*, vol. 8, no. 6, pp. 1–48, 2016.

[209] H. Mehrotra, R. Singh, M. Vatsa, and B. Majhi, “Incremental granular relevance vector machine: A case study in multimodal biometrics,” *Pattern Recognit.*, vol. 56, pp. 63–76, 2016.

[210] M. I. Ahmad, W. L. Woo, and S. Dlay, “Non-stationary feature fusion of face and palmprint multimodal biometrics,” *Neurocomputing*, vol. 177, pp. 49–61, 2016.

[211] D. Miao, M. Zhang, Haiqingli, Z. Sun, and T. Tan, “Bin-based weak classifier fusion of iris and face biometrics,” *2015 IEEE 7th Int. Conf. Biometrics Theory, Appl. Syst. BTAS 2015*, vol. 224, no. May 2016, pp. 105–118, 2015.

[212] Y. T. L. Wolf T. Hassner, “Similarity scores based on Background samples,” in *Asian Conference on Computer Vision. Springer, Berlin, Heidelberg*, pp. 88–97, 2009.

[213] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, “Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors,” *Proc. 6th Int. Conf. Mob. Comput. Appl. Serv.*, pp. 197–205, 2014.

[214] A. R. Syafeeza, S. S. Liew, and R. Bakhteri, “Convolutional Neural Network for Face Recognition with Pose and Illumination Variation,” *Int. J. Eng. Technol.*, vol. 6, no. 1, pp. 44–57, 2014.

[215] W. D. Rummel, “A Path Forward for NDE Reliability,” in *Proceedings of the 5th European-American Workshop on Reliability of ND*, pp. 1–10, 2013.

[216] M. Zeng, L. T. Nguyen, B. Yu, O. J. Mengshoel, J. Zhu, P. Wu, and J. Zhang, “Convolutional Neural Networks for Human Activity Recognition Using Mobile Sensors,” *Mob. Comput. Appl. Serv. (MobiCASE), 2014 6th Int. Conf.*, pp. 197–205, 2014.

[217] R. Jafri and H. R. Arabnia, “A Survey of Face Recognition Techniques,” *J. Inf. Process. Syst.*, vol. 5, no. 2, pp. 41–68, Jun. 2009.

[218] M. K. Bhowmik, D. Bhattacharjee, M. Nasipuri, D. K. Basu, and M. Kundu, “Quotient Based Multiresolution Image Fusion of Thermal and Visual Images Using Daubechies Wavelet Transform for Human Face Recognition,” *Int. J. Comput. Sci. Issues*, vol. 7, no. 3, pp. 18–27, 2010.

[219] X. Wu and J. Zhao, “Curvelet feature extraction for face recognition and facial expression recognition,” *2010 Sixth Int. Conf. Nat. Comput.*, no. Icnc, pp. 1212–1216, Aug. 2010.

[220] A. Majumdar and A. Bhattacharya, “A Comparative Study in Wavelets , Curvelets and Contourlets as Feature Sets for Pattern Recognition,” *Int. Arab J. Inf. Technol.*, vol. 6, no. 1, pp. 47–51, 2009.

[221] E. Candes, L. Demanet, D. Donoho, and Y. Lexing, “Fast Discrete Curvelet Transforms,” *Multiscale Model. Simul.*, vol. 5, no. 3, pp. 861–899, 2006.

[222] M. Choi, R. Young, and M. Kim, “The curvelet transform for image fusion,” *Int. Soc. Photogramm. Remote Sens.*, vol. 35, no. 2004, pp. 59–64, 2004.

[223] S. Arivazhagan, L. Ganesan, and T. G. Subash Kumar, “Texture classification using Curvelet Statistical and Co-occurrence Features,” in *18th International Conference on Pattern Recognition (ICPR’06)*, pp. 938–941, 2006.

[224] G. G. Bhutada, R. S. Anand, and S. C. Saxena, “Edge preserved image enhancement using adaptive fusion of images denoised by wavelet and curvelet transform,” *Digit. Signal Process.*, vol. 21, no. 1, pp. 118–130, Jan. 2011.

[225] Y. Li, Q. Yang, and R. Jiao, “Image compression scheme based on curvelet transform and support vector machine,” *Expert Syst. Appl.*, vol. 37, no. 4, pp. 3063–3069, Apr. 2010.

[226] M.-S. Chen and S.-D. Lin, “Image fusion based on curvelet transform and fuzzy logic,” in *mage and Signal Processing (CISP), 2012 5th International Congress on. IEEE*, pp. 1063–1067, 2012.

[227] Y. Ruihong, T. Liwei, W. Ping, and Y. Jiajun, “Image Denoising Based on Curvelet Transform and Continuous Threshold,” *2010 First Int. Conf. Pervasive Comput. Signal Process. Appl.*, pp. 13–16, Sep. 2010.

[228] Y.-C. Lee and C.-H. Chen, “Face Recognition Based on Digital Curvelet Transform,” *2008 Eighth Int. Conf. Intell. Syst. Des. Appl.*, pp. 341–345, Nov. 2008.

[229] J. Xie, “Face Recognition Based on Curvelet Transform and LS-SVM,” in *Proceedings of the International Symposium on Information Processing*, pp. 140–143, 2009.

[230] R. Lopes and N. Betrouni, “Fractal and multifractal analysis: a review.,” *Med. Image Anal.*, vol. 13, no. 4, pp. 634–49, Aug. 2009.

[231] B. Mandelbrot, “The Fractal Geometry of Nature,” *Libr. Congr. Cat. Publ. Data, United States Am.*, 1983.

[232] B. Mandelbrot, “Self-affinity and fractal dimension,” in *PHYSICA SCRIPTA*, vol. 32, pp. 257–260, 1985.

[233] T. Hsu and K.-J. Hu, “Multi-resolution Texture Segmentation Using Fractal Dimension,” *2008 Int. Conf. Comput. Sci. Softw. Eng.*, pp. 201–204, 2008.

[234] O. S. Al-Kadi, “A multiresolution clinical decision support system based on fractal model design for classification of histological brain tumours,” *Comput. Med. Imaging Graph.*, vol. 41, no. 2015, pp. 67–79, Jun. 2015.

[235] Z. Zhu, J. Gao, and H. Yu, “Face Detection Based on Fractal and Complexion Model in the Complex Background,” *2010 Int. Work. Chaos-Fractal Theor. Appl.*, pp. 491–495, Oct. 2010.

[236] K. Lin, K. Lam, and W. Siu, “Locating the human eye using fractal dimensions,” in *Proceedings International Conference on Image Processing (Cat. No.01CH37205)*, pp. 1079–1082, 2001.

[237] M. H. Farhan, L. E. George, and A. T. Hussein, “Fingerprint Identification Using Fractal Geometry,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 4, no. 1, pp. 52–61, 2014.

