

Ph.D. Thesis

**IMPROVED BIOMETRIC AUTHENTICATION SYSTEM
USING MULTIMODAL CUE INTEGRATION**

Submitted to

THE COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY

in partial fulfilment of the requirement for the award of the degree of

Doctor of Philosophy

by

BINSU C. KOVOOR

Under the guidance of

Dr. K. Poulose Jacob & **Dr. Supriya M.H.**
Professor, Computer Science *Associate Professor, Electronics*

DEPARTMENT OF COMPUTER SCIENCE
FACULTY OF TECHNOLOGY
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
KOCHI – 682 022, INDIA

JANUARY 2015

**IMPROVED BIOMETRIC AUTHENTICATION SYSTEM USING MULTIMODAL
CUE INTEGRATION**

Ph.D. Thesis in the field of Biometrics

Author

BINSU C. KOVOOR

Asst. Professor

Division of Information Technology

School of Engineering

Cochin University of Science and Technology

Kochi– 682 022, India

e-mail: binsu@cusat.ac.in

Research Advisors

Dr. K. Poullose Jacob

(Supervising Guide)

Professor, Computer Science

Cochin University of Science and Technology

Kochi– 682 022, India

e-mail : kpj@cusat.ac.in

Dr. Supriya M.H.

(Co Guide)

Associate Professor, Electronics

Cochin University of Science and Technology

Kochi– 682 022, India

e-mail : supriya@cusat.ac.in

JANUARY 2015

Dedicated to.....

My Parents, Husband & Kids



FACULTY OF TECHNOLOGY
COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY
KOCHI – 682 022, INDIA

Certificate

This is to certify that this thesis entitled, *Improved Biometric Authentication System Using Multimodal Cue Integration* is a bonafide record of the research work carried out by *Ms. Binsu C. Kovoov* under our supervision in the Department of Computer Science, Cochin University of Science and Technology. All the relevant suggestions and modifications suggested by the audience during the pre-synopsis seminar and recommended by the Doctoral Committee of the candidate have been incorporated in the thesis. The results presented in this thesis or parts of it have not been presented for the award of any other degree(s).

Dr. K. Poulouse Jacob
(Supervising Guide)
Professor, Computer Science

&

Dr. Supriya M.H.
(Co Guide)
Associate Professor, Electronics

Kochi - 682022
26th January 2015

Declaration

I hereby declare that the work presented in this thesis entitled *Improved Biometric Authentication System Using Multimodal Cue Integration* is a bonafide record of the research work carried out by me under the supervision of Dr. K. Poullose Jacob, Professor, Department of Computer Science, Cochin University of Science & Technology & Dr Supriya M.H., Associate Professor, Department of Electronics, Cochin University of Science & Technology. The results presented in this thesis or parts of it have not been presented for the award of any other degree(s).

BINSU C. KOVOOR
Ph.D Reg No. 3439

Kochi – 22
26th January 2015

Acknowledgements

I wholeheartedly express my sincere gratitude to my guide Prof. (Dr) K. Poulose Jacob, Pro-Vice Chancellor of Cochin University of Science and Technology for his guidance, encouragement and support throughout my career at CUSAT and during the course of this work. .

I have been deeply indebted in the preparation of this thesis to my co-guide, Dr. Supriya M.H, Associate Professor, Department of Electronics, CUSAT, whose patience and kindness, as well as her academic experience, has been invaluable to me. I am very much thankful to her for spending so much time and motivating me throughout the period.

I am thankful to Prof.(Dr.) Sumam Mary Idicula, Director, Department of Computer Science, CUSAT, for granting me permission to continue the research in the Department. I must also thank the Department of Electronics and Department of Computer Science, CUSAT for providing me with all the facilities and a very congenial environment.

My sincere thanks to Dr.A.Unnikrishnan, former NPOL Associate Director, Dr. David Peter, Registrar, CUSAT and Dr. G. Santhosh Kumar, Associate Professor, Department of Computer Science, for their suggestions and support. I thank all the staff members in the Cochin University of Science and Technology for their encouragement throughout the program.

I would like to thank my friends Suraj Kamal, Satheesh Chandran C. and Shameer K. Muhammed for extending their helping hand when I desperately needed. I also extend my sincere thanks to the faculties in Division of Information Technology for their valuable suggestions and constant support and encouragement rendered to me. I would like to express my deepest appreciation to Sri.Subash Chandrabose M.R, Group Head, Transducer Division, NPOL, for his brilliant comments and suggestions during the preparation of my thesis.

I thankfully bear in mind the sincere co-operation and support I received from the library and administrative staff of the School of Engineering.

It is beyond words to express my gratitude to my parents and family members especially to my husband Dr Job Thomas and my kids Gary and Jenifer for their sacrifice in connection with preparation of my thesis. I am sure I could not have completed this great task without their support and cooperation.

Finally, I bow my head with devotion and reverence before the Almighty, who makes everything a reality

BINSU C. KOVOOR

26th January 2015.

Table of Contents

<i>Declaration</i>	<i>vii</i>
<i>Acknowledgements</i>	<i>ix</i>
<i>Table of Contents</i>	<i>xi</i>
<i>List of Figures</i>	<i>xv</i>
<i>List of Tables</i>	<i>xix</i>
<i>Abbreviations</i>	<i>xxi</i>
CHAPTER 1	1
INTRODUCTION	1
1.1 Introduction	1
1.2 Biometric Modes	3
1.3 Performance of a Biometric System	4
1.3.1 False Acceptance Rate (FAR) or False Match Rate (FMR)	4
1.3.2 False Rejection Rate (FRR) or False Non-Match Rate (FNMR).....	5
1.3.3 Receiver Operating Characteristic (ROC)	5
1.3.4 Equal Error Rate (EER) or Crossover Error Rate (CER).....	5
1.3.5 Failure To Enroll Rate (FTER).....	5
1.3.6 Failure To Capture Rate (FTCR).....	6
1.4 Challenges in Biometrics	6
1.4.1 Accuracy.....	6
1.4.2 Scalability	7
1.4.3 Security and Privacy.....	8
1.5 Multibiometric Systems	9
1.5.1 Multimodal Biometric Fusion Scenarios	10
1.5.2 Multimodal Biometric Architecture.....	12
1.5.3 Design Issues in Multibiometrics	14
1.5.4 Levels of Fusion	16
1.5.5 Challenges in Multibiometrics System Design.....	20
1.6 Motivation and phases of the work	22
1.7 Organization of the Thesis	24
1.8 Summary	26
CHAPTER 2	27
REVIEW OF LITERATURE	27
2.1 Introduction	27

2.2	Speech	28
2.3	Face	37
2.4	Signature	46
2.5	Iris	56
2.6	Multimodal.....	63
2.7	Summary	81
CHAPTER 3		82
METHODOLOGY.....		82
3.1	Introduction.....	82
3.2	System Overview	84
3.3	Data Acquisition.....	85
3.4	Unimodal Processing.....	87
3.4.1	Speech Processing	87
3.4.2	Face Processing	87
3.4.3	Signature Processing.....	88
3.4.4	Iris Processing	89
3.4.5	Multimodal Fusion	89
3.4.6	Development of knowledge base.....	90
3.5	Performance Analysis	91
3.6	Summary	92
CHAPTER 4		93
SPEAKER RECOGNITION.....		93
4.1	Introduction.....	93
4.2	Dimensions of Difficulty.....	96
4.3	Speaker Recognition Techniques.....	97
4.3.1	Front- End Processing	97
4.3.2	Cepstral Features	99
4.3.3	Speaker Modeling.....	105
4.3.4	Decision logic.....	108
4.4	Results and Discussions	109
4.5	Performance analysis of speaker recognition system.....	114
4.6	Summary	116
CHAPTER 5		118
FACE RECOGNITION		118
5.1	Introduction.....	118
5.2	Face Recognition Algorithms	120
5.2.1	Template-based Face Recognition.....	121
5.2.2	Appearance-based Face Recognition.....	122
5.2.3	Model-based Face Recognition	123
5.3	Linear Subspace Techniques.....	125
5.3.1	Principal Component Analysis	125

5.3.2	Recognition Procedure from PCA Model.....	135
5.4	Results and Discussions	136
5.5	Summary	140
CHAPTER 6		141
SIGNATURE IDENTIFICATION.....		141
6.1	Introduction.....	141
6.2	Signature Identification Methodology.....	144
6.2.1	Data Acquisition.....	144
6.2.2	Preprocessing.....	145
6.2.3	Feature Extraction.....	147
6.2.4	Classification	151
6.3	Enrolment and Identification.....	151
6.4	Results and Discussions	152
6.4.1	Effect of static features	152
6.4.2	Effect of dynamic features.....	156
6.5	Summary	158
CHAPTER 7		159
IRIS RECOGNITION SYSTEM.....		159
7.1	Introduction.....	159
7.1.1	Iris Anatomy.....	160
7.1.2	Iris as a Biometric.....	162
7.2	Iris recognition system.....	163
7.2.1	Iris Pre-processing	165
7.2.2	Iris localization or segmentation.....	165
7.2.3	Iris Normalization and Unwrapping	169
7.2.4	Feature Extraction.....	170
7.2.5	Matching.....	176
7.3	Iris Database.....	177
7.4	Results and Discussions	178
7.4.1	F ₁ Score	180
7.5	Summary	182
CHAPTER 8		183
MULTI-MODAL BIOMETRIC SYSTEM.....		183
8.1	Introduction.....	183
8.2	Score Level Fusion	184
8.2.1	Non-weighted Fusion	184
8.3	User-Dependent Fusion Strategy	185
8.3.1	Score Normalization Technique	186
8.4	Multimodal Technique of identification.....	189
8.4.2	User-Dependent Decision.....	193
8.5	Results and Discussions	194
8.5.1	Calculation of weights for fusion.....	195

8.5.1.1	Calculation of minimum Authorised and maximum Unauthorised Score	196
8.6	System Performance and Evaluation	198
8.7	Summary	201
CHAPTER 9	203
CONCLUSIONS AND FUTURE WORK	203
9.1	Highlights of the Thesis.....	203
9.1.1	Need and Requirement of a Biometric Authentication System	204
9.1.2	Preparation of a State-of-the-art Literature.....	205
9.1.3	Feature Vector Based Authentication System	205
9.1.4	Extraction of Feature Vector Element from Speech.....	206
9.1.5	Extraction of Feature Vector Element from Face	206
9.1.6	Extraction of Feature Vector Element from Signature	207
9.1.7	Extraction of Feature Vector Element from Iris	207
9.1.8	Authentication System Based on Multi-modal Approach	208
9.1.9	Towards Improving the Performance of the Prototype.....	208
9.2	Future Scope for Research	209
9.3	Summary	209
References	211
List of Publications	233
Subject Index	235

List of Figures

Figure 1.1 Enrolment Phase of the Biometric System	3
Figure 1.2 (a) Verification (b) Identification modes in a biometric system	4
Figure 1.3 Architecture for several classifier combinations	13
Figure 1.4 Fusion Levels in multimodal biometric fusion.....	16
Figure 1.5 Flow of information in a match score level fusion scheme.....	19
Figure 2.1 ROC curves for unimodal and multimodal systems [127].	67
Figure 2.2 ROC curves for unimodal and multimodal systems [130].	68
Figure 2.3 ROC curves for user-specific thresholds and user-specific	69
Figure 2.4 ROC curves for unimodal and multimodal systems [132].	70
Figure 3.1 Block diagram of the proposed multimodal biometric system.	90
Figure 4.1 Block Diagram of a Speaker Identification System.....	95
Figure 4.2 Speaker Recognition Technique.....	97
Figure 4.3 Speech process models	101
Figure 4.4 The Mel-scale.....	101
Figure 4.5 Mel-Frequency filter bank on a linear frequency (Hz) scale	102
Figure 4.6 Procedure for vector quantization	107
Figure 4.7 Conceptual diagram illustrating VQ codebook formation.	109
Figure 4.8 Speech signal of Speaker 1.....	110
Figure 4.9 Power spectral analysis of speech signal of Speaker 1	111
Figure 4.10 Hamming Window	112
Figure 4.11 Power spectrum of speaker 1 modified through mel-spaced filter bank	113
Figure 4.12 MFCCs corresponding to speaker1 through 5th and 6th filters.....	113
Figure 4.13 Codebooks and MFCCs corresponding to speaker 1 and 2.....	114
Figure 4.14 Influence of Cepstral & Spectral features.....	115
Figure 4.15 Success Rate of Cepstral and Spectral features.....	116
Figure 5.1 Comparison of various biometrics based on MRTD compatibility.....	120
Figure 5.2 Classification of face recognition techniques	121
Figure 5.3 Multiview faces overlaid with labeled graphs.....	123
Figure 5.4 3D morphable face model [201]	124
Figure 5.5 Overview of the face recognition system	128
Figure 5.6 Subset of Training face database	129
Figure 5.7 Schematic representation of mapping of single	130

Figure 5.8 Mean image of face database.....	131
Figure 5.9 Eigen faces of the Training Image Set, which is a subset.....	133
Figure 5.10 Sample reconstructed image of the original input image.....	135
Figure 5.11 Root mean square error for reconstructed images.....	137
Figure 5.12 Number of Eigen faces vs Eigen value variation.....	138
Figure 5.13 Execution time with and without covariance matrix.....	139
Figure 5.14 PCA Recognition Rate	139
Figure 6.1 Architecture for an online signature identification system.....	144
Figure 6.2 Signature acquisition device -WACOM Bamboo pad.....	145
Figure 6.3 Signature before and after rotation.	147
Figure 6.4 The real and imaginary part of Gabor Filters	152
Figure 6.5 Gabor Filters.....	153
Figure 6.6 Amplitude Response of Gabor features.....	153
Figure 6.7 Phase Response of Gabor filters.....	154
Figure 6.8 Influence of the size of the normalisation box on false rejection and acceptance rates against threshold.	155
Figure 6.9 Equal error rate against size of the normalization box.	156
Figure 6.10 Average verification time versus size of the normalization box.	156
Figure 6.11 Example of a signature, its x and y plot as well as its normalised x plot against time.....	157
Figure 6.12 Effect of dynamic feature on the success rate.	157
Figure 7.1 Diagrammatic view of anatomy of an eye adapted from [218].....	161
Figure 7.2 Block diagram for an Iris Recognition System	164
Figure 7.3 Iris Segmentation Stages.....	166
Figure 7.4 Unwrapping: Daugman’s Rubber Sheet Model	169
Figure 7.5 Unwrapped Normalized iris image	170
Figure 7.6 Feature Extraction Stages.....	171
Figure 7.7 Histogram Equalized image	171
Figure 7.8 Canny edge-detected image	173
Figure 7.9 Subset of UBIRIS eye database used for training.....	177
Figure 7.10 Detection of iris edge using various operators	179
Figure 7.11 Decision making in iris biometric system.	179
Figure 7.12 Comparison of edge detection operators	181
Figure 8.1 Mode of operation of a multimodal identification system	190

Figure 8.2 Authorized-Unauthorized frequency distribution curve	191
Figure 8.3 Authorised-Unauthorised Distribution Curve for the Biometric traits...	195
Figure 8.4 Authorised-Unauthorised Distribution Curve for trimodal fusion.....	196
Figure 8.5 Authorised-Unauthorised Distribution Curve for possible fusion of the Biometric traits for bimodal.....	197
Figure 8.6 Authorised-Unauthorised Distribution Curve for the Tetramodal.....	198
Figure 8.7 Width of the confusion zone.	200
Figure 8.8 Chart of Success Ratio of the biometric identification system	202

List of Tables

Table 2.1 Comparison of multimodal biometric systems	64
Table 2.2 Characteristics of the five modalities [141].....	73
Table 3.1 Acquisition method and specifications	85
Table 6.1 Details of the number of signature images.....	152
Table 7.1 EER of different operators.....	178
Table 7.2 Performance measures for F_1 score.....	180
Table 8.1 Weights calculated for each trait in all the possible fusions	193
Table 8.2 Success Ratio (SR) of Biometric identification system	201

Abbreviations

ASR	- Automatic Speech Recognition
DWT	- Discrete Wavelet Transform
EER	- Equal Error Rate
FAR	- False Acceptance Rate
FFT	- Fast Fourier Transform
FMR	- False Match Rate
FNMR	- False Non Match Rate
FRR	- False Rejection Rate
FTCR	- Failure To Capture Rate
FTER	- Failure To Enrol Rate
FV	- Feature Vector
GAR	- Genuine Acceptance Rate
GMM	- Gaussian Mixture Model
GWT	- Gabor Wavelet Transform
HD	- Hamming Distance
LDA	- Linear Discriminant Analysis
LPC	- Linear Prediction Coefficients
MAD	- Median Absolute Deviation
MD	- Mahalanobis Distance
MFCC	- Mel Frequency Cepstral Coefficients
MLP	- Multi Layer Perceptron Network
PCA	- Principal Component Analysis
PDF	- Probability Distribution Function
ROC	- Receiver Operating Characteristics
SNR	- Signal to Noise Ratio
SVM	- Support Vector Machine
VQ	- Vector Quantization
ZCR	- Zero Crossing Rate

CHAPTER 1

INTRODUCTION

The dependence of society on the usage of information technology for everyday tasks makes the establishment of the identity of a person in a reliable and time-efficient manner a matter of paramount importance. Biometrics is an efficient technology with great possibilities in the area of security system development for official and commercial applications. In unimodal biometrics, there are challenges such as noisy data, restricted degree of freedom, intra-user variability, non-universality, spoof attack and unacceptable error rate. It has been identified that there is a need to combine multiple traits to address these challenges. This chapter touches upon the discipline of biometrics and its evolution towards multimodal biometrics. The chapter describes the modes of operations, performance measures and limitations of a unimodal biometric system. It also investigates the key issues in multimodal biometric systems along with various architectures for information integration.

1.1 *Introduction*

In the current scenario where the society is becoming increasingly dependent on the usage of information technology for everyday tasks, the establishment of the identity of a person in a reliable and time-efficient manner becomes a matter of paramount importance. Biometrics, a measurable distinctive physical characteristic or personal trait such as fingerprint, face, iris, voice, gait etc. can be used to identify an individual

or to verify the claimed identity of an individual and has gained considerable attention these days due to its robustness as well as reliability. Biometric authentication [1] automatically achieves the establishment of the identity of an individual based on the physiological or behavioural biometric trait rather than knowledge based or possession based method.

A biometric trait cannot be easily lost, forgotten, shared or forged and is a strong link between a person and his/her identity. Since in biometric systems, the presence of an individual at the time of authentication is a necessity, prevention of unauthorised access as well as false refusal claims can be achieved. The biometric systems which have two phases, the enrolment as well as the verification/identification phase, comprise four main modules [2], [3], *viz.* a sensor, feature extractor, matcher and decision components.

In the enrolment phase, which is the process of registering information in a knowledge base, the sensor obtains the biometric data from an individual and the feature extractor module extracts significant information regarding the acquired biometric sample leading to the generation of a feature vector. The feature vector, being unique for each person becomes the template for authentication. The enrolment phase is as shown in Figure 1.1. The enrolment can be for either verification or positive identification, which is referred to as positive enrolment or can be for negative identification, referred to as negative enrolment.

Proper authentication is often ensured with biometrics so that only authorised people have access while unauthorised people are prevented. The matcher module, a part of the authentication phase, compares the extracted feature set of the person-seeking authorisation to the template in the knowledge base. Based on the degree of similarity between the

template and the query, the decision module resolves the identity of the user.

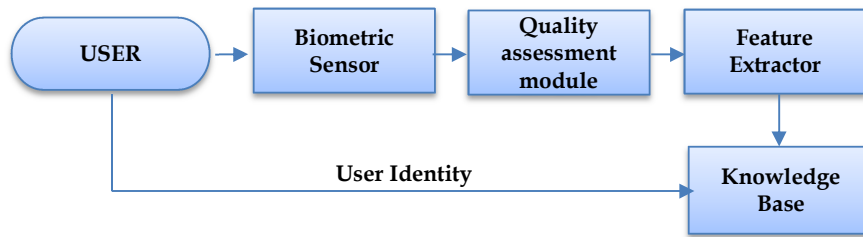


Figure 1.1 Enrolment Phase of the Biometric System

Even though the biometric identification systems out-perform their peer technologies, the unimodal biometric systems have to contend with a variety of problems, namely, noisy data, intra-class variations, restricted degrees of freedom, non-universality and spoof attacks. Deploying a multimodal biometric system can address the limitations by integrating the evidences presented to it by multiple sources of information.

1.2 Biometric Modes

A biometric system operates either in verification or identification mode. In verification, the input is compared with the template of the claimed identity, while in identification the user's biometric trait is compared with templates of all the persons enrolled in the database. The block diagram depicting the two modes of operation is as shown in Figure 1.2.

When used in verification mode, a good degree of similarity identifies the claimant as genuine, else is rejected as impostor. In the identification mode, when the query is having a good degree of similarity with any one of the templates in the database, the query is authenticated in the positive enrolment while found ineligible in the negative enrolment.

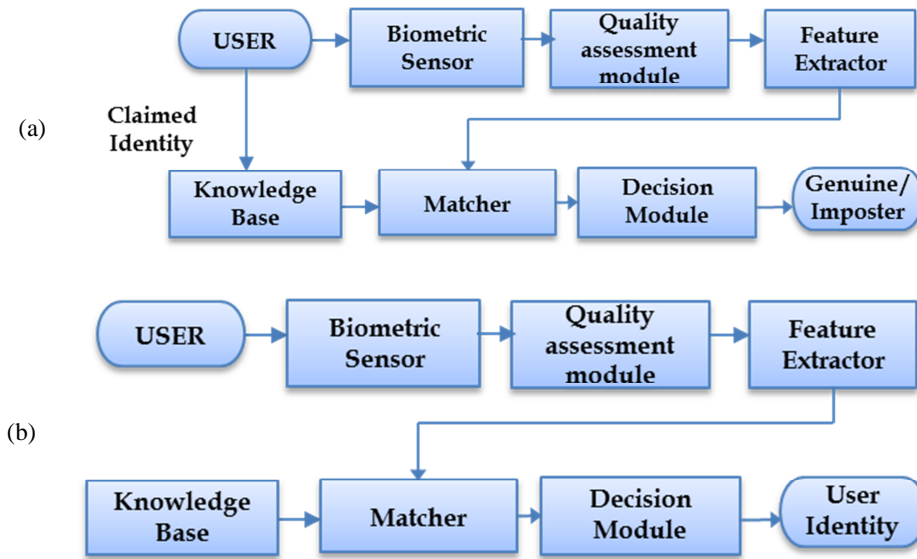


Figure 1.2 (a) Verification (b) Identification modes in a biometric system

1.3 Performance of a Biometric System

The performance metrics [4] used to quantify the performance of a biometric system are

- False Match Rate (FMR)
- False Non Match Rate (FNMR)
- Receiver Operating Characteristic (ROC)
- Equal Error Rate (EER)
- Failure to Enroll Rate (FTER)
- Failure to Capture Rate (FTCR)

1.3.1 False Acceptance Rate (FAR) or False Match Rate (FMR)

FAR is the probability that the system incorrectly matches the input pattern to a non-matching template in the database. It measures the percentage of invalid inputs, which are incorrectly accepted. In case of similarity scale, if the person is an imposter in reality, but the matching score is higher than the threshold, then he/she is treated as genuine. This

increases the FAR, which thus also depends upon the threshold value. An FMR of 0.01% implies that on an average, 1 in 10,000 impostor attempts are likely to succeed.

1.3.2 False Rejection Rate (FRR) or False Non-Match Rate (FNMR)

FRR is the probability that the system fails to detect a match between the input pattern and a matching template in the database. A measurement of the percent of valid inputs, which are incorrectly rejected, is evaluated by this parameter. A 10% FNMR indicates that on average, 10 in 100 genuine attempts do not succeed

1.3.3 Receiver Operating Characteristic (ROC)

The ROC plot is a visual characterization of the trade-off between the FAR and the FRR. In general, the matching algorithm performs a decision based on a threshold, which determines how close to a template the input needs to be for it to be considered a match. If the threshold is reduced, there will be fewer false non-matches but more false accepts. Conversely, a higher threshold reduces the FAR but increases the FRR.

1.3.4 Equal Error Rate (EER) or Crossover Error Rate (CER)

EER is the rate at which both acceptance and rejection errors are equal. The value of the EER can be easily obtained from the ROC curve. The EER is a quick way to compare the accuracy of devices with different ROC curves. In general, the device with the lowest EER is the most accurate.

1.3.5 Failure To Enroll Rate (FTER)

FTER is the rate at which attempts to create a template from an input is unsuccessful. Low quality inputs normally result in high FTER.

1.3.6 Failure To Capture Rate (FTCR)

Within automatic systems, the probability that the system fails to detect a biometric input when presented correctly is referred to as Failure to Capture Rate.

1.4 Challenges in Biometrics

Accuracy, scalability and usability contribute to the main challenges faced in the implementation of a biometric system. An ideal biometric system should be highly accurate and secure, convenient to use, and easily scalable to a large population and hence the major obstacles that hinder the design and development of such a system are currently the thrust areas of research.

1.4.1 Accuracy

A biometric system rarely encounters 100 percent match between the captured biometric trait of the user and the template in the knowledge base. The accuracy of a biometric system [5] is affected by:

- Noisy sensor data
- Non-universality
- Inter-user similarity
- Lack of invariant representations

The defective or improperly maintained sensors can cause noise in the acquired biometric traits. Typically, failure of camera focus can result in blurring of face and iris image leading to a noisy sensor data which can bring in significant reduction in the Genuine Accept Rate (GAR) of a biometric system [6], [7].

A biometric trait becomes universal if and only if every individual in the target population is able to submit it as a means for identification.

But it is often seen that all the biometric traits are not truly universal; for example, a fingerprint verification system cannot enrol an individual with hand related disabilities, manual workers with many cuts on their fingerprint etc. Similarly, people having eye abnormalities and diseases cannot provide good quality eye images for automatic recognition. Non-universality leads to high FTER and FPCR in a biometric system.

The features extracted from the biometric traits of two persons may not be distinct in some cases, which will lead to inter-user similarity. The similarity between the biometric samples from two different individuals leads to an increase in FMR.

Use of different sensors during enrolment and verification can result in variations in the biometric samples. The ambient environmental conditions, aging of a person, appearance of wrinkles, presence of facial hair etc can also bring in variations in the representation of intra-user trait. There will be a decrease in the GAR of a biometric system due to intra-user variations.

1.4.2 Scalability

When the authentication is performed by matching the query with a single template, as in the case of a verification system, the size of the database is irrelevant. In an identification system, the query needs a sequential comparison with all the templates stored in the database, often resulting in a reduction in the throughput and an increase in the false match rate. Scaling can be achieved by a process known as filtering or indexing where the database is classified based on external or internal factors and the search is restricted to a smaller fraction of the database thereby resulting in the minimisation of errors. Some examples of the external factors are gender, age etc. and that of the internal are fingerprint, iris pattern etc.

1.4.3 Security and Privacy

The major issue encountered in biometric systems is the protection of the template of a user trait stored in the database. It is possible that data obtained during biometric enrolment will be used in ways for which the enrolled individual has not consented. In addition, stolen biometric templates can affect the security of the system in several ways like gaining fake access using enrolled voice by an imposter. Other typical examples of circumventing a biometric system can be by dodging a fingerprint biometric system by constructing fake fingers [8], easy execution of spoofing of behavioural traits such as signature [9] and voice [10].

One of the major concerns for biometric verification and identification systems is privacy protection. When the biometric templates are stored in an unprotected database leaving free access to any intruder, this can result in hacking of the knowledge base. Since easy modification of biometric traits is not possible unless it has been stored using cryptographic methods, the attack against the stored templates causes a major security and privacy threat.

Another important issue associated with biometric systems is the privacy of biometric traits. Privacy is the ability to lead your life free of intrusions, to remain autonomous and to control access to own private information. The systematic privacy concerns that come into play are unintended functional as well as application scope and covert recognition. In unintended functional scope, collectors might glean additional (possibly statistical) personal information from scanned biometric measurements whereas in unintended application scope, strong biometric identifiers such as fingerprints allow the possibility of unwanted identifications. Since biometric characteristics are not secrets, it is often possible to obtain a

biometric sample, such as a person's face, without that person's permission. This permits covert recognition of previously enrolled people. Consequently, the biometric recognition can deny the privacy of those who desire to remain anonymous in any particular situation.

1.5 Multibiometric Systems

Multimodal or Multibiometric utilizes more than one physiological or behavioural characteristic for enrolment, verification, or identification purpose [4]. This implies integration of the evidences from multiple sources of biometric information in order to authenticate the identity of an individual. The identification process becomes more reliable when multiple traits are being accounted, thereby alleviating the limitations of unimodal system [11]. Some of the advantages of multibiometric systems over unibiometric systems are as follows.

- The overall accuracy of a biometric system improves significantly by combining the evidences obtained from different biometric sources using an effective fusion scheme.
- A multibiometric system addresses the non-universality problem thereby achieving a reduction in FTER and FPCR. For example, if the enrolment of a person in iris recognition systems is inhibited due to eye diseases, he can still be identified using other biometric traits such as face, fingerprint etc.
- In a multimodal system, since the user enrolls using different traits, achievement of a degree of flexibility in user authentication is possible. Depending on the nature of the application, it is possible to use the system with a subset of these traits. For example, when face, voice and fingerprint are used to enrol into banking systems, it is possible to select one or two traits for authentication based on convenience.

- The effect of noisy data is reduced in multimodal system and this enables reliable determination of identity even if one of the biometric samples is noisy.
- Multibiometric systems help to search a large database in a computationally efficient manner. The voluminous database is minimised by eliminating non-probable personnel using less accurate models and fine tuning is then used to perform final identification tasks.
- Since simultaneous spoofing of multiple biometric traits is improbable, they are more resistant to spoof attacks. Further, a multibiometric can also check the liveness of the users by acquiring a subset of traits in some random order.

Some of the disadvantages associated with multibiometric systems are that they are more expensive as well as need more time for enrolment. Expensiveness arises since it requires more resources for computation and storage while the additional time requirement can cause inconvenience to the user. There can also be an accuracy reduction if appropriate techniques are not followed for combining the evidences provided by different sources. With all these limitations, multibiometric systems offer features that are attractive and as a result, they can be used widely in security critical applications.

1.5.1 Multimodal Biometric Fusion Scenarios

A multimodal biometric verification system can be considered as a classical information fusion problem i.e. can be thought to combine evidences provided by different biometrics to improve the overall decision accuracy. Generation of multiple evidences is possible from the same biometric sample by distinct algorithms, which may differ in the manner

the biometric data is pre-processed, or in terms of the features extracted from the pre-processed data, or in the choice of the algorithm performing the matching of the biometric sample to one or more user templates. Repeated application of the same chain of processing to more than one biometric sample of the same modality could yield a set of candidate decisions. Extension of the range of decision sets is possible by collecting more than one type of biometric from the user. Integration of information provided by multi biometric systems thus becomes crucial for the efficient performance of the system and this raises a challenging problem of information fusion.

Information fusion in principle can be performed at data, feature or decision levels. Although there may be merits in fusing information at low levels, from the multi biometric system design point of view, it is most appealing to focus on the decision level fusion, as there will be integration of decisions by specialists in the respective biometric modalities. The logical consequence of this argument is that the fusion should be performed at the symbolic decision level where each expert has already determined the user's most likely identity. Some form of voting would then be sufficient to resolve any conflicts of opinions from a given set of experts.

Multiple modalities or Multimodal biometrics, Multiple sensors, Multiple algorithms, Multiple instances and Repeated instances are some of the common fusion scenarios [12], [13]. In multiple modalities or multimodal biometrics, the biometric traits are extracted from two or more modalities using single or multiple sensors. In multiple sensor fusion scenario, different sensors obtain the same instance of a biometric trait. The verification of face image based on an image captured via two sources like static digital image and video frame is an example of multiple sensor fusion scenario.

In multiple algorithms, two or more different algorithms process a single sample captured by a single sensor. The face recognition verification system utilising feature-based and view-based approaches is an example of multiple algorithm fusion scenario. In multiple instances, the system is built using a number of biometric samples from different instances of the same biometric trait. The use of left and right iris images for identity authentication is an example of multiple instances fusion scenario.

In repeated instances, the same sensor acquires the same biometric modality several times. The sequential frame capture of facial images is an example of repeated instances fusion scenario.

1.5.2 Multimodal Biometric Architecture

The sequence of acquisition and processing of biometric samples has a significant impact on the time required for enrolment and authentication. A multimodal system operates in one of the three different operational modes: serial, parallel or hierarchical [14] as shown in Figure 1.3.

1.5.2.1 Serial mode

In serial or cascade mode, examination of each modality is performed before the investigation of the next modality. There can be a reduction in the overall recognition duration, as not all the multiple biometric traits need be captured at the same time. Sometimes a decision can be arrived at, even before acquiring all the traits.

1.5.2.2 Parallel mode

In this mode of operation, processing of the information from multiple modalities is handled concurrently, independently and all at

once. Then the combination of results yields the final classification decision.

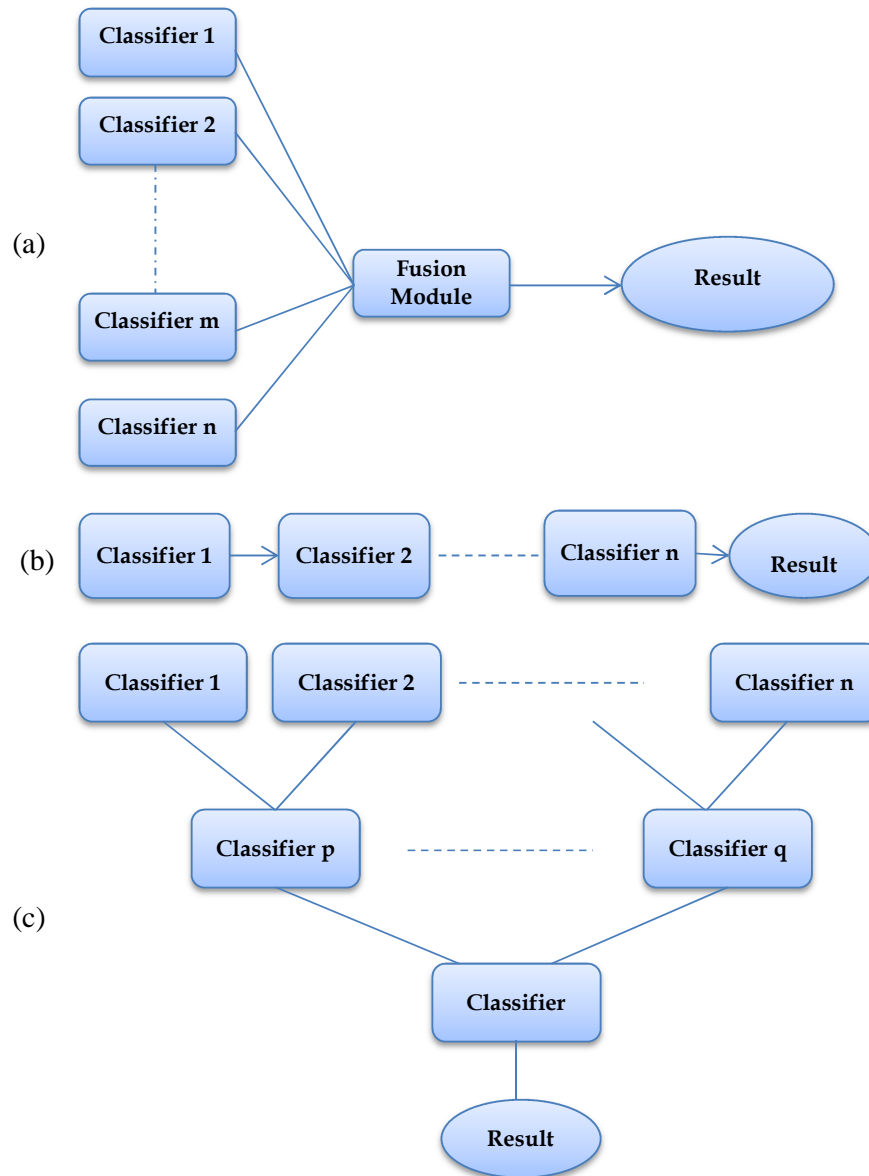


Figure 1.3 Architecture for several classifier combinations
(a) Parallel (b) Serial (c) Hierarchical adapted from [14]

1.5.2.3 Hierarchical mode

In this operational mode, the individual classifiers are combined in a treelike structure. This mode is preferred when a large number of classifiers are expected.

Most of the current multimodal biometric systems operate either in the serial mode or in the parallel mode. The serial mode is computationally efficient, whereas the parallel mode is more accurate [14]. The application requirement determines the choice of the system architecture. A sequential multibiometric system is used in user-friendly and low security applications, whereas a parallel multibiometric system is preferred in security-critical applications.

1.5.3 Design Issues in Multibiometrics

The major factors under consideration in the design of a multibiometric system are the biometric sources, sequence of capturing biometric information and biometric fusion.

The biometric sources include multiple sensors, multiple biometric and processing systems. The factors affecting the selection of a biometric sensor are its ease of use and maintenance, the size of acquisition area, the resolution or the acquisition noise, its reliability and physical robustness, its dynamic range or the time it needs to acquire a sample etc. The factors affecting a processing system are the data format, algorithms for data processing, speed constraints, data compression techniques, which may degrade the sample or template quality etc. The factors proposed by Maltoni *et al.* [15] for the selection of a biometric are:-

- Universality
- Distinctiveness

- Permanence
- Collectability
- Performance
- Circumvention
- Acceptability

Universality means that every person using the system should possess that trait. Distinctiveness means that the traits should be sufficiently different for individuals. Permanence relates to the manner in which the biometric resists aging and other variance over time. Collectability relates to the ease of acquisition for measurement. Performance refers to the accuracy, speed and robustness of the technology used, as well as the operational and environmental factors that affect them. Circumvention relates to the resistance to be fooled or copied and finally acceptability relates to the degree of social and personal acceptability.

The designer needs to decide the sequence (serial or parallel) in which the multiple sources of information are required and processed. The application scenario decides the selection of appropriate acquisition and processing architecture.

Another main design issue is the integration of different biometric sources, referred to as biometric fusion, as it has a good impact on the performance of the system. The classifications of the fusion scheme are sensor level, feature level, score level and decision level. The choice of fusion depends on the type of information from the biometric sources namely, raw biometric samples, feature sets, match score and decision labels.

The trade-off between the additional cost and performance improvement is the major concern in the design of a multibiometric system.

The cost depends on the number of sensors employed, the time for acquisition and processing, the gain in performance by reduction in FAR/FRR, storage and computational requirements as well as perceived convenience to the user [16].

1.5.4 Levels of Fusion

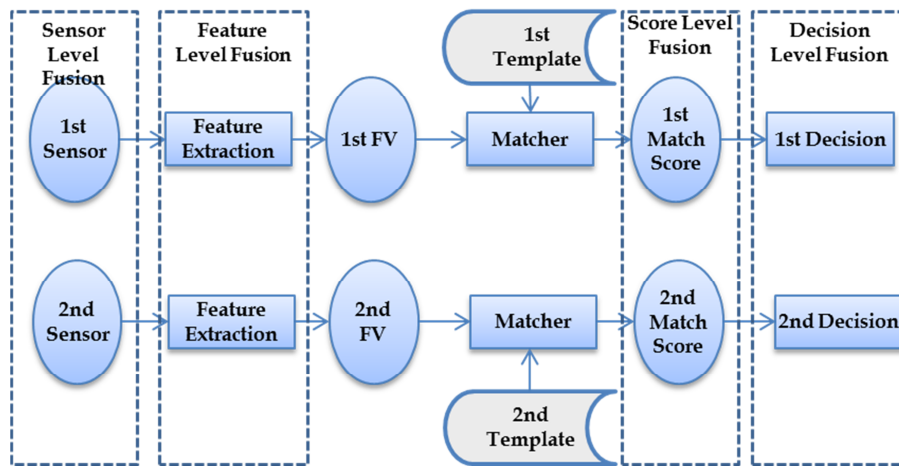


Figure 1.4 Fusion Levels in multimodal biometric fusion

The main classifications of the fusion scheme are sensor level, feature level, score and decision level as demonstrated in Figure 1.4. The type of information to be fused decides the major issues associated with the design of a multibiometric system. Broad classifications based on levels of fusion are fusion prior to matching and fusion after matching.

The sensor module utilizes the raw biometric data and possesses high information content. The information decreases by subsequent processing from the sensor module to the decision module. In the feature level fusion, the information available from different sources may not be compatible. For example, the feature vectors of fingerprint minutia and eigen face coefficients are not compatible for direct fusion. In most of the multibiometric systems, it is easy to access and combine the match score

generated by different biometric matchers. Therefore, match score level fusion offers the best tradeoff in terms of information content and ease of fusion. In the design of a multibiometric system, the type of information that is to be fused is one of the main issues.

1.5.4.1 Fusion Prior to Matching

The biometric information fusion prior to matching is either at the sensor level or at the feature level.

Sensor Level fusion

In the sensor level fusion, the raw data from the sensors are fused [17]. Either the raw data can be the same biometric trait acquired from multiple compatible sensors or multiple instances of the same biometric trait obtained using a single sensor. In sensor level fusion, the multiple data must be compatible and correspondence between points must be estimated reliably.