[238] G. Hinton, “A Practical Guide to Training Restricted Boltzmann Machines A Practical Guide to Training Restricted Boltzmann Machines,” *Computer (Long. Beach. Calif).*, vol. 9, no. 3, p. 1, 2010.

[239] G. E. Hinton, “Training products of experts by minimizing contrastive divergence,” *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, 2002.

[240] H. Larochelle and Y. Bengio, “Classification using discriminative restricted Boltzmann machines,” in *Proceedings of the 25th International Conference on Machine Learning*, pp. 536–543, 2008.

[241] G. E. Hinton, S. Osindero, and Y.-W. Teh, “A Fast Learning Algorithm for Deep Belief Nets,” *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[242] B. Abibullaev, J. An, S. H. Jin, S. H. Lee, and J. Il Moon, “Deep Machine Learning—A New Frontier in Artificial Intelligence Research,” *Med. Eng. Phys.*, vol. 35, no. 12, pp. 1811–1818, 2013.
[243] J. Liu, C. Fang, and C. Wu, “A Fusion Face Recognition Approach Based on 7-Layer Deep Learning Neural Network,” *J. Electr. Comput. Eng.*, vol. 2016, pp. 1–7, 2016.
[244] P. Fousek, S. Rennie, P. Dognin, and V. Goel, “Direct product based deep belief networks for automatic speech recognition,” *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. ICASSP 2013*, pp. 3148–3152, 2013.
[245] H. Lee, P. Pham, Y. Largman, and A. Ng, “Unsupervised feature learning for audio classification using convolutional deep belief networks,” *Adv. neural Inf. Process. Syst. Conf.*, pp. 1096–1104, 2009.
[246] R. Sarikaya, G. E. Hinton, and A. Deoras, “Application of deep belief networks for natural language understanding,” *IEEE Trans. Audio, Speech Lang. Process.*, vol. 22, no. 4, pp. 778–784, 2014.
[247] P. Viola, O. M. Way, and M. J. Jones, “Robust Real-Time Face Detection,” *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.
[248] R. Minhas, A. Adeel Mohammed, and Q. M. Jonathan Wu, “Shape from focus using fast discrete curvelet transform,” *Pattern Recognit.*, vol. 44, no. 4, pp. 839–853, 2011.
[249] D.-R. Chen, R.-F. Chang, C.-J. Chen, M.-F. Ho, S.-J. Kuo, S.-T. Chen, S.-J. Hung, and W. K. Moon, “Classification of breast ultrasound images using fractal feature.,” *Clin. Imaging*, vol. 29, no. 4, pp. 235–45, 2005.
[250] C. Chen, J. S. Daponte, and M. D. FOX, “Fractal Feature Analysis and Classification in Medical Imaging,” *IEEE Trans. Med. Imaging*, vol. 8, no. 2, pp. 133–142, 1989.
[251] S. Liu, “An Improved Differential Box-Counting Approach to Compute Fractal Dimension of Gray-Level Image,” in *2008 International Symposium on Information Science and Engineering*, pp. 303–306, 2008.
[252] M. Long and F. Peng, “A Box-Counting Method with Adaptable Box Height for Measuring the Fractal Feature of Images,” in *Radioengineering*, pp. 208–213, 2013.
[253] J. Li, Q. Du, and C. Sun, “An improved box-counting method for image fractal dimension estimation,” *Pattern Recognit.*, vol. 42, no. 11, pp. 2460–2469, Nov. 2009.
[254] Y. Sun, X. Wang, and X. Tang, “Deep learning face representation from predicting 10,000 classes,” *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 1891–1898, 2014.
[255] Y. Yin, L. Liu, and X. Sun, “SDUMLA-HMT: A multimodal biometric database,” *Chinese Conf. Biometric Recognition, Springer-Verlag Berlin Heidelb.*, pp. 260–268, 2011.
[256] P. J. Phillips, Hyeonjoon Moon, S. A. Rizvi, and P. J. Rauss, “The FERET evaluation methodology for face-recognition algorithms,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.
[257] W. Gao, B. Cao, S. Shan, D. Zhou, X. Zhang, D. Zhao, W. Gao, B. Cao, S. Shan, D. Zhou, X. Zhang, and D. Zhao, “The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations,” *IEEE Trans. Syst. Man. Cybern.*, vol. 38, no. May, pp. 1–24, 2004.

[258] G. B. Huang, M. Mattar, T. Berg, and E. Learned-miller, “Labeled Faces in the Wild : A Database for Studying Face Recognition in Unconstrained Environments,” in *Technical Report 07-49, University of Massachusetts, Amherst*, pp. 1–14, 2007.

[259] A. Campilho and M. Kamel, “Restricted Boltzmann Machines for Gender Classification,” in *International Conference Image Analysis and Recognition. Springer*, pp. 274–281, 2014.

[260] D. Maturana, D. Mery, and A. Soto, “Learning discriminative local binary patterns for face recognition,” *2011 IEEE Int. Conf. Autom. Face Gesture Recognit. Work. FG 2011*, pp. 470–475, 2011.

[261] S. U. Hussain, B. Triggs, S. U. Hussain, B. Triggs, V. Recognition, Q. Patterns, S. Lazebnik, P. Perona, Y. Sato, and C. S. Eccv, “Face Recognition using Local Quantized Patterns,” *Br. Machive Vis. Conf. Sep 2012, Guildford, United Kingdom*, pp.1-11, 2012.

[262] H. Tan, Z. Ma, and B. Yang, “Face recognition based on the fusion of global and local HOG features of face images,” *IET Comput. Vis.*, vol. 8, no. 3, pp. 224–234, 2014.

[263] S. Liao, D. Shen, and A. C. S. Chung, “A markov random field groupwise registration framework for face recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 657–669, 2014.

[264] Z. Lei, D. Yi, and S. Z. Li, “Local gradient order pattern for face representation and recognition,” *Proc. - Int. Conf. Pattern Recognit.*, pp. 387–392, 2014.

[265] N. Farajzadeh, K. Faez, and G. Pan, “Study on the performance of moments as invariant descriptors for practical face recognition systems,” *IET Comput. Vis.*, vol. 4, no. 4, pp. 272–285, 2010.

[266] D. Maturana, D. Mery, and S. Alvaro, “Face Recognition with Decision Tree-Based Local Binary Patterns,” *Eur. Conf. Comput. Vis.*, pp. 469–481, 2004.

[267] Y. Yan, H. Wang, C. Li, C. Yang, and B. Zhong, “A Novel Unconstrained Correlation Filter and Its Application in Face Recognition,” in *International Conference on Intelligent Science and Intelligent Data Engineering. Springer Berlin Heidelberg*, p. pp 32-39, 2013.

[268] J. Hu, J. Lu, X. Zhou, and Y. P. Tan, “Discriminative transfer learning for single-sample face recognition,” *Proc. 2015 Int. Conf. Biometrics, ICB 2015*, pp. 272–277, 2015.

[269] G. B. Huang and E. Learned-miller, “Labeled Faces in the Wild : Updates and New Reporting Procedures,” *Univ. Massachusetts, Amherst, UM-CS-2014-003,* pp. 1–14, 2014.

[270] A. Ouamane, M. Bengherabi, A. Hadid, and M. Cheriet, “Side-Information based Exponential Discriminant Analysis for Face Verification in the Wild,” in *11th IEEE International Conference and Workshops on*, pp. 1–6, 2015.

[271] O. Barkan, J. Weill, L. Wolf, and H. Aronowitz, “Fast high dimensional vector multiplication face recognition,” *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1960–1967, 2013.

[272] T. Hassner, S. Harel, E. Paz, and R. Enbar, “Effective face frontalization in unconstrained images,” *2015 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 4295–4304, 2015.

[273] M. R. Islam, “Feature and score fusion based multiple classifier selection for iris recognition,” *Comput. Intell. Neurosci.*, vol. 2014, 2014.

[274] W. Aydi, “Improved Masek approach for iris localization,” *ICM 2011 Proceeding. IEEE*, pp. 1–5, 2011.
[275] R. Hentati, M. Hentati, and M. Abid, “Development a New Algorithm for Iris Biometric Recognotion,” *Int. J. Comput. Commun. Eng.*, vol. 1, no. 3, pp. 283–286, 2012.
[276] R. H. Abiyev and K. I. Kilic, “Robust Feature Extraction and Iris Recognition for Biometric Personal Identification,” *Biometric Syst. Des. Appl. InTech*, 2011.
[277] A. Das and R. Parekh, “Iris Recognition using a Scalar based Template in Eigen-space,” *Int. J. Comput. Sci. Telecommun.*, vol. 3, no. 5, pp. 3–8, 2012.
[278] Q. Tian, H. Qu, Z. Lanfang, and R. Zong, “Personal Identity Recognition Approach Based on Iris Pattern,” *State art Biometrics*, pp. 1–326, 2011.
[279] M. Boyd, D. Carmaciu, F. Giannaros, T. Payne, and W. Snell, “Iris Recognition,” Imperial College London, 2010.
[280] L. D. and D. Yu, “Deep Learning Methods and Applications,” *Signal Processing*, vol. 28, no. 3, pp. 198–387, 2013.
[281] T. Pellegrini, “Comparing SVM, Softmax, and shallow neural networks for eating condition classification,” in *Sixteenth Annual Conference of the International Speech Communication Association*,, pp. 899–903, 2015.
[282] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout : A Simple Way to Prevent Neural Networks from Overfitting,” *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.
[283] A. Krizhevsky, Ii. Sulskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Adv. Neural Inf. Process. Syst.*, pp. 1–9, 2012.
[284] T. Atherton and D. Kerbyson, “The coherent circle Hough transform,” in *Proceedings of the British Machine Vision Conference*, pp. 1–10, 1993.
[285] M. A. S. P. Vasconcelos, “On the Radon Transform of Sampled Functions,” *PhD thesis, Engineeering Sci. Harvard Univ.*, pp. 1–115, 2003.
[286] M. V. Ginkel, C. L. Luengo Hendriks, and L. J. Van Vliet, “A short introduction to the Radon and Hough transforms and how they relate to each other,” in *Delft University of Technology*, pp. 1–11, 2004.
[287] S. Ding, H. Zhu, W. Jia, and C. Su, “A survey on feature extraction for pattern recognition,” *Artif. Intell. Rev.*, vol. 37, no. 3, pp. 169–180, May 2011.
[288] Y. Jihua, H. Dan, X. Guomiao, and C. Yahui, “An Advanced BPNN Face Recognition Based On Curvelet Transform and 2DPCA,” *8th Int. Conf. Comput. Sci. Educ. (ICCSE)*, pp. 1019–1022, 2013.
[289] H. Khalajzadeh, M. Mansouri, and M. Teshnehlab, “Face Recognition using Convolutional Neural Network and Simple Logistic Classifier,” *Soft Comput. Ind. Appl. Springer Int. Publ.*, pp. 197–207, 2014.
[290] R. Collobert and J. Weston, “A unified architecture for natural language processing: Deep neural networks with multitask learning,” *Proc. 25th Int. Conf. Mach. Learn.*, pp. 160–167, 2008.
[291] L. G. Hafemann, L. S. Oliveira, P. R. Cavalin, and R. Sabourin, “Transfer Learning between Texture Classification Tasks using Convolutional Neural Networks,” in *Neural Networks (IJCNN), 2015 International Joint Conference on. IEEE*, pp. 1–7, 2015.

[292] H. El Khiyari and H. Wechsler, “Face Recognition across Time Lapse Using Convolutional Neural Networks,” *J. Inf. Secur.*, vol. 7, no. 3, pp. 141–151, 2016.

[293] G. E. Dahl, “Deep Learning Approaches to Problems in Speech Recognition, Computational Chemistry, and Natural Language Text Processing,” *PhD thesis, Dep. Comput. Sci. Univ. Toronto*, pp. 1–92, 2015.

[294] R. Salakhutdinov and G. Hinton, “Semantic hashing,” *Int. J. Approx. Reason.*, vol. 50, no. 7, pp. 969–978, 2009.

[295] D. C. Ciresan, U. Meier, J. Masci, and J. Schmidhuber, “Multi-Column Deep Neural Network for Traffic Sign Classification,” *Neural Networks*, vol. 32, pp. 333–338, 2012.

[296] Y. Bengio, “Learning Deep Architectures for AI,” *Found. Trends® Mach. Learn.*, vol. 2, no. 1, pp. 1–127, 2009.

[297] R. Zeng, J. Wu, Z. Shao, L. Senhadji, H. Shu, R. Zeng, J. Wu, Z. Shao, L. Senhadji, and H. Shu, “Quaternion softmax classifier,” *Electron. Lett. IET*, vol. 50, no. 25, pp. 1929–1930, 2015.

[298] T. a. Camus and R. Wildes, “Reliable and fast eye finding in close-up images,” *Object Recognit. Support. by user Interact. Serv. Robot.*, vol. 1, 2002.

[299] A. Zaim, M. Quweider, J. Scargle, J. Iglesias, and R. Tang, “A robust and accurate segmentation of iris images using optimal partitioning,” in *Proceedings - International Conference on Pattern Recognition*, pp. 578–581, 2006.

[300] T. H. Min and R. H. Park, “Comparison of eyelid and eyelash detection algorithms for performance improvement of iris recognition,” *Proc. - Int. Conf. Image Process. ICIP*, pp. 257–260, 2008.

[301] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7. pp. 629–639, 1990.

[302] R. Duda, P. Hart, and D. Stork, “Patterns Classification,” *John Wiley Sons,* 2012.

[303] J. Duchi, E. Hazan, and Y. Singer, “Adaptive Subgradient Methods for Online Learning and Stochastic Optimization,” *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, 2011.