Feature Level fusion

The different feature sets extracted from multiple biometric sources are combined in feature level fusion. The fusion can be applied to the extraction of different features from the same modality or different modalities to construct a joint feature vector. Appropriate feature normalization, selection and reduction techniques must be applied before combining the extracted features into one single feature vector [18]. The concatenation of the feature vectors extracted from fingerprints and palm print modalities is an example of a feature-level based system.

The feature level fusion is difficult to achieve in some cases. It is difficult to fuse two incompatible feature sets. The concatenation of feature vectors may lead to ‘curse of dimensionality’ problem [19]. This reduces the accuracy of the system. It is not possible to access features sets in the

commercial biometric systems due to proprietary reasons. The success in the use of feature level fusions is limited due to the constraints mentioned above.

1.5.4.2 Fusion After Matching

Fusion after the matching stage can be divided into Fusion at the match score level, fusion at the rank level and fusion at the decision level.

Score Level Fusion

The measure of similarity between the input and database template is the match score. These match scores are integrated in order to achieve the final recognition decision and is carried out at the match score level.

The score level fusion is also referred to as decision, confidence, expert or opinion level fusion. In the score level fusion, it is possible to combine scores obtained from the same biometric trait or different ones using one or more classifiers. The divisions of fusion level are combination and classification. In the former approach, the separate matching scores are gathered to yield one score, which is used to make the final decision. In the latter approach, the input matching scores are considered as input features for a two-class pattern recognition problem, where the subject is classified as legitimate or not. The classifier presents a distance measure or a similarity measure between the input feature vector and the templates previously stored in the database.

Normalization must be carried out prior to score level fusion. The range of match scores may be $[-1, +1]$ for one trait and $[0, 100]$ for another trait. The match scores after minmax normalization are combined using the sum of scores fusion rule. Figure 1.5 shows the flow of information in a match score level fusion scheme.

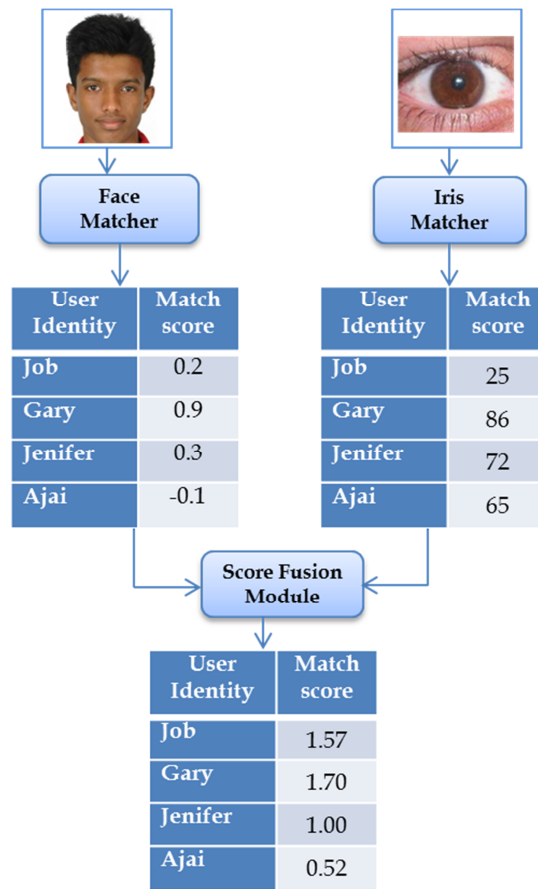


Figure 1.5 Flow of information in a match score level fusion scheme

The match scores generated by individual matchers may not be homogeneous. For example, the output of one matcher may indicate a dissimilarity measure whereas that of another may reveal a similarity measure. In dissimilarity measure, a smaller distance indicates a better match whereas in similarity, a larger distance indicates a better match. This discrepancy should be considered in the match score level fusion.

Rank Level Fusion

In this, a rank is assigned to the top matching identities. Match scores are sorted in a decreasing order, which is the output of rank level

fusion. There are three methods to combine the ranks assigned by different matchers [20]. In the highest rank method, each possible identity is assigned the best (minimum) of all ranks computed by different systems. Ties are broken randomly to arrive at a strict ranking order and the final method utilises the sum of the ranks assigned by the individual systems to a particular identity in order to calculate the fused rank. The logistic regression method is a generalization of the Borda count method where a weighted sum of the individual ranks is used. The weights are determined using logistic regression.

Decision Level Fusion

The commercial biometric matchers provide access only to the final recognition decision. In such cases, the feasible solution is decision level fusion. In the literature, the methods used for this fusion include ‘AND’ and ‘OR’ rules [21], majority voting [22], weighted majority voting [23], Bayesian decision fusion [24] and Dempster Shafer theory of evidence [24].

1.5.5 Challenges in Multibiometrics System Design

The design of a multibiometric system is a difficult task even though it offers several advantages such as large population coverage, greater security and flexibility. The challenging problem in a multibiometric system is that it is difficult to predict the optimal sources of biometrics information and the optimal fusion strategy for a particular application. The reasons for these difficulties are:

- **Heterogeneity of information sources**

The fusion at sensor level and feature level is difficult due to the incompatibility or heterogeneity of information content. For

example, in multibiometric systems it may be impossible to fuse either the raw images or the features extracted from them.

- **Fusion Complexity**

The complexity of the fusion algorithm may nullify the advantages of fusion. For instance, fusion at the sensor or feature levels involves additional processing complexities such as registration and design of new algorithms to match the fused data. Further, the raw data from the sensor and the extracted feature sets are usually corrupted by various types of noise (e.g., background clutter in a face image, spurious minutiae in fingerprint minutiae set etc.). Hence, fusion at the sensor and feature levels may not lead to any performance improvement.

- **Varied discriminative ability**

The unique information provided by each biometric source can be different. Simple fusion rules utilising equal weightage for matchers having higher and lower accuracy may yield lower performance for the combined system. Hence, it is also necessary to assign appropriate weights to the different biometric sources based on their information content.

- **Correlation between sources**

There may be some statistical dependency among different biometric sources. For example, the speech and lip movement of a user are physically related traits. Similarly, multiple samples of the same biometric trait are correlated. In general, it is expected that the fusion

of dependent evidences provides a large improvement in accuracy compared to the fusion of correlated sources.

1.6 *Motivation and phases of the work*

Biometrics is an efficient technology with great possibilities in the area of security system development for official and commercial applications. The biometric systems, employing a single biometric trait, referred to as unimodal biometric systems, face lots of challenges such as noisy data, restricted degree of freedom, intra-user variability, non-universality, spoof attack and unacceptable error rate.

A need has been identified to combine multiple traits to address the challenges presented in the earlier section. In the work carried out, multiple biometric traits are combined to enhance the performance of automatic user authentication systems. The development of the multimodal biometric system that utilizes both behavioural as well as physiological traits viz. speech, face, signature and iris is discussed in detail in this thesis.

Development of a personal authentication system using speech as the behavioral biometric modality utilizing a hybrid speaker recognition system that incorporates both spectral and cepstral information was the first phase of the prototype development. The performance improvement due to the use of spectral features along with the cepstral features has been identified. The spectral features extracted were spectral centroid, spectral range, spectral rolloff and spectral flux while the cepstral features comprise the Mel Frequency Cepstral Coefficients (MFCCs).

Development of a biometric system using face, which is considered as the essential, natural and unique characteristics to identify a particular person, was the second stage. The eigen face approach has been used to

extract features from the face. The Principal Component Analysis (PCA) technique has been employed for the reduction of dimensions of the feature vector. The computational time efficiency has been compared for the system with and without reduced dimensionality. The Euclidean distance between the eigen face coefficients of the template and the detected face has been utilized as the measure of identification.

Development of a system that utilizes signature, since it has been a traditionally accepted longstanding method of identification, was the next phase of the research. The static signature image alone was insufficient for identification of a forged signature. The improvement of the system by using dynamic features of the signature has been established in this work. The static features have been derived using Gabor filters and dynamic features, namely, average stroke velocity and area in X- and Y- global axis have been considered for the system. The reason to select these features is that they are simple to compute with minimum pre-processing effort.

Development of an identification system using iris trait, which is least affected by genetic developmental process was the last phase of the prototype development. The most suitable edge detection technique has been identified for iris texture extraction. The minimum feature based representation has been used for complex iris image and Haar wavelets have been utilized for extracting iris code.

The performance of a bimodal, tri-modal and tetra-modal biometric has been evaluated separately. The performance of the system with trait specific score level fusion technique has been compared with that of a system with equal weight score fusion technique. The effect of user dependencies found at the score level in individual systems has been identified. It is expected that the results of the present work will help a

system designer to choose the appropriate combination of the traits within the scope of this study to achieve the specified level of performance.

1.7 Organization of the Thesis

The thesis is organized into nine chapters. A brief description of each of these chapters is given below:

- The first chapter introduces the discipline of biometrics and its evolution towards multimodal biometrics. The chapter describes the modes of operations, performance measures and limitations of a unimodal biometric system. It also investigates the key issues in multimodal biometric systems along with various architectures for information integration.
- In the second chapter, a review of work carried out in the fields related to the unimodal biometrics such as voice, face, signature and iris available in open literature is discussed. It also summarizes the related works of multimodal biometrics and the common approach of information integration in multimodal systems.
- The third chapter discusses on the methodology for the proposed multimodal biometric system. The logic behind choosing speech, face, signature and iris as biometric traits for this system is described. It also gives an overview of the data acquisition method, development of the knowledge base, the feature component selected for each trait, the fusion strategy and the parameters used for performance evaluation.
- The fourth chapter focuses on the results of authenticating speaker recognition using MFCC and Vector Quantization. The techniques

and procedures involved in extracting the various feature components for speech are highlighted in this chapter. The incorporation of additional spectral features in the feature vector is found to improve the overall performance of the system.

- In the fifth chapter, face recognition using Eigen faces is explained. The chapter also discusses the computational time efficiency of PCA estimated through covariance matrix with and without the use of dimensionality reduction.
- In chapter six, the suitability and effectiveness of combining the Grid based Gabor Wavelet Transform approach and dynamic features in a signature biometric system are demonstrated. The optimum size of the grid is determined based on the analysis of EER and average verification time.
- In chapter seven, segmentation for iris authentication using canny operator is investigated. The effect of the eyelids and eyelashes is minimized by trimming the area of the iris above the upper and below the lower boundaries of the pupil. The Haar wavelet transform is extracted from the iris image and used to improve the recognition accuracy.
- Chapter eight deals with the possibility of fusing the information contained in the voice signal, face image, signature image and iris image for the purpose of personal identification at the match score level. The novel fusion strategy used in this work is also discussed. The results of the performance analysis of a multimodal identification system developed in this work are also presented in this chapter.

- Finally, chapter nine contains the summary of the work, and the directions for future work.

1.8 Summary

The biometrics has recently become a significant part of any efficient person authentication solution. The advantage of using biometric traits is that they cannot be stolen, shared or even forgotten. A background on biometric and multimodal biometric is discussed in this chapter. A majority of the currently used biometric systems referred as unibiometric systems utilise a single biometric feature. They have several limitations, regardless of the significant advances in the field of biometrics. . The limitations of unibiometric systems can be alleviated by integrating various biometrics into one unit. Multimodal biometric systems are those which utilize more than one physiological or behavioural biometric trait for authentication purposes. The biometric data can be fused at various levels, namely, sensor, feature, matching score and decision. The challenges in the biometric systems and research needs in the current scenario are reviewed in this chapter.

CHAPTER 2

REVIEW OF LITERATURE

This chapter is devoted to the review of the research work reported in open literature in the areas of speaker recognition, face recognition, handwritten signature recognition, iris recognition, multimodal biometrics, normalization methods, different fusion techniques, etc. The consolidated results by various researchers on the success rates of the different methods are discussed. This chapter also covers the recent trends in the fusion technology of various biometric traits.

2.1 Introduction

Multimodality forms the core of human-centric interfaces, extending the access to a diverse number of users and in different usage contexts. The security that safeguards proper access to computers, communication networks and private information thus becomes an issue of prime importance in our everyday lives. The classical user authentication relies on tokens and passwords that may be easily lost or forgotten. Use of biometric authentication can overcome this problem as it verifies the user's identity based on his/her physiological or behavioural characteristics such as facial features, voice and fingerprints. User authentication should be transparent to human-computer interaction to maximize usability. In this regard, multimodal human inputs to the computer offer multiple biometric information sources for user authentication.

External conditions and variabilities often affect the performance in biometric verification due to the mismatched conditions between enrolment

and verification sessions, e.g. handset/microphones for recording speech, cameras for capturing facial images and fingerprint readers. In addition, the user's speech may vary according to the ambient noise conditions, speaker's health or speaking styles. The user's facial images may vary due to changes in backgrounds, illuminations, head positions and expressions. While none of the biometrics alone can guarantee absolute reliability, they can reinforce one another when used jointly to maximize the verification performance. This motivates multimodal authentication, where decisions based on individual biometrics are fused.

2.2 Speech

Shaneh *et al.* (2009) [25] designed a system to recognize voice using MFCC and Vector Quantization (VQ) technique. The feature extraction is carried out using MFCC algorithm where the cepstral coefficients are calculated on a mel frequency scale. VQ (vector quantization) method is used to minimize the computation time by reducing the amount of data. In the feature matching stage, Euclidean distance is applied as the similarity criterion. The changes in the shape of the human vocal tract cause variations in resonant frequencies and formants. Utilizing this phenomenon, the voice features of each utterance are extracted. In the training session, the user repeatedly utters the same word five times. The system then recognizes the user after receiving the utterance twice in each testing session.

Zhou *et al.* (2006) [26] developed a speaker identification system based on adaptive discriminative vector quantization. Adaptive Discriminative Vector Quantization technique for speaker identification (ADVQSI) is found to have improved performance in comparison with conventional VQ techniques. The ADVQSI technique takes advantage of

the inter-speaker variations between each individual speaker and all the speakers in the speaker identification group. In the training mode, for each speaker, the speech feature vector space is divided into a number of subspaces on the basis of inter-speaker variations. Then, an optimal discriminative weight is adaptively acquired for each speaker and each subspace. The template differences between different speakers are maximised. In the test mode of ADVQSI, discriminative weighted average VQ distortion is used as the similarity measurement between speakers' templates and each testing waveform. The speaker whose template leads to the highest similarity score will be identified.

Shi, Huang *et al.* (2009) [27] studied the use of MFCCs and Support Vector Machines (SVMs) for text-dependent speaker verification. The MFCCs are extracted from the password spoken by the user. These MFCCs are then normalized and used as the speaker features for training a claimed speaker model via SVM. Experiments were conducted on the Aurora-2 database with various orders of MFCCs. The experimental results indicated the average accuracy rate of the text-dependent speaker verification system based on the 22nd-order MFCC and SVM to be 95.1 percent.

Jabloun *et al.* (1999) [28] developed a speech recognition system based on multirate signal processing and the Teager energy operator. The speech signal is divided into nonuniform subbands in mel-scale using a multirate filterbank. After estimating the Teager energies of the subsignals, computation of the feature vector by log-compression and inverse discrete cosine transform (IDCT) is carried out. The feature parameters derived in this study seems robust even in the presence of car engine noise. A continuous density Hidden Markov Model (HMM) based speech recognition system with five states and three Gaussian mixture densities is used. The recognition performances of the Teager energy operator feature

parameters are evaluated using the TI-20 speech database of TI-46 Speaker Dependent Isolated Word Corpus after incorporating various types of additive noise.

Revathi *et al.* (2009) [29] presented the effectiveness of perceptual features and iterative clustering approach in a speaker recognition system. Clustering models are utilized and the performance of the system is evaluated using isolated digits and continuous speeches from TI digits_1, TI digits_2 and TIMIT database. The statistical analysis results of F-ratio and by using χ^2 distribution are discussed. The accuracies of speaker identification using isolated digit recognition and continuous speech recognition are found to be 91 and 99.5 percent respectively. The performance of the speaker verification system is evaluated and its equal error rate is found to be 9 percent.

Shafran *et al.* (2003) [30] proposed a HMM based classifier to identify the speaker. SVMs with rational kernels are used to identify the content of the speech. Ming *et al.* (2007) [31] investigated the problem of speaker recognition in noisy conditions assuming absence of information about the noise. The multicondition model training and missing-feature theory are combined to model noise with unknown temporal-spectral characteristics. Multicondition training is carried out using simulated noisy data and the missing-feature theory is applied to refine the compensation by ignoring noise variation outside the given training conditions. The algorithm was tested using two databases with simulated and realistic noisy speech data. The first database is a redevelopment of the TIMIT database by rerecording the data in the presence of various types of noise. The second database is a handheld-device database collected in realistic noisy conditions. The new model is compared to baseline systems and is observed to achieve lower error rates.

Togneri *et al.* (2011) [32] presented speaker identification in both noisy as well as noiseless environments. In clean environments, a high recognition rate of 95 percent on 64 speakers is achieved. The Gaussian Mixture Model- Universal Background Model (GMM-UBM) and GMM-SVM systems are found to be robust when confronted with limited training data. It is found that the missing data methods can compensate against arbitrary disturbances and remove environmental mismatches. The significance of combined approaches involving bottom-up estimation and top-down processing is reviewed.

Campbell (1997) [33] designed an automatic speaker recognition system. The information, theoretic shape measure and Line Spectral Pair (LSP) frequency features are used to discriminate between the speakers. The divergence shape is interpreted geometrically as the shape of an information-theoretic measure. The LSP frequencies are found to be effective features in this divergence-shape measure. The high-quality telephone-bandwidth is used to collect the speech of 80 seconds in real world office environments. The accuracy of the speaker-identification system is found to be 98.9 percent.

Cui and Alwan (2005) [34] proposed a feature compensation algorithm based on polynomial regression of utterance. The bias between the clean and noisy speech features is approximated by a set of polynomials. The feature vector is derived using expectation-maximization (EM) algorithm under the maximum likelihood (ML) criterion. In Automatic Speaker Recognition (ASR), Signal to Noise Ratio (SNR) for the speech signal is first estimated and the noisy speech features are then compensated using regression polynomials. The compensated speech features are decoded via acoustic HMMs trained with clean data.

Lee *et al.* (1990) [35] developed a speaker recognition system, SPHINX, which accounts for large vocabulary and continuous speech. It is based on discrete HMMs with Linear Predictive Coding (LPC) derived parameters. It provides speaker independence by incorporating multiple codebooks of fixed-width parameters into HMM.

Chulhee *et al.* (2003) [36] proposed a method to minimize the loss of information during the feature extraction stage. This is achieved by optimizing the parameters of the mel-cepstrum transformation. Typically, the mel-cepstrum is obtained by critical band filters whose characteristics play an important role in converting a speech signal into a sequence of vectors. The performance of the mel-cepstrum is analyzed by changing the parameters of the filters such as shape, center frequency and bandwidth. Experiments with Korean digit words show that the recognition rate is improved by 4 to 7 percent.

Reynolds *et al.* (2000) [37] described the major elements of GMM used for speaker verification system. The system is built around the likelihood ratio test for verification. The Bayesian adaptation technique is used to derive speaker models from the UBM and score normalization is employed to improve the verification performance. The representative performance benchmarks and system behavior experiments on Speaker Recognition Evaluations are presented.

Furui (1981) [38] developed an automatic speaker verification system for telephone speech data. The operation of the system is based on a set of functions of time obtained from acoustic analysis of a fixed and sentence-long utterance. Cepstrum is extracted by means of LPC analysis successively throughout an utterance and the frequency response distortions introduced by transmission systems are removed. The time functions are

expanded by orthogonal polynomial representations and the overall distance between registration and the stored reference is calculated. This is accomplished by a new time warping method using a dynamic programming technique. A decision is made to accept or reject an identity claim, based on the overall distance. The reference functions and the decision thresholds are updated for each customer. Several sets of experimental utterances, which include male and female utterances recorded over a conventional telephone connection, are used for the evaluation of the system. Results of the experiment indicated that a verification error rate of one percent or less could be obtained even if the reference and test utterances are subjected to different transmission conditions.

Markel and Davis (1971) [39] analysed a database consisting of unconstrained extemporaneous speech of 36 hours recorded over a period of more than three months from 17 speakers. The experiments demonstrated that a monotonic increase in the probability of correct identification and a monotonic decrease in the equal error probability for speaker verification could be obtained using long-term average feature vectors. The accuracy of the system is found to be 98 percent when tested with text independent speech data (without linguistic constraints) of 39 seconds.

Soong *et al.* (1985) [40] used VQ codebook for characterizing the short-time spectral features of a speaker. A set of codebooks is used to recognize the identity of an unknown speaker. The minimum distance (distortion) classification rule is used. A series of speaker recognition experiments is performed using a 100 talker (50 male and 50 female) telephonically recorded database consisting of isolated digit utterances. The effects on performance of different system parameters such as codebook

sizes, the number of test digits, phonetic richness of the text and difference in recording sessions are studied. The speaker identification accuracy of the system is found to be 98 percent for ten random but different isolated digits.

Lin *et al.* (1994) [41] explored the use of an array of microphones to capture speech under adverse acoustic conditions in an automatic speaker identification system. The system is evaluated using reverberated speech signals, generated by a computer model of room acoustics and transduced by different simulated microphone arrays. The influence of vector quantization techniques, codebook size and the order of cepstrum coefficients on the performance of the speaker identification system is evaluated. It is found that the 2-D matched-filter microphone arrays are capable of producing high speaker identification scores in a hostile acoustic environment such as multipath distortion and competing noise sources.

Reynolds *et al.* (1995) [42] used GMM for robust text-independent speaker identification. The individual Gaussian components of a GMM are shown to represent some general speaker-dependent spectral shapes that are effective for modeling speaker identity. The proposed model is found to yield high identification rates for short utterances from unconstrained conversational speech. A telephone speech database of 49 speakers is used for testing the system. The algorithmic issues such as initialization, variance limiting and model order selection are examined. The system performance is compared with other speaker modeling techniques such as uni-modal Gaussian, VQ codebook, tied Gaussian mixture and radial basis functions. The Gaussian mixture speaker model is found to provide an identification accuracy of 96.8 percent when clean speech utterances of 5 seconds are used and an accuracy of 80.8 percent for telephone speech utterances of 15 seconds.

Maleh *et al.* (2000) [43] proposed a speech/music classification technique based on the line spectral frequencies (LSFs). Two different classification methods such as quadratic Gaussian and nearest neighbour are used. The feature vector for a one-second window is calculated and the correct classification rate is obtained as 90.7 percent. When LSF features are used in conjunction with Zero Crossing Rate (ZCR), the performance of the system is found to be 94.8 percent.

Meindo and Neto (2003) [44] carried out audio segmentation, classification and clustering. The symmetric Kullback-Liebler, KL2 is used for audio segmentation and it is calculated over Perceptual Linear Prediction (PLP) coefficients extracted from an audio signal. The same features are used for the purpose of speech/non-speech classification. For analysis window of 0.5 seconds, a correct classification rate of around 92.6% is obtained.

Panagiotakis and Tziritas (2004) [45] developed a speech/music discriminator based on RMS and Zero-crossings. A correct classification rate of about 95 percent is obtained in this study.

Zhang and Kuo (1999) [46] proposed a system that classifies audio recordings into basic audio types using simple audio features such as the energy function, average zero crossing rate and spectral peak track. An accuracy rate of more than 90 percent is obtained for this audio classification system.

Sourjya S *et al.*(2014) [47], in the paper demonstrated the significance of stochastic feature compensation methods for robust speaker verification in noisy environment. The performance of these data-driven methods was studied for speaker verification on the NIST-2003-SRE

database, in different simulated noisy environment. The algorithms based on joint GMM modeling of clean and noisy data (e.g., Stereo based Stochastic Mapping (SSM), Trajectory based stochastic Mapping (TRAJMAP)) were found to outperform well-known algorithms like Stereo Piece-Wise Linear Compensation for Environment (SPLICE) and Multivariate Model Based Cepstral Normalisation (MMCN) in terms of EER and minimum Decision Cost Function (minDCF) metrics of speaker verification. The overall best performance was observed in case of the TRAJMAP algorithm, which thereby suggests significance of dynamic feature correlation and robustness of long term utterances towards background noise.

Mohan A *et al.*(2014) [48] presented an experimental study that investigated acoustic modelling configurations for speech recognition in the Indian languages – Hindi and Marathi. The experimental study was performed using data from a small vocabulary agricultural commodities task domain that was collected for configuring spoken dialogue systems. Two acoustic modelling techniques for mono-lingual ASR were compared namely – the conventional Continuous Density HMM (CDHMM) and the Subspace GMM (SGMM) acoustic modelling technique. The SGMM mono-lingual models were seen to outperform their CDHMM counterparts when there is insufficient acoustic training data. The best performing system at 77.77% is obtained when the Hindi language states are weighted at 0.8. At this point, an improvement of 1.57% absolute is seen with respect to the SGMM baseline of 76.2%.

Dufour R *et al.*(2014) [49], proposed an analysis of various acoustic and linguistic features extracted from an automatic speech recognition processing in order to characterize and detect spontaneous speech segments from large audio databases. This classification method, performed at the

segment level, allowed the system to associate a class of spontaneity to each speech segment. It is then extended to the classification process by using a probabilistic contextual tag-sequence model that takes into consideration information of surrounding segments: the classification becomes a global process. This method improved the results: 73.0% precision in the detection of high spontaneous speech segments, with a 73.5% recall measure, and a 66.8% precision and a 69.6% recall on prepared speech segments.

Sahidullah M. *et al.*(2012) [50], proposed a block based MFCC computation schemes for speaker recognition in noisy condition. The feature extraction schemes using non-overlapped and overlap block transformation are analytically formulated. The experimental evaluation is performed on standard databases, and this shows that formant specific block transformations perform better. The EER of 13.8266% and minDCF of 5.9546% are achieved.

2.3 Face

Phillips *et al.* (2005) [51] described the challenge problem and data corpus in respect of face recognition system. The baseline performance and preliminary results on natural statistics of facial imagery are presented in this paper. The Face Recognition Grand Challenge (FRGC) problem consists of experiments, which include performance measurement on still images, 3D imagery, multi-still imagery as well as between 3D and still images. The influence of images taken with and without controlled lighting as well as background is to be studied.

Prakash *et al.* (2008) [52] proposed a human face segmentation using the elliptical structure of the human head. It makes use of the

information present in the edge map of the image. In this approach, the eigen values of the covariance matrix representing the elliptical structure are used. Additional parameters are also used to identify the centre and orientation of the face. The Circular Hough Transform (CHT) is used to evaluate the elliptical parameters. Sparse matrix technique is employed to perform CHT, as it squeezes zero elements, and has only a small number of non-zero elements, thereby having an advantage of less storage space as well as computational time. Neighborhood suppression scheme is used to identify the valid Hough peaks. The accurate positions of the circumference pixels for occluded and distorted ellipses are identified using Bresenham's Raster Scan Algorithm, which uses the geometrical symmetry properties. This method does not require the evaluation of tangents for curvature contours, which are very sensitive to noise. It is found that increasing the training set improves the performance of the system.

Kim *et al.* (2002) [53] proposed Kernel Principal Component Analysis (KPCA), which is a nonlinear extension of conventional PCA technique. The input space is mapped into a feature space via nonlinear mapping and the principal components in the feature space are computed. By adopting a polynomial kernel, the principal components are computed within the space spanned by high-order correlations of input pixels making up a facial image, thereby resulting in a good performance. The error rate reported for the proposed method is 2.5 percent for 20 simulations.

Demirel and Anbarjafari (2008) [54] proposed a high performance pose invariant face recognition system based on the probability distribution functions (PDF) of pixels in different color channels. The PDFs of the equalized and segmented face images are used as statistical feature vectors. The Kullback–Leibler distance (KLD) between the PDF of a given face and that in the database is computed and the minimum is selected for

recognition. Feature vector fusion (FVF) and majority voting (MV) methods are employed to combine the feature vectors obtained from different color channels in HSI (Hue Saturation and Intensity) and YCbCr color spaces. The proposed system is tested on the FERET and the Head Pose face databases. The recognition rate obtained using FVF approach for FERET database is 98.00% compared with 94.60% and 68.80% for MV and PCA based face recognition techniques, respectively.

Spies and Ricketts (2000) [55] developed a face recognition system using Fourier spectra analysis. Recognition is achieved by finding the closest match between feature vectors containing the Fourier coefficients at selected frequencies. In Fourier spectra of facial image analysis, every pixel in the image contributes to each value of its spectrum. The Fourier spectrum is a plot of the energy against spatial frequencies, which is related to the spatial relations of intensities in the image. In the case of face recognition, this translates to distances between areas of particular brightness, such as the overall size of the head, or the distance of the eyes. Higher frequencies describe finer details, which are less useful for identification of a person. The Euclidean distances are calculated between feature vectors with entries that are the Fourier Transform values at specially chosen frequencies. The faces are recognized by finding the minimum Euclidean distance between a newly presented face and all the training faces. It is found that as low as 27 frequencies yield an accuracy rate of 98 percent. Moreover, this small feature vector combined with the efficient Fast Fourier Transform (FFT) makes this system extremely fast.

Temdee *et al.* (1999) [56] presented a frontal view face recognition method by using fractal codes. The fractal codes are determined from the edge pattern of the face region covering eyebrows, eyes and nose. In this recognition system, the obtained fractal codes are fed as inputs to a

backpropagation neural network for identifying an individual. The system performance is tested on the ORL face database. The correct recognition rate of the proposed system is found to be 85 percent in the ORL face database.

Wiskott *et al.* (1997) [57] presented a geometrical local feature based system for face recognition. Elastic Bunch Graph Matching (EBGM) is used in this system. The faces are represented by labeled graphs, based on a Gabor Wavelet Transform (GWT). Image graphs of new faces are extracted by an Elastic Graph Matching process and compared by a simple similarity function. The phase information is used for accurate node positioning and object-adapted graphs are used to handle large rotations in depth. The image graph extraction is based on the bunch graph, which is constructed from a small set of sample image graphs. In contrast to many neural-network systems, no extensive training for new faces or new object classes is required in this system. The system inhibits most of the variance caused by position, size, expression and pose changes by extracting concise face descriptors in the form of image graphs. In these image graphs, some predetermined points on the face (eyes, nose, mouth etc.) are described by sets of wavelet components (jets). The recognition rate of the architecture is found to be 98 percent when tested using a gallery of 250 individuals.

Chung *et al.* (1999) [58] suggested the combined use of PCA and Gabor Filters in face recognition system. Gabor Filters are used to extract facial features from the original image on predefined fiducial points and PCA is used to classify the facial features optimally. The combined system is suggested to overcome the shortcomings of a standalone PCA. When raw images are used as a matrix of PCA, the eigenspace cannot reflect the correlation of facial features well because the original face images have deformation due to in-plane and in-depth rotation, illumination and contrast

variation. Gabor Filters are used to overcome these problems. The experimental results show an improvement in recognition rates of 19 and 11 percent when compared to conventional PCA methods in SAIT dataset and Olivetti dataset respectively.

Joo *et al.* (1999) [59] suggested the use of Radial Basis Function (RBF) Neural Networks for the extraction of discriminant eigenfeatures of a face image. Hybrid learning algorithm is used to decrease the dimension of the search space in the gradient method, which is crucial for the optimization of high dimension problems. The facial features are extracted by both the PCA and LDA methods and the Hybrid learning algorithm is used to train the RBF Neural Networks. The experimental results on the ORL database image of Cambridge University show an error rate of 1.5 percent, which is a tremendous improvement over the best existing result of 3.83 percent.

Pan *et al.* (2005) [60] developed a 3D face recognition system using the facial shape. The Region of Interest (ROI) from an image is automatically extracted and it is flattened using isomorphic mapping to get the relative depth image. The eigen face vectors are computed for the recognition.

Belhumeur *et al.* (1997) [61] demonstrated that the Fisherface error rates are lower than those of the Eigen face technique. The PCA is used to reduce the dimensionality of the input space and the Linear Discriminant Analysis (LDA) is applied to the PCA reduced space in order to perform the real classification. However it has been demonstrated in the recent works of Chen *et al.* (2000) [62] that the discriminant together with the redundant information gets discarded by combining PCA and LDA. Hence, it is suggested to apply the LDA directly on the input space in some cases.

Chen *et al.* (2005) [63] developed a feature vector generation scheme based on multi-class mapping of Fisher score for face recognition. The HMM method is employed to model the classes of face images. A set of Fisher scores is calculated through partial derivative analysis of the parameters estimated in each HMM. These Fisher scores are further combined with some traditional features such as log-likelihood and appearance based features. The feature vectors exploit the strengths of both the local and holistic features of human face. LDA is applied to analyze these feature vectors for face recognition. Experimental results on a public available face database are used to demonstrate the viability of this scheme.

Graf *et al.* (1995) [64] developed a method to locate the facial features and faces in gray scale images. After band pass filtering, morphological operations are applied to enhance the regions with high intensities that have certain shapes (e.g., eyes). The histogram of the processed image typically exhibits a prominent peak. Based on the peak value and its width, adaptive threshold values are selected in order to generate two binarized images. The connected components are identified in both the binarized images to locate the areas of candidate facial features. Combinations of such areas are then evaluated with classifiers to determine whether and where a face is present. The proposed system is tested with head-shoulder images of 40 individuals and with five video sequences where each sequence consists of 100 to 200 frames.

Han *et al.* (1998) [65] developed a morphology-based technique to extract eye-analogue segments for face detection. It is suggested that eyes and eyebrows are the most salient and stable features of human face and thus, useful for detection. The eye-analogue segments are defined as edges on the contours of eyes. The morphological operations such as closing, clipped difference and thresholding are applied to extract pixels at which

the intensity values change significantly. These pixels become the eye-analogue pixels in this approach. The labeling process is performed to generate the eye-analogue segments and these segments are used to guide the search for potential face regions with a geometrical combination of eyes, nose, eyebrows and mouth. The candidate face regions are verified by neural network techniques. The experiments demonstrated that the accuracy rate is 94 percent when tested with a set of 122 images.

Amit *et al.* (1998) [66] presented a method for shape detection and applied it to detect frontal-view faces in still intensity images. Detection is carried out in two stages namely, focusing and intensive classification. Focusing is based on spatial arrangements of edge fragments extracted from a simple edge detector using intensity difference. From a set of 300 training face images, particular spatial arrangements of edges, which are more common in faces than backgrounds, are selected using an inductive method developed by Amit *et al.* (1998) [66]. Meanwhile, the Classification and Regression Trees (CART) algorithm proposed by Breiman *et al.*(1984) [67] is applied to develop a classification tree from the training images. Given a test image, the regions of interest are identified from the spatial arrangements of edge fragments. Each region of interest is then classified as face or background using the learned CART tree. The experimental results on a set of 100 images from the Olivetti data set of Samaria, (1994) [68] indicated a false positive rate of 0.2 percent per 1,000 pixels and a false negative rate of 10 percent.

Augusteijn and Skufca (1993) [69] developed a method that infers the presence of a face through the identification of face-like textures. The textures are computed using second-order statistical features proposed by Haralick *et al.* (1973) [70] for sub images of 16 x16 pixels. Three types of features are considered, namely, skin, hair, and others. A cascade

correlation neural network developed by Fahlman and Lebiere (1990) [71] for supervised classification of textures is used. A Kohonen self-organizing feature map [72] is employed to form clusters for different texture classes. To infer the presence of a face from the texture labels, the occurrences of hair and skin textures are suggested.

Dai and Nakano (1996) [73] applied Space Gray Level Dependence (SGLD) model to face detection, where color information is incorporated with the face-texture model. Using the face-texture model, a scanning scheme for face detection in color scenes is designed and the skin colour regions in the face areas are enhanced. One advantage of this approach is that it can detect faces, which are not upright or have anomalies such as beards and glasses. The detection rate is found to be 98 percent for a test set of 30 images with 60 faces.

Adini *et al.* (1997) [74] suggested that the differences between images of one face under different illumination conditions are greater than the differences between images of different faces under the same illumination conditions. A method is proposed to identify and compensate for the effect of lighting conditions in various face recognition systems.

Zhao and Chellappa (2000) [75] used a generic 3-D surface of a face, together with a varying albedo reflectance model and a Lambertian physical reflectance model to compensate for both the lighting and head orientation. A recognition system based on linear discriminant analysis is developed.

Cutler (1996) [76] showed that infrared images can be successfully applied to decrease the error rate. An artificial infrared light source is used to reduce the effect of external light sources. The proposed system is

suggested to be used in security applications such as site access. However, the use of such a light source is not always practical, particularly if the camera is far from the subject.

Pentland *et al.* (1994) [77] developed an eigen face system by including multiple viewing angles of a person's face. A modular eigenspace system is incorporated to improve the overall performance of the system subjected to varying lighting conditions.

Liu H.D *et al.* (2014) [78] proposed two methods based on Local Histogram Specification (LHS) to preprocess face images under varying lighting conditions. A high-pass filter is applied on a face image to filter the low frequency illumination. Then, local histograms and local histogram statistics are learned from normal lighting images. The LHS is applied on the entire image. By fusing Local Binary Patterns (LBP), Gabor and Monogenic Binary Code – Orientation (MBC-O) features get recognition rates over 76%.

Gaidhane V. H *et al.* (2014) [79] proposed a technique for face recognition based on the polynomial coefficients, covariance matrix and algorithm on common eigen values. In the proposed approach the identification of similarity between human faces is carried out without computing actual eigen values and eigen vectors. In this approach, a companion matrix, which is obtained using the polynomial coefficients, represents the features of images. A symmetric matrix is calculated using the proposed approach and the nullity of such symmetric matrix is used as a similarity measure for classification. The numerical value of nullity is zero for dissimilar images and distinctly large for similar images. The recognition rate of 98.00% is achieved.

Zhou C. *et al.*(2013) [80], proposed a method based on PCA image reconstruction and LDA for face recognition. The inner-classes covariance matrix for feature extraction is used as generating matrix and then eigen vectors from each person is obtained, then the constructed images are obtained. The residual images are computed by subtracting reconstructed images from original face images. Furthermore, the residual images are applied by LDA to obtain the coefficient matrices. The two classifiers, minimum distance classifier and SVM are used. The simulation experiments illustrate on the ORL face database. When the minimum distance classifier is used, the average recognition rate is 97.48%, and while using SVM, the average recognition rate is 97.74%.

2.4 Signature

Guo *et al.*(1997) [81] proposed a model-based segmentation approach for the verification of static (off-line) signature images and the detection of forgeries. The segmentation involves identification of junction points and recovery of strokes consistent with the model. For verification, a questioned signature is segmented based on the edge information. The features of the segments such as width, direction and type (loop, retrace etc.) are obtained. In the matching process, the correspondence between the test image and the model is established. The quantified one-to-one correspondence is used for detecting forgeries. A significant feature of the approach is that a segment-wise correspondence between the model and the questioned signatures is obtained. This enables to examine both the global and local features of the questioned signature.

Justino *et al.* 2001 [82] used a grid segmentation scheme to collect the features of the signature image. HMM is employed for identifying intrapersonal and interpersonal variations of signature models. Both static

and pseudodynamic features are used to characterize the signal. The pixel density in each of the grid cells called the Extended-Shadow-Code (ESC) is obtained as the static feature. A signature skeleton image projected into the grid is used to determine the predominate stroke slant in each cell. The axial slant is recorded as the pseudodynamic feature of the signature. A set of codebooks for each feature is generated using Vector Quantization process, based on the k-means algorithm. The HMM signature model adapted to each writer is generated and it has shown promising results for random forgery identification.

Mizukami *et al.* (1999) [83] proposed a method using an extracted displacement function originally proposed by Fang *et al.* (2003) [84]. The method consists in minimization of a function, defined as the weighted sum of a squared Euclidean distance between two signatures and a penalty term for the smoothness of the displacement function. In order to avoid stopping at local minima, the signatures are transformed into coarse images by Gaussian filtering technique. The performance of the autoassociator neural network based on the constructive cascade correlation architecture (CASCOR) method is compared with a multilayer perceptron network (MLP) with back propagation. The set of 12 features is extracted by the method of moments and PCA. Skeleton image produces six moments and edge images give signature global slopes. From the pressure response, the pressure threshold and pressure factor are extracted. While using PCA, first forty eigen values and eigen vectors are used as the feature vector. The test results indicated that CASCOR performs significantly better than MLP in the signature verification tasks of simple and random forgeries.

Sato and Kogure (1982) [85] proposed the use of Dynamic Programming Matching (DPM) to align the shape of signatures consisting of pen-down strokes. After having normalized the data with respect to

translation, rotation, trend and scale, the signature is aligned. The DPM is also used to compute the alignment of the pressure function and to calculate a measure of the difference in writing motion. The classification of the signature is carried out based on three measurements such as the residual distance between shapes after time alignment, the residual distance between pressure functions and the distance between writing motions.

Parizeau and Plamondon (1990) [86] evaluated the use of DPM for signature verification by aligning either the position, velocity or acceleration in the horizontal or vertical direction. The complete signing trajectories consisting of both pen-down and pen-up strokes are used.

Hastie *et al.* (1991) [87] obtained a statistical model of signatures that allows for variations in the speed of writing as well as affine transformations. DPM is used to find the correspondence between speed signals of pairs of signatures. The distance measure provided by DPM is used as the classification parameter. During training, the signature with the lowest distance to all the others is chosen as the reference and its speed signal is used to perform letter segmentation. All other signatures are also segmented into letters by using velocities and accelerations of the individual strokes. Pen-up strokes are merged with the pen-down strokes in the pre-processing phase.