[304] R. Fergus and M. Zeiler, “Visualizing and Understanding Convolutional Networks,” *Eur. Conf. Comput. Vision, Springer Int. Publ.*, pp. 818–833, 2015.

[305] M. Oquab, M. Oquab, I. Laptev, J. S. Learning, T. Mid-level, M. Oquab, and L. Bottou, “Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks,” *IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014.

[306] V. Turchenko and A. Luczak, “Creation of a Deep Convolutional Auto-Encoder in Caffe,” in *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2017 9th IEEE International Conference*, pp. 651–659, 2017.

[307] Chinese Academy of Science—Institute of Automation, CASIA Iris Image Database Version 1.0 (CASIA-IrisV1). http://biometrics. dealtest.org/dbDetailForUser.do?id=3, [accessed in Apr. 2016].

[308] Chinese Academy of Science—Institute of Automation, CASIA Iris Image Database Version 3.0 (CASIA-IrisV3). http://biometrics.dealtest.org/dbDetailForUser.do?id=3, [accessed in Apr. 2016].

[309] A. Kumar and A. Passi, “Comparison and combination of iris matchers for reliable personal authentication,” *Pattern Recognit.*, vol. 43, no. 3, pp. 1016–1026, 2010.

[310] H. Raida, M. Boussellmi, and M. Abid, “Study the iris segmentation method for biometric pattern recognition,” *2011 3rd Int. Conf. Next Gener. Networks Serv. NGNS’2011*, pp. 108–111, 2011.

[311] A. Uhl and P. Wild, “Weighted adaptive Hough and ellipsopolar transforms for real-time iris segmentation,” *Proc. - 2012 5th IAPR Int. Conf. Biometrics, ICB 2012*, pp. 283–290, 2012.

[312] A. A. Ghali, S. Jamel, F. Wang, J. Han, and C. Model, “International Conference on Recent Trends in Physics 2016 (ICRTP2016),” *J. Phys. Conf. Ser.*, vol. 755, no. 2017, p. 11001, 2017.

[313] M. Shamsi, P. Saad, and A. Rasouli, “Iris segmentation and normalization approach,” *J. Teknol. Mklm.*, vol. 19, no. 2, pp.88-101, 2008.

[314] L. Masek, “Recognition of human iris patterns for biometric identification,” *Bachelor Thesis, Eng. degree Sch. Comput. Sci. Softw. Eng. Univ. West. Aust.*, pp. 1–53, 2003.

[315] M. Mahlouji and A. Noruzi, “Human Iris Segmentation for Iris Recognition in Unconstrained Environments,” *Int. J. Comput. Sci. Issues*, vol. 9, no. 1, pp. 149–155, 2012.

[316] G. Xu, Z. Zhang, and Y. Ma, “Automatic iris segmentation based on local areas,” *Proc. - Int. Conf. Pattern Recognit.*, vol. 4, no. c, pp. 505–508, 2006.

[317] R. Yew, F. Ng, Y. H. Tay, and K. M. Mok, “An Effective Segmentation Method for Iris Recognition System,” *Eng. Technol.*, pp. 548–553, 2008.

[318] M. M. Ibrahim, J. S. Soraghan, and N. A. Manap, “Iris Localisation Using Fuzzy Centre Detection (FCD) Scheme and Active Contour Snake,” *J. Teknol. (Sciences Eng.*, vol. 69, no. 6, pp. 73–77, 2014.

[319] S. R. B. Kulkarni, R. S. Hegadi, and U. P. Kulkarni, “ROI based iris segmentation and block reduction based pixel match for improved biometric applications,” *Advances in Computing, Communication, and Control- Springer, Berlin, Heidelberg*, pp. 548–557, 2013.

[320] K. Wang and Y. Qian, “Fast and accurate iris segmentation based on linear basis function and Ransac,” *Proc. - Int. Conf. Image Process. ICIP*, vol. 2, pp. 3205–3208, 2011.

[321] I. U. Nkole and G. Bin Sulong, “An Enhanced Iris Segmentation Algorithm Using Circle Hough Transform,” in *Proc. of the IEEE international conference on Digital Signal and Image processing*, pp. 1–7, 2012.

[322] P. Wild, H. Hofbauer, J. Ferryman, and A. Uhl, “Segmentation-level fusion for iris recognition,” in *In: 14th International Conference of the Biometrics Special Interest Group (BIOSIG 2015)*, pp. 1–6, 2015.

[323] H.-L. Wan, Z.-C. Li, J.-P. Qiao, and B.-S. Li, “Non-ideal iris segmentation using anisotropic diffusion,” *IET Image Process.*, vol. 7, no. 2, pp. 111–120, 2013.

[324] P. Wild, H. Hofbauer, J. Ferryman, and A. Uhl, “Quality-based iris segmentation-level fusion,” *EURASIP J. Inf. Secur.*, vol. 2016, no. 1, p. 25, 2016.

[325] D. Pérez, C. Fernández, C. Segura, and H. Pericás, “New approaches for iris boundary localization,” in *in Libro de actas de las VI Jornadas de Reconocimiento Biométrico de Personas: Las Palmas de Gran Canaria*, pp. 64–72, 2012.

[326] C. Valentina, R. Hartono, T. Tjahja, and A. Nugroho, “Iris Localization using Circular Hough Transform and Horizontal Projection Folding,” in *Proceedings of International Conference of information Technology and Applied Mathematics Jakarta, Indonesia*, pp. 64–68, 2012.

[327] C. Fernandez, D. Perez, C. Segura, and J. Hernando, “A novel method for low-constrained iris boundary localization,” in *Proceedings - 2012 5th IAPR International Conference on Biometrics, ICB 2012*, pp. 291–296, 2012.

[328] A. R. Chowdhury, T.-Y. Lin, S. Maji, and E. Learned-Miller, “One-to-many face recognition with bilinear CNNs,” *IEEE Winter Conf. Appl. Comput. Vis.*, pp. 1–9, 2016.

[329] Y. Chen, Y. Liu, X. Zhu, F. He, H. Wang, and N. Deng, “Efficient iris recognition based on optimal subfeature selection and weighted subregion fusion,” *Sci. World J.*, vol. 2014, no. 2, pp.1-19, 2014.

[330] S. S. Chowhan, U. V. Kulkarni, and G. N. Shinde, “Iris Recognition Using Modified Fuzzy Hypersphere Neural Network with different Distance Measures,” *IJACSA - Int. J. Adv. Comput. Sci. Appl.*, vol. 2, no. 6, pp. 130–134, 2011.

[331] D. R. Shashi Kumar, K. B. Raja, R. K. Chhootaray, and S. Pattnaik, “PCA based Iris Recognition using DWT,” *Int. J. Comput. Technol. Appl.*, vol. 2, no. August, pp. 884–893, 2011.

[332] A. A. Khan, S. Kumar, and M. Khan, “Iris Pattern Recognition using Support Vector Machines and Artificial Neural Networks,” *Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng.*, vol. 2, no. 12, pp. 2208–2211, 2014.

[333] A. A. Kerim and S. J. Mohammed, “New Iris Feature Extraction and Pattern Matching Based on Statistical Measurement,” *Int. J. Emerg. Trends Technol. Comput. Sci.*, vol. 3, no. 5, pp. 226–231, 2014.

[334] R. M. Da Costa and A. Gonzaga, “Dynamic Features for Iris Recognition.,” *IEEE Trans. Syst. Man. Cybern. B. Cybern.*, vol. 42, no. 4, pp. 1072–1082, 2012.

[335] H. Zhang and X. Guan, “Iris recognition based on grouping KNN and Rectangle Conversion,” *ICSESS 2012 - Proc. 2012 IEEE 3rd Int. Conf. Softw. Eng. Serv. Sci.*, pp. 131–134, 2012.

[336] K. Roy, P. Bhattacharya, and C. Y. Suen, “Iris recognition using shape-guided approach and game theory,” *Pattern Anal. Appl.*, vol. 14, no. 4, pp. 329–348, 2011.

[337] A. D. Rahulkar and R. S. Holambe, “Partial iris feature extraction and recognition based on a new combined directional and rotated directional wavelet filter banks,” *Neurocomputing*, vol. 81, pp. 12–23, 2012.

[338] S. Minaee, A. Abdolrashidi, and Y. Wang, “Iris Recognition Using Scattering Transform and Textural Features,” *Signal Process. Signal Process. Educ. Work. (SP/SPE), IEEE*, pp. 37–42, 2015.

[339] M. H. Abhiram, C. Sadhu, K. Manikantan, and S. Ramachandran, “Novel DCT based feature extraction for enhanced Iris Recognition,” *Proc. - 2012 Int. Conf. Commun. Inf. Comput. Technol. ICCICT 2012*, pp. 1–6, 2012.

[340] M. Vatsa, R. Singh, and A. Noore, “Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing,” *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 38, no. 4, pp. 1021–1035, 2008.

[341] T. W. Ng, T. L. Tay, and S. W. Khor, “Iris recognition using rapid Haar wavelet decomposition,” *Signal Process. Syst. (ICSPS), 2010 2nd Int. Conf.*, vol. 1, pp. V1-820-V1-823, 2010.

[342] C. C. Tsai, H. Y. Lin, J. Taur, and C. W. Tao, “Iris recognition using possibilistic fuzzy matching on local features,” *IEEE Trans. Syst. Man, Cybern. Part B Cybern.*, vol. 42, no. 1, pp. 150–162, 2012.

[343] M. Y. Shams, A. S. Tolba, and S. H. Sarhan, “A Vision System for Multi-View Face Recognition,” *Int. J. CIRCUITS, Syst. SIGNAL Process.*, vol. 10, pp. 455–461, 2016.

[344] S. Vignesh, M. P. K. V. S. N. L, and S. S. Channappayya, “Face Image Quality Assessment for Face Selection in Surveillance Video using Convolutional Neural Networks,” *IEEE Glob. Conf. Signal Inf. Process.*, pp. 577–581, 2015.

[345] V. Khryashchev, I. Nenakhov, A. Lebedev, and A. Priorov, “Evaluation of face image quality metrics in person identification problem,” *9th Conf. Open Innov. Assoc.*, pp. 80–87, 2016.

[346] J. Yu, K. Sun, F. Gao, and S. Zhu, “Face biometric quality assessment via light CNN,” *Pattern Recognit. Lett.*, vol. 0, pp. 1–8, 2017.

[347] A. Abhyankar and S. Schuckers, “Iris quality assessment and bi-orthogonal wavelet based encoding for recognition,” *Pattern Recognit.*, vol. 42, no. 9, pp. 1878–1894, 2009.

[348] Z. Peng and J. Wu, “The research on adaptive fast iris capture and online iris image quality assessment algorithm,” *Optik (Stuttg).*, vol. 126, no. 24, pp. 5971–5978, 2015.

[349] S. Bharadwaj, M. Vatsa, and R. Singh, “Biometric quality: a review of fingerprint, iris, and face,” *EURASIP J. Image Video Process.*, vol. 2014, no. 1, p. 34, 2014.

[350] J. Liu, Z. Sun, and T. Tan, “Distance metric learning for recognizing low-resolution iris images,” *Neurocomputing*, vol. 144, pp. 484–492, 2014.

[351] R. Connaughton, A. Sgroi, K. W. Bowyer, and P. Flynn, “A cross-sensor evaluation of three commercial iris cameras for iris biometrics,” *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 90-97, 2011.

[352] L. Xiao, Z. Sun, R. He, and T. Tan, “Coupled feature selection for cross-sensor iris recognition,” *IEEE 6th International Conference on Biometrics: Theory, Applications and Systems, BTAS 2013*. pp. 1-6, 2013.

[353] M. Baqar, A. Ghani, A. Aftab, S. Arbab, and S. Yasin, “Deep Belief Networks for Iris Recognition Based on Contour Detection,” in *International Conference on Open Source Systems and Technologies (ICOSST) Deep*, pp. 72–77, 2016.

Appendix A

Graphical User Interface

As mentioned earlier that the code of all proposed approaches presented in chapters 3, 4 and 5 was written to run in MATLAB R2015a and later versions. The main window of the proposed hybrid multimodal biometric identification system consists of 3 buttons (See Fig. A.1). The system user can choose one of these multimodal biometric systems based on the biometric traits selected by the user at the identification point, as follows:

1. **Multi-biometric Face Recognition System:** When the system user presses the button a new window of the proposed multi-biometric face recognition system will appear, as shown in Fig. A.2. From the right side of this window, the user will be able to choose which recognition task to implement (either verification task or identification task), and he/she will be able to choose which approach to be used for encoding the face image (e.g., *Curvelet-Fractal* approach, MDFR framework, Multiple Algorithms). Here, Multiple Algorithms option refers to the fusion between the *Curvelet-Fractal* approach and MDFR framework. For face verification task, the user needs to press the '**Load Face Image**' button in order to load face images. A file chooser will appear twice and the user has to select two different face images and after that the '**Face Verification**' panel will show up to visualize the selected face images along with their detected face regions, as shown in Fig. A.2. By pressing the '**Face Matching**' button, the selected face recognition approach will be implemented and a message box will show up to tell if these images are belong to the same person or not. The elapsed time during this process in millisecond will be shown as well. Furthermore, the user is able to choose different fusion method from the Pop-up menu to make the final decision. This Pop-up menu is active only when Multiple Algorithms is selected to make the final decision. On the other hand, when the face identification task is selected the user has to load a face image after pressing the '**Load Face Image**' button. Next, the '**Face**

Identification' panel will appear to visualize the selected face image along with the detected face region, as shown in Fig. A.3. As mentioned before, the system user able to choose which face recognition approach to be used for encoding the face image and also which fusion approach to make the final decision. To establish the person's identity, the user needs to press '**Face Recognition**' button. If the person is registered in the system's database a message box will show up to tell that the person is found and the '**Personal Information**' panel will appear containing different personal information on the identified person (e.g., Full Name, Person ID, Date of Birth, etc.). Otherwise, if the person is not identified by the system a message box will appear to tell that the person is not found. Finally, there are other buttons like '**Face Detection**' button to visualize the detected facial features, '**Add New User**' button to add the person's image and his personal information to the system's database, and '**Curvelet-Fractal Training**' and '**MDFR Training**' buttons are used to automatically train the classifiers of the *Curvelet-Fractal* approach and MDFR Framework, respectively, after the face images of the new registered person have been added.