Nalwa (1997) [88] parameterized the pen-down strokes of the signature using arc length instead of time; a number of characteristic functions such as the coordinates of the center of mass, torque and moments of inertia are computed using a sliding computational window and a moving coordinate frame. A simultaneous DPM over arc-length of all these characteristic functions for the two signatures under comparison is taken as the measure of similarity for classification.

Hairong *et al.* (2005) [89] proposed a novel off-line Chinese signature verification method based on support vector machine. The method uses both static and dynamic features. The static features include moment features and 16-direction distribution whereas the dynamic features include gray distribution as well as strokewidth distribution. The support vector machine is used to classify the signatures. The experiments on real data sets revealed an average error rate of 5 percent, which is obviously satisfactory.

Kalera *et al.* (2004) [90] used a combination of Gradient, Structural and Concavity features to extract the significant features of a signature. The global, statistical and geometrical features of the signature are obtained. The Bayes and k-nearest neighbor classifiers are used in the online signature verification and identification domains respectively. It is claimed that an accuracy of 78.1 percent for verification and 93.18 percent for identification can be obtained for the system.

Munich *et al.* (2003) [91] proposed a new camera based biometric, visual signature identification system. The importance of the parameterization of the signatures to achieve good classification results independent of the variations in the position of the camera with respect to the writing surface is discussed. It showed that affine arc length parameterization is better than Euclidean arc length. The verification error rate of the system is found to be 4 percent on skilled forgeries and 1 percent on random forgeries, better than camera-based biometrics.

Jayasekara *et al.* (2006) [92] proposed a signature recognition method based on the fuzzy logic and genetic algorithm (GA) methodologies. It consists of two phases; the fuzzy inference system training using GA and the signature recognition. The feature extraction

process follows selective pre-processing. The projection profiles, contour profiles, geometric centre, actual dimensions, signature area, local features and the baseline shift are considered as the feature set in this study. The input feature set is divided into five sections and five separate fuzzy subsystems are used. The results are combined using a second stage fuzzy system. The fuzzy membership functions are optimized using the GA. The genuine signatures, random forgeries, skilled forgeries of a particular signature and different signatures are used for training. The optimized recognition system is used to identify the signature. It is found that the signature recognition accuracy rate of the system is about 90 percent for genuine signatures, 77 percent for random forgeries and 70 percent for skilled forgeries.

Vargas *et al.* (2009) [93] represented information about pressure distribution from a static image of a handwritten signature as histogram. The pseudo-cepstral coefficients are calculated from the histogram. Finally, the unique minimum-phase sequence is estimated and used as the feature vector for signature verification. The optimal number of pseudo-coefficients is estimated for best system performance. Experiments are carried out using a database containing signatures from 100 individuals. The robustness of the analyzed system for simple forgeries is tested out with a Least Square Support Vector Machine (LS-SVM) model. The experimental results showed that using pseudo-coefficients, the equal error rate is found to be 6.2 percent.

Mohamadi (2006) [94] presented a Persian static signature identification system using PCA and MLP neural network. In training phase, PCA is used to construct some eigen vectors based on training database images. In test phase, the eigen value of each eigen vector from a new signature image is extracted using PCA. These eigen values are used

as features and are fed to a MLP classifier. For the experiment, 20 classes of Persian signatures comprising 10 signatures for training and 10 signatures for testing are used. Identification rate of the system is reported to be 91.5 percent.

Sigari and Pourshahabi (2006) [95] investigated signature identification and verification using signal-processing approaches. The performances based on Discrete Cosine Transform (DCT), Hough transform, Radon transform and Gabor Wavelet Transform (GWT) are compared. It is reported that the GWT system performs better than the other systems. GWT is used as the feature extractor and Euclidean distance as the classifier in both identification as well as verification. A virtual grid is placed on the image of the signature and some coefficients are computed by GWT on each point of the grid. A Persian signature database is used for experimentation. Identification and equal error rates are reported to be 99.5 percent and 15 percent respectively.

Ozgunduz *et al.* (2005) [96] presented an off-line signature verification and recognition system using global, directional and grid features. SVM is used in order to verify and classify the signatures. For recognition, the signatures are accounted as a multi-class problem of one-against-all SVM. The performance of the system is compared with MLP and the results showed that SVM has better performance than MLP.

Martinez *et al.* (2006) [97] presented an efficient offline human signature recognition system based on SVM and the results are compared with an MLP system. Two approaches are used in both the cases:- In the first approach, each feature vector is constructed using a set of global geometric and moment-based characteristics from each signature while in the second approach, the feature vector is constructed using the bitmap of

the corresponding signature. Signature database containing 228 signatures in 38 classes is used for performance analysis. Results showed that the recognition rate of SVM is 71 percent while that of MLP is only 47 percent.

Coetzer *et al.* (2004) [98] presented an offline signature verification system using Discrete Radon transform as feature extractor and HMM as classifier. A database containing 924 English signatures of 22 writers is used. The experimental results showed that EER is 18 percent for skilled forgery and 4.5 percent for casual forgery.

Fakhlai *et al.* (2011) [99] proposed a new offline Persian signature recognition system based on the contourlet transform (CT). SVM is utilized as a tool to evaluate the performance of the proposed method. In the proposed method, the first signature image is normalized by size and the image is enhanced to remove the noise. After pre-processing, the signature image is divided into four regions and contourlet coefficients are computed on each region. The histogram of orientation and the direction of each region are computed. The histograms are fed to a layer of SVM classifiers as the feature vector. The Persian dataset consisting of 400 genuine images and 200 forgery images is used for testing the performance of the system. The recognition rate is found to be 98 percent.

Yang *et al.* (1995) [100] used HMM to model the sequence of normalized angles along the trajectory of the signature. For the individual HMM, the Baum-Welch algorithm was used for estimating the parameters of the HMM during training and testing. A single HMM having left-to-right skip topology with 6states models each signature. Each individual contributed 16 signatures, eight of them are used for training and the rest eight kept for testing. The model is tested on a database of 496 signatures

gathered from 31 users. The FAR and EER of the system are found to be 6.45 and 1.18 percent respectively.

Shafiei and Rabiee (2003) [101] proposed a system based on segmenting each signature and identifying perceptually important points. A vector consisting of seven features, four dynamic and three static, is computed for each segment. The feature vector is scale and displacement invariant. The resulting vectors are used for training an HMM to achieve signature verification. A database consisting of 622 genuine signatures and 1010 forgery signatures collected from a population of 69 users is used for testing the proposed system. The False Acceptance Rate is found to be 4 percent and False Rejection Rate is 12 percent.

Lee (1996) [102] investigated the use of three neural network approaches for classifying signatures. Bayes Multilayer Perceptrons (BMP), Time Delay Neural Networks (TDNN) and Input Oriented Neural Networks (IONN) are the various approaches considered in this study. The input to the neural networks is a sequence of instantaneous absolute velocities extracted from the spatial coordinates. Consequently, the database used consists of 1000 genuine signatures from only one user and 450 skilled forgeries from 18 trained forgers. The back propagation algorithm is used for network training. In this experiment, BMP is found to provide the lowest misclassification error rate, which is equal to 2.67 percent.

Lejtman *et al.* (2001) [103] applied wavelets and back-propagation neural network together for the on-line signature verification purpose. The feature vector comprises functions such as the pen pressure, x and y velocity and the angle of pen movement. The Daubechies-6 wavelet transform with 16 coefficients is applied to compress the feature vector.

The end coefficients are given as the input to a neural network. The FRR is found to be 0.05 percent and False Acceptance Rate is less than 0.1 percent when a database of 922 genuine and forged signatures gathered from 41 persons is used.

Zanuy (2007) [104] studied the performance of VQ, K-Nearest Neighbor(KNN), Dynamic Time Wrapping (DTW) and HMM. A database of 330 users, which includes 25 skilled forgeries performed by five different impostors, is used. The experimental results showed that the first proposed combination of VQ and DTW outperformed the other combination of DTW and HMM. The minimum detection cost function value is found to be 1.37 percent for random forgeries and 5.42 percent for skilled forgeries.

Nanni and Lumini (2008) [105] proposed an on-line signature verification system using Linear Programming Descriptor (LPD) classifier. The information is extracted as time functions of various dynamic properties of the signatures. The discrete 1-D wavelet transform (WT) is performed on the extracted features and the Discrete Cosine Transform (DCT) is used to reduce the approximation coefficients vector obtained by WT to a feature vector of a given dimension. Test using all the 5000 signatures from the 100 subjects of the SUBCORPUS-100 MCYT Bimodal Biometric Database yielded an Equal Error Rate of 3 to 4 percent in the skilled forgeries and 1 percent in random forgeries.

Roy S. *et al.*(2014) [106], presented a grid based, contour based and area based approach for signature verification. Intersecting points and centroids of two equal half of the signature is being calculated and then those centroids are connected with a straight line and the angles of these intersecting points with respect to the centroids connecting lines are

calculated. In this paper simple forgery and skilled forgery both are considered, simple forgery case produce a low FAR but skilled forgery case produces 11-20% FAR. on the other hand FRR was reduced to 7-19%.

Kumar R. *et al.*(2012) [107], proposed a writer-independent off-line signature verification scheme based on surroundedness features extracted from the binary image of signatures. The feature set based on surroundedness property of a signature represented both shape and texture attributes of the signature. Two classifiers namely, multilayer perceptron and support vector machine are implemented and tested on two publicly available database namely, GPDS300 corpus and CEDAR signature database. The best accuracy of the proposed system may go up to 93.46%.

Vargas J.F *et al.*(2011) [108] , described a method for conducting off-line handwritten signature verification. It works at the global image level and measures the grey level variations in the image using statistical texture features. The co-occurrence matrix and local binary pattern are analysed and used as features. A histogram is also processed to reduce the influence of different writing ink pens used by signers. Genuine samples and random forgeries have been used to train an SVM model and random and skilled forgeries have been used for testing it. The experimental results for skilled forgeries show that using grey level information achieves a reasonable system performance of EER, 16.27%.

Shukla A.K *et al.*(2014) [109], process the scanned hand written signature and it is verified on the following parameters such as Permissible boundary, Hand pressure, Euclidian distance, Center of cylinder generated from minimum spanning tree, Delaunay triangulation of the signature, Angle between base line and center of gravity. The Cost of Delaunay Triangulation gives FAR, 5.25%.

2.5 Iris

Daugman (2004, 1993) [110],[111], applied Gabor wavelets filtering to encode the iris regions and extract the phase information of iris textures to create a 2048 bit (256 bytes) of iris template. Only phase information is used for recognizing irises because amplitude information is not very discriminating, and it depends upon extraneous factors such as imaging contrast, illumination and camera gain. The benefit that arises from the fact that phase bits are set also for a poorly focused image, even if based only on random CCD thermal noise, is that different poorly focused irises never become confused with each other when their phase codes are compared. The Hamming Distance (HD) is used to compare the stored iris template with the claimed iris. Since the search database contains 1 million different iris patterns, it is only necessary for the HD match criterion to adjust downwards from 0.33 to 0.27 in order to maintain still a net false match probability of 10^{-6} for the entire database. The complete execution time of this 2-D focus assessment algorithm, implemented in C using pointer arithmetic, operating on a (480 x640) image, is 15 ms on a 300-MHz RISC processor.

Wildes *et al.* (1997) [112] proposed another iris recognition system that decomposed the distinctive spatial characteristics of the iris into four levels Laplacian pyramid and used a normalized correlation for matching.

Boles and Boashash (1998) [113], detected zero crossings of one-dimensional dyadic wavelet transform with various resolution levels over concentric circles on the iris. Both the position and magnitude information of zero-crossing representations were used to measure the similarity between the recognition and enrolment images. The effect of noise on the classification ability of the algorithm was tested using images corrupted

with varying degrees of white Gaussian noise. It uses only a few selected intermediate resolution levels for matching, thus making it computationally efficient as well as less sensitive to noise and quantization errors.

Ma *et al.* (2004) [114] proposed an iris texture analysis method based on using multi-channel Gabor filtering to capture both global and local details in the iris. Ma *et al.* considered the characteristics of the iris as a sort of transient signals and identified the local sharp variation points as iris features. This method regards the texture of the iris as a kind of transient signals and uses the wavelet transform to process such signals. The local sharp variation points, good indicators of important image structures, are extracted from a set of intensity signals to form discriminating features. The basic idea is that local sharp variation points, denoting the appearance or vanishing of an important image structure, are utilized to represent the characteristics of the iris. The whole procedure of feature extraction includes two steps: The first step is, a set of one-dimensional intensity signals is constructed to effectively characterize the most important information of the original two-dimensional image. In the second step, using a particular class of wavelets, a position sequence of local sharp variation points in such signals is recorded as features. A fast matching scheme based on exclusive OR operation is also presented to compute the similarity between a pair of position sequences. The performance of the algorithm is very high and the EER is only 0.09% for different session comparisons. In particular, if one and only one false match occurs in 1 000 000 trails, the false nonmatch rate is less than 1.60%.

Tan *et al.* (2009) [115] presented an efficient and robust algorithm for noisy iris image segmentation in the context of non-cooperative and less-cooperative iris recognition. The main contributions are summarized as follows. Firstly, a novel region-growing scheme (namely, eight-neighbor

connection based clustering) is proposed to cluster the whole iris image into different parts. The genuine iris region is then extracted with the assistance of several semantic priors, and the non-iris regions (e.g. eyelashes, eyebrow, glass frame, hair etc.) are identified and excluded as well, which greatly reduces the possibility of mis-localizations on non-iris regions. Secondly, an integrodifferential constellation is introduced to accelerate the traditional integrodifferential operator, meanwhile enhancing its global convergence ability for pupillary and limbic boundary localization. Thirdly, a 1-D horizontal rank filter and an eyelid curvature model are adopted to tackle the eyelashes and shape irregularity, respectively, during eyelid localization. Finally, the eyelash and shadow occlusions are detected via a learned prediction model based on intensity statistics between different iris regions. Extensive experiments on the challenging UBIRIS iris image databases resulted in an accuracy rate of 99.9 percent.

Proenca *et al.* (2010) [116] announced the availability of the UBIRIS.v2 database. It is a database that contains a multisession iris image, which singularly contains the data captured in the visible wavelength, at a distance and on-the-move.

Pillai, J.K. *et al.* (2011) [117] in their paper, proposed a unified framework based on random projections and sparse representations, that can simultaneously address the ability to handle unconstrained acquisition, robust and accurate matching as well as privacy enhancement without compromising security. The proposed system quality measure can handle segmentation errors and a wide variety of possible artifacts during iris acquisition. It also demonstrates how the proposed approach can be easily extended to handle alignment variations and recognition from iris videos, resulting in a robust and accurate system. The proposed approach includes enhancements to privacy and security by providing ways to create

cancelable iris templates. Results on public data sets show significant benefits of the proposed approach.

Lim *et al.* (2001) [118] proposed an iris recognition system, which includes a compact representation scheme for iris patterns by the 2-D wavelet transform, a method of initializing weight vectors, and a method of determining winners for recognition in a competitive learning method like Learning Vector Quantization (LVQ). Two methods—Gabor transform and Haar wavelet transform, which are widely used for extracting features, were evaluated. From this evaluation, it is found that Haar wavelet transform has better performance than that of Gabor transform. Lim *et al.* (2001)[118] used 2D Haar wavelet transform to decompose the iris image into four levels and quantized the fourth-level high-frequency information to form an 87-bit code. Secondly, the Haar wavelet transform was used for optimizing the dimension of feature vectors in order to reduce processing time and space. With only 87 bits, an iris pattern could be presented without any negative influence on the system performance. Lastly, the accuracy of a classifier, a competitive learning neural network, was improved by proposing an initialization method of the weight vectors and a new winner selection method designed for iris recognition. The researchers improved the efficiency and accuracy of the proposed system by using a modified competitive learning neural network. With these methods, the recognition performance could be increased to 98.4%.

Donald M *et al.* (2007) [119], described an approach to human iris recognition based on the 1D Discrete Cosine Transform (DCT). The work was motivated by the near-optimal decorrelating properties of the DCT compared to the Karhunen-Loeve transform, and the results achieved indicate the good performance of the approach in which there are no False Accepts/Rejects on the CASIA and Bath data sets used. The method as

implemented also has low complexity, making it superior to the other methods evaluated in terms of both speed and accuracy. It demonstrated the use of novel patch encoding methods in capturing iris texture information, proposed the worst-case (nearest nonmatch) EER as a new practical metric for evaluating systems and investigated better classifier designs for wider interclass separability. EER is predicted to be as low as 2.59×10^{-4} on the available data sets. Statistical analysis has also been carried out to find the best models for matching and nonmatching probability distributions in order to predict the worst-case equal error rates, where no failures occur.

Kang and Park (2007) [120] suggested that conventional iris cameras have small depth-of-field (DOF) areas, so input iris images can easily be blurred, which can lead to lower recognition performance. In the paper, it is proposed that by using a constrained least square restoration filter, where noise is considered, the performance can be greatly enhanced. Experimental results showed that the iris recognition errors when using the proposed restoration method were greatly reduced as compared to those results achieved without restoration or those achieved using previous iris-restoration methods. Because the algorithm estimated the MTF parameters and the weight value of the noise-regularization terms based on camera optics and focus scores, the processing time was greatly reduced. In addition, it was able to reduce the recognition errors by defocusing with the proposed restoration algorithm. The total processing time was 98 ms, which is much faster than conventional iterative image restoration methods.

Zhou *et al.* (2009) [121] proposed a method to upgrade the traditional iris recognition system to work on the non-ideal situation using a video-based image approach. This method will quickly eliminate poor-quality images, evaluate the segmentation accuracy and measure if the iris image has sufficient feature information for recognition. In addition, the

system combines the segmentation and quality measure scores together to predict the system performance. ICE and IUPUI databases have been used to test and validate the proposed method. The research results show that the proposed method can work effectively and objectively. The experimental results show that the evaluation score is independent from the quality score and is an important component to improve the accuracy. The combination of segmentation and quality scores is highly correlated with the recognition accuracy and can be used to improve and predict the performance of iris recognition systems. The proposed method has also been applied to the commercialized system (LG IrisAccess 4000), without intervening in the operation of the existing system. The experimental results using MBGC 2008 NIR-iris still images and MBGC 2008 NIR face videos show that the proposed system can predict the system performance accurately.

Rakvic *et al.* (2009) [122], presented a more direct and parallel processing alternative by using field-programmable gate arrays (FPGAs), offering an opportunity to increase the speed and potentially alter the form factor of the resulting system. In particular, the portions of iris segmentation, template creation and template matching are parallelized on an FPGA-based system, with a demonstrated speedup of 9.6, 324 and 19 times, respectively, when compared to a state-of-the-art CPU-based version. Furthermore, the parallel algorithm on the FPGA also greatly outperforms the calculated theoretical best Intel CPU design.

Chou *et al.* (2010) [123], proposed a non-orthogonal view iris recognition system comprising a new iris imaging module, an iris segmentation module, an iris feature extraction module and a classification module. A dual-charge-coupled device camera was developed to capture four-spectral (red, green, blue and near infrared) iris images which contain useful information for simplifying the iris segmentation task. An intelligent

random sample consensus iris segmentation method is proposed to robustly detect iris boundaries in a four-spectral iris image. In order to match the iris images acquired at different off-axis angles, a circle rectification method is proposed to reduce the off-axis iris distortion. The rectification parameters are estimated using the detected elliptical pupillary boundary. An iris descriptor, which characterizes an iris pattern with multi scale step/ridge edge-type maps, is also proposed. The edge-type maps are extracted with the derivative of Gaussian and the Laplacian of Gaussian filters. The iris pattern classification is accomplished by edge-type matching which can be understood intuitively with the concept of classifier ensembles. Experimental results show that the equal error rate of the approach is only 0.04% when recognizing iris images acquired at different off-axis angles within $\pm 30^\circ$.

Kekre *et al.* (2010) [124] proposed an iris recognition system based on vector quantization. The proposed system does not need any pre-processing and segmentation of the iris. Linde-Buzo-Gray (LBG), Kekre's Proportionate Error Algorithm (KPE) & Kekre's Fast Codebook Generation Algorithm (KFCG) have been tested for the clustering purpose. From the results, it is observed that KFCG requires 99.79% less computations as that of LBG and KPE. Further, the KFCG method gives best performance with an accuracy of 89.10%, outperforming LBG that gives accuracy around 81.25%. The performance of individual methods is evaluated and presented in this paper.

Dong *et al.*(2011) [125], in their paper, presented a novel personalized iris matching strategy based on weight map. The weight map of each iris class is learned based on intra-class iris matches among many registered templates. This weight map is updated and stabilized with increase in the number of training images. It proposes a personalized iris

matching strategy using a class-specific weight map learned from the training images of the same iris class. The weight map can be updated online during the iris recognition procedure when the successfully recognized iris images are regarded as the new training data. The weight map reflects the robustness of an encoding algorithm on different iris regions by assigning an appropriate weight to each feature code for iris matching. Such a weight map trained by sufficient iris templates is convergent and robust against various noises. Extensive and comprehensive experiments demonstrate that the proposed personalized iris matching strategy achieves much better iris recognition performance than uniform strategies, especially for poor quality iris images.

Tan T *et al.*(2012) [126], described a scheme for matching noisy iris images under visible lighting. It consists of image preprocessing, feature extraction and matching, and multi-modal fusion. In image preprocessing, a decision level fusion method is proposed to localize limbic and pupillary boundaries using the original iris images and the corresponding mask images. For feature representation and matching, multiple cues, including ordinal measures, color histogram, texton representation, and semantic information, are adopted for noisy iris image matching. In multimodal fusion, a score level fusion strategy is used to combine the four matching scores into the final dissimilarity measure. The UBIRIS.v2 database training dataset are used for testing.

2.6 Multimodal

A number of multimodal biometric systems proposed in the literature are presented in the following sections. Table 2.1 presents a comparison of these multimodal systems in terms of the design parameters and recognition performance.

Table 2.1 Comparison of multimodal biometric systems

Authors	Traits used	Accuracy	Architecture	Fusion level	Fusion Strategy
Hong and Jain(1998) [127]	Fingerprint, Face	FRR: 4.4% at 0.1% FAR	Serial	Matching score/Rank	Bayes
Jain, Hong, and Kulkarni (1999) [129]	Fingerprint, Voice, Face	FRR: 3% (15%) at 0.1% FAR	Parallel	Matching score	Neyman–Pearson
Jain and Ross (2002) [131]	Fingerprint, Handgeometry, Face	FRR: 4% (18%), at 0.1% FAR	Parallel	Matching score	Weighted sum score
Ross and Jain (2003) [132]	Fingerprint, Hand geometry, Face	FRR: 1% (18%) at 0.1% FAR	Parallel	Matching score	Sum score, Decision trees, Linear discriminant function
Snelick <i>et al.</i> (2003)[133]	Fingerprint, Face	FRR: 5% (18%) at 0.1% FAR	Parallel	Matching score	Sum score, Min score, Max score, Sum rule, Product rule
Snelick <i>et al.</i> (2005)[135]	Fingerprint, Face	FRR: 1% (3.3%)at 0.1% FAR	Parallel	Matching score	Sum score, Min score, Max score, Weighted sum score
Brunelli and Falavigna (1995)[137]	Voice, Face	FRR: 1.5% at 0.5% FAR (FRR: 8% at 4% FAR)	Parallel	Matching score/ Rank	Geometric weighted average / HyperBF
Bigun <i>et al.</i> (1997)[138]	Voice, Face	FRR: 0.5% at < 0.1% FAR (FRR: 3.5% at < 0.1% FAR)	Parallel	Matching score	Model based on Bayesian Theory

Comparison of multimodal biometric systems (Contd..)					
Authors	Traits used	Accuracy	Architecture	Fusion level	Fusion Strategy
Verlinde and Chollet (1999) [140]	Voice, Face	TER: 0.1% (TER: 3.7%)	Parallel	Matching score	k-NN, Decision tree, Logistic regression
Chatzis <i>et al.</i> (1999) [141]	Voice, Face	FRR: 0.68% at 0.39% FAR (FRR: 0.0% at 6.70% FAR)	Parallel	Matching score	Fuzzy k-means, Fuzzy vector quantization, Median radial basis function
Ben-Yacoub <i>et al.</i> (1999)[142]	Voice, Face	TER: 0.6% (TER: 1.48%)	Parallel	Matching score	SVM, Multilayer perceptron, C4.5 decision tree, Fisher's linear discriminant, Bayesian
Frischholz and Diechmann (2000)[148]	Voice, Lip Movement, Face	N/A	Parallel	Matching score/ Abstract	Majority voting, Weighted-sum score
Wang <i>et al.</i> (2003) [150]	Iris, Face	TER: 0.27% (TER: 0.3%)	Parallel	Matching score	Sum score, Weighted sum score, Fisher's linear discriminant, Neural network
Kale <i>et al.</i> (2005) [156]	Gait, Face	ROA: 97% (ROA: 93%) for cascade mode	Cascade /Parallel	Matching score (probabilities)	Sum rule, Product rule

Hong and Jain (1998) [127] proposed an identification system that integrates face and fingerprint modalities. After noting that the face recognition is relatively fast but not very reliable, and fingerprint recognition is reliable but slow (hence not very feasible for database retrieval), the authors have cited that these two modalities can be combined to design a system that can achieve both high performance and acceptable response time. Their face-recognition module was based on the eigen faces method [128] and an elastic-string minutiae-matching algorithm [129] was used for fingerprint recognition. The multimodal system operates by finding the top n identities using the face-recognition system alone and then verifying these identities using the fingerprint-verification subsystem. Hence, this system had a serial architecture. Impostor distributions for fingerprint and face subsystems were estimated and used for selecting at most one of the n possible identities as the genuine identity, hence the system did not always correctly retrieve an identity from the database. In their experiments, Hong and Jain [127] used a database of 1500 images from 150 individuals with 10 fingerprints each.

The face database contained a total of 1132 images of 86 individuals, resampled to size 92×112 . A total of 64 individuals in the fingerprint database were used as the training set, and the remaining 86 users were used as the test set. Virtual subjects were created by assigning an individual from the fingerprint database to an individual from the face database consistently. The face-recognition system retrieved the top five ($n = 5$) matches among the 86 individuals and the fingerprint system provided the final decision. FRR of unimodal face and fingerprint systems as well as the multimodal system were 42.2%, 6.9% and 4.4%, respectively, at the FAR of 0.1%. Figure 2.1 shows the associated ROC curves. These results

indicate that a multimodal system can significantly improve the performance of a face-recognition system.

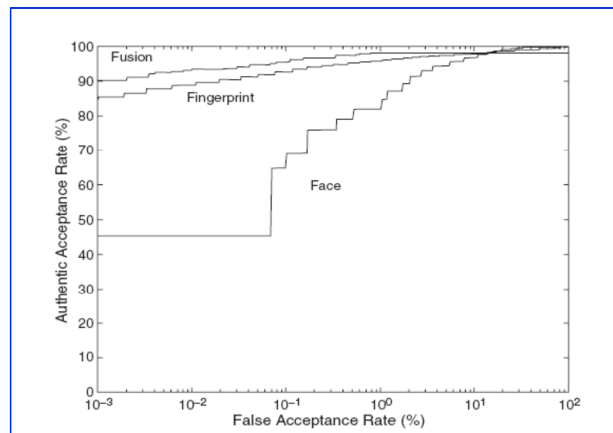


Figure 2.1 ROC curves for unimodal and multimodal systems [127].

Jain *et al.* (1999) [129] combined face, fingerprint and speech modalities at the matching-score level. This specific set of traits was chosen because these traits are frequently used by law enforcement agencies. The parallel fusion scheme submits the matching scores corresponding to these three modalities as inputs to the Neyman–Pearson decision rule to arrive at the verification result. The face-recognition subsystem was based on the eigen faces approach and the fingerprint verification was based on minutiae-based elastic-string matching [130].

Linear-prediction coefficients (LPC) were extracted from the speech signal and modeled using a HMM. The speaker verification was text dependent (four digits, 1, 2, 7 and 9 were used). The training database consisted of 50 individuals, each one providing 10 fingerprint images, 9 face images and 12 speech samples. The test database consisted of 15 fingerprint images, 15 face images and 15 speech samples collected from 25 individuals. The fused system attained nearly 98% genuine-acceptance rate (GAR) at an FAR of 0.1%. This translates to nearly 12% improvement

in GAR over the best individual modality (fingerprint) at 0.1% FAR. The associated ROC curves are shown in Figure 2.2.

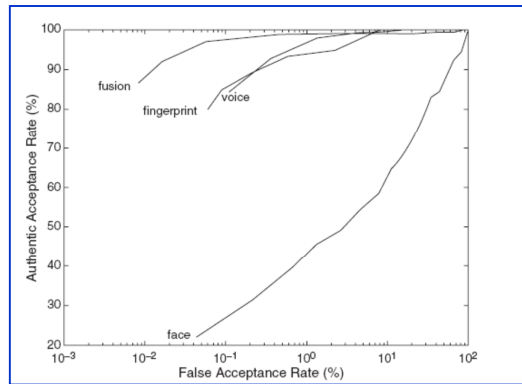


Figure 2.2 ROC curves for unimodal and multimodal systems [130].

Jain and Ross (2002) [131] proposed algorithms for estimating user-specific decision thresholds and weights associated with individual matchers for a face–fingerprint–hand geometry based parallel fusion system. The face module used the eigen face approach. The algorithm, minutiae-based elastic-string matching [129], was used for fingerprint verification. The hand-geometry subsystem [130] used 14 features comprising lengths and widths of fingers and palm widths at several locations of the hand. The user-specific thresholds for each modality were computed with the help of cumulative impostor scores. The weights for individual modalities were found by an exhaustive-search algorithm: all three weights were varied over the range (0, 1) with increments of 0.1, and the best combination resulting in the smallest total error rate (sum of false accept and false reject rates) was selected for each user. The database used in these experiments had 50 users, 40 of them provided 5 samples of each biometric and 10 users provided around 30 samples. One-third of the samples were used in the training phase, while the remaining was used in

the testing phase. At an FAR of 0.1%, user-specific thresholds resulted in nearly 2% GAR improvement; at the same operating point, user-specific weights improved the GAR by nearly 4%. Figure 2.3 shows these performance improvements.

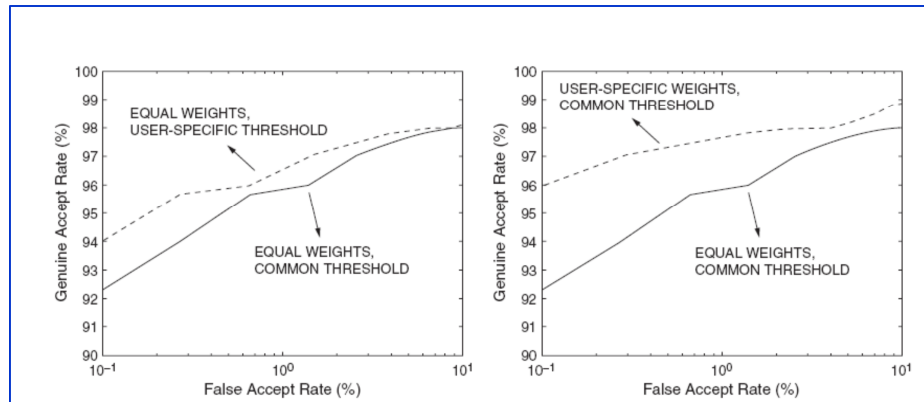


Figure 2.3 ROC curves for user-specific thresholds and user-specific weights in a multimodal system [131]

Ross and Jain (2003) [132] further investigated the effect of different fusion strategies on the multimodal system, proposed in Jain and Ross, 2002. They employed three methods of fusion, namely, sum rule, decision trees and linear discriminant function. The simple sum fusion outperformed the other two methods, resulting in nearly 17% GAR improvement at an FAR of 0.1%. Figure 2.4 shows the associated ROC curves.

Snelick *et al.* (2003) [133] fused the matching scores of commercial face and fingerprint verification systems. They considered the min-max, z-score, MAD (median absolute deviation) and tanh techniques for normalizing the matching scores. In the fusion stage, they investigated the sum-score, min-score, max-score, sum-of probabilities and the product-of-probabilities rules. The database consisted of 1005 individuals, each one

providing 2 face and 2 fingerprint images. The face images were selected from the FERET database [134], but the authors did not provide any information about the characteristics of the fingerprint images. Their results showed that while every normalization method resulted in performance improvement, the min-max normalization outperformed the other methods. Further, the sum score fusion gave the best performance among all the fusion methods considered in this study. At 0.1% FAR value, the min-max normalization followed by the sum-score fusion rule resulted in a GAR improvement of nearly 13% compared to the best performing individual modality (fingerprint) at the same operating point. In addition, the authors reported a considerable decrease in the number of falsely rejected individuals (248 for face, 183 for fingerprint and 28 for multimodal system), indicating that the multimodal systems have the potential to increase user convenience by reducing false rejects, as well as reducing false acceptances.

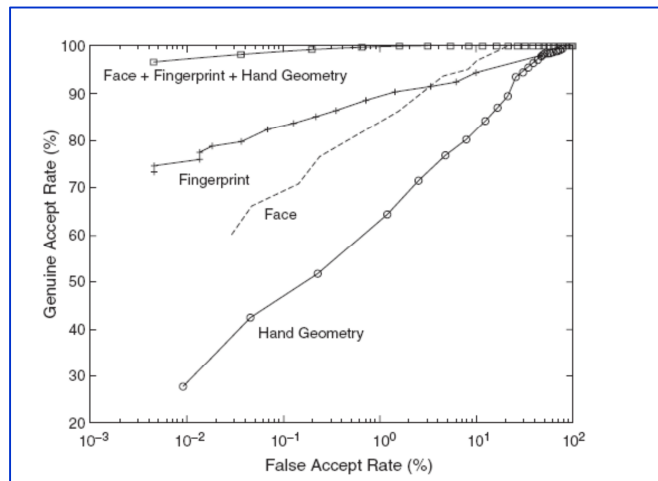


Figure 2.4 ROC curves for unimodal and multimodal systems [132].

Snelick *et al.* (2005) [135] again used commercial face and fingerprint systems from four vendors in a parallel matching-score fusion framework. They experimented with several normalization and fusion

techniques to study their effect on the performance for a database of 972 users. A number of normalization techniques, including min-max, z-score, and tanh schemes were considered along with a novel adaptive normalization technique. The adaptive normalization scheme transforms the min-max normalized scores with the aim of increasing the separation between the genuine and impostor score distributions. The fusion techniques considered in this study included the sum of scores, max score, min score, matcher-weighted sum rule and weighted sum of scores using user-specific weights. The authors used the relative accuracy of individual matchers, indicated by their EER, to determine the matcher weights. Their user-weighting scheme made use of the wolf-lamb concept [136], originally proposed in speaker-recognition community. The set of weights for each user was found by considering the chance of false accepts for the respective (user, matcher) pairs. The results indicated that the min-max and adaptive normalization techniques outperformed the other normalization methods, while the sum score, max score and matcher-weighted sum score outperformed the other fusion methods. The multimodal system had nearly 2.3% GAR improvement at an FAR of 0.1%.

Brunelli and Falavigna (1995) [137] presented a person-identification system combining acoustic and visual (facial) features. A rejection option was also provided in the system using two different methods. A HyperBF network was used as the rank/measurement level integration strategy. The speaker-recognition subsystem was based on vector quantization of the acoustic-parameter space and included an adaptation phase of the codebooks to the test environment. Two classifiers were used for static and dynamic acoustic features. Face identification was achieved by analyzing three facial components: eyes, nose and mouth. The basic template matching technique was applied for face matching. Since the

matching scores obtained from the different classifiers were non-homogeneous, the scores were normalized based on the corresponding distributions. The normalized scores were combined in two different ways: a weighted geometric average and a HyperBF network. The acoustic and visual-cue-based identification achieved 88% and 91% correct recognition rates individually, while the fusion achieved 98% accuracy.

Bigun *et al.* (1997) [138] introduced a new model based on Bayesian theory for combining the matching scores of face and voice-recognition systems. Experiments on the M2VTS database [139], showed that their model resulted in higher accuracy than the sum score rule. For a false-acceptance rate of less than 0.1%, the Bayesian model accepted 99.5% of the genuine users. This was substantially better than the accuracy of the unimodal face and speaker recognition systems that were reported to be 94.4% and 96.5%, respectively.

Verlinde and Chollet (1999) [140] formulated the multimodal verification as a classification problem. The inputs were the matching scores obtained from the individual modalities and the output was a label belonging to the set {reject, accept}. The KNN classifier using vector quantization, the decision-tree-based classifier and the classifier based on a logistic regression model were applied to this classification problem. The modalities were based on profile face image, frontal face image and speech. The experiments were carried out on the multimodal M2VTS database [139] and the total error rate (sum of the false accept and false reject rates) of the multimodal system was found to be 0.1% when the classifier based on a logistic regression model was employed. The total error rates of the individual modalities were 8.9% for profile face, 8.7% for frontal face and 3.7% for speaker verification. Hence, the multimodal system was more accurate than the individual modalities by an order of magnitude.

Chatzis *et al.* (1999) [141] used classical algorithms based on k -means clustering, fuzzy clustering and median radial basis functions (MRBFs) for fusion at the matching score level. Five methods for person authentication that were based on gray-level and shape information of face image as well as voice features were explored. Among the five modalities, four used the face image as the biometric and the remaining one utilized the voice biometric. Table 2.2 shows the algorithms used in this work along with the features used and the rates of genuine as well as false acceptance. Each algorithm provided a matching score and a quality metric that measures the reliability of the matching score. Results from the five algorithms were concatenated to form a 10-dimensional vector. Clustering algorithms were applied on this 10-dimensional feature vector to form two clusters, namely, genuine and impostor. The M2VTS database was used to evaluate the fusion algorithms. Clustering of the results obtained from Morphological Dynamic Link Architecture (MDLA), Gabor Dynamic Link Architecture (GDLA), Profile Shape Matching (PSM) and Speech authentication based on HMM (MSP) algorithms by the k -means method had the best genuine accept rate of 99.32% at an FAR of 0.39%.

Table 2.2 Characteristics of the five modalities [141]

Algorithm	Features	GAR (%)	FAR (%)
Morphologicaldynamic link architecture (MDLA)	Gray-level and Shape	91.9	10.4
Profile shape matching (PSM)	Shape	84.5	4.6
Gray-level matching (GLM)	Gray-level	73.7	1.3
Gabor dynamic link architecture (GDLA)	Gabor features	92.6	3.7
Hidden markov models (MSP)	Speech	100	6.7

Ben-Yacoub *et al.* (1999) [142] considered the fusion of different modalities as a binary classification problem, i.e., accepting or rejecting the identity claim. A number of classification schemes were evaluated for combining the multiple modalities, including SVM with polynomial kernels, SVM with Gaussian kernels, C4.5 decision trees, multilayer perceptron, Fisher linear discriminant and Bayesian classifier. The experiments were conducted on the XM2VTS database [143] consisting of 295 subjects. The database included four recordings of each person obtained at one-month intervals. During each session, two recordings were made: a speech shot and a head rotation shot. The speech shot was composed of the frontal face recording of each subject during the dialogue. The two modalities utilized in the experiments were face image and speech. The face recognition was performed by using elastic graph matching (EGM) [144]. Two different approaches were used for speaker verification. Asphericity measure [145] was used for text-independent speaker verification and HMM were used for text-dependent speaker verification. The total error rate of 0.6% achieved by the Bayesian classifier was significantly lower than the total error rate of 1.48% achieved by the HMM based speaker recognition system, which was the best individual modality in terms of total error rate.

Fierrez-Aguilar *et al.* (2003) [146] developed a multimodal approach including a face verification system based on a global appearance representation scheme, a minutiae-based fingerprint verification system and an online signature verification system based on HMM modeling of temporal functions. The scores were combined by means of SVM classifiers, from which user-independent and user-dependent strategies were applied at the score level. Results indicated that appropriate selection of parameters for the learning-based approach delivered better verification

performance than the rule-based approach. The EERs of the unimodals of face, online signature and fingerprint verification systems were 10%, 4% and 3%, respectively. Results showed that the Sum Rule reduced the EER to 0.5% and the Radial Basis Function (RBF) SVM fusion strategy reduced the EER to 0.3% and 0.05% respectively for the user-independent and user-dependent fusion strategies.

Kumar *et al.* (2003) [147] proposed a multimodal approach for palmprint and hand geometry images. Two schemes of fusion were applied, one at the feature level by concatenating the feature vectors, and the other at the matching score level by max rule. Only the fusion approach at the matching score level outperforms the unimodal systems. The multimodal approach obtained an FAR of 0% and an FRR of 1.41%, while the best unimodal approach in this study, the palmprint-based verification system, obtained an FAR of 4.49% at an FRR of 2.04%.

Frischholz and Diechmann (2000) [148] developed a commercial multimodal identification system utilizing three different modalities: face, voice and lip movement. Unlike other multimodal biometric systems, this system not only included the static features such as face images, but also a dynamic feature, namely, the lip movement. The face was located in an image using an edge-based Hausdorff distance metric. The lip movement was calculated by the optical-flow approach. The synergetic computer [149] was used as the learning classifier for the “optical” biometrics, namely, face and lip movement. Vector quantization was applied for acoustic biometric-based recognition. The input sample was rejected when the difference between the highest and the second highest matching scores was smaller than a given threshold. The sum rule and majority voting served as the two fusion strategies according to the security level of the application. The proposed system was tested on a database of 150 subjects

for three months and the false-acceptance rate was reported to be less than 1%. However, the corresponding genuine-acceptance rate was not reported.