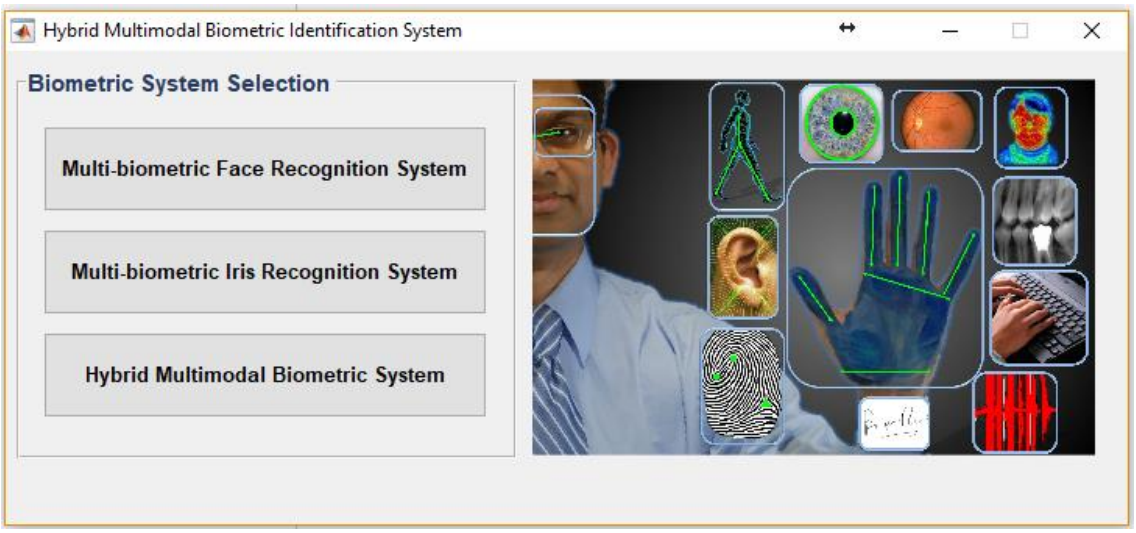


Figure A.1: The main GUI window of the proposed hybrid multimodal biometric identification system.

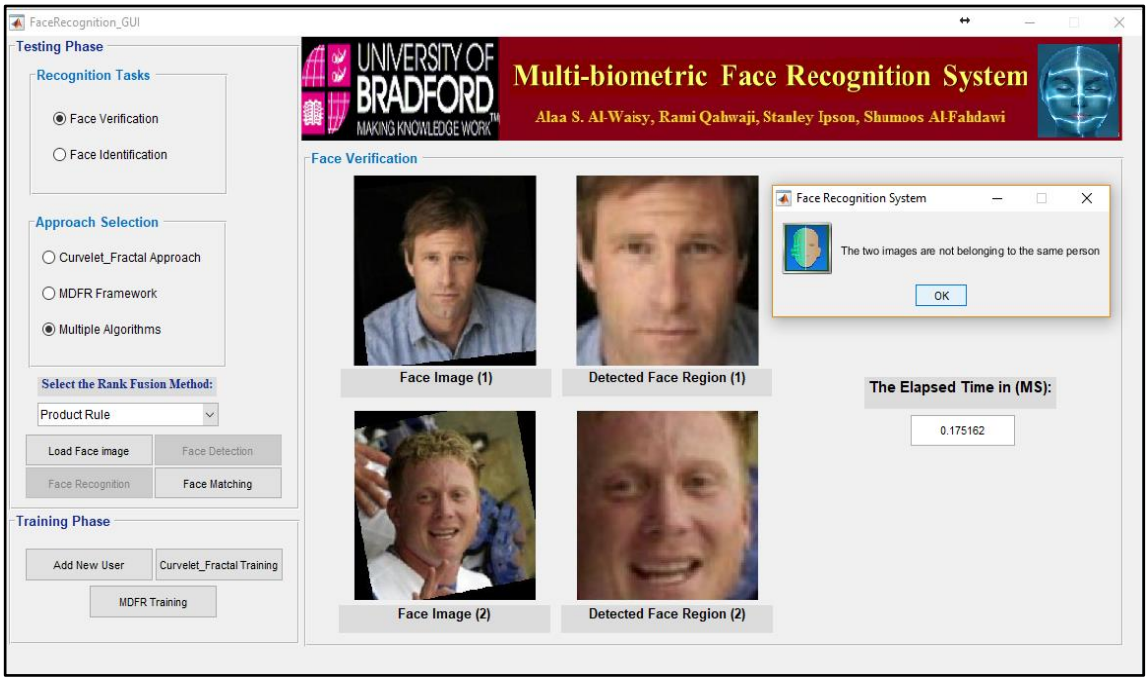


Figure A.2: The main GUI window of the proposed multi-biometric face recognition system operates in the verification mode.