Wang *et al.* (2003) [150] studied the usefulness of combining face and iris biometric traits in identity verification. Iris-recognition systems generally have a relatively high failure-to-enroll rate [151], so using face as an additional biometric trait can reduce the FTE rate of the multimodal system. Further, some of the commercial iris-acquisition equipments can also capture the face image of the user. Therefore, no additional sensor is required for obtaining the face image along with the iris image. The authors used the eigen face approach for face recognition and developed a wavelet-based approach that identifies local variations in the iris images [152]. Fusion was carried out at the matching-score level using strategies like the sum rule, the weighted-sum rule, Fisher's discriminant-analysis classifier and neural network classifier using radial basis functions (RBFNN). Both matcher weighting and user-specific weighting of the modalities were attempted for fusion using the weighted-sum rule. Fusion using learning-based methods like weighted-sum rule, discriminant analysis and RBFNN was found to perform better in terms of their ability to separate the genuine and impostor classes. Since the iris-recognition module was highly accurate (total error rate of 0.3%), the error rate was not reduced significantly after fusion.

Metallinou *et al.* (2008) [153] recognised the emotional information by facial and vocal modalities using Bayesian classifier and SVM. Kumatani *et al.* (2007) [154] presented a method for modelling the state transition between product HMMs for audio visual automatic speech recognition (AV-ASR). Kaur *et al.* (2010) [155] described a bimodal system which combines speech and signature. A study of the product of

likelihoods fusion, sum fusion and z-norm fusion is performed in this paper.

A view-invariant gait-recognition algorithm [156] and a probabilistic algorithm for face recognition [157] were employed by Kale *et al.* (2005) [158] to build an integrated recognition system that captures a video sequence of the person using a single camera. They explored both cascade and parallel architectures. In the cascaded system, the gait-recognition algorithm was used as a filter to prune the database and pass a smaller set of candidates to the face-recognition algorithm. In the parallel architecture, the matching scores of the two algorithms were combined using the sum and product rules. Experiments were conducted on the NIST database consisting of outdoor face and gait data of 30 subjects. No recognition errors were observed when the multimodal biometric system operated in the parallel mode. In the cascade mode, the rank-one accuracy was 97% and the number of face comparisons was reduced to 20% of the subjects in the database.

A recent trend in multimodal biometrics is the combination of 2D and 3D facial informations. Beumier and Acheroy (2000) [159], Wang *et al.* (2002) [160] and Chang *et al.* (2003) [161] have proposed systems that employ fusion of 2D and 3D facial data. Lu and Jain (2005) [162] proposed an integration scheme to combine the surface matching and appearance-based matching for multiview face recognition. All these studies show that the multimodal 2D–3D face recognition can achieve a significantly higher accuracy compared to the current face-recognition systems operating on either 2D or 3D information alone.

Xiuquin (2008) [163] proposed a multimodal biometric system using face and ear at feature level. Kernel discriminant analysis was

employed as the feature extraction method to obtain the features of face and ear independently and then concatenate the two feature vectors to form a single feature vector. Rattani *et al.* (2009) [164] proposed a multimodal biometric system of iris and face in which Scale Invariant Feature Transform (SIFT) features of individual modalities were extracted and concatenated to form the fused feature space.

Ross and Govindarajan (2004) [165] proposed a multimodal biometric system utilising face and hand geometry at feature level. Face was represented using PCA and LDA while 32 distinct features of hand geometry were extracted and then concatenated to form a fused feature. After that, Sequential Feed Forward Selection (SFFS) was employed to select the most valuable features from the fused feature space.

Toh *et al.* (2004) [166] integrated fingerprint, hand geometry and voice biometrics using weighted-sum-rule based match-score-level fusion. They addressed the multimodal decision fusion problem as a two-stage problem: learning and decision. They introduced a reduced multivariate polynomial model to overcome the tedious recursive learning problem in multimodal biometrics in order to achieve good decision accuracy. Four global and local learning as well as decision paradigms were suggested and explored to observe their decision capability. The four learning and decision paradigms were investigated, adopting the reduced polynomial model for biometric decision fusion. Experiments showed that the local learning alone could improve ERRs of about 50%. They have noticed that the local decision can be improved once the threshold settings are appropriately selected for each user.

Veeramachaneni *et al.* (2005) [167] proposed an approach supporting highly secure systems that utilize multisensor fusion to improve

the security level of a system by combining biometric modalities. An algorithm is presented that adaptively derives the optimum Bayesian fusion rule as well as individual sensor operating points in a system. The evolutionary nature of adaptive, multimodal biometric management (AMBM) allows it to react in pseudo real time to changing security needs as well as user needs. Error weights are modified to reflect the security and user needs of the system. The AMBM algorithm selects the fusion rule and sensor operating points to optimize the system performance in terms of accuracy. Cost of false acceptance is a weighting parameter that is used to adaptively control the system's performance in real time. The AMBM algorithm uses the given error costs and searches through the space of all possible rules and the sensor operating points. The sensor operating point is defined by a decision threshold that determines the sensor's FAR and FRR.

Lupu *et al.* (2007) [168], in their paper presented the use of multimodal biometrics in order to identify or to verify a person who wants to start the engine of a car. First of all, a fingerprint sensor was posted on the car's door, one on the steering wheel, a camera for iris recognition was developed on the car's main mirror, and finally a microphone for voice recognition was included. There are two possibilities: if the person is identified as the car owner or a known user, then he/she can take control over the car; if it is an intruder, the car can announce the security service or the police using a complex GPRS system.

Yuan S *et al.*(2013) [169], proposed a multimodal biometric system consisting of face, fingerprints and palmprints, based on the characteristics of multi-dimensional in optical technique. This system combines the optical encryption with multimodal biometric authentication technique, changes one-to-many matching into one-to-one matching, so the matching time is reduced significantly. With the aid of the encryption system and biometric

verification, the verification keys are difficult to be forged and even the keys are lost or stolen, they are useless for other person. In addition, the standard biometric templates are generated real- timely by the verification keys owned by legal user so that they are unnecessary to be stored in a database. The results show that the FAR is 0%, but the FRR is 12%.

Huang Z. *et al.*(2013) [170] developed a robust face and ear based multimodal biometric system using Sparse Representation (SR), which integrates the face and ear at feature level, and can effectively adjust the fusion rule based on reliability difference between the modalities. The Sparse Coding Error Ratio (SCER) is utilized to develop an adaptive feature weighting scheme for dynamically reducing the negative effect of the less reliable modality. A recognition rate of 97.837% is achieved.

Emerich S *et al.*(2013) [171], presented a set of features for a biometric system based on speech and on-line signature. The feature vector is nonhomogeneous and it comprises using TESPAP DZ coefficients, wavelet energy coefficients and also some additional features resulted from the time domain analysis in the case of speech. A feature selection procedure is then applied to reduce the feature vector dimension. A modified symbols alphabet for the TESPAP DZ method is presented. Experimental results were reported using the SVC2004 database for signature and a bimodal database, for on-line signature and speech. A feature level fusion strategy was adapted.

Poh N. *et al.*(2013) [172], investigated a relatively new fusion strategy that is both user-specific and selective. In user-specific, each user in a biometric system has a different set of fusion parameters that have been tuned specifically to a given enrolled user. In selective, only a subset of modalities may be chosen for fusion. The reason for this is that if one

biometric modality is sufficiently good to recognize a user, fusion by multimodal biometrics would not be necessary.

2.7 Summary

In this chapter, a review of work carried out in the fields related to the unimodal biometrics such as voice, face, signature and iris available in open literature is discussed. It also summarizes the related works of multimodal biometrics and the common approach of information integration in multimodal systems.

CHAPTER 3

METHODOLOGY

This chapter addresses the methodology that primarily involves the extraction of feature vectors of different biometric traits and the compilation of the knowledge base, which forms the backbone of the authentication system. In the proposed system, the main goal is to evaluate the performance of the multimodal biometric system based on score level fusion using user-dependent weighted fusion method, over the unimodal biometric system. The different biometric traits from which features are extracted are voice, face, signature and iris since they include both physiological as well as behavioral type. In order to form the feature vector of speech, features such as spectral centroid, spectral flux, spectral rolloff and MFCC coefficients have been extracted while feature vector of face has been formed using their eigen vectors. A combination of static as well as dynamic features acquired from signature and binary code using Haar wavelet generated from iris constitute their feature vector respectively. The four score list is combined using the user specific weighted score level fusion approach to find out a consensus score of the identities and the identity at the top of the consensus score list will be identified as the desired identity.

3.1 Introduction

The methodology of the proposed multimodal biometric system based on user dependent weighted fusion utilizing biometric traits of voice,

face, signature and iris is addressed in this chapter. Since no single biometric can be said to be the best, the need for a multimodal biometric system is one of the leading areas of research in the current era. In a multimodal system identifying and selecting proper biometric traits is one of the main tasks. The appropriate biometric type for a given application depends on many factors including the type of biometric system operation (identification or verification), perceived risks, types of users, and various need for security. A single biometric trait may not meet these requirements needed by different applications as each biometric trait has its own associated advantages and limitations.

In the proposed system, the main goal is to evaluate the performance of the multimodal biometric system based on score level fusion using user-dependent weighted fusion method over the unimodal biometric system and other multimodal systems. The biometric traits under consideration are voice, face, signature and iris as these includes both physiological as well as behavioural type.

The methodology primarily involves the extraction of feature vectors of different biometric traits and the compilation of the knowledge base, the backbone of the authentication system. The enrolment phase involves the acquisition of the biometric traits as well as extraction and compilation of feature vector. Inorder to form the feature vector of speech, features such as spectral centroid, spectral flux, spectral rolloff and MFCC coefficients have been extracted while feature vector of face has been formed using their eigen vectors. A combination of static as well as dynamic features acquired from signature and binary code using haar wavelet generated from iris constitute their feature vector respectively.

The identification phase recognizes speech and face by measuring the Euclidean distance between the query template and the knowledge base template. However for iris, the Hamming distance between the codes generated from the test iris and the iris code in the knowledge base is calculated. For signature identification, Mahalanobis distance between the test signature template and knowledge base template is measured.

Although information fusion in a multimodal system can be performed at various levels, integration at the matching score level is the most common approach due to the ease in accessing and combining the scores generated by different identifiers. We find a consensus score of the identities and the identity at the top of the consensus score list will be identified as the desired identity.

3.2 System Overview

The main steps involved in the development of the prototype are as follows.



- **Acquisition of biometric trait data**
- **Development of knowledge base**
- **Enrolment/training of multimodal biometric system**
- **Identification/testing of multimodal biometric system**
- **Performance analysis**
 - **FAR, FRR & EER**

In the enrolment phase, the knowledge base is prepared from the features extracted from biometric data acquired using dedicated equipment. The test database is used for the performance analysis and the FAR, FRR and EER are determined for various combinations of the traits used in this study.

3.3 Data Acquisition

The speech, face, signature and iris data are acquired using different acquisition device. The Table 3.1 shows the acquisition devices and their specifications.

Table 3.1 Acquisition method and specifications

Biometric Trait	Acquisition method	Specifications
<i>Speech</i>	 <p>Microphone MIC 01A</p>	Sensitivity : 60dB
		Impedence :2K ohms
		Frequency :100Hz-11kHz
		Features :360° Rotatable
		Direction : Omni-directional
		Output Imp : 2K ohms
		Connector : 3.5mm
		Type :CMOS
		Megapixels :15MP
		USB :2.0 port
<i>Face</i>	 <p>Logitech HD Pro Webcam C920</p>	Optical resolution: :15MP
		Focal Length :3.67mm
		Diagonal Field of View : 78 deg
		Frame rate :1080@30fps

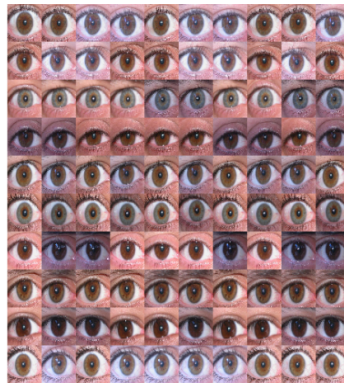
Signature



Wacom Baboo Pad USB Pen Pad without LCD Display

Dimensions - W x D x H	:141.4 x 166.5 x 4.5-15.7mm
Weight (excl. cable)	:145 g
Capturing Active Area : W x H	:109 x 67 mm (16:10)
Capturing Rate.	: 200 points per sec
Resolution	:508 lpi
Coordinate accuracy	: ± 0.5 mm / 0.02 inch
Reading height	:16 mm (0.63")
Data Interface	:USB, also available wireless
Power Source	:Via USB

Iris



UBIRIS.v1

Camera Model	:Nikon E5700
Software	:E5700v1.0
Color Representation	:RGB
Focal Length	:71 mm
Exposure Time	:1/30 sec.
ISO Speed	:ISO-200
Images Width	:2560 pixels
Height	:1704 pixels
Horizontal Resolution	:300 dpi
Vertical Resolution	:300 dpi
Bit Depth	:24
Format	:JPEG

3.4 Unimodal Processing

3.4.1 Speech Processing

In speech, preprocessing is performed by slicing the voice data into different records of fixed length of 30 milliseconds. To remove undesirable undulations, the frames are windowed using a Hamming window. The windowing also smoothen the edges of each data record, reduce spectral distortion, discontinuities or abrupt changes at the end. This windowed voice data are transformed into the frequency domain. The feature vectors are generated using Spectral and Cepstral features of voice data. The spectral features used in this study are Spectral centroid (Brightness), Spectral range (Bandwidth), Spectral roll-off and Spectral flux. MFCCs are taken as Cepstral features. The MFCC are computed by taking the Fourier transform of the windowed signal. The powers of the spectrum obtained above is mapped onto Mel scale using triangular overlapping windows. The logs of the powers at each of the Mel frequencies are taken and the discrete cosine transform of the list of Mel log powers is computed. The MFCCs are the amplitudes of the resulting spectrum. The feature vectors thus obtained are coded using vector quantization. In vector quantization the LBG algorithm is used to generate codebooks for each speaker. The decision is performed from the Euclidean distance, which is computed between the feature components of the unknown target and that of the various templates in the knowledge base.

3.4.2 Face Processing

For improved face recognition performance, the acquired face image is preprocessed starting with the image size normalization so that acquired image size is modified to a default image size. As the next step of preprocessing, histogram equalization is performed on dark or bright

images so that face image quality is enhanced and some important facial features become more apparent. While median filtering is used for reducing the noises in the images. The eigen face approach is used for face recognition. In the Eigen face approach, PCA is used to derive the low dimensional representation of faces by applying it to a representative dataset of images representing faces. The system is implemented by projecting face images onto a feature space that spans the significant variations among known face images. These significant features are called Eigen faces. After enrolling the face images, the recognition phase is performed by projecting the test face image into the Eigen space. The Euclidean distance is measured between the unknown face image and the mean of all the known face images in the Eigen space.

3.4.3 Signature Processing

Following the preprocessing steps of binarization, dilation, thinning rotation invariance and scaling invariance, static features are extracted from the acquired signature image. Binarization converts an image of up to 256 gray levels to a black and white image while dilation fills the gaps and broken necks and the pixels from the outside edges are removed by thinning. The rotation invariance aligns the main inertia axis of the signature with the horizontal axis and the scaling invariance, normalizes the original signature in size to preserve the aspect ratio of the signature. In signature recognition the feature vector is the combination of static and dynamic features. While the static features are generated using 2D Gabor filter, the dynamic features is the x- and y- stroke average velocity. The feature vector of the signature image is stored as the template in the knowledge base. In the identification phase, the matching score of the test template and the training templates are derived. Mahalanobis distance is used for calculating the matching score of signature image.

3.4.4 Iris Processing

In iris recognition system, the pre-processing stage requires the localization of the iris which generally involves the detection of the edge of the iris as well as that of the pupil. Since varying levels of illumination can result in dimensional inconsistencies between eye images due to the stretching of the iris, normalization needs to be performed so that iris region is transformed to have fixed dimensions. After unwrapping the normalized iris region into a rectangular region, it is encoded using Haar wavelets to generate the iris code. In the recognition stage, Hamming distance is used for comparison of the iris code, the most discriminating feature of the iris pattern, with the existing iris templates.

3.4.5 Multimodal Fusion

The match scores at the output of the individual trait may not be homogeneous and need not be on the same numerical range. Due to these reasons, it is necessary to normalize the scores of different traits before combining the scores. The min-max normalization technique is employed in this work to normalize the matching score. The fusion technique employed in this work is on the basis of the different weights assigned to each biometric trait. These different weights are computed based on the Equal Error Rate. The fused similarity score is obtained using user-dependent fusion method. This score is compared to a decision threshold in order to accept or reject the identity claim. In the decision module, an individual is recognized by searching the templates of all the users in the database for a match. One-to-many comparison is carried out to accept the identity of an individual or reject if the individual is not enrolled in the system database. The Figure 3.1 shows overall block diagram of the multimodal biometric system based on user dependent fusion.

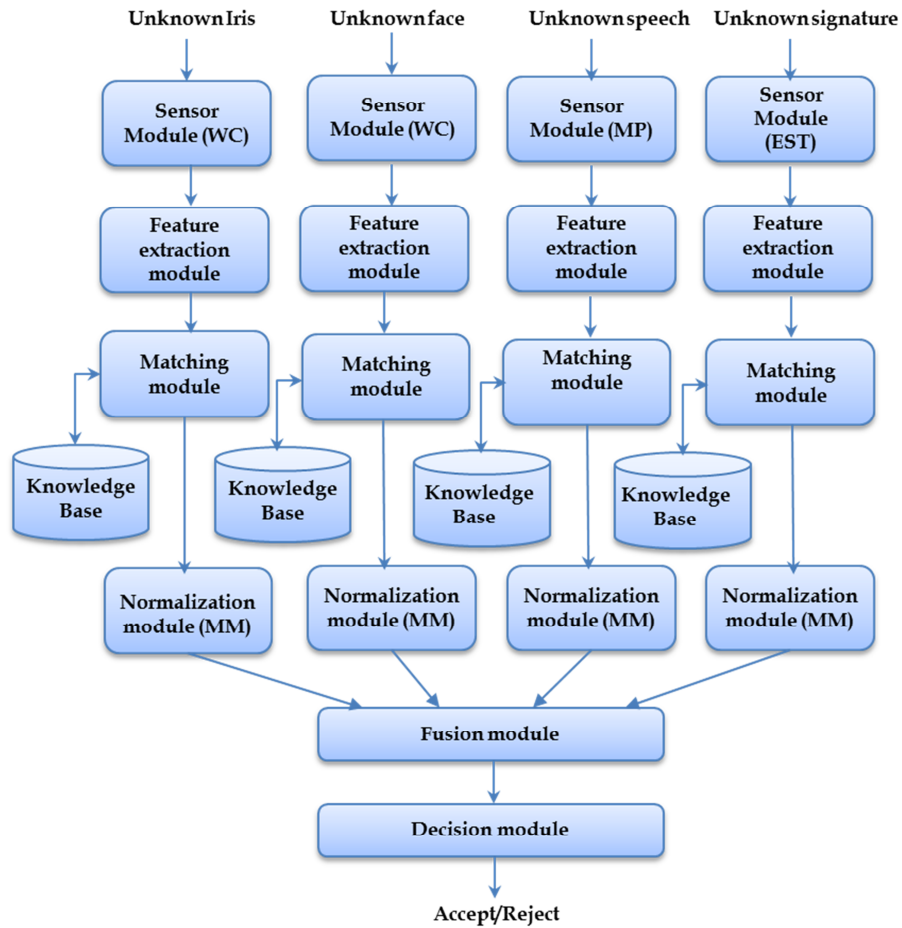


Figure 3.1 Block diagram of the proposed multimodal biometric system.

3.4.6 Development of knowledge base

Training data: 4000 (=100x4x10) biometric data, ten samples each of four traits, namely, speech, face, signature and iris were collected from 100 users for training. After pre-processing, the collected data, features were extracted and the feature vector was stored in the knowledge base.

Testing data: 2000 (=100x4x5) traits each of registered and unregistered users are used for testing. In the case of registered candidates, two samples acquired newly of each trait and three samples from the knowledge base are used. The five samples acquired afresh from the unregistered users formed part of the testing data. The equal number of traits of registered and unregistered users minimizes the bias in the performance analysis. The newly acquired data for the registered users used in the testing simulates the possibility of accounting for the variation in the ambient conditions.

3.5 Performance Analysis

The performance of the proposed biometric system is analysed. The type of decision made by a biometric system for an individual is either genuine or impostor [4] and for each type of decision of genuine and impostor case, there are two possible outcomes, true or false. Therefore, all together, there are four possible outcomes, namely, genuine acceptance (genuine individual is accepted), false rejection (genuine individual is rejected), genuine rejection (impostor is rejected) and false acceptance (impostor is accepted). The genuine distribution and the impostor distribution ascertains the confidence associated with different decisions and can be established by minimizing two error rates, namely, False Accept Rate (FAR) and False Reject Rate (FRR).

By definition, FAR, is “the probability of an impostor being accepted as a genuine individual” [4]. That is, in a biometric authentication system, the FAR is the rate of number of impostor falsely accepted over the total number of enrolled people for a predefined threshold.

FRR, is defined as “the probability of a genuine individual being rejected as an impostor” [4]. That is, in a biometric authentication system,

the FRR is the rate of number of genuine people falsely rejected over the total number of enrolled people for a predefined threshold.

FAR and FRR can be changed by a significant amount depending on the threshold used in the system. On usage of a lower threshold in a similarity based biometric matching system, FAR will be higher and the FRR will be lower while a higher threshold results in lower FAR and higher FRR. The Genuine Acceptance Rate (GAR) is another approach used to measure the accuracy of a biometric system. The rate of number of people accepted in the case of genuine users and rejected in the case of impostors over the total number of enrolled people for a predefined threshold defines GAR. Mathematically, genuine acceptance can be determined by subtracting the number of falsely rejected people from the total number of genuine people. Yet another evaluation criterion is the Equal Error Rate (EER), which refers to that point in a ROC curve where the FAR equals to the FRR and a lower EER value thus indicates better performance.

3.6 Summary

In this chapter, the methodology for the proposed multimodal biometric system as well as the rationale for choosing speech, face, signature and iris as biometric traits has been covered. The data acquisition methods of each traits as well as, the processing of each biometric traits is explained briefly. Identification of the various components of the feature vector as well as combination of outcomes of unimodal matchers were through user specific fusion methods is discussed. Development of knowledge base of the biometric traits and the parameters used for the performance analysis is discussed. The details of the development of uni-, bi-, tri- and tetra-modal biometric system is described in the next chapters.

CHAPTER 4

SPEAKER RECOGNITION

This chapter highlights the technology involved in extracting the feature components of speech required for generating the identification clues. During the speaker enrolment phase, speech samples that contain the discriminating features are collected from the speakers and feature vectors are generated to form the knowledge base of the model. In the recognition phase, the feature vectors extracted from the unknown person's utterance are compared with the knowledge base of the model to find the similarity score, for the purpose of decision making. Since accuracy of identification is highly dependent on the type and number of features used, feature selection is of great significance. Features are computed from the spectrogram on a frame-by-frame basis and relates directly to some perceptual characteristics of sound, such as loudness, pitch etc. This chapter also touches upon a more systematic approach for computing the cepstral coefficients achieved by estimating the Mel Frequency Cepstral Coefficients (MFCC). The performance analysis of the system was carried out and it was found that along with MFCC, the incorporation of additional spectral features in the feature vector improved the overall performance of the system.

4.1 Introduction

Speaker recognition or voice recognition is the task of recognizing people from their voices. Fundamental technology behind a speaker

recognition biometric system is that information extracted from the speech signal is unique leading to signal characterisation. This information can be represented using a number of different feature sets. Biometric system extracts features from speech, model them and use them to recognize the person from his/her voice.

Speaker recognition can be achieved through the acoustic features of speech that differs among individuals. These acoustic patterns reflect both anatomy (e.g., size and shape of the throat, mouth etc.) and learned behavioural patterns (e.g., voice pitch, speaking style etc.) [173]. Since both anatomy as well as the behavioural pattern differs between individuals, these acoustic patterns can be analysed for speaker recognition and its classification.

The various technologies used to process and store voiceprints include many complex technologies like Frequency Estimation, Hidden Markov Models, Pattern Matching Algorithms, Neural Networks, Matrix Representation and Decision Trees [174]. Speaker verification purposes also use anti-speaker techniques such as cohort models and word models.

Biometrics security system based on voice data is an emerging reliable method and is being highly popularized. Voice is a very intuitive behavioural and ubiquitous biometric which can be captured by modern personal computers. Further, it requires no expensive special hardware other than a microphone.

The two main phases of speaker identification are the training phase or enrolment phase and the testing phase or identification phase. During the speaker enrolment phase, the model is trained with feature vectors generated from speech samples collected from the speakers as they are

containing the discriminating features. Recognition phase is comparison of the feature vectors extracted from the unknown person's utterance with the model in the system database to find the similarity score, for decision-making. Feature selection is of great importance in speech recognition, as accuracy is highly dependent on the type and number of features used. Features generated from the spectrogram of the speech signal on a frame-by-frame basis relates directly to some perceptual characteristics of sound, such as loudness, pitch etc. [33]. The block diagram of a speaker identification system is shown in Figure 4.1.

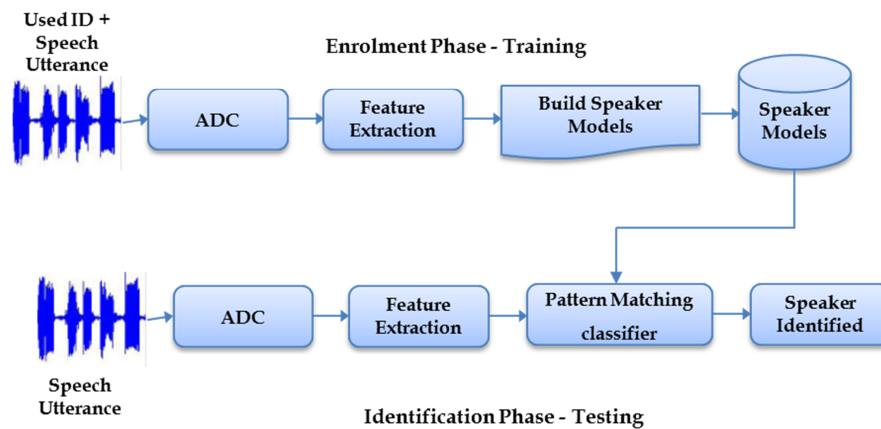


Figure 4.1 Block Diagram of a Speaker Identification System

A speaker recognition system can be categorised into text-dependent or text-independent. In the text-dependent, a known utterance is presented to the recognizer, while in text-independent case, the text being spoken is unknown. The identification of a speaker using text independent utterance is more complex than that utilising text dependent utterance. In this work, text independent utterance has been utilised to identify a speaker.

4.2 Dimensions of Difficulty

The sources of error lead to a broad classification of dimensions of difficulty in a speaker recognition system. The errors in a system can be associated either with the speaker or with the technical component of the system. The typical dimensions of difficulty in speaker recognition are Intra-Individual Variations, Voice Disguise and Imitations and Technical Error Sources

- Intra-Individual Variations

Variations in a speaker's voice can happen not only due to aging and other physiological changes but also by his/her physical as well as mental states [175]. A phonetically balanced training data containing instances of all the sounds of the language in different contexts is one of the major challenges in the consistent performance of a speaker recognition system.

- Voice Disguise and Imitations

A deliberate changing of one's voice to hide his/her identity referred to as Voice disguise and mapping of one's voice to sound like another speaker referred to as imitation or mimicry results in degradation of the performance of a speaker recognition system [176].

- Technical Error Sources

The environmental noise can be added up with speech wave when recorded with a microphone or telephone handset and reverberation can add a delayed version of the original signal to the recorded signal [177]. Poor quality microphones introduce nonlinear distortions in the true speech spectrum [178] and the A/D converter may introduce additional distortions

in the original signal. The coding of speech degrades the performance of speaker recognition significantly [179],[180]. The sources of technical errors are mismatch in the environmental acoustics, amount of background noise, microphone type and recording quality [181],[182]. Technical errors arise when the user speaks training utterances in a clean environment while using the recognition system in a noisy environment.

4.3 Speaker Recognition Techniques

Speaker recognition techniques, as shown in Figure 4.2, are composed of the following modules.

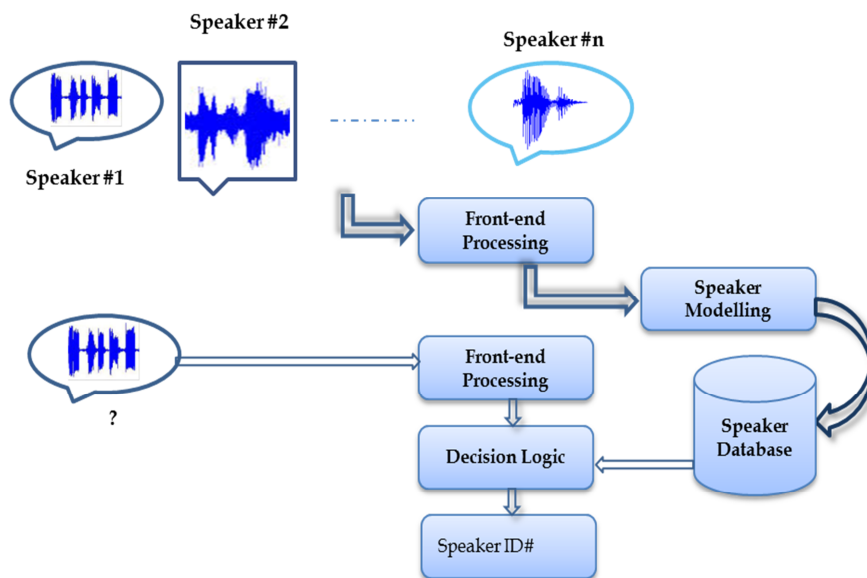


Figure 4.2 Speaker Recognition Technique

4.3.1 Front- End Processing

In front end processing, the spectral features Spectral Centroid, Spectral Range, Spectral Rolloff, Spectral Flux and cepstral features

evaluated as MFCCs from the signal spectrum forms the feature set. The spectrogram on a frame-by-frame analysis gives the spectral features.

4.3.1.1 Spectral Centroid

The spectral centroid, which may also be referred to as the spectral brightness, gives an indication of the spectral shape and is defined as the amplitude-weighted average or centroid of the spectrum [183],[184]. It is a simple, yet efficient parameter, estimated by summing together the product of each frequency component of the spectrum and its magnitude, which is further normalized by dividing with the sum of all the magnitudes. Thus the spectral centroid SC is given by

$$SC = \frac{\sum_{k=0}^{N/2-1} f_k S_k}{\sum_{k=0}^{N/2-1} S_k} \quad (4.1)$$

where S_k is the magnitude spectrum of the k^{th} frequency component f_k and N is the record size.

4.3.1.2 Spectral Range

The spectral range or bandwidth refers to the range of frequencies that are present in the signal. It is computed using the spectral magnitude weighted average of the difference between each frequency component and the centroid, SC . Thus the spectral range, SR is expressed as

$$SR = \frac{\sum_{k=0}^{N/2-1} S_k |f_k - SC|}{\sum_{k=0}^{N/2-1} S_k} \quad (4.2)$$

4.3.1.3 Spectral Rolloff

Another spectral feature, which gives a measure of the spectral shape, is the spectral rolloff and is defined as the frequency below which 85% of the magnitude distribution of the signal is concentrated [183].

i.e. $RO = \text{Minimum}(R)$, such that

$$\sum_{k=0}^R S_k \geq 0.85 \sum_{k=0}^{N-1} S_k \quad (4.3)$$

4.3.1.4 Spectral Flux

This is a measure of the amount of local spectral change. This is defined as the squared difference between the normalized magnitude spectra of successive frames.

$$\text{Flux} = \sum (norm_f[i] - norm_{f-1}[i])^2 \quad (4.4)$$

where $norm_f$ is the magnitude spectrum of the current frame, scaled to the range 0 to 1 and $norm_{f-1}$ is the normalised magnitude spectrum of the previous frame. Spectral flux is a measure of how quickly the power spectrum of the signal is changing and is computed by comparing the power spectrum of one frame with that of the previous frame.

4.3.2 Cepstral Features

The extracted features using spectral estimation of the speech signal alone cannot always perform reliable identification especially, in the presence of external noise and varying environmental parameters. To make the identification process more robust and reliable, additional feature components are incorporated by exploiting the other unexplored features of the noise sources. A variety of signal processing applications use the collection of nonlinear techniques known as cepstral analysis, which is capable of yielding potential features that, can aid in the process of

identification. One of the important properties of the cepstrum is that it is a homomorphic transformation, in which the output is a superposition of the input signals.

The spectrum of a waveform consists of two components, the slowly varying part, referred to as the filter or spectral envelope and the rapidly varying part, referred to as the source or harmonic structure. Separation of these two components can be achieved by taking the cepstrum, an anagram of the word spectrum. The cepstrum is defined as the inverse Fourier transform of the log magnitude Fourier spectrum of the signal and is said to be in the quefrequency domain, an anagram of frequency [185]. The cepstral values are stored as discrete components referred to as the cepstral coefficients, where the n^{th} cepstral coefficient is the amplitude of the n^{th} component along the quefrequency axis.

4.3.2.1 Mel Frequency Cepstral Coefficients

A more systematic approach for extracting the cepstral features makes use of the estimation of Mel Frequency Cepstral Coefficients (MFCC), which is a measure of the perceived harmonic structure of sound [186], [187]. A Mel is a psychoacoustic unit of frequency which relates to the human perception and is approximated using the expression

$$m = 2595 \log_{10} \left[1 + \frac{f}{700} \right] \quad (4.5)$$

where f is the frequency in Hz. The spectrum can be transformed into a spectrum emphasized at Mel intervals using Mel filter banks, which is a row of triangular filters overlapping at Mel-spaced intervals [188]. The cepstrum of this transformed spectrum yields MFCCs.

The term MFCC was coined by Davis and Mermelstein in 1980 [187] when they combined nonuniformly spaced filters with discrete cosine transform (DCT) [188] as a front-end algorithm for Automatic Speech Recognition System.

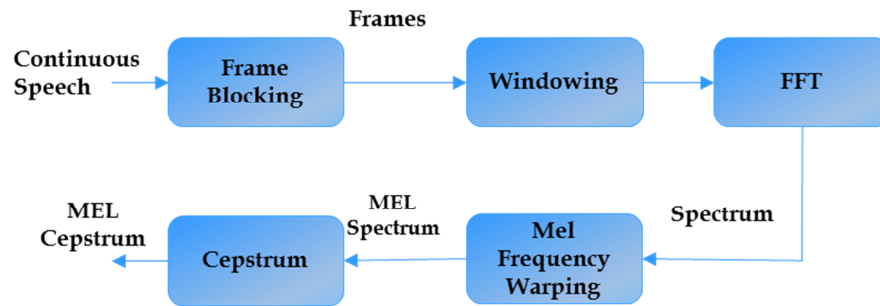


Figure 4.3 Speech process models

The spectral decomposition approach, as shown in Figure 4.3, is used to improve the accuracy of operation.

4.3.2.2 MFCC Estimation - Mel frequency scaling and Cepstrum

The cochlea of the inner ear acts as a spectrum analyser and hence researchers have undertaken psychoacoustic experimental work to derive frequency scales that attempt to model the natural response of the human perceptual system. The perceptual attributes of sound signal at different frequencies are not entirely simple or linear in nature.

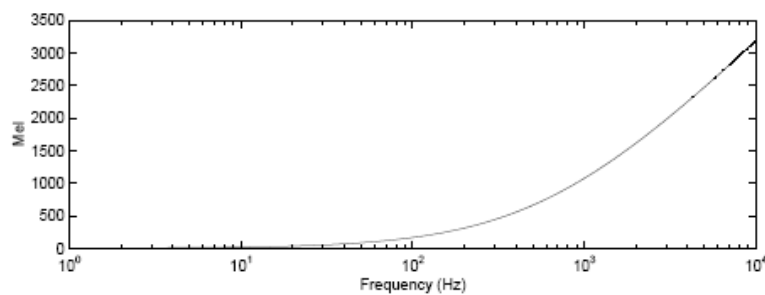


Figure 4.4 The Mel-scale

The cochlea in the auditory system acts as if it is made up of overlapping filters having bandwidths equal to the critical bandwidth. The frequency scaling is used to map the linear frequencies into human perception. Mel frequency scale is a perceptually motivated scale, which is linear below 1 kHz and logarithmic at higher frequencies as shown in Figure 4.4.

The Mel scale more closely models the sensitivity of the human ear than a purely linear scale and provides for greater discriminatory capability between audio segments. The Mel-scale frequency analysis has been widely used in current speech recognition systems.

The Discrete Fourier Transform (DFT) of an input signal is given by Eq.(4.6).

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp(-j2\pi nk / N) \quad k = 0, 1, 2, \dots, N - 1 \quad (4.6)$$

where $x(n)$ is the input signal. The Mel-frequency filter bank [188], [189] comprising of p filters can be represented as shown in Figure 4.5.

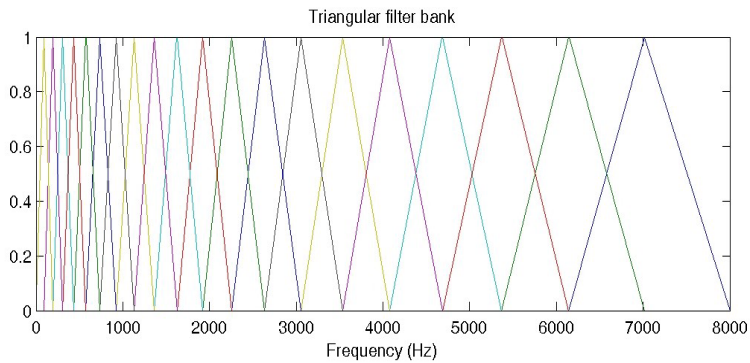


Figure 4.5 Mel-Frequency filter bank on a linear frequency (Hz) scale

The energy in each band is given by m_j ($j=1, 2, \dots, p$), and is computed as detailed in Eq.(4.7). The Fourier Transform magnitude coefficient is multiplied by the corresponding filter gain and the results are accumulated.

$$m_j = \sum_{k=0}^{N-1} |X(k)|^2 H_j(k) \quad 0 \leq j \leq p \quad (4.7)$$

where $H_j(k)$ is the transfer function of j^{th} filter and $X(k)$ is given by Eq.(4.6). The Mel-frequency cepstrum is the discrete cosine transform [188] of the p filter outputs and is represented as in Eq. (4.8).

$$c_i = \sqrt{\frac{2}{N}} \sum_{j=1}^p m_j \cos\left(\frac{\pi i j}{p}\right) \quad (4.8)$$

where c_i is the i^{th} MFCC coefficient.

MFCCs are extracted from the noise data waveforms. The key difference between MFCCs and cepstral coefficients lies in the process involved in extracting the characteristics of a noise signal [190]. With a sampling frequency of 11025 Hz, a filter bank of 40 equal area filters, which cover the frequency range [50, 6400] Hz is implemented. The centre frequencies of the first 13 of them are linearly spaced in the range [50, 1000] Hz with a step of 73.077 Hz and that of the next 27 are logarithmically spaced in the range [1001, 6400] Hz with a $\log\text{Step} = 1.071103$. The $\log\text{Step}$ is computed using Eq.(4.9).

$$\log\text{Step} = \exp\left(\frac{\ln\left(\frac{f_{c40}}{1000}\right)}{\text{NumLogFilt}}\right) \quad (4.9)$$

where f_{c40} is 6400, which is the centre frequency of the last one in the logarithmically spaced filters. The NumLogFilt is equal to 27, which is the

number of logarithmically spaced filters. Each one of these equal area triangular filters is defined as in Eq.(4.10). In the equation, i takes the values $1, 2, \dots, p$ and stands for the i^{th} filter, f_{bi} is the width of the filter bank and is defined at $p + 2$ boundary points.

$$H_i(k) = \begin{cases} 0 & \text{for } k < f_{bi-1} \\ \frac{2(k - f_{bi-1})}{(f_{bi} - f_{bi-1})(f_{bi+1} - f_{bi-1})} & \text{for } f_{bi-1} \leq k \leq f_{bi} \\ \frac{2(f_{bi+1} - k)}{(f_{bi+1} - f_{bi})(f_{bi+1} - f_{bi-1})} & \text{for } f_{bi} \leq k \leq f_{bi+1} \\ 0 & \text{for } k > f_{bi+1} \end{cases} \quad (4.10)$$

Here, k is equal to $1, 2, \dots, N$ and corresponds to the k^{th} coefficient of the N point *DFT*. The boundary points f_{bi} are expressed in terms of the position. The filter bank given by Eq.(4.10) is normalised in such a way that the sum of the coefficients for every filter is equal to one. Thus, the i^{th} filter satisfying the Eq. (4.11) is used.

$$\sum_{k=1}^N H_i(k) = 1 \text{ for } i = 1, 2, \dots, p \quad (4.11)$$

The equal area filter bank given by Eq.(4.10) is employed for generating the log-energy output. The MFCCs are obtained by performing Discrete Cosine Transform on the logarithm of Mel-spectral coefficients [191]. Of the many MFCCs, only the first 20 coefficients of each frame are considered, since most of the features of the noise source can be extracted from these coefficients. The use of DCT minimizes distortion in the frequency domain and results in high computation efficiency, since an N -point DCT can be evaluated using a symmetric $2N$ -point FFT.