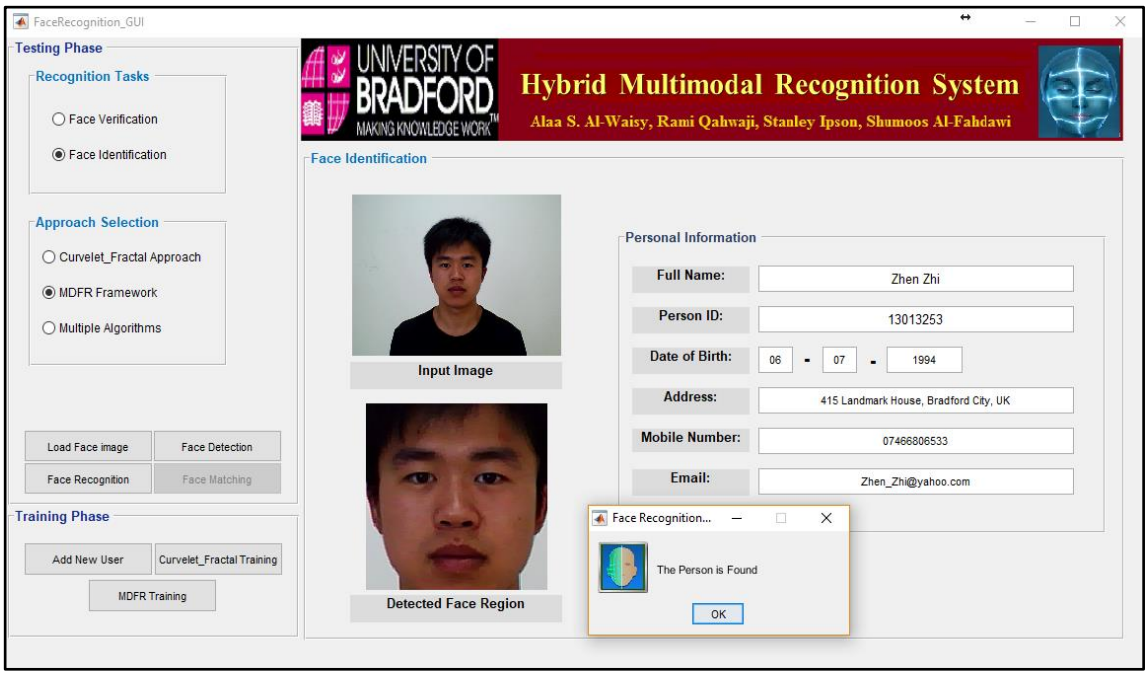


Figure A.3: The main GUI window of the proposed multi-biometric face recognition system operates in the identification mode.

2. **Multi-biometric Iris Recognition System:** When the system user presses the button a new window of the proposed multi-biometric iris recognition system will appear, as shown in Fig. A.4. From the right side of this window, the user will be able to choose which iris trait will be used to establish the person's identity (e.g., Right iris image, Left iris image, or both irises images). If the last option was selected from '**Iris Trait Selection**' panel the system allows to the user to select different fusion method to make the final decision on the person's identity who his/her irises images are under the consideration. Firstly, the user needs to press the '**Load Iris Image'** button to load either one iris image or two based on the selected option from the '**Iris Trait Selection**' panel. If one of the first two options in the from the '**Iris Trait Selection**' panel was selected the '**Iris Localization Stages**' panel will appear to visualize the input image, as shown in Fig. A.4. By pressing the '**Iris Localization**' button, the detected iris boundaries and the detected eyelids and eyelashes will be visualized next to the input image. The result of the iris normalization can be obtained by pressing the '**Iris Normalization**' button. To recognize the person's identity the user needs to press the '**Iris Matching**' button. If the person correctly recognized by the system the '**Personal Profile**' panel will appear containing a personal photo of the person, a message box tells that the person is found, and a '**Personal Information**' panel contains all his/her personal information (See Fig. A.5). Otherwise, if the person is not registered in the database, a message box will appear to tell that the person is not found. The elapsed time will be visualized in edit box below the '**Personal Information**' panel. If the last option was selected from '**Iris Trait Selection**' panel, the '**Right & Left Iris Images**' panel will show up to visualize the input irises images along with their detected iris regions, as shown in Fig. A.6. Finally, the person's identity can be established in the same procedure described above by pressing the '**Iris Matching**' button.

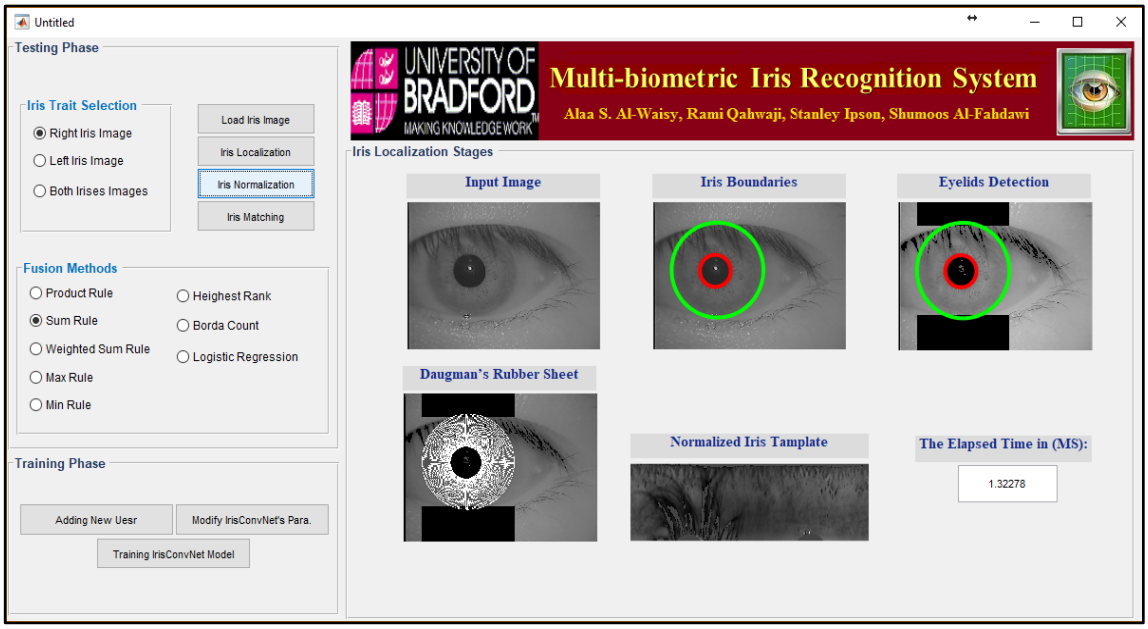


Figure A.4: The main GUI window of the proposed multi-biometric iris recognition system when only one of the iris image is selected.