4.3.2.3 Window function

When the spectral analysis techniques like the FFT are applied to the segments as a whole, it behaves as if it is operating on a data signal

waveform that is zero just before the segment and then abruptly jumps to the signal during the segment and then back to zero when the segment ends. This introduces significant distortion of the signal and warrants the need for windowing.

The windowing operation removes the undesirable undulations and smoothens the edges of each data record. It reduces spectral distortions, discontinuities or abrupt changes at the end points. More specifically, if the original signal level is $s(i)$ at time i , then the windowed signal can be represented as $s(i) * W(i)$ where $W(i)$ is the window function. In this study, the windowing function is performed by Hamming window, defined by Eq. (4.12).

$$W(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (4.12)$$

When the cepstral coefficients are extracted, it is observed that the MFCCs for different records vary over a wide range of values. Hence, the optimum set of values for the cepstral coefficients are to be synthesized by a technique referred to as vector quantization [191].

4.3.3 Speaker Modeling

In the speaker modeling module, the dimension of feature data is reduced by vector quantization technique. The vector quantization is a lossy data compression method based on the principle of block coding. It codes the values from a multidimensional vector space into values in a discrete subspace of lower dimension. In this work, the LBG (Linde, Buzo, Gray) design algorithm [192] for vector quantization is used for trimming the cepstral coefficients to the nearest value.

The MFCC matrix is vector quantized by passing each column of this matrix through a vector quantizer. In vector quantization, the columns of the MFCC matrix are taken as source vectors, which will generate the quantized code vectors comprising of the various cepstral coefficients at different frequencies. If the source vectors are k-dimensional, then X_m can be represented by Eq.(4.13).

4.3.3.1 Vector Quantization and Optimization with LBG

The MFCC matrix is vector quantized by passing each column of this matrix through a vector quantizer. In vector quantization, the columns of the MFCC matrix are taken as source vectors, which will generate the quantized code vectors comprising of the various cepstral coefficients at different frequencies. If the source vectors are k-dimensional, then X_m can be represented by Eq.(4.13).

$$X_m = (x_{m1}, x_{m2}, \dots, x_{mk}) \text{ where } m = 1, 2, \dots, M. \quad (4.13)$$

Let P be the number of code vectors, which are synthesized from the M source vectors. The average of the entire elements in a row of the MFCC matrix is computed and a code vector is generated.

$\{C\} = \{c_1, c_2, \dots, c_P\}$ represents a set of k-dimensional code vectors with c_p given by $(c_{p1}, c_{p2}, \dots, c_{pk})$ where $p = 1, 2, \dots, P$. P represents the number of code vectors that are to be synthesized from the source vectors and specified at the time of initialization.

The LBG algorithm requires an initial codebook containing one code vector obtained by evaluating the row wise mean of the MFCC matrix. The initial code vector is split into two column vectors by adding and subtracting an error term [193]. From these column vectors, the

minimum distance to the various columns of the MFCC matrix is computed by the Euclidean distance technique using the Eq.(4.14).

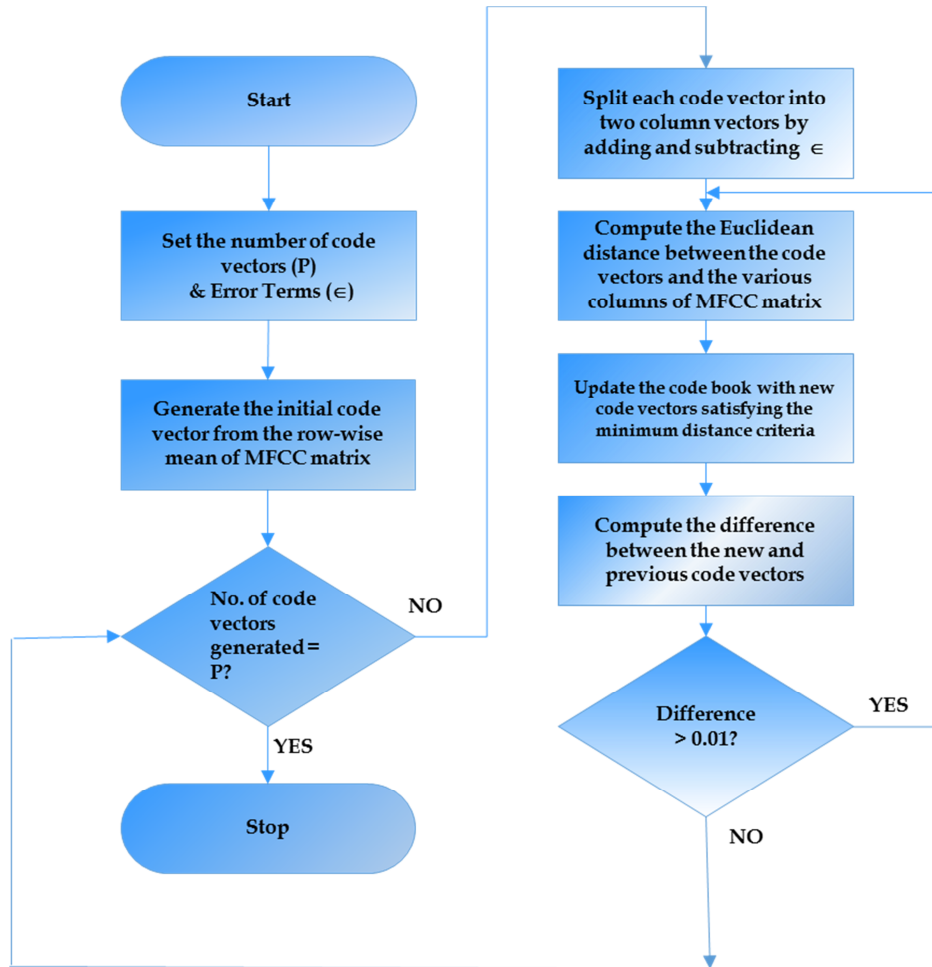


Figure 4.6 Procedure for vector quantization

$$D = \sqrt{\sum_{m=1}^M (\vec{X}_m - \vec{Y}_m)^2} \quad (4.14)$$

where D is the Euclidean distance, \vec{X}_m is the source vector and \vec{Y}_m is the code vector. The stipulated procedure for trimming the cepstral coefficients using vector quantization is illustrated in the flowchart shown in Figure 4.6. The column vectors of the MFCC matrix are identified by using this

minimum distance criterion. The index and new code vectors are generated from the average values of the feature vector.

The new code vectors, so generated, are compared with the previous code vectors. If the difference is greater than the error term, minimum distances between these code vectors and the MFCC matrix is again computed by the same method. The codebook is updated until the difference is less than the error term. The process of splitting the code vectors is continued until the number of iterations is equal to that specified at the time of initialisation.

4.3.4 Decision logic

A conceptual diagram to illustrate the recognition process is shown in Figure 4.7. In the figure, only two speakers and two dimensions of the acoustic space are shown [194]. The circles refer to the acoustic vectors from speaker 1 while the triangles represent the vectors from speaker 2. In the training phase, a speaker-specific VQ codebook is generated for each known speaker by clustering his/her acoustic training vectors. In Figure 4.7, black circles and triangles represent the resultant codewords or centroids for speaker 1 and 2, respectively. The distance from a vector to the closest codeword of a codebook is called VQ-distortion [195]. VQ distortion is the Euclidean distance between the two vectors and is given by Eq. (4.14). In the recognition phase, an input utterance of an unknown voice is “vector-quantized” using each trained codebook and the total VQ distortion is computed. The speaker in the database corresponding to the VQ codebook with the smallest total distortion is identified.

4.4 Results and Discussions

Hundred users have been considered for the training of the proposed system. Altogether, 1000 sound records have been utilized for the training purpose. For the testing purpose, in addition to the sound records of 100 users who had been considered for training, sounds of 100 new unregistered users have been employed. Hence the prototype accounts for the attempts of unregistered users in the system.

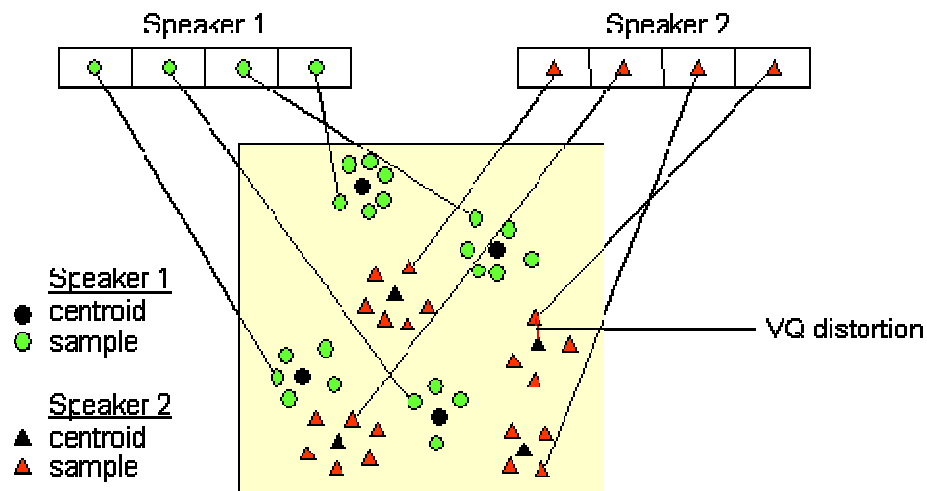


Figure 4.7 Conceptual diagram illustrating VQ codebook formation.

The speech signal of speaker1 is plotted in Figure 4.8. The signal characteristics are found to be stationary when examined over a short period ranging between 5 and 100 milliseconds. However, over longer time intervals, the magnitudes of the signal characteristics change, reflecting the different speech sounds. Therefore, in this study, short-time spectral analysis has been used to characterize the speech signal. The MFCC representing the speech signal has been utilized for the speaker recognition task.

In the frame-blocking step, the continuous speech signal is blocked into frames of N samples, with adjacent frames being separated by M , where M is less than N . The first frame consists of the first N samples; the second frame begins M samples after the first frame, and overlaps it by $N-M$ samples. Similarly, the third frame begins $2M$ samples after the first frame and overlaps it by $N-2M$ samples. This process continues until the whole speech is accounted for within one or more frames.

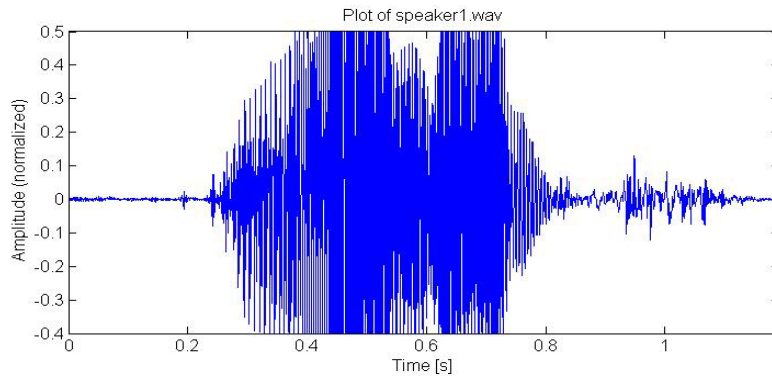


Figure 4.8 Speech signal of Speaker 1

The values of N and M are chosen as 256 and 85 respectively. The magnitude of $N = 256$ is equivalent to ~ 30 millisecond windowing. Frame blocking of the speech signal is done sufficiently for a short period of time because its characteristics are found to be fairly stationary. The magnitude of N is taken as 256 in order to compromise between the time resolution and frequency resolution. These time and frequency resolutions can be observed from the corresponding power spectrum of speech files, as shown in Figure 4.9. In each case, frame increment M is taken as $N/3$.

The resolution of time is found to be high when N is equal to 128. Furthermore, each frame lasts for a very short period. The analysis result shows that the signal for a frame doesn't change its nature (Figure 4.9). On

the other hand, there are only 64 distinct frequency samples which indicates that the frequency resolution is poor. When N is equal to 512, there is good frequency resolution and 256 different values are obtained. The number of frames in this case is relatively small, which means that the resolution in time is strongly reduced. When the value of N becomes 256, the number of frames is relatively small, which reduces the computing time. For N equal to 256, there is a compromise between the resolution in time and the resolution in frequency. Hence, $N = 256$ is adopted in this study.

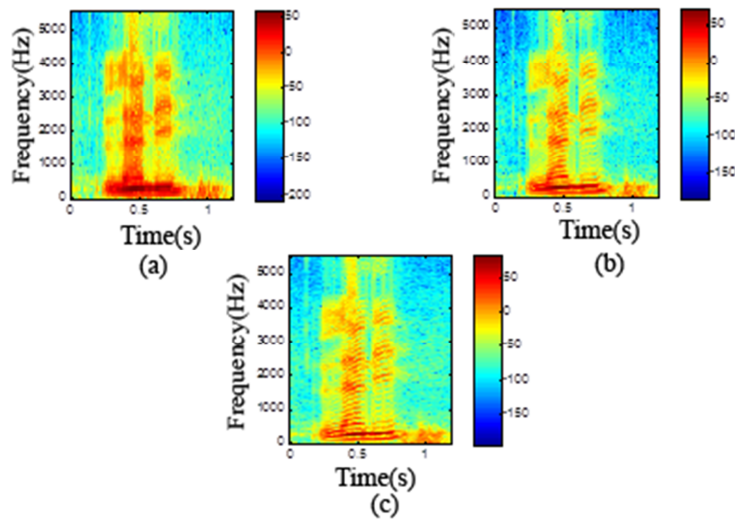


Figure 4.9 Power spectral analysis of speech signal of Speaker 1
(a) $M= 43, N= 128, \text{Frames} = 303$
(b) $M= 85, N= 256, \text{Frames} = 152$
(c) $M= 171, N= 512, \text{Frames} = 74$

The next step in processing is the windowing of individual frame which is done to minimize the signal discontinuities at the beginning and the end of each frame. The Hamming window is used for windowing operations and its plot is given in Figure 4.10.

The next step is the determination of FFT. Each frame of N samples from the time domain is converted into the frequency domain using FFT. The result of this step is often referred to as spectrum or periodogram. The next step is the mel-frequency wrapping. The mel-spaced filterbank, which operates with triangle-shaped windows, is applied in the frequency domain, and the results are given in the Figure 4.11.

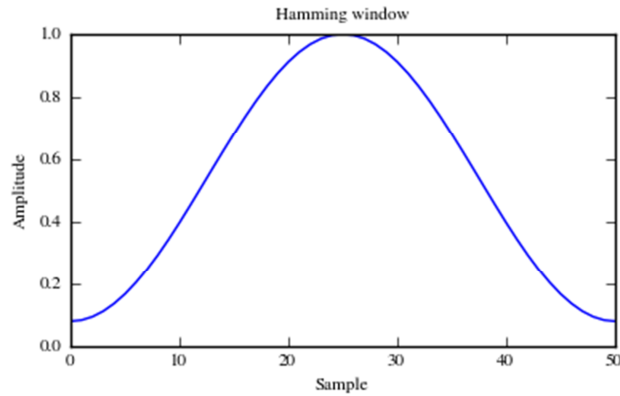


Figure 4.10 Hamming Window

In the final step, the log Mel spectrum is converted back to time domain. The result is called the MFCC. The cepstral representations of the speech spectrum provides a good representation of the local spectral properties of the signal. The Mel spectrum coefficients, which are real numbers, can be converted to the time domain using DCT. The first component of MFCC calculated from DCT is excluded since it represents the mean value of the input signal, which carries little information about the specific speaker.

The resulting acoustic vectors of MFCCs corresponding to fifth and sixth filters are plotted in the Figure 4.12. The vectors corresponding to two signals overlap in some regions. Certain regions seem to be used

exclusively by one or the other speaker. This is used to distinguish the different speakers. The vectors don't form actual clusters, but there are areas where their density is higher. The figure shows only a two-dimensional plot whereas the actual vector contains 20 dimensions.

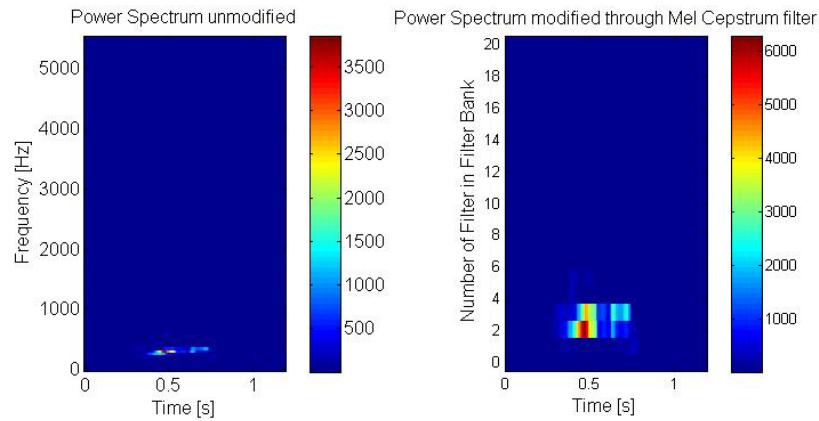


Figure 4.11 Power spectrum of speaker 1 modified through mel-spaced filter bank

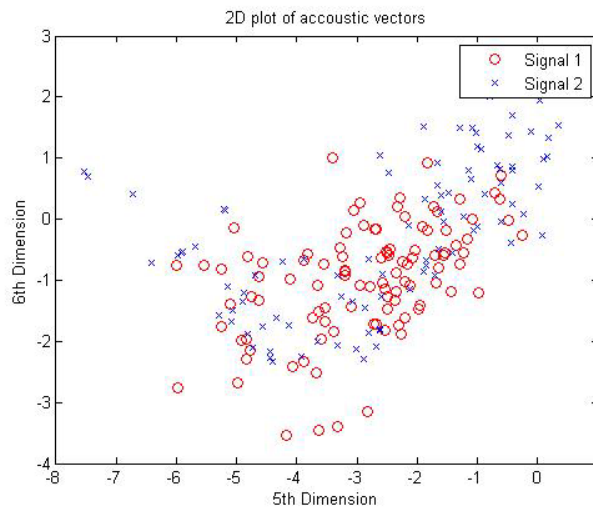


Figure 4.12 MFCCs corresponding to speaker1 through 5th and 6th filters.

For each speech frame of 30 milliseconds, a set of MFCCs is computed. These are the result of cosine transform on the logarithm of short-term power spectrum expressed on a mel-frequency scale. This set of coefficients is called an acoustic vector. Therefore, each input utterance is transformed into a sequence of acoustic vectors. The speaker specific codebook is formed using LBG VQ algorithm. The resultant codebooks along with the MFCCs are shown in Figure 4.13.

The speech signals corresponding to the speakers in the training folder are compared with the speech files of the same speakers in the testing folder. The Euclidean distance is small when the test data matches with the template in the database.

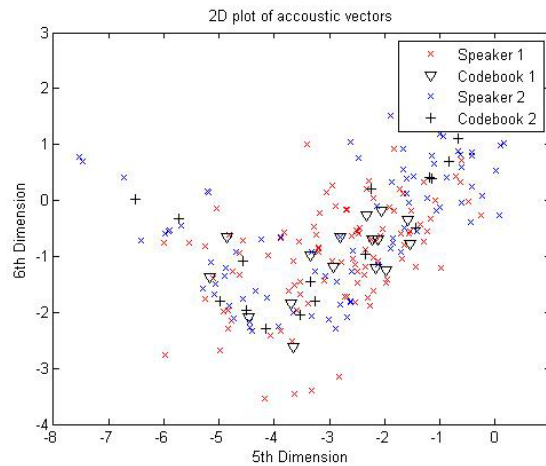


Figure 4.13 Codebooks and MFCCs corresponding to speaker 1 and 2.

4.5 Performance analysis of speaker recognition system

The system is tested to find out the False Acceptance Rate (FAR) and False Rejection Rate (FRR). False acceptance means acceptance of

impostors whereas false rejection means rejection of a true claimant. These parameters are used to evaluate the performance of the system.

The false acceptance and false rejection rates of the speaker and face unimodal systems are computed and shown in Figure. 4.14. In speaker unimodal system, the optimum value of threshold is found to be 1.2. The magnitudes of FAR and FRR are found to be 20% at the optimum value of threshold when cepstral features alone is considered.

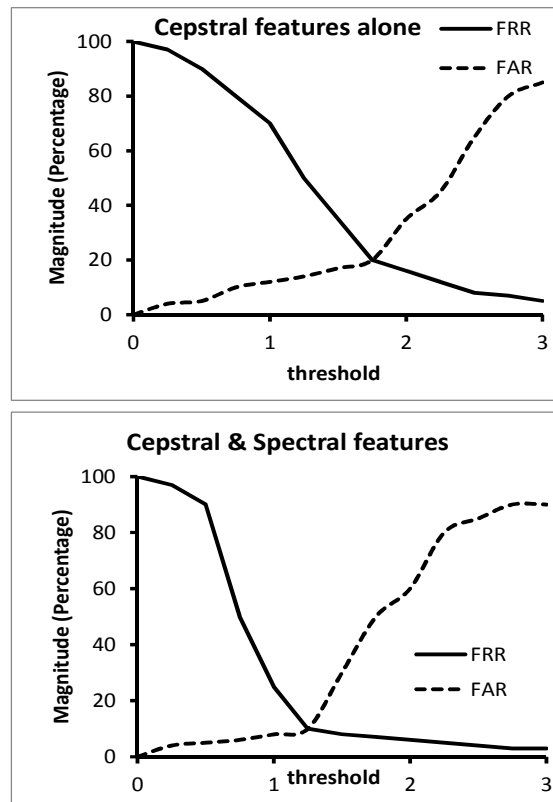


Figure 4.14 Influence of Cepstral & Spectral features

It is found from the literature that Shaneh *et al.* (2009) [25] and Shi. Huang *et al.* (2009) [27] considered MFCC alone as the feature

vector for modelling their speaker recognition system. In the present study, in addition to MFCC, other features like spectral rolloff, spectral centroid and spectral flux are used for constructing the feature vector. The speaker recognition system developed in this study is found to be more robust with distinct feature matrix. The success rate with MFCC alone is 80%. It is found that the success rate increases by 10% after the inclusion of spectral features such as spectral centroid, spectral flux, spectral rolloff and spectral range as shown in Figure 4.14. The success rate stays at 90% after including the other spectral features like number of peaks, spectral crest etc. This result is as shown in the Figure 4.15. Hence, in the proposed system, 20 MFCC coefficients and four spectral features are considered..

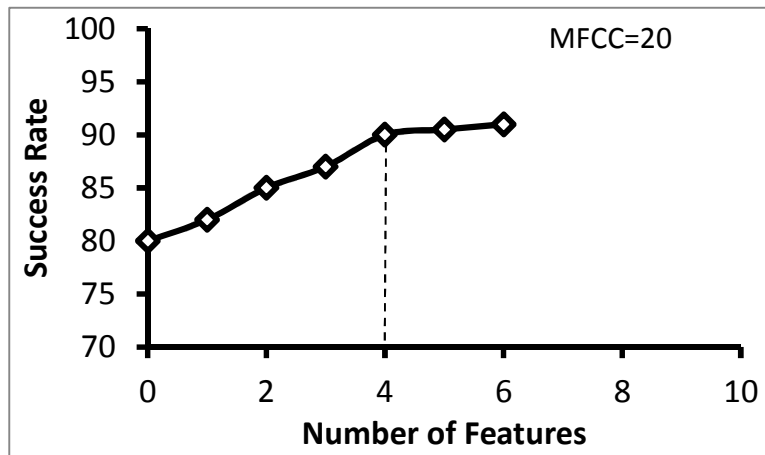


Figure 4.15 Success Rate of Cepstral and Spectral features

4.6 Summary

The techniques and procedures involved in extracting the various feature components for speech are highlighted in this chapter. The classical and a parametric model based power spectral estimators for extracting the

spectral features are presented in this chapter. It is observed that the parametric model based estimators give better results for short data segments and yield better frequency resolutions than conventional estimators. This chapter also describes the concepts of cepstral analysis which belongs to an area of signal processing known as homomorphic analysis. The cepstral coefficients are computed by estimating the MFCCs, which is a measure of the perceived harmonic structure of sound.

The performance analysis of speaker recognition system is also carried out. The incorporation of additional spectral features in the feature vector is found to improve the overall performance of the system.

CHAPTER 5

FACE RECOGNITION

In this chapter, human face has been identified as a key to security and can be used for a wide variety of applications in both law enforcement as well as non-law enforcement. Facial recognition records the spatial geometry of distinguishing features of the face. Face images are identified based on Eigen face approach using PCA. PCA extracts only the components with the largest magnitudes and the dimensionality reduction removes the unwanted information. The face structure is precisely decomposed into uncorrelated components known as eigen faces and will be stored as a 1-D array which is actually a weighted sum of the components. In eigen face approach, after the dimensionality reduction of the face space, the distance is measured between the image under consideration and the template. When the distance is less than some set threshold value, then it is identified as a known face.

5.1 Introduction

Due to the easiness in collection of data, face recognition, which started to evolve as early as 1936 focusing just on still images, is a widely accepted biometric. Decades of research efforts have brought out feasible, machine recognition based techniques that use computer to work more systematically even for video images and is used in a number of applications including crowd surveillance, criminal identification, access to

entry etc. The state-of-the-art face recognition techniques have reached a certain stage of maturity but are still limited to specific environments with constraints like illumination change or pose variation.

In general, the human recognition system utilizes a broad spectrum of stimuli, obtained from many senses viz. visual, auditory, tactile etc., in individual or collective manner for the purpose of recognition. For contextual knowledge, the surroundings play an important role implying that holistic and feature information are crucial for the perception and recognition of faces. When dominant features are present, holistic descriptions may not be used. For example, hair, face outline and mouth are determined to be important for perceiving and remembering front view faces and when it comes to the side view, nose plays a significant role. The original look of the face changes with the variations in hairstyle, wearing spectacles, facial hair like beards, aging etc. As face is quite complex, a single change in a feature of the face can alter its look considerably making face recognition a really complex task involving visual techniques [130].

Face recognition is a part of a wide area of pattern recognition technology [196]. The process includes mainly three tasks - acquisition, normalization and recognition. The term acquisition refers to the detection and tracking of face-like image patches in a static scene. Normalization is the segmentation and alignment of face images while recognition is the representation and modelling of face images as identities as well as the association of novel face images with known models. Automation of face recognition algorithms mostly deals with digital image processing, which is a quite complex field that poses many problems.

The machine recognition of face from stills is an active research area spanning several disciplines such as image processing, pattern recognition and PCA computer techniques. Although humans seem to recognise faces in cluttered scenes with relative efficiency, machine

recognition is a much more complex task. Face recognition from a single image is a challenging task because of variability in scale, location, orientation and pose. Face localization is an invariable step towards the process of face detection, which aims to determine the image position of single faces with the assumption that an input image contains only one face. The face recognition compares an input image against a database and in turn, reports matched cases.

5.2 Face Recognition Algorithms

Face recognition systems received considerable attention in recent years, both from industry and the research community. Among the popular biometric technologies, facial features scored the highest compatibility in a Machine Readable Travel Documents (MRTD) system [197] and the details are given in Figure 5.1. There has been considerable amount of research in this field and many techniques have been proposed and implemented.

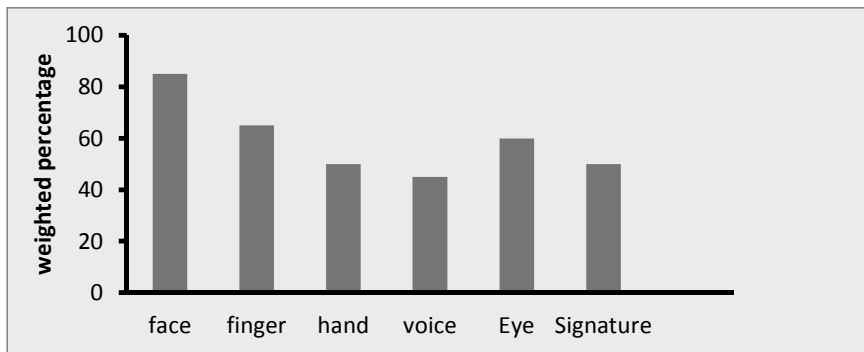


Figure 5.1 Comparison of various biometrics based on MRTD compatibility

The four main categories of current face recognition techniques are shown in Figure 5.2. They are

- *Appearance based*, which uses holistic texture features.

- *Model-based*, which employs shape and texture of the face, along with the 3D depth information.
- *Template-based* face recognition.
- Techniques using *neural networks*.

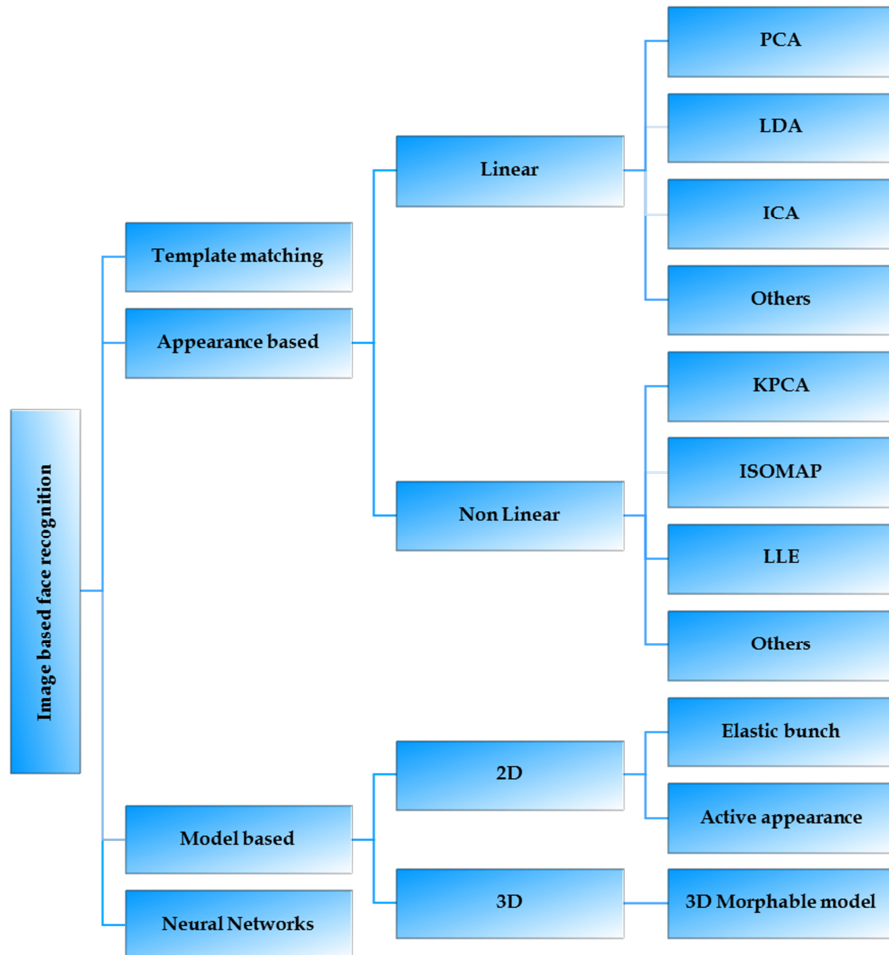


Figure 5.2 Classification of face recognition techniques

5.2.1 Template-based Face Recognition

Template matching uses pixel intensity information, either as original gray-level or as processed data to highlight specific aspects. The template can be either the entire face or regions corresponding to general

feature locations such as the eyes or the mouth. Cross correlation of test images with all the training images identifies the best match.

Brunelli and Poggio (1993) [137] compared feature and template-based methods directly with the same database of frontal face views. In this work, feature-based templates of mouth, eyes and nose were automatically detected. The test results indicated that the template-based techniques outperform feature-based techniques.

5.2.2 Appearance-based Face Recognition

The faces are stored as two-dimensional intensity matrices. Each image is a point in a high-dimensional vector space. In order to identify the different faces, depending on the application of interest, an efficient and effective representation of feature space is derived and for a given test image, the similarity between the stored prototypes and the test image is carried out in this space. This technique can be either Linear (subspace) Analysis and Nonlinear (manifold) Analysis.

5.2.2.1 Linear (Subspace) Analysis

PCA [128], ICA [198] and LDA [61][199] are classical linear subspace analysis techniques used in face recognition. Each classifier has its own representation of high-dimensional face vector spaces called basis vectors. Some statistical considerations forms the backbone for the basis vectors and feature vector is obtained, after projecting the faces on to this. The matching score between the projected test image and the projected training images is calculated and face identification is carried, based on the magnitude of matching score.

5.2.2.2 Nonlinear (Manifold) Analysis

The nonlinear manifold is more complicated than linear models. Actually, linear subspace analysis approximates this nonlinear manifold.

Direct nonlinear manifold modelling schemes helps in learning this nonlinear manifold. Kernel Principal Component Analysis (KPCA) is a nonlinear analysis technique.

5.2.3 Model-based Face Recognition

This approach uses a model formed from a prior knowledge of the facial features of the face performs the recognition. The tool developed by Wiskott *et al.* (1997) [57] utilizes the elastic bunch graph matching technique. Cootes *et al.* (1998) [200] integrated both shape, texture and developed a new technique called the 2D morphable face model, which measures the face variations. The 3D morphable face model is one of the latest developments in which a true 3D structure of the human face surface helps in recognition. The model-based approach usually involves four steps:

- Model development
- Model fitting
- Development of feature vector
- Matching

5.2.3.1 The Feature-based Elastic Bunch Graph Matching Technique

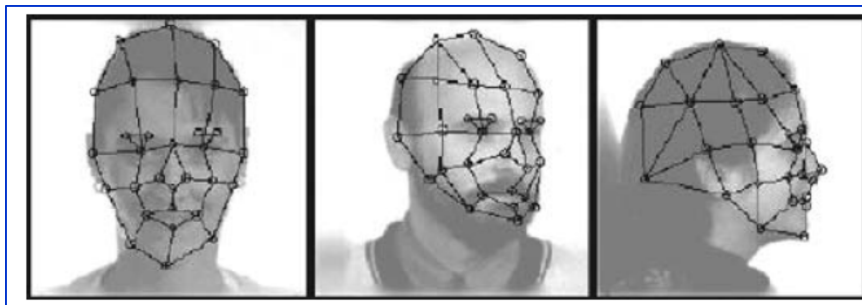


Figure 5.3 Multiview faces overlaid with labeled graphs

All human faces share a similar topological structure. Wiskott *et al.* (1997) [57] presented a general class recognition

method for classifying members of a known class of objects. Faces represented as graphs, with nodes positioned at fiducial points and edges labelled with 2D distance vectors, are shown in Figure 5.3.

5.2.3.2 The 3D Morphable Model

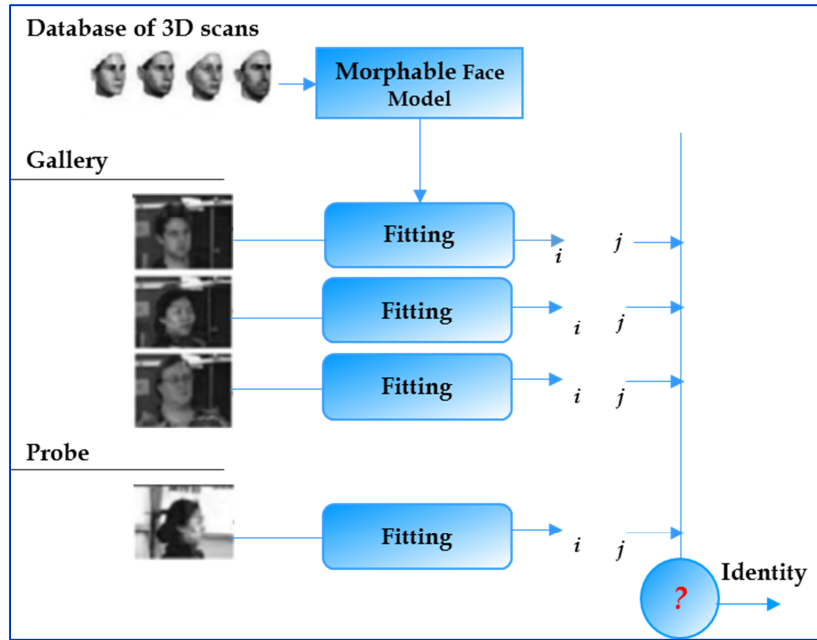


Figure 5.4 3D morphable face model [201]

A labelled graph is a set of nodes connected by edges, with nodes labelled with jets and edges labelled with distances. Each node contains a set of 40 complex Gabor wavelet coefficients, including both phase and magnitude, commonly known as ‘jet’. The edges encode the geometry of an object, while nodes encode gray-value distribution patch-wise. A face bunch graph is a stack-like structure that combines graphs of individual sample faces. A human face can be represented as a surface lying in 3D space which implies that the 3D model is better for representing faces especially while handling facial variations such as pose and illumination. Blanz *et al.* (2003, 2002) [201],[202] proposed a method based on a 3D

morphable face model that encodes the shape and texture in terms of the model parameters, the details of which are in Figure 5.4.

5.3 Linear Subspace Techniques

Linear subspace technique helps in the dimensionality reduction of facial features. Since a large number of feature variables in the database are to be analysed in a face recognition system, it is likely that subsets of variables are highly correlated with each other. The accuracy of classifiers suffers when highly correlated variables are used. One of the key steps in face recognition is finding ways to reduce dimensionality without sacrificing accuracy.

PCA, LDA and Independent Component Analysis (ICA) [203] fall under the broad class of linear transformations that transform a number of possibly correlated variables into a smaller number of variables. The objective is to reduce the dimensionality or the number of variables in the dataset, but retain most of the original variability in the data. Linear subspace techniques forms the feature extraction technique used to reduce or remove redundant or irrelevant information from the data.

The reduction of dimensionality in face recognition is more complex when the important measurement of feature for a given application is unknown. In the work, PCA is the linear subspace technique used for dimensionality reduction and Euclidean distance calculates the matching score between training and test image.

5.3.1 Principal Component Analysis

Principal Component Analysis (PCA) is a constructive statistical technique with many application areas such as face recognition, image compression, security access control, criminal identification, law enforcement etc. and is a common technique for finding patterns in data of

high dimension. In communication theory, it is known as the Karhunen-Loeve transform [204]. The main idea is to find a set of M orthogonal vectors in the data space that account for maximum possible variance of data. Projecting the data from their original N -dimensional space onto the M -dimensional subspace spanned by these vectors then undergoes a dimensionality reduction that often retains most of the inherent information in the data. The first principal component is along the direction with the maximum variance and the second principal component is constrained to lie in the subspace perpendicular to the first. Within that subspace, it points in the direction of maximal variance. Then, the third principal component is taken along the maximum variance direction in the subspace perpendicular to the first two, and so on. PCA is the best known dimensionality-reduction tool that helps to reduce a large dataset to a smaller set while retaining the information in the larger set simultaneously. The divide-and-conquer method minimizes the inherent complexity of dealing with a large problem in a given time. Similarly, the dimension of difference information embodied in large covariance matrix is brought down using PCA to enhance the subsequent computations in face identification. Eigen face approach utilises this technique to compute the variations in similarities of the faces in the database and project them onto a face-space. PCA is a technique used to remove the correlated fraction of the input data and provides an insight into the information content of the input facial image data, emphasizing the significant features. These features are not related to the conventional notion of facial features such as eyes, nose and mouth. The algorithm is a decomposition algorithm based on principal component analysis that finds the vectors, which best account for the distribution of facial images within the entire face database.

The main reason for using the approach for face recognition is actually the outcome of the dilemma of what aspects of the face are to be considered for identification - whether the face is treated as a uniform pattern or the positions of features are adequate etc. Depending too much on feature representation is not sufficient to support robust face recognition because it causes problems when the image is degraded by noise or when the features are occluded. This led to the concept of using the approach, which treats the whole face as a uniform pattern for face recognition. The eigen face approach is the first real successful demonstration of automation of human face recognition. PCA is used to derive the low dimensional representation of faces in the eigen face approach by applying it to a dataset of images representing faces. The system implementation is achieved by projecting the face images onto a feature space that spans the significant variations among known face images. These significant features are called eigen faces. However, these features do not really represent the individual facial features such as eyes, nose and ears. They just capture the image points that cause meaningful variations among the faces in the database and in turn differentiates them.

The eigen face approach is a simple and effective algorithm that can be applied on test images unaffected by illumination changes, provided that the faces are recorded under similar illumination conditions. Formally, eigen faces are the principal components of the distribution of faces or the eigen vectors of the covariance matrix from the set of face images. A linear combination of eigen faces represents each face exactly by using eigen vectors corresponding to the largest eigen values. Eigen vectors are the coordinates that define the direction of the axes, whose lengths are given by the eigen values. To account for different lighting conditions, modular eigen space approach, which is less sensitive to appearance changes when

compared to the standard eigen face method, is used. The evaluation of eigen values and eigen vectors is a unique matrix operation.

Recognition of images using PCA involves three basic steps. The first step is creation of transformation matrix Ω using training images, while the next is projection of the training images are projected onto the matrix Ω . Finally, identification of the test images by projecting them onto the subspace and comparing them with the trained images in the subspace domain. Figure 5.5 gives an overview of the face recognition system.

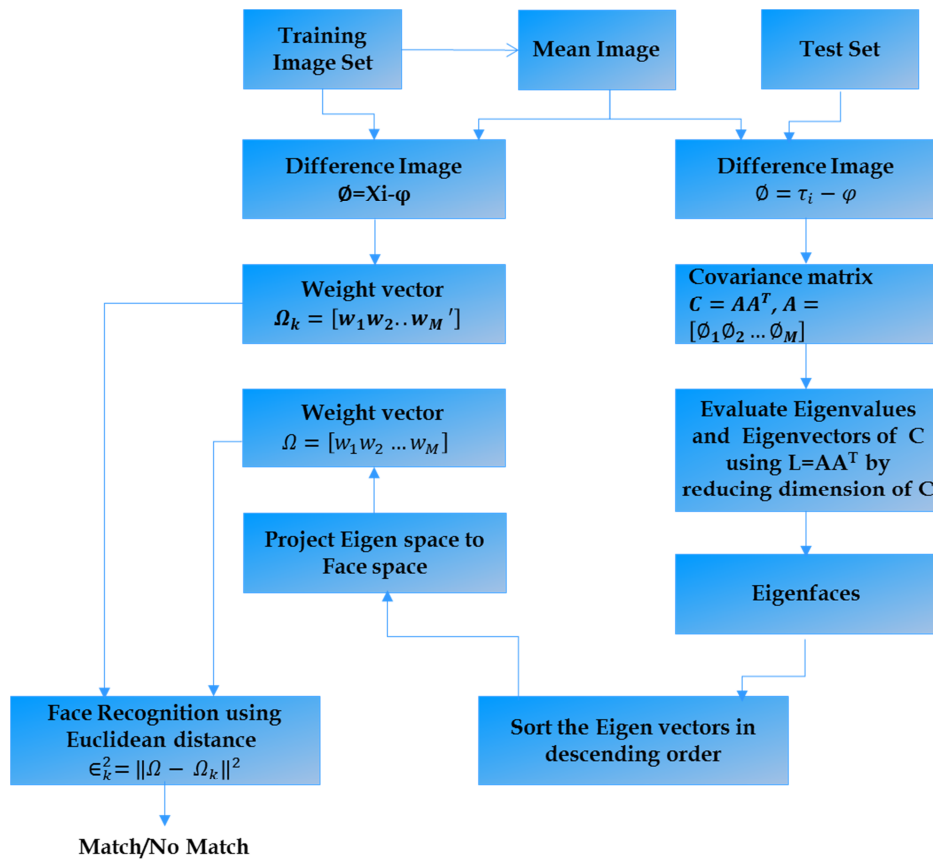


Figure 5.5 Overview of the face recognition system

5.3.1.1 Creating Matrix Ω

The steps like removing the mean, computing of the covariance matrix, eigen values and eigen vectors are carried out to compute matrix Ω .

Removing the mean:

Each of the training images is first mean adjusted. A subset of training face database is shown in Figure 5.6. The mean image is subtracted from each training image. A two-dimensional image is defined as a function of $i_{(x,y)}$, the intensity of the image at any pair of coordinates (x, y) and the details are shown in Figure 5.7. Every image I_i is expressed as a matrix of intensity at every pixel in the chosen resolution $m \times n$ and is given in Eq. (5.1). These M images forms 1-D vectors X_i as in Eq.(5.2). Thus, the whole face is considered as a collection of pixels and coded by many vector components, arranged sequentially by concatenating one to the other. The face space F consists of all the M images of X_i in the training as given in Eq.(5.3).



Figure 5.6 Subset of Training face database

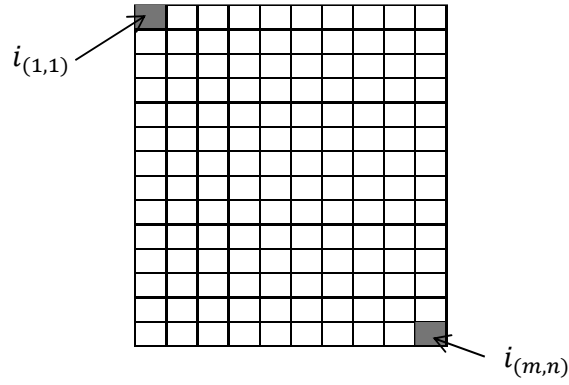


Figure 5.7 Schematic representation of mapping of single image into intensity matrix.

$$I_i = \begin{bmatrix} i_{(1,1)} & \cdots & i_{(1,n)} \\ \vdots & \ddots & \vdots \\ i_{(m,1)} & \cdots & i_{(m,n)} \end{bmatrix} \quad (5.1)$$

$$X_i = \begin{bmatrix} i_{(1,1)} \\ \vdots \\ i_{(m,n)} \end{bmatrix} \quad (5.2)$$

$$F = (X_1 \quad \dots \quad X_M) \quad (5.3)$$

The mean image φ as shown in Figure 5.8, is the average information of all the images representing the mean value of every pixel in N-dimensional vector as in Eq. (5.4) and Eq. (5.5).

$$\varphi = \frac{1}{M} \sum_{i=1}^M X_i \quad (5.4)$$

$$\varphi = \begin{bmatrix} \varphi_{(1,1)} \\ \vdots \\ \varphi_{(m,n)} \end{bmatrix} \quad (5.5)$$

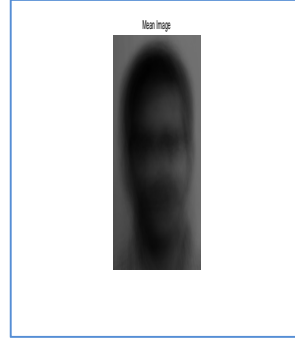


Figure 5.8 Mean image of face database

The difference image ϕ , representing the deviation from the average image φ , is given by Eq. (5.6).

$$\phi = X_i - \varphi \quad (5.6)$$

The matrix A is developed by assembling the values of ϕ 's and is given in Eq. (5.7).

$$A = \begin{bmatrix} \phi_{(1,1)} & \cdots & \phi_{(1,M)} \\ \vdots & \ddots & \vdots \\ \phi_{(mn,1)} & \cdots & \phi_{(mn,M)} \end{bmatrix} \quad (5.7)$$

Compute the covariance matrix (C):

The main idea behind the eigen face technique is to exploit the similarities among various images. Separation of average information and deviation from the mean will be the first step in this approach. Eigen faces are extracted from the difference information. For this purpose, the covariance matrix C with dimension $N \times N$ is computed as in Eq. (5.8), where N is equal to the product of m and n .

$$C = \frac{1}{M} \sum_{i=1}^M \phi_i \phi_i^T = AA^T \quad (5.8)$$

Where, A is of the dimension $N \times M$. This covariance matrix dimension is normally large and full eigen vector calculation is done with

practical difficulty in dimension. Without loss of generality of the whole training set, it is possible to reduce the dimensionality of the covariance matrix. The reduced dimensionality matrix L , of size $M \times M$, is given by Eq.(5.9).

$$L = A^T A = \frac{1}{M} \sum_{i=1}^M \phi_i^T \phi_i \quad (5.9)$$

Computing the eigen values and eigen vectors of C:

The eigen vectors of the covariance matrix C are computed by using the matrix L . Then the eigen vector ρ_i and the eigen value λ_i of L are obtained by solving the characteristic equation of eigen value problem $|L - \lambda I| = 0$.

$$L \cdot \rho_i = \lambda_i \cdot \rho_i \quad (5.10)$$

Substituting the value of L in Eq. (5.10),

$$A^T A \cdot \rho_i = \lambda_i \cdot \rho_i \quad (5.11)$$

Multiplying both sides of Eq. (5.11) by A ,

$$A \cdot A^T A \cdot \rho_i = A \cdot \lambda_i \cdot \rho_i \quad (5.12)$$

Since λ_i is a scalar quantity, Eq. 5.12 can be rearranged as:

$$A \cdot A^T A \cdot \rho_i = \lambda_i \cdot A \rho_i \quad (5.13)$$

Let

$$\mu_i = A \rho_i \quad (5.14)$$

Then from Eq. (5.8) and (5.14),

$$C \mu_i = \lambda_i \mu_i \quad (5.15)$$

where μ_i and λ_i are the M eigen vectors and eigen values of C respectively. In practice, a subset M' of face space M is sufficient for face reconstruction because the subspace of eigen faces can be treated as the basis for the face space.

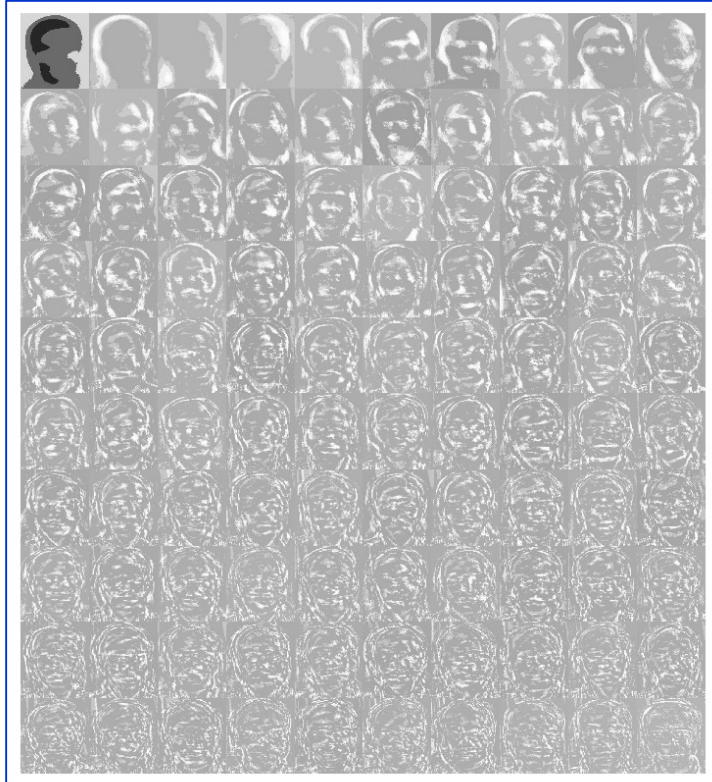


Figure 5.9 Eigen faces of the Training Image Set, which is a subset of the face database.

The original face can be represented as a linear combination of these M' vectors. The remaining $(N - M')$ eigen vectors that are associated with eigen values play insignificant role in reconstruction computation. Figure 5.9 displays the eigen faces computed with the above equation. The eigen faces are the vectors that describe faces in the face space. These vectors are perceived as limited points in the N -dimensional space. Each face is recorded as a matrix of dimension 640×480 in gray scale mode.

5.3.1.2 Reconstruction of image from PCA model

The weight (w_k) is simply the dot product of each image with each of the eigen vectors as in Eq.(5.16).

$$\mathbf{w}_k = \boldsymbol{\mu}_k^T \cdot \phi_i = \boldsymbol{\mu}_k^T \cdot (\mathbf{X}_i - \boldsymbol{\varphi}), \quad k = 1, 2, \dots, M' \quad (5.16)$$

where $\boldsymbol{\mu}_k$ is the k^{th} eigen vector of the covariance matrix, ϕ_i is the i^{th} difference image, \mathbf{X}_i is the i^{th} image and $\boldsymbol{\varphi}$ is the mean image.

All the weights are converted in the form of a matrix ($\boldsymbol{\Omega}$) with dimension $M' \times 1$ as in Eq.(5.17).

$$\boldsymbol{\Omega} = [\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \dots, \mathbf{w}_{M'}]^T \quad (5.17)$$

Reconstructed image X_R is obtained by multiplying the weight matrix ($\boldsymbol{\Omega}$) of the unknown image with the eigen vector matrix ($\boldsymbol{\mu}$) and adding the mean face image ($\boldsymbol{\varphi}$) to it as in Eq. (5.18). The trained images are projected onto the eigen face space and the weight of each eigen vector is evaluated.

$$\begin{aligned} X_R &= \boldsymbol{\mu} \cdot \boldsymbol{\Omega} + \boldsymbol{\varphi} \\ &= [\boldsymbol{\mu}_1 \ \boldsymbol{\mu}_2 \ \dots \ \boldsymbol{\mu}_{M'}] \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_{M'} \end{bmatrix} + \boldsymbol{\varphi} \\ &= \sum_{i=1}^{M'} \boldsymbol{\mu}_i \mathbf{w}_i + \boldsymbol{\varphi} \end{aligned} \quad (5.18)$$

The reconstructed image X_R is shown in Figure 5.10.

As seen from Eq.(5.18), reconstruction of the face image is obtained by adding each eigen face along with its weight to the mean face of the training set. Error estimation is carried out with the number of eigen values and the Root Mean Square (RMS) error of the reconstructed image using Eq. (5.19).

$$Error = \|\mathbf{X}_R - \mathbf{X}\| \quad (5.19)$$

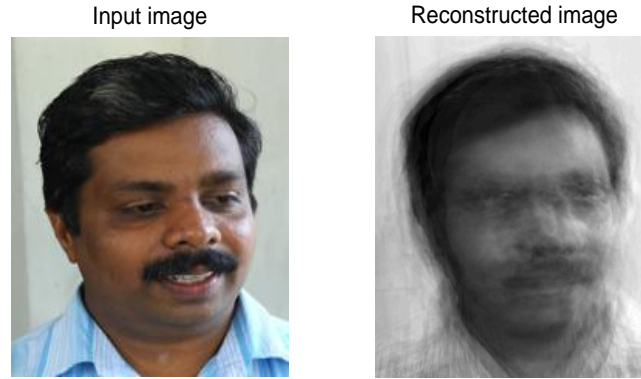


Figure 5.10 Sample reconstructed image of the original input image.

5.3.2 Recognition Procedure from PCA Model

The test image is also projected onto the same facespace and its weight is calculated. The weight vector defined in Eq. (5.20) is used for face recognition. To recognize a new test image (\mathbf{X}_T), its weight (\mathbf{w}_{Ti}) is evaluated first by multiplying the eigen vector (μ_i) of the covariance matrix (C) with difference image ($\mathbf{X}_T - \boldsymbol{\varphi}$).

$$\mathbf{w}_{Ti} = \boldsymbol{\mu}_i^T (\mathbf{X}_T - \boldsymbol{\varphi}) \quad (5.20)$$

Now the weight matrix of the test image ($\boldsymbol{\Omega}_T$) becomes as in Eq. (5.21).

$$\boldsymbol{\Omega}_T = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{M'}]^T \quad (5.21)$$

Then the Euclidean distance ϵ_k between weight matrices of the unknown image ($\boldsymbol{\Omega}_T$) and each face class ($\boldsymbol{\Omega}_i$) is obtained from Eq. (5.22).

$$\epsilon_k^2 = \|\boldsymbol{\Omega}_T - \boldsymbol{\Omega}_i\|^2 \quad (5.22)$$

Calculating the Euclidean distance between two data points involves computing the square root of the sum of the squares of the

differences between corresponding values of weight. In order to identify a test image, the distance between the image data and training data is computed. Then, this distance is compared with a threshold value. Threshold value is the maximum allowable distance between the test and training image for confirmation.

5.4 Results and Discussions

Observations have revealed the fact that the root mean square error increases as the training set members differ from each other with more variation. This is because of the addition of the mean face image. So, when there is a lot of variation among the members of the training set, the mean face image becomes cluttered, which in turn increases the root mean square error. Since identification is a pattern recognition task, accurate reconstruction of the image is not necessary. A smaller set of faces with the maximum variation is sufficient for identification. Sirovich and Kirby (1987) [205] evaluated a limited version for an ensemble of 115 images and found that about 40 eigen faces were sufficient for a very good description of the set. Turk and Pentland (1991) [128] observed that for a training set of 16 face images, 7 eigen faces were used to identify a face. Also, Tat Jun Chin and David Suter (2004) [206] have come with an inference from their experiments that 8 eigen faces were enough to account for more than 90% of the variations among a training set of 20 images. The inference drawn from the graph in Figure 5.11 reveals that as the number of eigen faces increases, the root mean square error decreases. This means that the root mean square error is inversely proportional to the number of eigen faces in the training set. RMS error for two hundred images is found to be 0.2082 where as for the other eight hundred images, the root mean square error is approximately the same. The error graph in Figure 5.11 is plotted

by taking the number of eigen faces on the X-axis for 1000 eigen faces and root mean square error on the Y-axis.

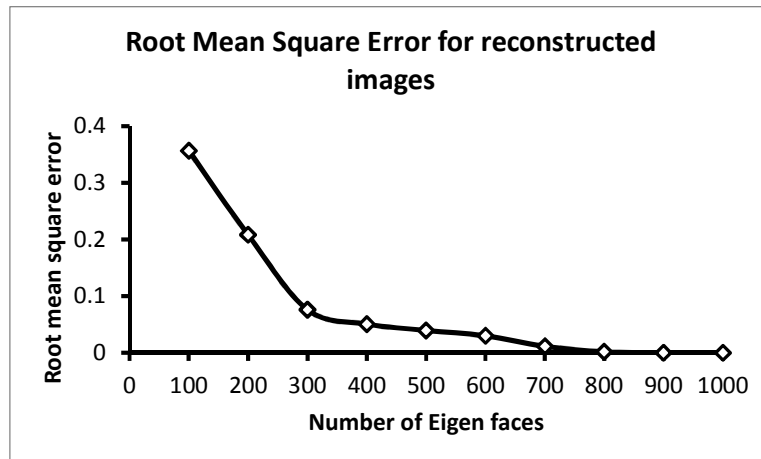


Figure 5.11 Root mean square error for reconstructed images

It means that the first two hundred eigen faces provide the principal components with prominent information and the next eigen faces contain insignificant information for the identification process. This is almost 20% of the entire training set, which implies that such a small number of eigen faces is enough for face recognition. In a database of 200 face images, just 40 images of the training set would suffice for face recognition. From this, the root mean square error reduces to be around 2% of the training image set. The graph represented in Figure 5.11 is a solid example of heuristic implementation of the above propositions.

5.4.1.1 Analysis of Eigen values

Eigen faces represent prominent features of face images in the training set. The highest eigen value of an eigen face is treated as the principal component of that face. The vector representing the highest eigen values of every eigen faces ($\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_M$) in the facespace is plotted in Figure 5.12. Figure 5.12 shows a training set of 1000 eigen faces with the

associated principal components (one from each eigen face). The first eigen face with its associated eigen value as 183.5726 is the first principal component of the entire facespace. The 50th eigen face has the highest eigen value variation of 78.9040, which is approximately half of the first. The 250th eigen face is having 16.8500 as the principal component, which is only one eleventh of the first one.

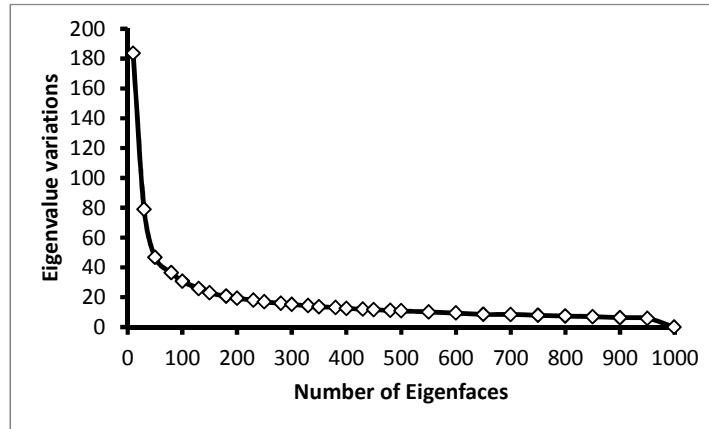


Figure 5.12 Number of Eigen faces vs Eigen value variation

The variation of execution time (for the collection of 1000 images for the entire face database) versus the number of eigen faces is illustrated in the Figure 5.13. The figure shows timing plots for both the cases- with covariance matrix reduction and without it. The graph is plotted with the number of eigen faces for 1000 images on the X-axis at an interval of 100 and the execution time on the Y-axis. It is clear from the graph that with covariance matrix reduction the timing for face recognition is around 34 seconds while without the reduction, it is around 176 seconds. This shows that without covariance matrix reduction, the execution time for the face recognition algorithm is almost five times that of with covariance matrix reduction. For 10 eigen faces, the execution time with covariance matrix

reduction is observed to be 6.1 seconds while it is 30.1 seconds without covariance matrix reduction. Similarly, for 250 eigen faces, it is 11.3 seconds with covariance matrix reduction and 55.8 seconds without covariance matrix reduction. For 500 eigen faces, it is 19.4 seconds with covariance matrix reduction and 101.01 seconds without covariance matrix reduction. For 1000 images, the execution time becomes 34.2 seconds with covariance matrix reduction and 171.2 seconds without covariance matrix reduction.

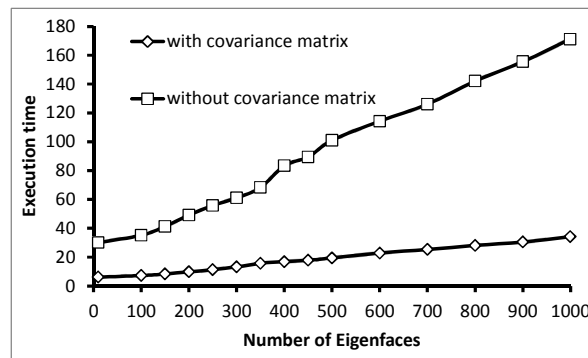


Figure 5.13 Execution time with and without covariance matrix

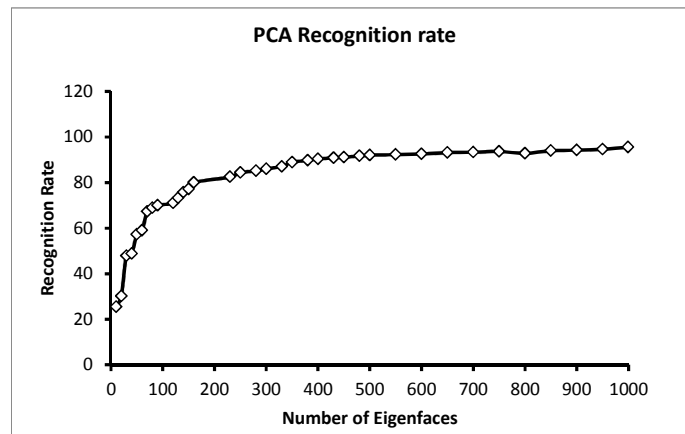


Figure 5.14 PCA Recognition Rate

The rate of recognition versus the number of Eigen faces is shown in Figure 5.14. It is observed from the graph that the recognition rate for

PCA is 25.455 with an eigen face. As the number of faces increases to five, the recognition rate becomes 57.2310. As the number of eigen faces increases further, the recognition rates are found to increase constantly and uniformly with a little difference from there until they reach the maximum value of 91.5480. It is noticed from the figure that 500 eigen faces are sufficient to reach a recognition accuracy of 89.0100%.

5.5 Summary

The problem of face recognition in a large database and a model that uses PCA approach have been proposed. The proposed PCA is found to result in a recognition efficiency of 91.548% when tested for 1000 training images. The computational time efficiency of the present PCA is estimated through covariance matrix with and without dimensionality reduction.

The computational time with reduced dimensionality is found to be 34.2 seconds (for 1000 face images) where as it is found to increase by five times i.e. 171.2 seconds for the case without dimensionality reduction. Hence, the reduction of dimensionality results in reduced computation time. The proposed model is tested using faces with pose variations. The limitation of the present work is that the training and testing images have been assumed to be of the same dimension, where as the real world images differ in dimensionality. Similarly, gray scale and static images have been used for the present evaluation procedure but they differ very much from the realistic environment. Since the proposed PCA algorithm does not possess a recognition accuracy of 100% due to its inherent disadvantage of selecting prominent but holistic features, this algorithm can be combined with any other algorithm like support vector machine, fisher faces etc. in future to achieve 100% recognition efficiency.

CHAPTER 6

SIGNATURE IDENTIFICATION

In this chapter, identity of a person based on user's handwritten signature is discussed. A signature, widely accepted as a means of identity authentication in legal and commercial transactions, is remarked as a consistent non-invasive authentication procedure by the majority of the users, thereby overcoming some of the privacy issues. The distinct features are extracted from the image of the signature trajectory captured by electronic signature tablet, after pre-processing. In signature recognition, the feature vector, which forms the signature template in the knowledge base, is selected as a combination of static and dynamic features. The static features are generated using 2D Gabor filter while the dynamic features under consideration are the x and y stroke as well as the average velocity in x and y directions. Mahalanobis Distance (MD) computed based on correlation between two signatures is used to verify the similarity of images. When the Mahalanobis distance between the feature vectors of enrolled and test signature is smaller, the similarity between the compared signatures is higher.

6.1 Introduction

Signature is a socially accepted method already in use in bank and credit card transactions. Based on the method used to capture the signatures, handwritten signature biometrics system is divided into two

categories, namely, offline and online [207]. Off-line or static is the analysis of features extracted from scanned images of handwritten signatures while on-line or dynamic is the analysis of handwritten signatures captured via digitizing tablets or other electronic devices, which capture the trajectory, pressure and velocity of handwriting. The offline method does not verify the physical presence and relies only on features, extracted from the scanned signature image. Online signatures can be captured using a variety of input devices such as digitizing tablets [208], specially designed pens, hand gloves [209] and tracking-camera.

Signature authentication is the process of verifying the identity of a person based on the user's handwritten signature. Recognizing people by their handwritten signature involves intense research [207], [208] because of the following reasons [104]:-

- Signature is resistant to fraudulent access attempts. Even though, hypothetically, no person writes his/her signature the same each time, in practice, it is very difficult to forge the dynamic data (such as speed, pen-up movement, pressure, etc.) for every digitized signature point.
- Signature has been widely accepted as a means of legal and commercial transactions identity authentication [87]. Signatures have played a historical role in authenticating documents. Signature based authentication is a consistent non-invasive authentication procedure by the majority of the users, therefore, it can help in overcoming some of the privacy difficulties [102], [103].
- The main drawback of biometrics when compared with conventional methods is that many biometrics can be copied or forged [2], [15],[132]. Whereas it is always possible to obtain a new key or another

password, it is not possible to replace any biometric data [210]. The user may change a signature, while it is not possible to change fingerprint, iris or retina pattern.

However, signature authentication is still a challenging issue for a number of reasons.

- A signature reflects people's writing habits. However, some people may experience a lot of inconsistency between their signatures, mostly because of lack of signing habit. One possible solution to cope with this limitation is to acquire multiple signature instances during enrolment instead of relying on a single instance as well as, authentication under conditions similar to those practiced during enrolment.
- While each ordinary literate human being has his/her unique style of writing the signature, the signatures tend to evolve with time and the process of signing is influenced by the physical and emotional states of the signatories.

The main task of any signature verification task is to detect whether the signature is genuine or forged [211]. The two main categories of forgeries are casual or random forgeries and skilled or traced forgeries.

Casual or random forgeries are the attempts to recreate signature trajectory without prior knowledge of the signature style whereas skilled forgery is a suitable imitation of the genuine signature [212]. The skilled forgeries are more difficult to detect than random forgeries, as the characteristic features of a skilled forgery resemble closely those of the original signature. Thus, it is difficult to discriminate skilled forgeries from authentic signatures by an offline system.

6.2 Signature Identification Methodology

The four modules of Image acquisition, Preprocessing, Feature extraction and Enrolment and identification constitute the system developed in the study. The image of the signature trajectory is captured using electronic signature tablet while the preprocessing module is responsible for preconditioning of the signature image. The distinct features of the signature image extracted in the feature extraction module helps in developing the knowledge base of the signature while the identification module authenticates. An architecture for an online signature identification system is shown in Figure 6.1.

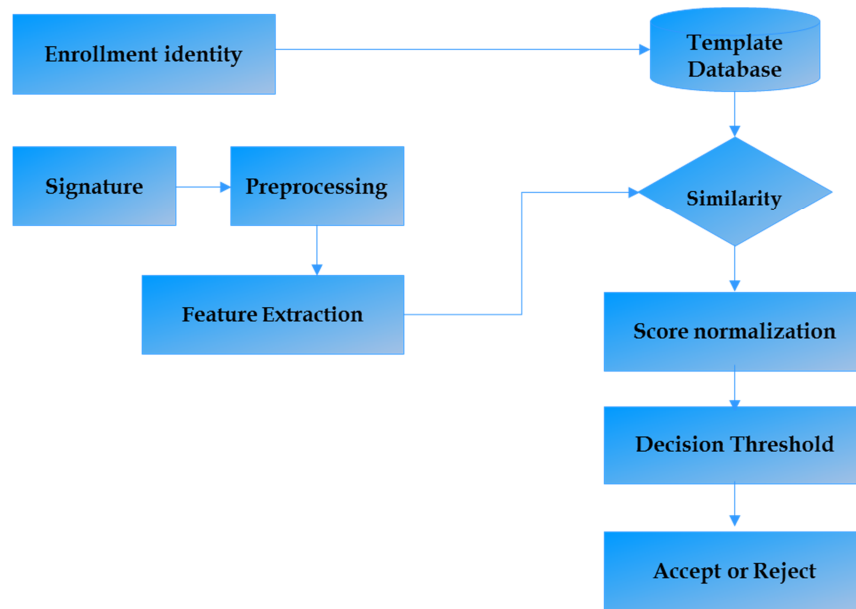


Figure 6.1 Architecture for an online signature identification system

6.2.1 Data Acquisition

Data acquisition is the process of recording the signature trajectory as well as dynamics and converting them to a digital form. In this study, the

signature information has been acquired using WACOM Bamboo pad graphic tablet as shown in Figure 6.2. When one signs on a graphic tablet, capturing of the location coordinates as well as the timing information tagged to each pair of the x and y coordinates occurs. Signatures captured using inking pens on the Bamboo Pad reflect the typical signing behaviour as users sign on paper in the way they are used to. The only difference is that there is a sensor underneath the paper digitizing data throughout the writing process. This capturing procedure requires no change in the signing behaviour of the user.



Figure 6.2 Signature acquisition device -WACOM Bamboo pad

6.2.2 Preprocessing

The images need to be preprocessed prior to processing. After binarizing the image, dilation is applied to fill the gaps and broken necks and the image is then thinned and the edges are pruned. The simplest way to use image binarization is to choose a threshold value, and classify all pixels with values above this threshold as white, and the remaining as black.

Thinning intends to reduce objects to the thickness of one pixel, generating a minimally connected axis that is equidistant from the object edges. Digital skeleton, generated by thinning algorithms, is often used to represent objects in a binary image for shape analysis and classification. The thinning process removes pixels from the outside edges of an object. The designed structuring elements find those edge pixels whose removal will not change the object's connectivity. On completion of thinning for the first pass with the eight structuring elements, the entire process for the second pass is repeated until no further changes occur. After thinning, pruning which is cleaning up of extra short noisy branches that appear in the skeleton of a pattern is taken up. The set of the short noisy branch is located using ends that have only one neighbouring pixel and moves along the path until touching another branch within a very limited number of pixels. The length of a noisy branch when related to the object characteristics and image size; is often given as at most three pixels. A branch junction is considered as the pixel having at least three neighbouring pixels. The rotation invariance of all the images is achieved by calculating the angle θ , of least second moment of the binary image S . In order to obtain some degree of invariance w.r.t. rotations, the main inertia axis of the signature is aligned with the horizontal axis. The signature when rotated about the centre of the area of the image eliminates the influence of skew angle of the signature. Finally, normalize the images in size to achieve scaling invariance preserving the aspect ratio of the signature. Figure 6.3 (a) shows the binary image of a genuine signature. Preprocessing involves finding the angle of least second moment, rotating the images as well as smoothing and thinning of the signature. The original image and the image after rotation are shown in Figure 6.3 (a) and (b) respectively.



Figure 6.3 Signature before and after rotation.

6.2.3 Feature Extraction

The discriminative power of the features in the reference set plays a major role in the entire identification process. It is important to find the features that are invariant with respect to slight changes in intra-class signatures. The features should be powerful enough to discriminate other signatures in the knowledge base.

6.2.3.1 Static Features

The features from signature images are extracted using Gabor Wavelet Transform (GWT) [213]. Gabor filters used extensively in image processing, texture analysis etc. are based on a sine or cosine wave as in Eq.(6.1). The cosine wavelets are the real parts of the wavelet and the sine wavelets are the imaginary parts of the wavelet, represented in Eq.(6.2) and (6.3) respectively. Therefore, a convolution with both the phases produces a complex coefficient.

Complex

$$g(x, y; \lambda, \theta, \varphi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(\frac{2\pi x'}{\lambda} + \varphi\right)\right) \quad (6.1)$$

Real

$$g(x, y; \lambda, \theta, \varphi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(\frac{2\pi x'}{\lambda} + \varphi\right) \quad (6.2)$$

Imaginary

$$g(x, y; \lambda, \theta, \varphi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \sin\left(\frac{2\pi x'}{\lambda} + \varphi\right) \quad (6.3)$$

In the two dimensional GWT, x' and y' are given in Eq.(6.4) and (6.5) respectively.

$$x' = x\cos\theta + y\sin\theta \quad (6.4)$$

$$y' = -x\sin\theta + y\cos\theta \quad (6.5)$$

where θ is the orientation of the normal to the parallel stripes of a Gabor function. The wavelet is rotated about its center using this parameter. The wavelet values from π to 2π are redundant due to symmetry. So, in most of the cases, θ is assumed to take values between 0 and π .

In Eq.(6.1), λ is the wavelength of the cosine wave. The gradual changes in the intensity of the image is characterised by a wavelet of large wavelength. The sharp edges and bars are represented by wavelets with short wavelengths.

φ given in Eq.(6.1) is the phase offset in the argument of the cosine factor of the Gabor function and is specified in degrees. Valid values are the real numbers between -180 and 180. The values 0 and 180 correspond to center-symmetric functions 'center-on' and 'center-off', respectively, while -90 and 90 correspond to anti-symmetric functions. All other cases correspond to asymmetric functions. If one single value is specified, one convolution per orientation will be computed and if a list of values (e.g. 0, 90 which is default) is given, multiple convolutions per orientation will be computed, one for each value in the phase offset list.

σ given in Eq.(6.1) is the radius of the Gaussian. The size of the Gaussian determines the fraction of the image influenced by convolution.

The computational effort becomes negligible when the convolution moves further from the center of the Gaussian. The radius of the Gaussian is proportional to the wavelength and is given by $\sigma = c\lambda$. This means that wavelets of different size and frequency are scaled versions of each other.

In Eq.(6.1), γ is the spatial aspect ratio of the Gaussian. It specifies the ellipticity of the support of the Gabor function. For $\gamma = 1$, the support is circular. For $\gamma < 1$, the support is elongated in orientation of the parallel stripes of the function. Default value is $\gamma = 0.5$.

The feature is extracted from the signature image by placing a virtual grid on the signature image and Gabor coefficients are computed on each point of the grid by convolution. Convolution between Gabor filter and a sub image around point (x, y) is calculated.

At each point of the virtual grid, 6 complex Gabor coefficients are computed corresponding to $\lambda \in \{2, 2\sqrt{2}\}$ and $\theta \in \{0, \frac{\pi}{4}, \frac{\pi}{2}\}$. Other Gabor filter parameters are assumed to take values of $\varphi \in \{0, \frac{\pi}{2}\}$, $\sigma = 2\lambda$ and $\gamma = 0.5$. This means that for each grid point, two frequencies in three orientations and two phases are computed. Therefore, for all the grid points of an image, Gabor coefficients are computed. The feature vector of the signature image consists of absolute values of the Gabor coefficients.

6.2.3.2 Dynamic Feature

The forging of the static image of the signature on a paper is easy. The forgers can reproduce the image (or shape) of a signature, but it is difficult to forge the motions that caused the image [85]. When a signature is captured with a signature tablet, the pen motions, which is dynamic in nature, are recorded. The features that are invariant with respect to slight

changes in intra-class signatures are used to discriminate other signature's classes. The dynamic features is broadly classified as local and global. Global features refer to the parameters extracted from a complete signature signal, such as average writing speed, total signing duration, number of pen-ups, number of strokes and standard deviation of the velocity and acceleration. Whereas, local features analyse signatures based on specific sampling points, such as the slope of the tangent at each point, velocity, the centre of mass and average speed within a stroke. A suitable set of global dynamic features [214] which is found to be reliable are used in our approach. Such features are simple to compute with a minimum pre-processing effort, and there is no need to maintain the original signatures once the features are extracted. The values in the output stream produced by the signature tablet are equidistant in time. It contains the x and y coordinates sampled at timestamp t and is represented as x(t) and y(t), respectively. At each sample point, the signature data as $S(t) = [x(t), y(t), \text{timestamp}(t)]$, $t = 1, \dots, N$, where N is total the number of samples of the signature trajectory along with the timestamp.

The explanation of dynamic features such as X-stroke, Y-stroke, average of velocity over all coordinates in the X-plane (S_{vx}) and average of velocity over all coordinates in the Y-plane (S_{vy}) are given in Eq.(6.6) and (6.7) respectively.

Average velocity in the X plane

$$S_{vx} = \frac{1}{N} \sum_{i=1}^{N-1} ((x_{i+1} - x_i) / (t_{i+1} - t_i)) \quad (6.6)$$

Average velocity in the Y plane

$$S_{vy} = \frac{1}{N} \sum_{i=1}^{N-1} ((y_{i+1} - y_i) / (t_{i+1} - t_i)) \quad (6.7)$$

6.2.4 Classification

Mahalanobis Distance (MD) computed based on correlation between two signatures verifies the similarity of images. When the Mahalanobis distance between the feature vectors of enrolled and test signature is smaller, the similarity between the compared signatures is higher. The signature raw matching score (s_{sg}) is the MD between two signatures. s_{sg} is computed as in Eq.(6.8).

$$S_{sg}(\bar{x}, \bar{y}) = \sqrt{(\bar{x} - \bar{y})^T S^{-1} (\bar{x} - \bar{y})} \quad (6.8)$$

where S is the covariance matrix, \bar{x} and \bar{y} denote the enrolled and test feature vector respectively..

6.3 Enrolment and Identification

Twelve signature samples each collected from hundred users forms the data set. Ten signature images of hundred users were used for training in enrolment module. Three types of impostors are equally likely in the case of signature identification, namely, unregistered, skilled forgery and random forgery signature images. The unregistered users are those who produce their own signature but not enrolled. Producing signature image of another person with and without prior knowledge is called skilled and random forgery respectively.

Additional hundred signature samples of skilled and random forgery were collected. A total of four hundred signature samples were used for testing the system. Out of four hundred, two hundred samples (100x2) from registered users, fifty each from unregistered, skilled, unskilled and random forgery were used. Thus, equal probability of all possible impostor attempts

was considered. The details of the number of signature images used at various stages are given in Table 6.1.

Table 6.1 Details of the number of signature images

Type of users	Number of samples	
	Training	Testing
Registered (100 persons)	100x 10 = 1000	100 x 2=200
Unregistered		50 x 1=50
• Skilled forgery	--	50 x 1=50
• Unskilled forgery	--	50 x 1=50
• Random forgery	--	50 x 1=50
TOTAL	1000	400

6.4 Results and Discussions

The effect of static and dynamic features extracted from the signature are illustrated in the following sections

6.4.1 Effect of static features

The Gabor wavelet transform was used to find the feature vector of signature images. The feature vector at various points in the virtual grid on the signature image was computed.

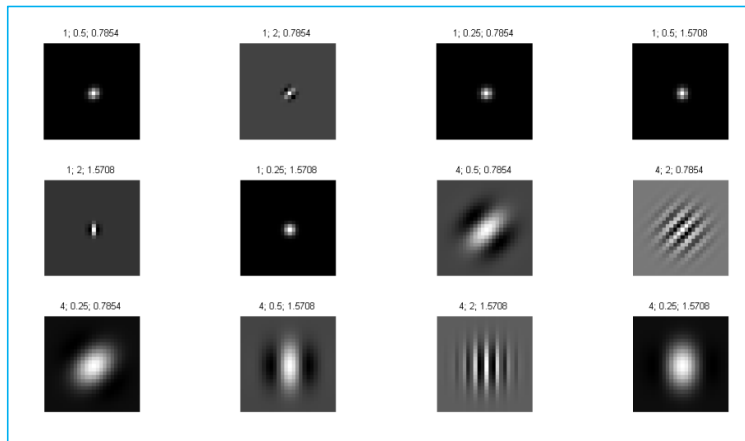


Figure 6.4 The real and imaginary part of Gabor Filters

The real and imaginary parts of gabor filters at various frequencies and orientations are shown in Figure 6.4 and Figure 6.5. The amplitude and phase response of a particular sample signature are shown in Figure 6.6 and Figure 6.7.