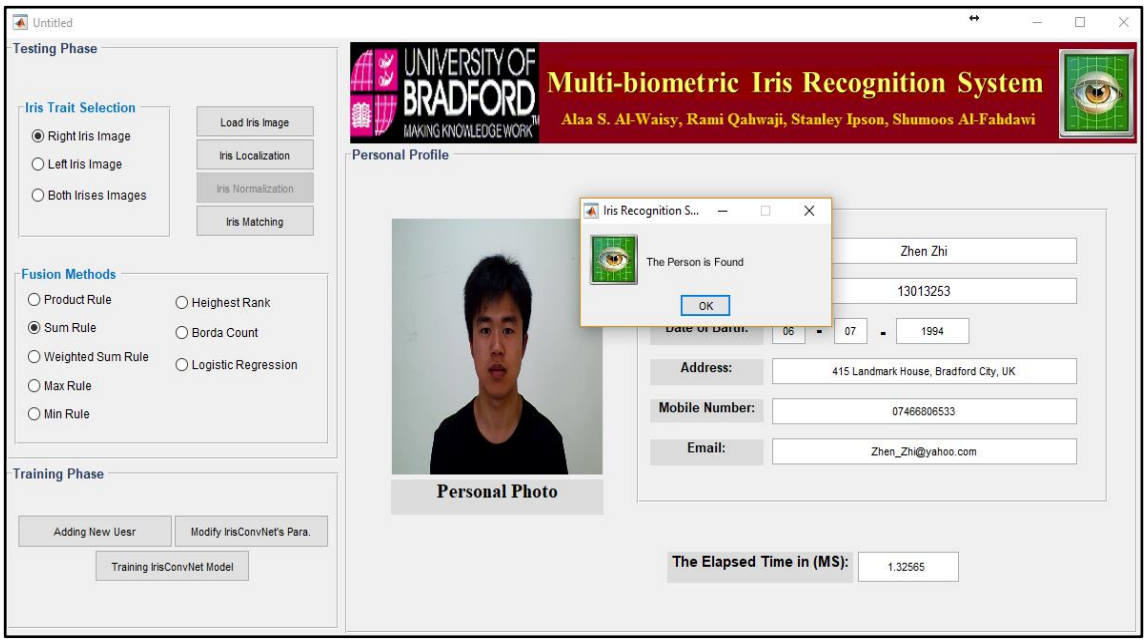


Figure A.5: The main GUI window of the proposed multi-biometric iris recognition system when the person is correctly identified by the system.

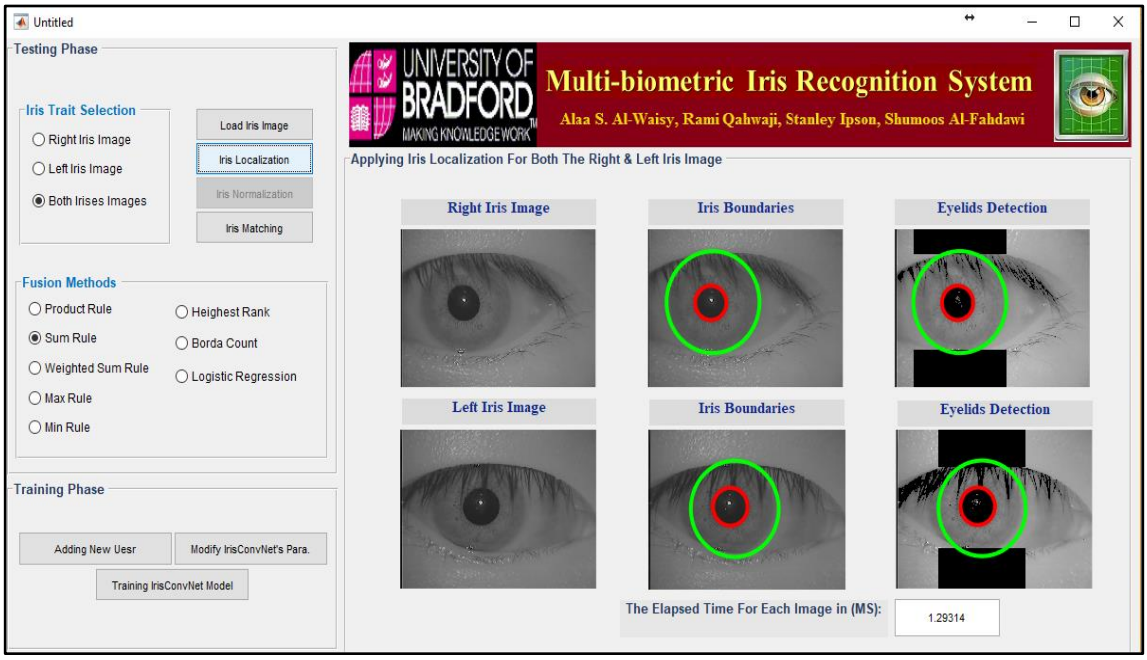


Figure A.6: The main GUI window of the proposed multi-biometric iris recognition system when both irises images of the same person are selected.

3. **Hybrid Multimodal Recognition System:** By pressing this button the main GUI window of the proposed hybrid multimodal recognition system will appear, as shown in Fig. A.7. As can be seen from this figure, the proposed system offers the user a high degree of flexibility to choose from the '**Biometric Trait Selection**' panel which biometric trait will be used at the identification point. Furthermore, the system user can choose different face recognition approaches and fusion method to make the final decision from the '**Face Approach Selection**' panel and '**Select the Fusion Method**' pop up menu, respectively. Finally, person's identity can be recognized by pressing the '**Person Recognition**' button.

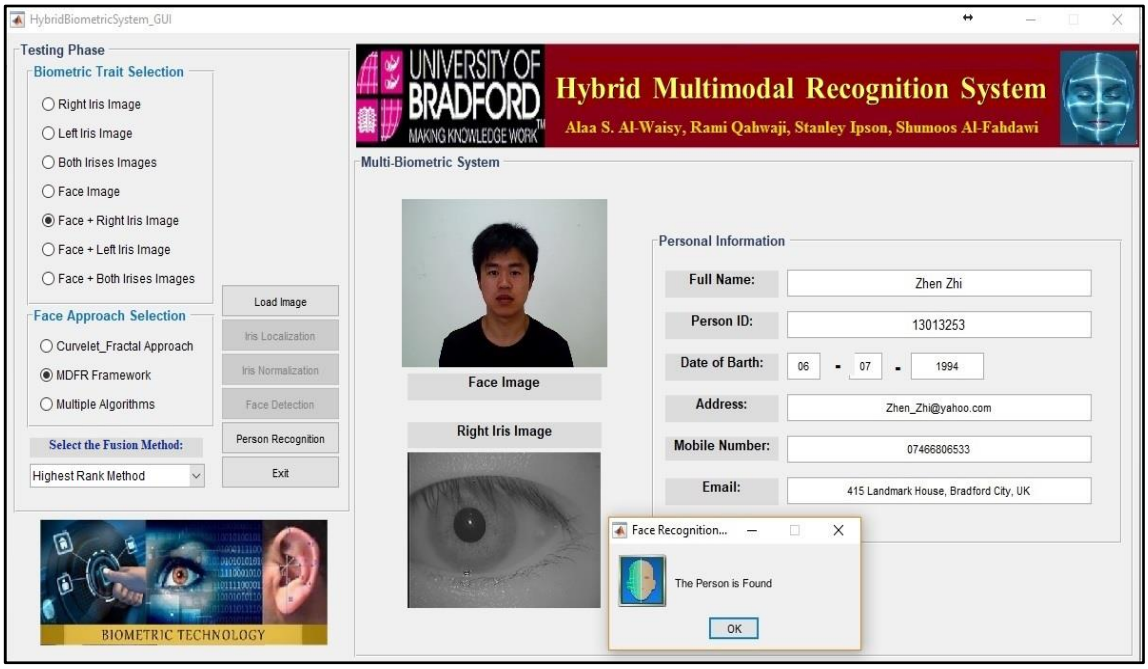


Figure A.7: The main GUI window of the proposed hybrid multimodal recognition system.