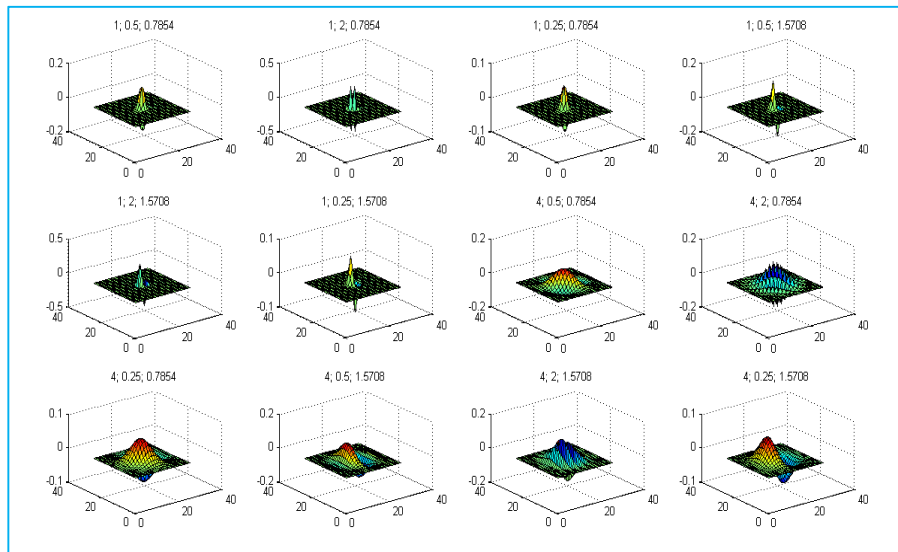


Figure 6.5 Gabor Filters



Figure 6.6 Amplitude Response of Gabor features

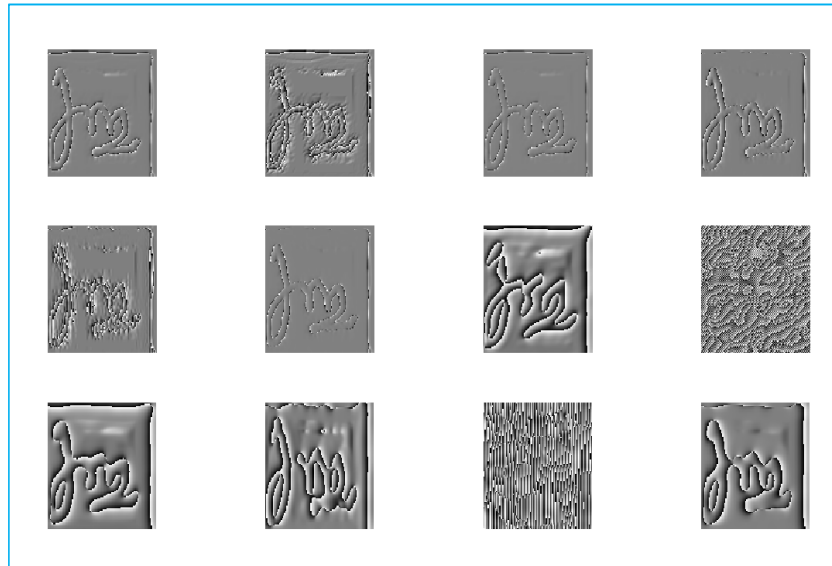


Figure 6.7 Phase Response of Gabor filters

The optimum number of grid points was identified by analysing the overall performance of the system in terms of Equal Error Rate (EER). The EER is defined as the rate at which FRR is equal to FAR, which corresponds to an optimum value of threshold for a given system. The normalized boxes of size 8 x 16, 16 x 32, 24 x 48, 32 x 64, 40 x 80, 48 x 96, 56 x 112 and 64 x 128 pixels were considered. The threshold factor in decision making for Mahalanobis Distance (MD) was varied between 0 and 3 in increments of 0.05.

The magnitudes of false rejection rate (FRR) and false acceptance rate (FAR) change against threshold for different grid sizes of bounding box were computed and are shown in Figure 6.8. As threshold increases, FRR reduces and FAR increases. The EER, which is the crossing point of FRR and FAR, was determined. The variation of EER corresponding to the size of the normalization box is presented in Figure 6.9. It is observed that EER decreases with increase in size of the grid. The variation in EER is found to be small when the grid is greater than 32 x 64. The average

verification time corresponding to different grid sizes was determined and is demonstrated in Figure 6.10. It is observed that the increase in average verification time of the system is significant when the size of the grid is greater than 32 x 64 pixels. Hence, the grid of size 32 x 64 pixels was used for this study.

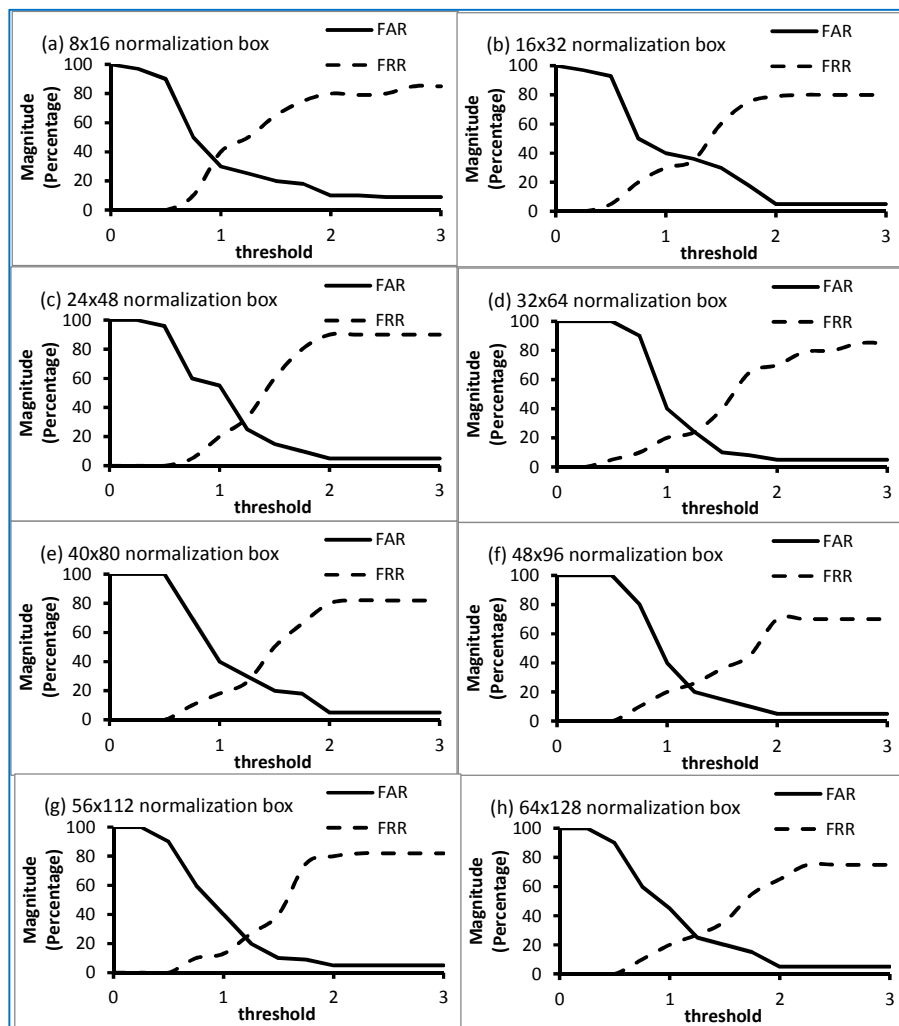


Figure 6.8 Influence of the size of the normalisation box on false rejection and acceptance rates against threshold.

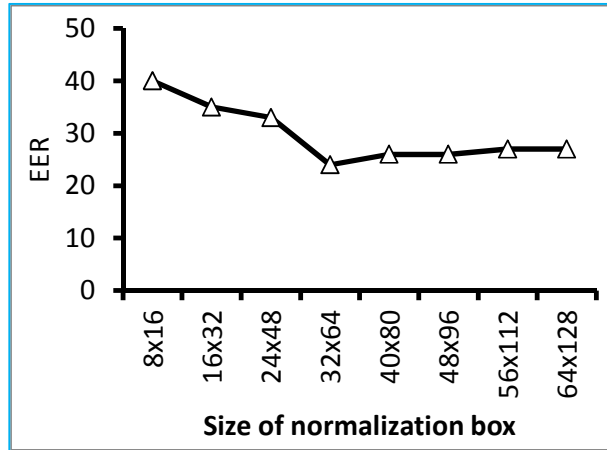


Figure 6.9 Equal error rate against size of the normalization box.

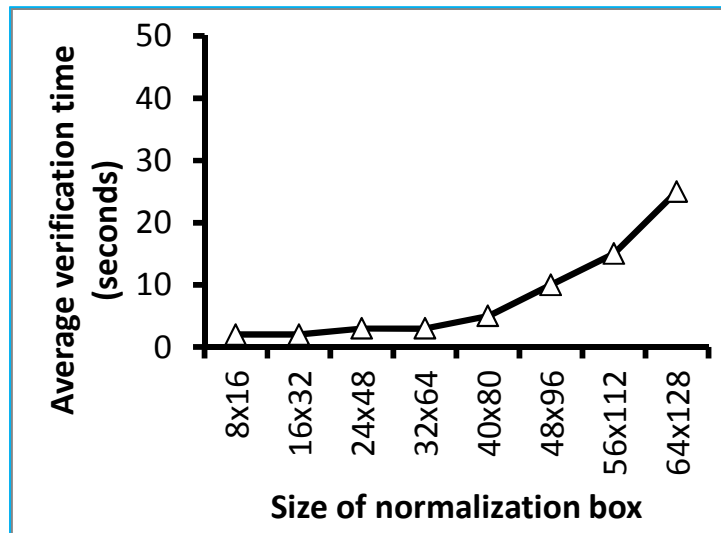


Figure 6.10 Average verification time versus size of the normalization box.

6.4.2 Effect of dynamic features

A sample of a dynamic signature, its x and y plot with respect to time as well as its normalised x versus time plot are shown in Figure 6.11. The success rate of the system with static features alone is 76 percent and

by the addition of dynamic features, the success rate is increased to 84 percent as shown in Figure 6.12.

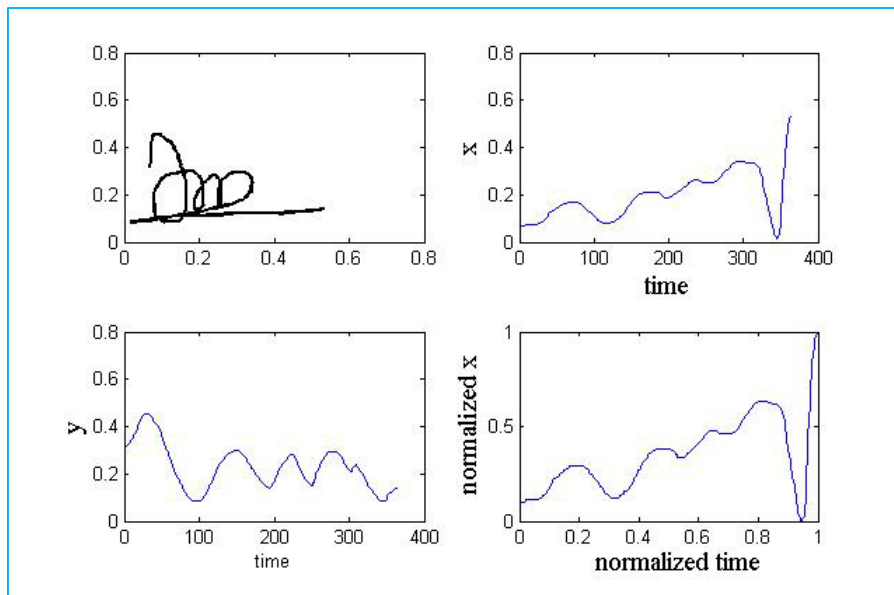


Figure 6.11 Example of a signature, its x and y plot as well as its normalised x plot against time.

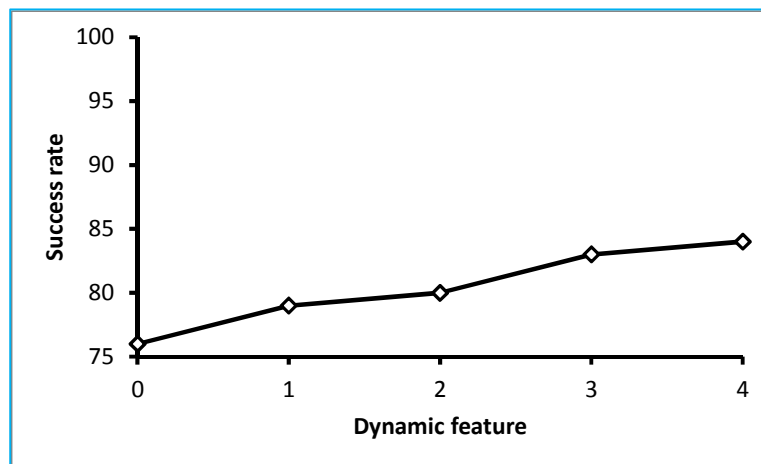


Figure 6.12 Effect of dynamic feature on the success rate.

6.5 Summary

A simple and effective signature identification method using grid based GWT and dynamic features is proposed. The performance of the signature identification is investigated. The optimum size of the normalised grid box is determined based on the analysis of equal error rate (EER) and average verification time. Three types of impostors, namely unregistered, skilled forged and random forged users, were accounted in the identification process. It is found that by including the dynamic features along with static features of the signature, the success rate is increased by 8 percent. The test results demonstrate the suitability and effectiveness of combining the grid based Gabor Wavelet Transform approach and dynamic features in a signature biometric system.

CHAPTER 7

IRIS RECOGNITION SYSTEM

This chapter, brings out the importance of iris in recognition systems as iris patterns are believed to be unique due to the complexity of the underlying environmental and genetic processes that influence its generation. These factors result in extraordinary textural patterns that are unique to each eye of an individual and are even distinct between twins. The pre-processing stage requires localization of the iris which generally involves the detection of the edge of the iris as well as that of the pupil. Since varying levels of illumination can result in dimensional inconsistencies between eye images due to the stretching of the iris, normalization needs to be performed so that iris region is transformed to have fixed dimensions. After unwrapping the normalized iris region into a rectangular region, it is encoded using Haar wavelets to generate the iris code. It was also observed that the Canny operator is best suited to extract most of the edges to generate the iris code for comparison. In the recognition stage, Hamming distance is used for the comparison of the iris code, the most discriminating feature of the iris pattern, with the existing iris templates.

7.1 Introduction

Among biometric technologies, iris-based authentication systems possess more advantages than other biometric technologies. Iris offers an

excellent recognition performance when used as a biometric. Iris patterns are unique due to the complexity of the underlying environmental and genetic processes that influence the generation of an iris pattern. These factors result in extraordinary textural patterns that are unique to each eye of an individual and even distinct between twins [110].

Iris is a delicate circular diaphragm that lies between the cornea and the lens of the human eye. The human iris pattern varies among different individuals. The iris is one of the most stable biometric [2], [4], [211] as it does not alter significantly during a person's lifetime. While Iris recognition is the most precise personal identification biometric, compared to other biometrics, such as fingerprints and face, iris-based authentication has a fairly short history of use. The idea of an automatic iris authentication procedure was conceptualized and patented by Flom and Safir in 1987 [215]. Most of the common approaches reported in the literature are based on iris code and integral-differential operators suggested by Daugman [216], [217].

7.1.1 Iris Anatomy

Iris is the coloured ring of tissue around the pupil through which light enters the interior of the eye [217]. It is located in front of the crystalline lens, and divides the anterior aqueous into the anterior and posterior chambers. The pigmented fibro vascular tissue known as stroma characterizes the iris. It's role is to help in regulating the amount of light that enters the eye. The iris made up of smooth muscle fibres known as sphincter and dilator, adjust the pupil size with the purpose of controlling the amount of light passing through the pupil. The sclera often referred to as white or white of the eye, is the outer white coat of connective tissue and blood vessels that surround the iris. Together with internal fluid pressure, it

maintains the eye shape and cares for its delicate internal components [218].

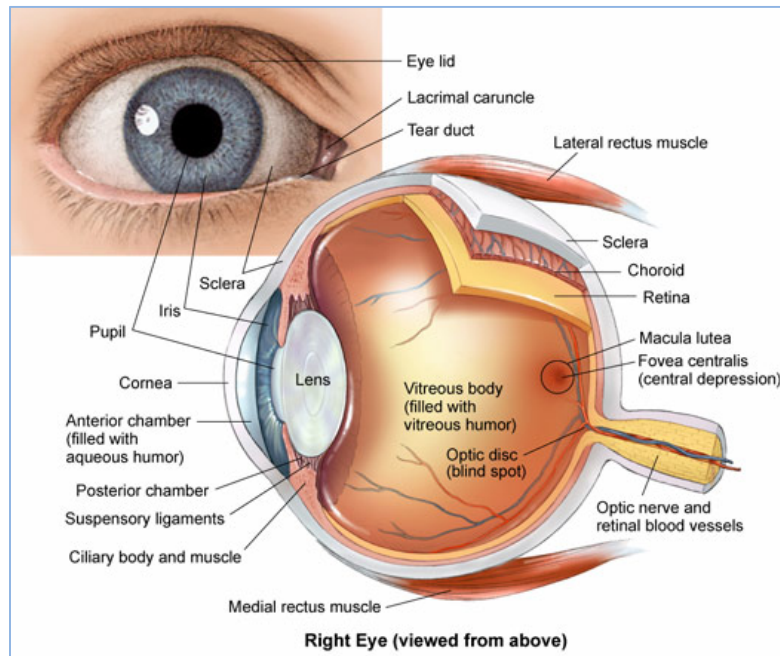


Figure 7.1 Diagrammatic view of anatomy of an eye adapted from [218].

A curved band of strong, clear tissue called the cornea covers the surface of the eye. It is the most powerful lens in the human eye's optical system. The cornea is a transparent window of the eye through which light passes. The transparency of the cornea is because, unlike most tissues in the body, it does not contain any blood vessels. However, the cornea receives its nourishment from the tears and aqueous humor in the chamber behind it. Figure 7.1 shows the anatomy of the eye. Iris naturally has a rich, distinctive and complex pattern of crypts, furrows, arching, collarette and pigment spots [217]. Each human iris has a distinctive texture which is believed to be determined randomly during the embryonic development of the eye [110]. Although iris colour can change based on the levels of

melanin concentration and distribution within the iris stroma, the appearance of an iris is relatively constant for most of a human's lifetime [219].

7.1.2 Iris as a Biometric

Iris recognition, a reliable method for identity authentication, plays an important role in many mission-critical applications such as access control and border checkpoints due to several reasons [113]:

- Iris is an internal organ of the eye, physically protected from external environment by the cornea. This makes it more consistent than fingerprints that are more susceptible to worn out due to age or manual labour.
- Iris starts to develop in the third month of gestation and the structures creating its pattern reach completion mainly by the eighth month [219] and does not vary throughout one's lifetime. Furthermore, the formation of iris depends on the initial environment of the embryo. Therefore, the texture patterns of the iris do not correlate with genetic determination. Consequently, even irises of genetically identical twins are extremely distinct. Actually, the left and the right irises of the same person are unique [220].
- Since iris is stable, iris-based technologies have demonstrated high levels of performance [221]. Moreover, surgical modification of the pigmentation and/or colour of the iris without unacceptable risk to damage the vision is also impossible.
- The physiological reaction of the iris to light sources provides one of the easiest liveness detection practices against spoofing attacks.
- Glasses or contact lenses rarely hinder Iris recognition efficacy [222]. In addition, the non-contact acquisition procedure used in

capturing iris images makes it more convenient than fingerprints, which mostly use optical touch based sensors.

- Among biometrics, iris has one of the smallest outlier populations, where few people cannot use or enrol using this technology.

Despite the aforementioned advantages of using iris recognition, the acquisition of satisfactory quality iris images for iris recognition is a critical yet challenging step [111]. It may act very poorly when deployed in real-time applications, especially for recognition at a distance. Besides, the iris is usually located at the back of a curved and reflecting surface, typically covered by eyelashes and partially occluded by eyelids.

7.2 Iris recognition system

Since the beginning of the iris recognition research, many different iris recognition systems have been developed [223], [224]. Perhaps the most successful and most well-known iris recognition algorithms, on which the state-of-the-art systems are based, are the algorithms developed by Professor John Daugman. The main stages of any typical iris recognition system include iris pre-processing, feature extraction and classification. Figure 7.2 illustrates the key phases of an iris recognition system based on the approach of Daugman [225].

The initial stage involves iris localization, iris normalization and image enhancement. The first step consists of localizing the iris area between the inner (pupillary) and the outer (limbic) boundaries, with the prior assumption that each border is either circular or elliptical. This process also obliges detection and removal of any specular reflection as well as eyelash or eyelid noise from the image prior to segmentation. In order to overcome the differences in the pupil size and the acquired images as well as to ensure consistency between eye images is ensured using

Daugman's Rubber Sheet method, which maps the original segmented iris region into a fixed length and dimensionless pseudo-polar coordinate system. The next step is to extract distinctive features from the iris texture pattern, with the intention of comparisons between templates.

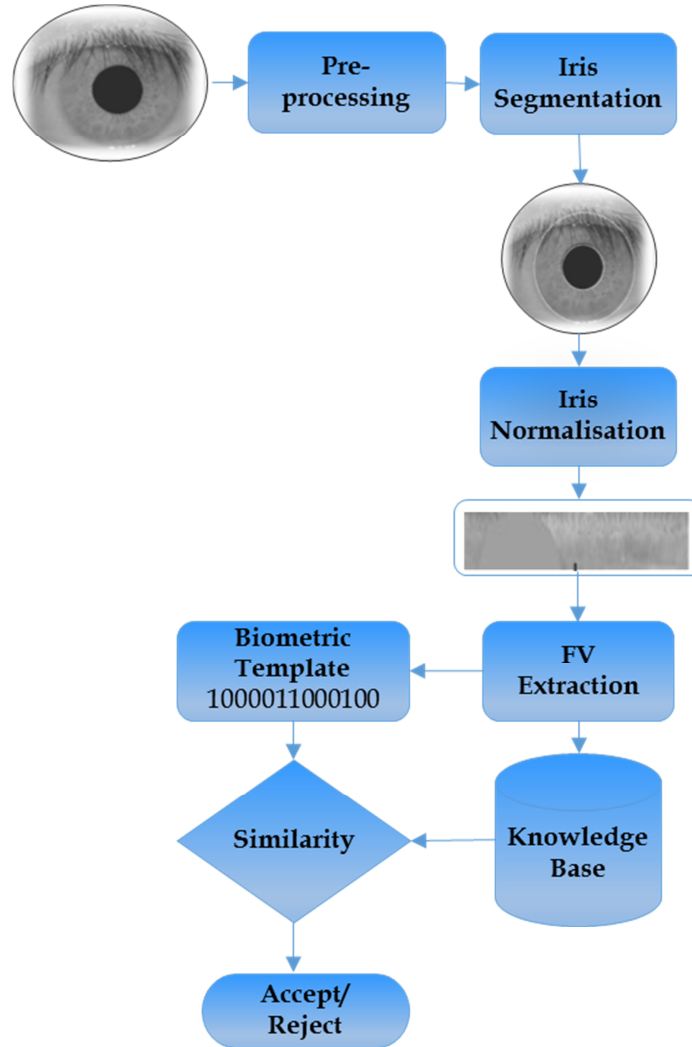


Figure 7.2 Block diagram for an Iris Recognition System

As stated before, segmentation plays a crucial role in the overall performance of the iris recognition system. The following sub-sections,

describes the proposed technique, starting with the detection of pupil as well as iris boundary regions, the isolation of eyelids together with eyelashes, feature extraction and finally matching the processed iris pattern with those in the knowledge base.

7.2.1 Iris Pre-processing

The primary step in iris pre-processing is to distinguish the iris texture from the input eye image. The first step in any iris recognition system is to localize the iris area between the inner (pupillary) and the outer (limbic) boundaries, usually with a prior assumption that each border is either circular or elliptical. Researchers have proposed different algorithms for iris detection [111],[112],[226],[227]. This process also obliges detection and removal of any specular reflections of illumination as well as eyelash or eyelid occlusions from the image prior to segmentation. Segmentation plays an essential role in the overall success of any iris recognition process, as image parts that are incorrectly considered as iris pattern data will eventually lead to poor recognition rates.

7.2.2 Iris localization or segmentation

Implementation of an iris biometric system needs the iris region to be isolated from other parts of the image. A ring defined by the iris/sclera (limbic) boundary and the iris/pupil (pupillary) boundary approximates the iris region. In iris localization, a technique is required to locate and isolate the iris region as well as to remove the eyelids and eyelashes [228]. The primary step in any iris recognition system is to localize the iris area between the inner and outer boundaries. The key steps [226] involved are Dimension Reduction, Pupil and Outer iris localization, Eyelids and Eyelashes as well as Boundary Detection. The different steps involved in iris segmentation are shown in Figure 7.3.

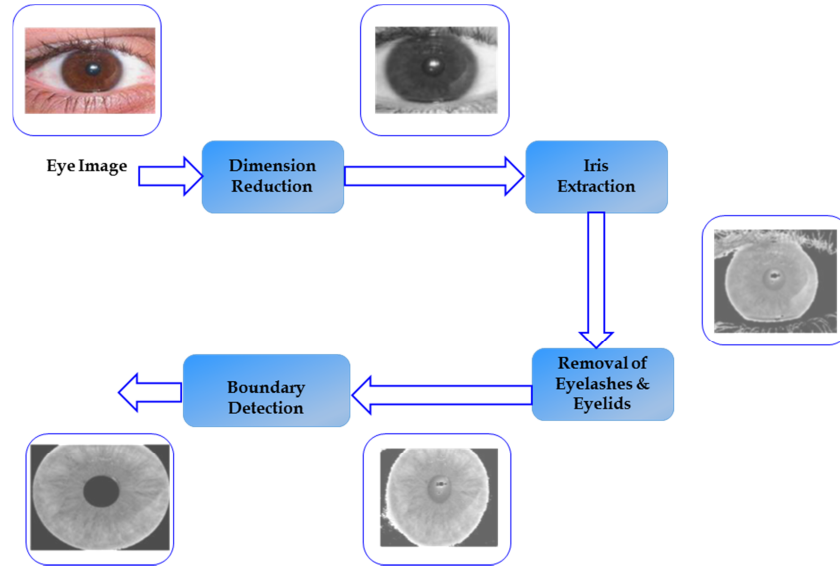


Figure 7.3 Iris Segmentation Stages

7.2.2.1 Dimension Reduction and Iris Extraction

The database used for the implementation had the image dimensions as 600 x 800 with coloured JPEG of bit depth 24, where 24 bits represent each pixel of the image. Therefore, to reduce the computational complexity, the iris image is first converted into grayscale reducing the bit depth to 8 and since it is the only region of interest the remaining region is removed. In the grayscale image, summation of the pixel values in the iris region when compared to that in the other regions will be less and so for extraction of the iris region, threshold technique based on the colour of the iris is used. The gray level values of the pixels for a dark iris will be lesser compared to that of a light iris. Two threshold values can be set to determine the iris region using the histogram of the eye image.

7.2.2.2 Removal of eyelids and eyelashes

After extracting iris from the input image, the unwanted information, such as eyelashes and eyelids, needs to be removed from it.

The Sobel operator [229], which performs a 2D spatial gradient measurement on an image, helps in detection of the edges of eyelids and eyelashes. An edge is characterised by a noticeable change in the intensity and the Sobel operator returns edges at those points where there is maximum gradient of the image. The two 3 x 3 masks G_x and G_y of the operator, are given in Eq.(7.1) and (7.2) respectively.

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (7.1)$$

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (7.2)$$

Applying the masks to the image and computing its magnitude as in Eq.(7.3), approximates the gradient.

$$G[I(i, j)] = |G_x| + |G_y| \quad (7.3)$$

To compute the gradient for the pixel $P_i(i, j)$ of the input image, I is given in Eq.(7.4).

$$I(i, j) = \begin{bmatrix} I(i-1, j-1) & I(i-1, j) & I(i-1, j+1) \\ I(i, j-1) & I(i, j) & I(i, j+1) \\ I(i+1, j-1) & I(i+1, j) & I(i+1, j+1) \end{bmatrix} \quad (7.4)$$

Eq.s (7.5) and (7.6) evaluates G_x and G_y respectively. Determination of the Gradient magnitude at each pixel and comparison with some threshold determines whether it is an edge pixel or not. Sobel operator is less sensitive to noise due to its large convolution masks.

$$G_x(i, j) = \begin{cases} [I(i-1, j+1) + 2 * I(i, j+1) + I(i+1, j+1)] \\ -[I(i-1, j-1) + 2 * I(i, j-1) + I(i+1, j-1)] \end{cases} \quad (7.5)$$

$$G_y(i, j) = \begin{cases} [I(i-1, j-1) + 2 * I(i-1, j) + I(i-1, j+1)] \\ -[I(i+1, j-1) + 2 * I(i+1, j) + I(i+1, j+1)] \end{cases} \quad (7.6)$$

7.2.2.3 Boundary Detection

For boundary detection, the centre pixel of the eyelash removed iris image is located and based on the centre co-ordinates of the pupil a circular strip is extracted. For detecting the inner and outer boundaries ($B_{in\ or\ out}$) of the iris, Integro-Differential operator [216], as defined in Eq.(7.7) is used.

$$B_{in\ or\ out} = \max_{(r, x_0, y_0)} \left| G_\sigma(r) * \frac{\partial}{\partial r} \oint_{r, x_0, y_0} \frac{I(x, y)}{2\pi r} ds \right| \quad (7.7)$$

where $I(x, y)$ is the original grayscale eye image. The parameter (r, x_0, y_0) represents a circle of radius r with centre coordinates (x_0, y_0) . The symbol $*$ denotes convolution and $G_\sigma(r)$ is a radial smoothing Gaussian function with centre r and standard deviation σ , defined as in Eq.(7.8).

$$G_\sigma(r) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(r-r_0)^2}{2\sigma^2}} \quad (7.8)$$

The assumption that both pupil and iris have circular boundaries justifies the usage of Gaussian filter for smoothing and integration operation along the iris circle. This method tries to find a circle in the eye image with maximum change in gray level difference with its neighbours. Due to significant contrast between iris and pupil regions, initially localization of the pupil boundary and then, using the same operator with difference radius and parameters, detection of the outer boundary is carried out. Applying the operator iteratively attains precise localization.

7.2.3 Iris Normalization and Unwrapping

On successful segmentation of iris regions from an eye images, the resulting iris regions will have dimensional inconsistencies, mainly due to stretching of the iris caused by pupil dilation from varying levels of illumination. The normalization process will produce iris regions, which have the same constant dimensions, so that two photographs of the same iris under different conditions will have the same characteristic features.

In fact, the homogenous rubber sheet model devised by Daugman remaps each point (x, y) within the iris region to a pair of polar coordinates (r, θ) , where r is in the interval $(0, 1)$ and θ is the angle $(0, 2\pi)$. The normalized iris region is then unwrapped into a rectangular region. Figure 7.4 illustrates the mechanism of Daugman's rubber sheet model.

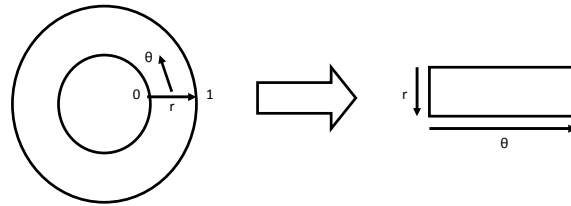


Figure 7.4 Unwrapping: Daugman's Rubber Sheet Model

The normalized remapping of iris region from Cartesian coordinates (x, y) to non-concentric polar representation is given by Eq.(7.9).

$$I(\mathbf{x}_{(r,\theta)}, \mathbf{y}_{(r,\theta)}) \rightarrow I_{(r,\theta)} \quad (7.9)$$

$$\text{where } \mathbf{x}_{(r,\theta)} = (\mathbf{1} - r) \mathbf{x}_p(\theta) + r \mathbf{x}_l(\theta)$$

$$\mathbf{y}_{(r,\theta)} = (\mathbf{1} - r) \mathbf{y}_p(\theta) + r \mathbf{y}_l(\theta)$$

where $I(x, y)$ is the iris region image, (x, y) is the original Cartesian co-ordinate, (r, θ) is the corresponding normalized polar co-ordinate,

(x_p, y_p) and (x_l, y_l) are the pupil and iris boundary respectively along the θ direction, as represented in equations from (7.10) to(7.13).

$$x_p(\theta) = x_{p0}(\theta) + r_p \text{Cos}\theta \quad (7.10)$$

$$y_p(\theta) = y_{p0}(\theta) + r_p \text{Sin}\theta \quad (7.11)$$

$$x_l(\theta) = x_{l0}(\theta) + r_l \text{Cos}\theta \quad (7.12)$$

$$y_l(\theta) = y_{l0}(\theta) + r_l \text{Sin}\theta \quad (7.13)$$

After getting the normalized polar representation of the iris region, this region is unwrapped by choosing a constant number of points along each radial line, irrespective of how much narrow or wide the radius is at a particular angle. The 2D array produced will have vertical dimensions as radial resolution and horizontal dimensions as angular resolution. Figure 7.5 shows the unwrapped normalized iris image.

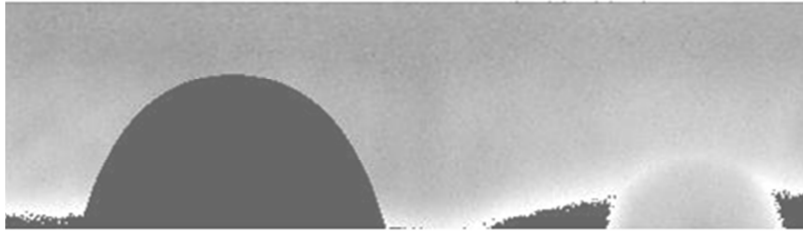


Figure 7.5 Unwrapped Normalized iris image

7.2.4 Feature Extraction

In order to provide an accurate recognition of individual, though the most discriminating information present in an iris pattern is extracted, only the significant features of the iris need to be encoded. In the feature extraction stage, histogram equalization is done initially to enhance the iris texture in the normalized image followed by extraction of iris texture using the canny edge detector [230]. For dimensionality reduction of the resulting 2D image, it is converted into a 1D energy signal by Vertical projection. Discrete wavelet transform applied to the resulting 1D energy signal results

in a set of low frequency and high frequency coefficients. High frequency coefficients can be discarded due to the lack of significant information while iris templates are selected from the low frequency coefficients, each of which has a dimension of 64 bytes. Figure 7.6 shows the different steps involved in the feature extraction stage.

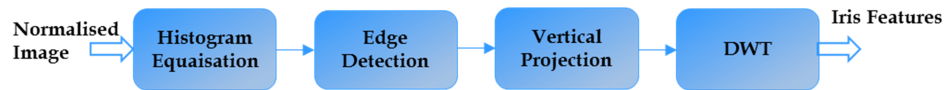


Figure 7.6 Feature Extraction Stages

7.2.4.1 Histogram Equalization

The intensities are better distributed on the histogram through a process known as histogram equalization, allowing for areas of lower local contrast to gain a higher contrast without affecting the global contrast. This is accomplished by effectively spreading out the most frequent intensity values in the image. Figure 7.7 shows the image with enhanced iris texture obtained after histogram equalization. The domes in the unwrapped image are due to the eyelid occlusion.

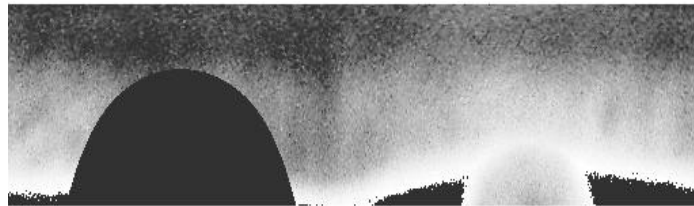


Figure 7.7 Histogram Equalized image

7.2.4.2 Edge Detection

Edge detection is performed to extract the iris texture from the histogram equalized image using available techniques such as Sobel,

Canny, Prewitt etc. It is observed that Canny edge detection technique is able to extract most of the iris texture from the enhanced image.

The Canny operator, an optimal edge detector, takes grayscale image as input producing an image showing the positions of tracked intensity discontinuities. The Canny operator works in a multi-stage process. After smoothing the image by Gaussian convolution, a simple 2-D first derivative operator (somewhat like the Roberts Cross) is applied to highlight regions of the image with high first spatial derivatives. Edges give rise to ridges in the gradient magnitude image. The algorithm then tracks along the top of these ridges and sets to zero all pixels that are not actually on the ridge top to give a thin line in the output, a process known as non-maximal suppression. The tracking process exhibits hysteresis controlled by two thresholds: T_1 and T_2 , with $T_1 > T_2$. Tracking can begin only at a point on a ridge higher than T_1 and it continues in both directions out from that point until the height of the ridge falls below T_2 . This hysteresis helps to ensure that noisy edges are not broken up into multiple edge fragments. The effect of the Canny operator is determined by three parameters - the width of the Gaussian kernel used in the smoothing phase and the upper as well as lower thresholds used by the tracker. Increasing the width of the Gaussian kernel reduces the detector's sensitivity to noise, at the expense of losing some of the finer details in the image. The localization error in the detected edges also increases slightly as the Gaussian width is increased. Usually, the upper tracking threshold can be set quite high and the lower threshold quite low for good results. Setting the lower threshold too high will cause noisy edges to break up while setting the upper threshold too low increases the number of spurious and undesirable edge fragments appearing in the output. The Gaussian smoothing in the Canny edge detector fulfils

two purposes: first, it can be used to control the amount of detail that appears in the edge image and second, it can be used to suppress the noise.

The Canny method [230] locates edges by looking for local maxima of the gradient of $I(i, j)$. The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds to detect strong as well as weak edges, and the weak edges are included in the output only if they are connected to strong edges. This method is therefore less likely to be fooled by noise compared to the others and more likely to detect true weak edges. Figure 7.8 shows the edge-detected image using canny operator.

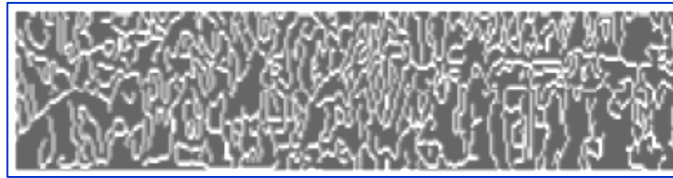


Figure 7.8 Canny edge-detected image

7.2.4.3 Vertical Projection

Vertical projection is a method used to convert the 2D signal to a 1D signal in order to reduce the system complexity. For vertical projection, energy of each row of the edge-detected image is calculated and is transformed into a row vector. The generalized form is given in Eq.(7.14).

The dimension of the normalized image is $m \times n$ and is taken as 128×512 . Hence, after vertical projection, its dimension becomes m , which is equal to 128.

$$\begin{bmatrix} x_{11} & \dots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \dots & x_{mn} \end{bmatrix} \rightarrow \left[\sum_{i=1}^n |x_{1i}|^2 \quad \dots \quad \sum_{i=1}^n |x_{mi}|^2 \right] \quad (7.14)$$

7.2.4.4 Discrete Wavelet Transform

At low data rates, the Discrete Cosine Transform (DCT), suffer from a *blocking effect* due to the unnatural block partition that is required in the computation. Other drawbacks include mosquito noise, a distortion that appears as random aliasing occurring close to the object's edges, and aliasing distortions. Due to the shortcomings of DCT, discrete wavelet transform (DWT) has become increasingly important. The main advantage of DWT is that it provides space-frequency decomposition of images, in comparison to the DCT and Fourier transform that only provide frequency decomposition. By providing space-frequency decomposition, the DWT allows energy compaction at the low-frequency sub-bands and the space localization of edges at the high-frequency sub-bands. Furthermore, the DWT does not present a blocking effect at low data rates. Wavelets are functions that integrate to zero, waving above and below the x-axis. Like sine and cosine in the Fourier transform, wavelets are used as the basis functions for signal and image representation. Such basis functions are obtained by dilating and translating a mother wavelet $c(x)$ by amounts s and τ , respectively as given in Eq.(7.15).

$$\Psi_{\tau,s}(x) = \left\{ \Psi \left(\frac{x-\tau}{s} \right), (\tau, s) \in R \times R^+ \right\} \quad (7.15)$$

The translation τ and dilation s allow the localization of wavelet transform in time and frequency. The discrete wavelet transform (DWT) decomposes the signal into mutually orthogonal set of wavelets [113]. The DWT of signal x is calculated by passing it through a series of filters. The samples are passed through a low pass filter with impulse response g resulting in a convolution as given in Eq.(7.16) and is simultaneously decomposed using a high pass filter h .

$$Y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k] g[n - k] \quad (7.16)$$

The 64 byte detail coefficients from the high pass filter and the 64 byte approximation coefficients from the low pass filter form the 1D signal of 128 bytes. The two filters are related to each other and are together known as a quadrature mirror filter. The Haar DWT [231],[232], the simplest of all wavelets have simpler boundary conditions, is orthogonal as well as symmetric and their minimum support property allows arbitrary grid intervals. In the proposed work Haar DWT is used since it works efficiently to detect the characteristics such as edges and corners. By using the quadrature mirror filter, wavelet coefficients from an $n \times n$ image I are computed as discussed below.

Filters H and G are applied on the rows of an image, splitting the image into two sub images of dimensions $n/2 \times n$ (half the columns) each. One of these sub images, $H_r I$ (where the subscript r denotes row), contains the low-pass information and the other, $G_r I$, contains the high-pass information. Next, the filters H and G are applied to the columns of both the sub images. Finally, four sub images with dimensions $n/2 \times n/2$ are obtained. Sub images $H_c H_r I$, $G_c H_r I$, $H_c G_r I$, and $G_c G_r I$ (where the subscript c denotes column) contain the low-low, high-low, low-high and high-high pass information, respectively. The same procedures are applied iteratively to the sub image containing the most low band information until the sub image's size reaches 1×1 . Therefore, the initial dimensions of the image are required to be powers of 2. In practice, it is not necessary to carry out all the possible decompositions until the size of 1×1 is reached. Usually, just a few levels are sufficient. After wavelet transform, a set of low frequency coefficients and high frequency coefficients, each of dimension 64 bytes, is obtained. After DWT, it is observed that the approximation coefficients

contain information while the detailed coefficients do not hold any information. Hence, approximation coefficients, with a dimension of 64 bytes, are selected as the feature vector and stored in the database. Since the wavelet functions are compact, the wavelet coefficients only measure the variations around a small region of the data array. This property makes wavelet analysis very useful in signal or image processing.

7.2.5 Matching

In the recognition stage, the features of the input eye image are compared with those of the images already stored in the database and if they match, the corresponding eye image is identified, otherwise it remains unidentified. Since a bitwise comparison is necessary, Hamming distance is chosen for identification.

7.2.5.1 Hamming Distance

The Hamming distance [111] gives a measure of how many bits are the same between two patterns. It is used for the comparison of iris templates in the recognition stage. Hamming distance D is given by Eq. (7.17).

$$D = \frac{1}{n} \sum_{k=1}^n x_k \oplus y_k \quad (7.17)$$

where x and y are the two bit patterns of the iris code while n indicates the number of bits. Hamming distance D gives out the number of disagreeing bits between x and y .

Ideally, the hamming distance between two iris codes generated for the same iris pattern should be zero. However, this will not happen in practice because normalization is not perfect. The larger the hamming distance (closer to 1), the more probable the two patterns are different

whereas the smaller the hamming distance (closer to 0) the more probable the two patterns are identical. By properly choosing the threshold upon which the matching decision is made, good iris recognition results with very low error probability can be obtained.

7.3 Iris Database

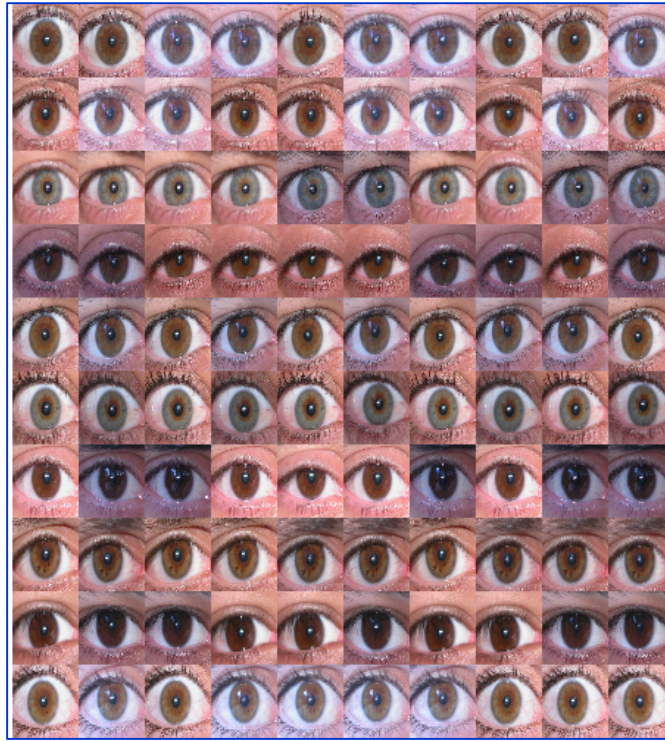


Figure 7.9 Subset of UBIRIS eye database used for training

The system has been tested against the UBIRIS database [225]. UBIRIS database includes 1877 images from 241 persons collected in two sessions. The images collected in the first photography session were low noise images. On the other hand, images collected in the second session were captured under natural luminosity factor, thus considering reflections, different contrast levels as well as luminosity and focus problems. Such

images might be a good model for realistic situations. Figure 7.9 shows a subset of the UBIRIS eye database used for training.

7.4 Results and Discussions

The data sets used for identification purpose were the hundred sets of eye images from the UBIRIS database, each set consisting of ten eye images of a person taken at different times. From each set, randomly selected ten eye images and their features were stored in the database. The 1000 images used for simulation are referred to as registered images since their features were stored in the database. The main challenge in identification was to recognize the other images in each set whose features were not stored. The unregistered images are the fifty images whose features were not stored in the database but were used to test the algorithm. The performance of iris acceptance algorithm is validated using F_1 score, precision and recall. Figure 7.10 displays iris edge detection using various edge detection operators. In Figure 7.11, it can be seen that the canny operator gives the lowest EER at a threshold of 0.4. The Table 7.1 shows the EER percent and its corresponding hamming distance.

Table 7.1 EER of different operators

Operators	EER (%)	Hamming distance
Canny	7.00	0.40
Sobel	19.25	0.45
Prewitt	16.50	0.50
Roberts	22.50	0.40
log	15.50	0.50
zerocross	14.50	0.45

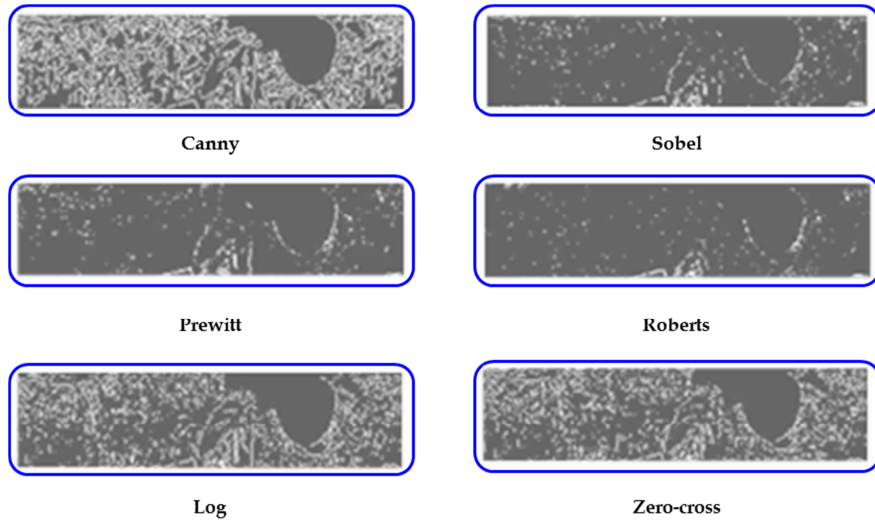


Figure 7.10 Detection of iris edge using various operators

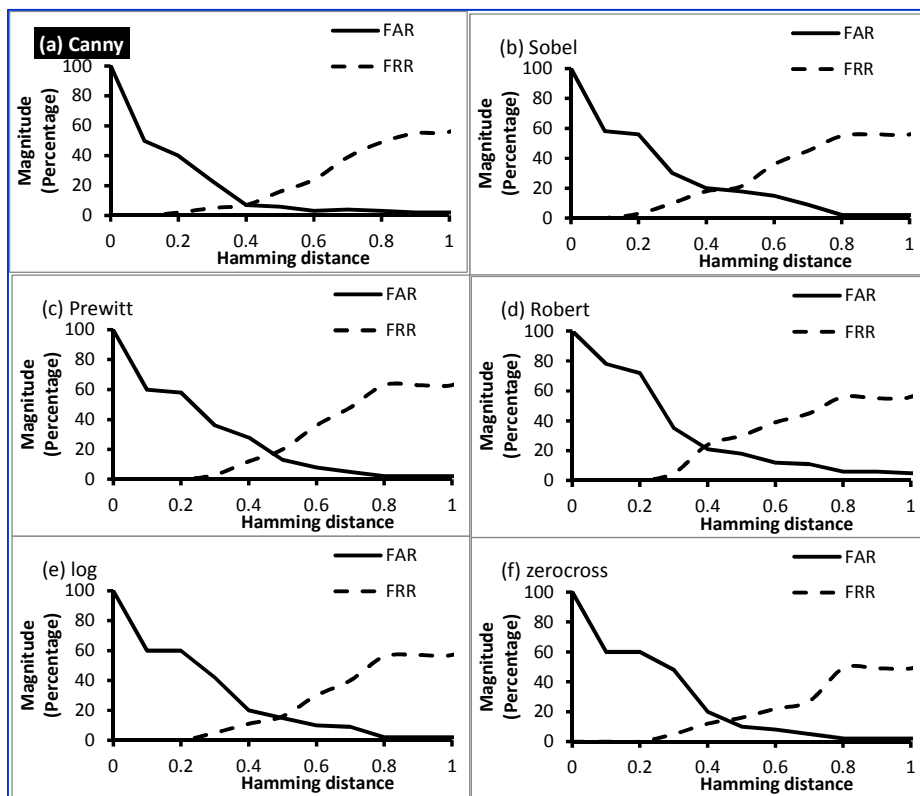


Figure 7.11 Decision making in iris biometric system.

7.4.1 F₁ Score

In pattern recognition and information retrieval with binary classification, precision, or positive predictive value, is the fraction of retrieved instances that are relevant, while recall or sensitivity is the fraction of relevant instances that are retrieved. Table 7.2 shows the condition and test outcome to calculate F1 score for performance measurement. Both precision as well as recall are therefore based on an understanding and measure of relevance. Precision can be seen as a measure of exactness or quality, whereas recall is a measure of completeness or quantity.

Table 7.2 Performance measures for F₁ score

		Condition		
		Condition positive(cp)	Condition negative(cn)	
Test outcome	Test outcome positive (top)	True positive(tp)	False positive(fp) (Type I error)	$precision = \frac{\sum tp}{\sum top}$
	Test outcome negative (ton)	False negative(fn) (Type II error)	True negative(tn)	Negative predictivity $= \frac{\sum tn}{\sum ton}$
		sensitivity $= \frac{\sum tp}{\sum cp}$	specificity $= \frac{\sum tn}{\sum cn}$	Accuracy $= \frac{\sum tp + \sum tn}{\sum Total\ population}$

In a classification task, the precision for a class is the ratio of the number of true positives to the total number of elements labelled as belonging to the positive class or say the sum of true positives and false positives. Recall is defined as the ratio of number of true positives to the total number of elements that actually belong to the positive class, say the sum of true positives and false negatives, which are items not labelled as

belonging to the positive class but should have been. In statistical analysis of binary classification, the F1 score (also F-score or F-measure) is a measure of a test's accuracy. It considers both the precision and the recall of the test to compute the score as in Eq.(7.18). Figure 7.12 shows the comparison of different edge detection operators using F₁ score.

$$F_1 = 2 \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (7.18)$$

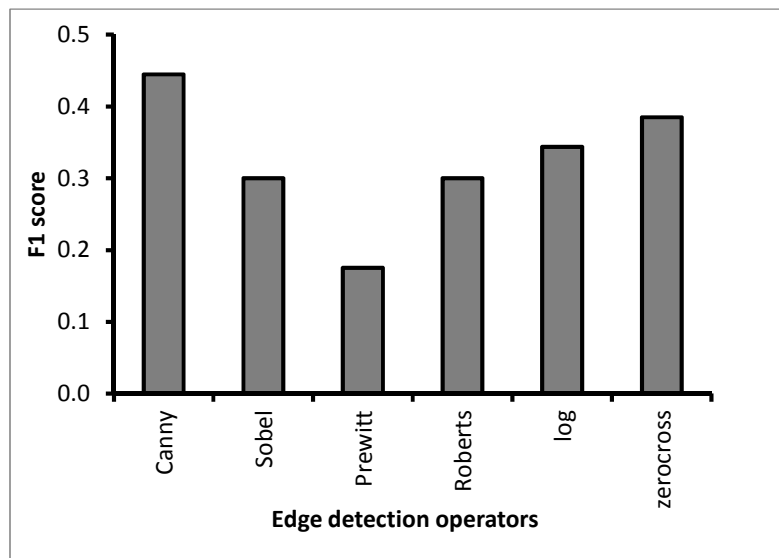


Figure 7.12 Comparison of edge detection operators

Using MATLAB, a comparison study between different classical operators like Canny, Sobel, Prewitt, Roberts, log and zero cross was done. The operators were applied to the enhanced normalized image. From the results, which reveal the performance of each of the operators, it is found that the Canny operator outperforms the others; in fact it is the only operator which is able to extract most of the iris texture. It is observed that an optimum result is obtained at a hamming distance threshold of 0.4 for the canny operator.

7.5 Summary

The iris images are segmented and projected into 1D signals by the process of vertical projection. The 1D signal features are then extracted by the Haar wavelet transform. The complexity of the feature extraction method for iris recognition is low and thus achieves a considerable computational reduction while maintaining good performance.

A low dimensional feature extraction algorithm has been developed and tested with thousand eye images from the database by varying the parameters such as dimension of the feature vector and hamming distance threshold. It is found that an optimum result is obtained at a hamming distance threshold of 0.4 and a feature vector dimension of 64 bytes. It is also observed that the Canny operator is best suited to extract most of the edges to generate the iris code for comparison. By this method, a recognition rate of 93% has been achieved.

CHAPTER 8

MULTI-MODAL BIOMETRIC SYSTEM

In this chapter, the implementation of a prototype system for user authentication using the multimodal biometrics utilizing four traits i.e., speech, face, signature and iris is discussed. The final decision is made by fusion at matching score level architecture, in which the feature vectors of query images are compared with the templates in knowledge base. Based on the proximity of the feature vector and template, each subsystem computes its own matching score. Finally, the individual traits are fused at matching score level using user dependent weighted sum of score technique and is then passed to the decision module. This system is tested on our knowledge base and the overall accuracy of the system is found to be more than 98%.

8.1 Introduction

The basic aim of pattern recognition is to devise automatic procedures that maximize the recognition performance. The comparison of different pattern recognition algorithms on the specific problem and selecting the best of them can lead to the required result. Some recognition errors committed by the best biometric trait approach may be resolved by the use of inferior trait methods. This observations motivated the interest in combining the different biometric traits and their classifiers, which results in a multi-modal biometric system. The performance of a unimodal system can drop significantly under noisy conditions while the multimodal systems

overcome these challenges by combining the evidence provided by a number of biometric traits.

8.2 Score Level Fusion

Score fusion methods can be divided into two categories, adaptive and non-adaptive fusion [233]. Non-adaptive fusion can employ fixed weights that is heuristically determined *a priori*, and place a higher weighting on the higher performing modalities. However, under adverse testing conditions, it is impossible to place a lower weighting on a degraded modality. Non-adaptive fusion can be without the use of any weights, i.e., equal weights are placed on each source of information, e.g., simple additions of scores. Adaptive fusion is required to alter the weights according to the confidence of each modality. This confidence will vary dynamically as the testing conditions for each modality vary.

8.2.1 Non-weighted Fusion

In non-weighted fusion, the classifier probabilities are fused from a purely theoretical level, without the use of weights. The commonly used fusion methodologies to combine multiple modalities at the matching scores level are the sum, product, min and max [234]. If S_i is the matching score from i^{th} modality, S represents the resulting fused score.

The **Simple Product Rule** combines the scores by multiplying all the individual scores as given in Eq.(8.1).

$$S = S_1 \times S_2 \times \dots \times S_n \quad (8.1)$$

The **Simple Sum Rule** combines the scores as a linear transformation as given in Eq.(8.2).

$$S = (a_1S_1 - b_1) + \dots + (a_nS_n - b_n) \quad (8.2)$$

where a_i and b_i represent the weight and bias, respectively, which can be specified by the user.

The **Simple Max Rule** is the maximum score from the different modalities as given in Eq.(8.3).

$$S = \text{Max} (S_1, S_2, \dots, S_n) \quad (8.3)$$

The **Simple Min Rule** is the minimum score from the different modalities as given in Eq.(8.4).

$$S = \text{Min} (S_1, S_2, \dots, S_n) \quad (8.4)$$

8.3 User-Dependent Fusion Strategy

In user-dependent fusion [235], the training scores of both authorized and unauthorized persons are employed. While in the first stage, the biometric samples of known users are employed for testing, in the second stage, the biometric samples of unknown users are used for testing. The results of the two sets of users are considered for determining the FAR and FRR of the system. The EER of the system and the corresponding threshold are determined. The general information provided by the pool of users can be exploited in user-dependent fusion schemes [236]. The three types of user dependent strategies are Global, Local and Adapted.

In global type, only the scores from the pool of users (both genuine and impostors) are used for training. In the local type, only the scores from the user at hand (both genuine and impostors) are used for training. In the adapted type, the scores from both the pool of users and the user at hand (both genuine and impostors) are used for training. In this study, adapted fusion strategy is used.

The match scores at the output of the individual trait may not be homogeneous and need not be on the same numerical range. The dissimilarity scores or similarity scores may be the outputs obtained by different traits. The statistical distribution of the match scores at the output of different traits may be different. Due to these reasons, it is necessary to normalize the scores of different traits before combining the scores.

8.3.1 Score Normalization Technique

Classifier scores can take many forms such as posteriors, likelihoods and distance measures. The scale, location and statistical distribution will vary across classifiers. Furthermore, the classifier scores may be heterogeneous, for example, a small distance measure indicates a good match, whereas a low posterior indicates a poor match; both similarity and dissimilarity scores can be given. Non-normalised scores cannot be integrated sensibly in their raw form, as it is impossible to fuse incomparable numerical scales. A score distribution with a relatively higher scale or dynamic-range will dominate the fused score. In order to combine these scores in an intelligible way at the score level, the score outputs from the various classifiers must first be normalised into a common domain. Various normalization methods exist in which essentially each method consists of changing the location and scale of the distribution. Examples include the min-max, Z-norm, decimal-scaling, Median-MAD and the *tanh* transformation.

8.3.1.1 Min-max Normalization

For normalizing the matching score the min-max normalization technique is employed in this work. Min-max normalization is a straightforward approach when compared to Z-score and tanh [233] normalization techniques. The minimum and maximum bounds of the

scores produced by a biometric trait are mapped to 0 and 1, respectively. The min-max technique is computed as in Eq. (8.5).

$$y = \frac{x - \min(S_x)}{\max(S_x) - \min(S_x)} \quad (8.5)$$

where x is the matching score before normalization, y is the matching score after normalization and S_x is the set of all possible matching scores generated by particular traits. Min-max normalization transforms all the scores into a common range [0, 1] and retains the original distribution of scores. The dissimilarity score can be transformed into similarity score by subtracting the normalized value from 1.

8.3.1.2 Decimal Scaling

If the scores of different classifiers exhibit disparate logarithmic scales e.g.[0, 1] and [0, 100], then the scores can be normalised using Eq.(8.6)

$$S'_i = \frac{S_i}{10^{\log_{10}(\max(S_i))}} \quad (8.6)$$

$S'_i \in [0,1]$, which essentially is a specific case of the min-max norm.

8.3.1.3 Z-norm

The zero-normalisation (Z-norm) method is the most common form of score normalisation. The list of N scores $\{S_i\}$, $i=1, \dots, N$ is transformed such that the new score distribution has zero mean and unity variance, i.e., the standard normal distribution. To carry out this, the location μ_s and scale σ_s parameters are calculated using Eqs. (8.7) and (8.8) respectively.

$$\mu_s = \frac{1}{N} \sum_{i=1}^N S_i \quad (8.7)$$

$$\sigma_s^2 = \frac{1}{N-1} \sum_{i=1}^N (S_i - \mu_s)^2 \quad (8.8)$$

where μ_s and σ_s are the mean and standard deviation of the pre-normalised score. Each score is normalised using Eq.(8.9).

$$S'_i = \frac{S_i - \mu_s}{\sigma_s} \quad (8.9)$$

8.3.1.4 Median-MAD

Robust statistical methods are employed to estimate the average or location parameter using the median. The median ignores the outlier score values at the distribution tails, which could be unfairly employed in skew μ_s above. The standard deviation or scale is estimated using the Median Absolute Deviation (MAD). The MAD value is calculated using $MAD = \text{median}(|S - \text{median}|)$. These form the median-MAD score normalization scheme as in Eq.(8.10).

$$S'_i = \frac{S_i - \text{median}}{MAD} \quad (8.10)$$

8.3.1.5 Tanh transformation

A hyperbolic tangent mapping is used to map a score distribution into the interval [-1, +1]; combined with a scaling and a shift results in the standard interval [0, 1]. For a specific classifier, given a list of N scores $\{S_i\}_{i=1 \dots N}$, the tangent normalised score [233] list is given in Eq.(8.11).

$$S'_i = \frac{1}{2} \left[\tanh \left(0.01 \frac{(S_i - \mu_{\tanh})}{\sigma_{\tanh}} \right) + 1 \right]; S'_i \in [0, 1] \quad (8.11)$$

where μ_{\tanh} and σ_{\tanh} respectively, are the Hampel tanh-estimator's values for the mean and standard deviation of $\{S_i\}$.

In conclusion, the basic min-max and z-norms are appropriate when the location and scale are known in advance. If these parameters are to be estimated on noisy training data, the robust *tanh* norm is more suitable. Also, different classifiers exhibit different score distributions, suggesting that a mixture of various normalization methods could be employed in a multi-modal system.

8.4 Multimodal Technique of identification

In a multimodal biometric system, various biometric traits are obtained from an individual. The use of multiple traits results in greater protection against fraudulent attacks. The different input traits are captured using separate sensors. The sensor module, feature extraction module, matching module, database and normalization module corresponding to each of the traits are different. The score of the individual trait is combined in the fusion module and passed to the decision module. The identity of the user is accepted or rejected based on the decision criteria with respect to the combined score in the decision module.

The acquisition sequence in a multimodal biometric system is the order in which the various biometric traits are obtained. In this study, the biometric trait is obtained sequentially. That means, each trait is independently acquired with a time interval between successive acquisitions and the corresponding algorithm process the acquired biometric information. The results acquired from each trait are combined

by an appropriate fusion scheme and the final decision is taken based on the combined results of all the traits. This approach improves the accuracy of the system. The mode of operation of multimodal identification is shown in Figure 8.1.

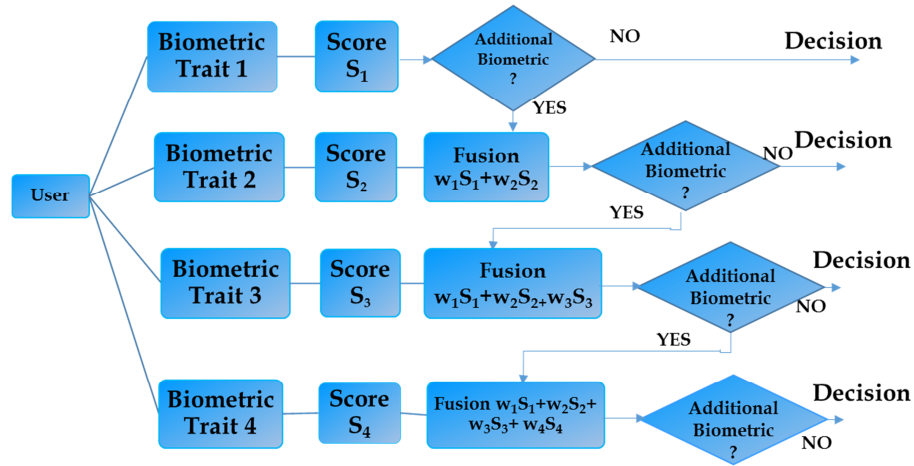


Figure 8.1 Mode of operation of a multimodal identification system

In this study, the system provides an option to select the number of biometrics depending on the availability, necessity and security requirements. At each stage, the system decides whether the individual is authorized or unauthorized. The authorized-unauthorized frequency distribution curve for unimodal and multi-modal system is shown in Figure 8.2. This curve shows the distribution of authorised and unauthorised match scores of a particular biometric trait. The system rejects the claimed identity if the score is less than ‘minA’, as given in Figure 8.2.

The reject zone essentially consists of two regimes, namely, confusion zone and true reject zone. The confusion zone is the overlap between true reject and accept zone. The regime corresponding to match score of unauthorized score less than ‘maxU’ and authorized score greater

than 'minA' is the confusion zone. Therefore, for a particular biometric trait, if the resulting score is greater than 'maxU', the system will recognize the person as authorized; similarly, if the resulting score is less than 'minA', the system will recognize the person as unauthorized. If the value is in between 'minA' and 'maxU', the decision is fuzzy. It is observed that the width of the confusion zone decreases, when the number of biometric traits is increased.

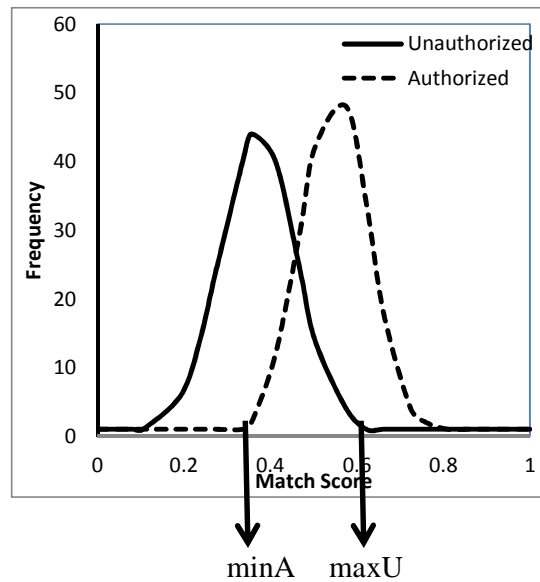


Figure 8.2 Authorized-Unauthorized frequency distribution curve

8.4.1.1 Fusion Strategy

Two fusion strategies considered in this study are equal weight for each traits and user dependent trait specific weight. The cases differ in computing the weight of the trait. In equal weight system, the weight for each trait in the system is taken as equal and hence, if 'n' traits are used, the weight of each trait will be '1/n'. In user dependent trait specific weight system, the average performance of each trait has been considered. The

EER corresponding to the ROC curve of the performance of all the users is used for the computation of the user dependent trait specific weight system. The different weight is considered for various traits for compensating the poor average performance of the particular trait. For example, in general, the success ratio of the signature trait is lower than the iris, which can be compensated in the system by considering higher weight for iris and lower weight for signature.

a) Equal weight fusion strategy

This is the simplest form of fusing various traits in which the equal importance is given to all traits considered. In this study, four traits are used. The weight of the j^{th} trait W_j to be used is $1/2$, $1/3$ and $1/4$ corresponding to bi-, tri- and tetra-modal system respectively. The fused score S_{fus} is calculated as in Eq.(8.12).

$$S_{fus} = \sum_{j=1}^n W_j \times S_j \quad (8.12)$$

where S_j is the match score of j^{th} trait. The range of the fused score is $[0, 1]$

b) User dependent trait specific weight fusion strategy

In this strategy, 1000 data corresponding to each trait in the knowledge base is used. The ROC curve for each trait is prepared and the EER is determined. The weight W_i for the i^{th} particular trait is calculated using the Eq.(8.13).

$$W_i = \frac{1/EER_i}{\sum_{j=1}^n (1/EER_j)} \quad (8.13)$$

where EER_j represents the Equal Error Rate for j^{th} trait and n represents the number of traits participating in fusion. The fused score is calculated as in Eq.(8.12). The weights corresponding to each trait calculated for 1000 training data are given in Table 8.1.

Table 8.1 Weights calculated for each trait in all the possible fusions

Multi-modal	Trait combinations	Face (f)	Iris (i)	Signature (s)	Voice (v)
Bi-modal	f+i	0.43750	0.56250		
	f+s	0.727273		0.272727	
	f+v	0.526316			0.473684
	i+s		0.774194	0.225806	
	i+v		0.588235		0.411765
	s+v			0.294118	0.705882
Tri-modal	f+i+s	0.375839	0.483221	0.140940	
	f+i+v	0.313901	0.403587		0.282511
	f+s+v	0.439560		0.164835	0.395604
	i+s+v		0.502092	0.146444	0.351464
Tetra-modal	f+i+s+v	0.280843	0.361083	0.105316	0.252758

8.4.2 User-Dependent Decision

Once a fused similarity score is obtained using user-dependent fusion method, this score is compared to a decision threshold in order to accept or reject the identity claim. In the decision module, an individual is recognized by searching the templates of all the users in the database for a match. One-to-many comparison is carried out to accept the identity of an individual or reject if the individual is not enrolled in the system database. The decision rule is stated as follows in Eq.(8.14).

$$Decision = \begin{cases} Accept & \text{if } S_{fus} \geq \tau_{fus} \\ Reject & \text{Otherwise} \end{cases} \quad (8.14)$$

where τ_{fus} is a threshold of fusion score and S_{fus} is a fused score.

8.5 Results and Discussions

In the multimodal system, four biometric traits, namely, speech, face, signature and iris are used. Face and iris are physiological biometric traits whereas signature and voice are behavioural biometric traits. A biometric authentication system makes decisions using the threshold and the matching score obtained by comparing the extracted feature of the query sample with the reference model of the claimed identity. During the decision making process, two types of error may happen: False Acceptance (FA) Error, which occurs when a system falsely accepts an impostor (a person claiming an identity other than his/her own) and False Rejection (FR) Error, which occurs when a system falsely rejects a client (a genuine user). In the literature, FA and FR errors are also referred to as False Match Error and False Non-Match Error, respectively. The normalized versions of FA and FR errors are often used and are known as False Acceptance Rate (FAR) and False Rejection Rate (FRR), respectively. They are defined as in Eqs. (8.15) and (8.16) respectively.

$$FAR = \frac{FA}{\text{Number of Unauthorised}} \quad (8.15)$$

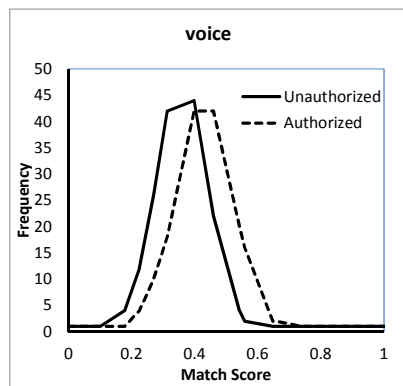
$$FRR = \frac{FR}{\text{Number of Authorised}} \quad (8.16)$$

It is observed that if the threshold is increased, the False Acceptance Rate (FAR) will decrease but the False Rejection Rate (FRR) will increase and vice versa. Hence, for a given biometric system, it is not possible to decrease both these errors simultaneously by varying the threshold. This has led to a threshold setting that produces Equal Error Rate (EER), a point

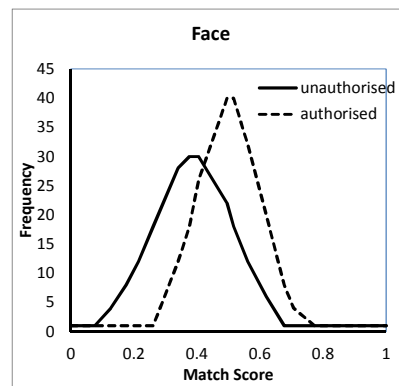
where FAR and FRR become equal on the training (validation) data set. The lower the EER, the better the system performance.

8.5.1 Calculation of weights for fusion

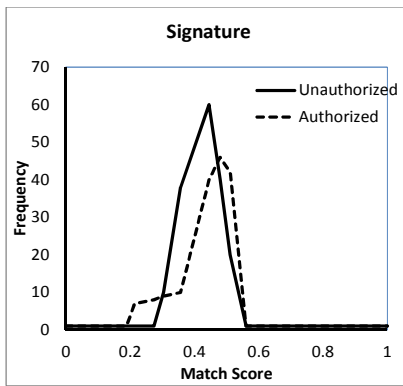
There are various possible fusions depending on the number of biometric traits. In this work, four biometric traits are used. Hence, there are 6 possible fusions of two traits, 4 fusions of 3 traits and 1 fusion of all the 4 traits. The weights assigned to different traits in all the possible fusions are shown in Table 8.1 given above.



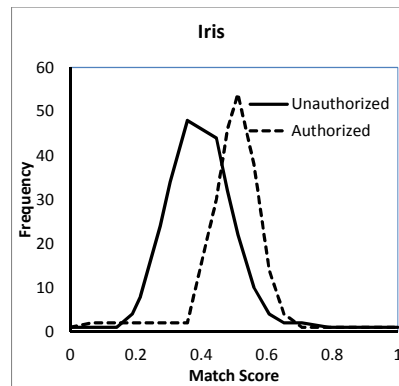
a) $\min A = 0.2273$, $\max U = 0.5923$



b) $\min A = 0.292912$, $\max U = 0.653596$



c) $\min A = 0.2358$, $\max U = 0.6428$

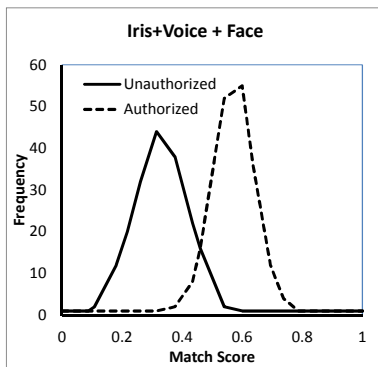


d) $\min A = 0.3042$, $\max U = 0.6315$

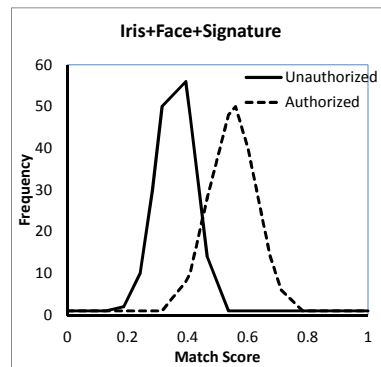
Figure 8.3 Authorised-Unauthorised Distribution Curve for the Biometric traits

8.5.1.1 Calculation of minimum Authorised and maximum Unauthorised Score

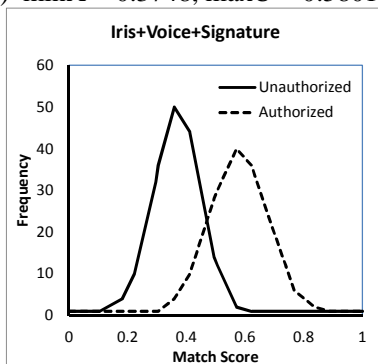
The authorized-unauthorized distribution curves for each of the four traits and for each of the 11 fusion results are used to calculate the minimum authorized and maximum unauthorized score. The authorized-unauthorized curves for all the traits, the unimodal, Figure 8.3, possible fusions of trimodal Figure 8.4, bimodal, Figure 8.5, as well as tetramodal Figure 8.6 has been plotted. The minimum authorized and maximum unauthorized score for each trait and their fusion results are shown in authorized-unauthorized distribution curve as minA and maxU, respectively.



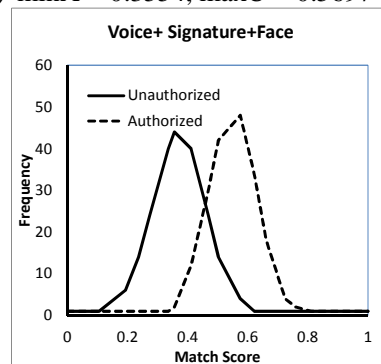
a) $\text{minA} = 0.3748, \text{maxU} = 0.5801$



b) $\text{minA} = 0.3554, \text{maxU} = 0.5697$

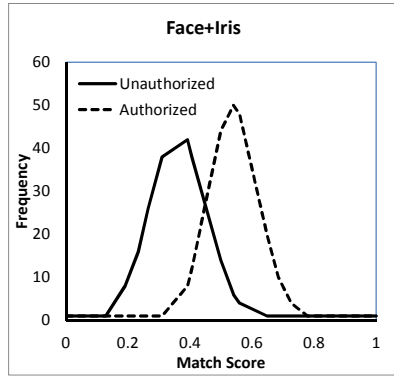


c) $\text{minA} = 0.3545, \text{maxU} = 0.5875$

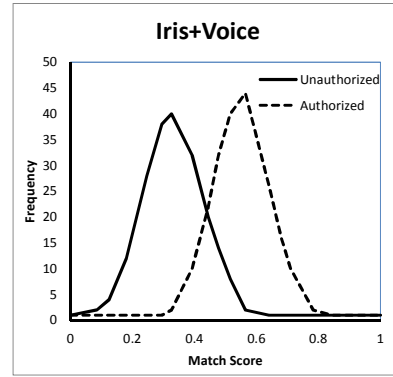


d) $\text{minA} = 0.3475, \text{maxU} = 0.5931$

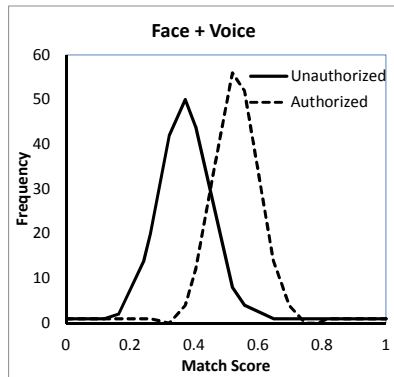
Figure 8.4 Authorised-Unauthorised Distribution Curve for trimodal fusion



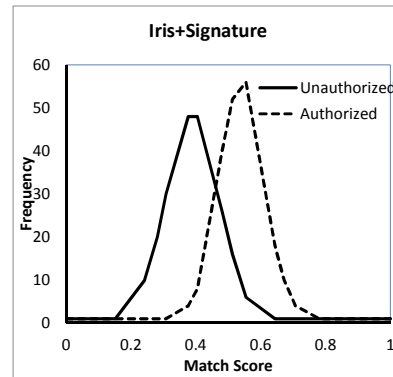
a) $\min A = 0.3476$, $\max U = 0.5990$



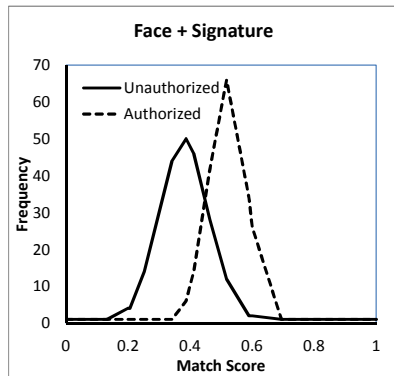
b) $\min A = 0.3371$, $\max U = 0.5890$



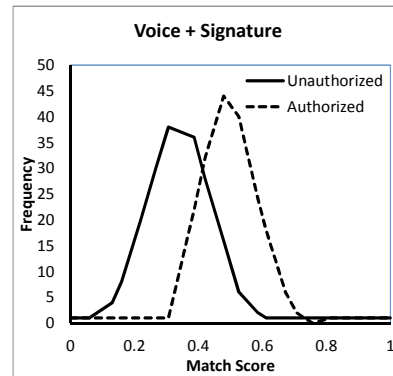
c) $\min A = 0.3278$, $\max U = 0.5906$



d) $\min A = 0.3416$, $\max U = 0.6088$



e) $\min A = 0.3317$, $\max U = 0.6044$



f) $\min A = 0.3015$, $\max U = 0.5942$

Figure 8.5 Authorised-Unauthorised Distribution Curve for possible fusion of the Biometric traits for bimodal

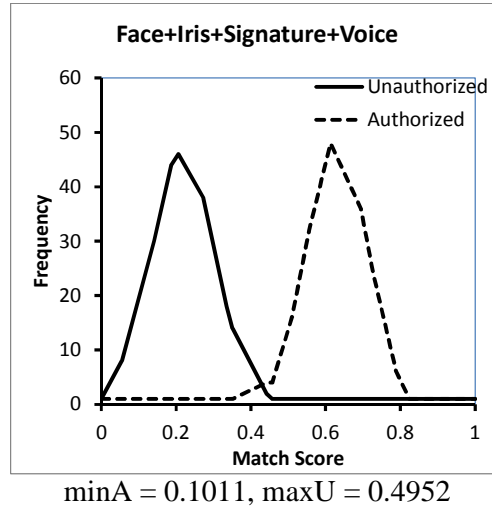


Figure 8.6 Authorised-Unauthorised Distribution Curve for the Tetramodal

8.6 System Performance and Evaluation

The performance of the multimodal biometric identification system is evaluated using width of the confusion zone as shown in Figure 8.4. The width of the confusion zone is the width between the control limit of the minimum authorized point (minA) and the maximum unauthorized point (maxU) in the authorized-unauthorized distribution curve. The width of the confusion zone is an indication of the effectiveness of the system. If the width is less, the system will be able to classify the test sample more correctly.

The authorised and unauthorised zones are distinct and the confusion in identifying the test sample is low in such cases. The system is more reliable and robust if the width of the confusion zone is lower. In general, the fusion of biometric traits in the identification system reduces the width of the confusion zone. The width of the confusion zone in the score-frequency response of a tetramodal system of speech, face, signature and iris is found to be as low as 0.1.

The performance of the multimodal biometric identification system is also evaluated using the Success Ratio (SR). Success is defined as the approval of a known person or the rejection of an unknown person. SR is found as given in Eq.(8.17).

$$SR = \frac{k_a + u_r}{k_a + k_r + u_a + u_r} \quad (8.17)$$

where k_a is the total number of acceptance of the known person, k_r is the total number of rejection of the known person, u_a is the total number of acceptance of the unknown person and u_r is the total number of rejection of the unknown person.

The success ratio of the user identity is recorded for unimodal, bimodal, trimodal and tetramodal biometric systems as given in Table.8.2. The success ratio of unimodal is found to be between 84 and 93 percent. The highest success ratio is found to be for iris and hence the iris system is relatively robust when compared to the other unimodal systems. The success ratio of the bimodal system is found to be from 88 to 94 percent. The maximum increase in the success ratio by fusing two traits is found to be 1 percent and is for the combination of face and iris traits. This may be attributed to various factors such as the quality of the sensor units, noise levels, algorithms used for extracting feature vectors, selected features etc.

The signature trait is found to have the lowest success ratio. This is attributed to the variations and noise associated with the signature.

Two fusion technologies are adopted in this study, namely equal weight and user dependent trait specific weight approach. In equal weight approach, the maximum success ratio of bimodal, trimodal and tetra modal

is found to be 91, 94 and 96 percent. In general, success ratio of multimodal fusion system is greater than the unimodal system.

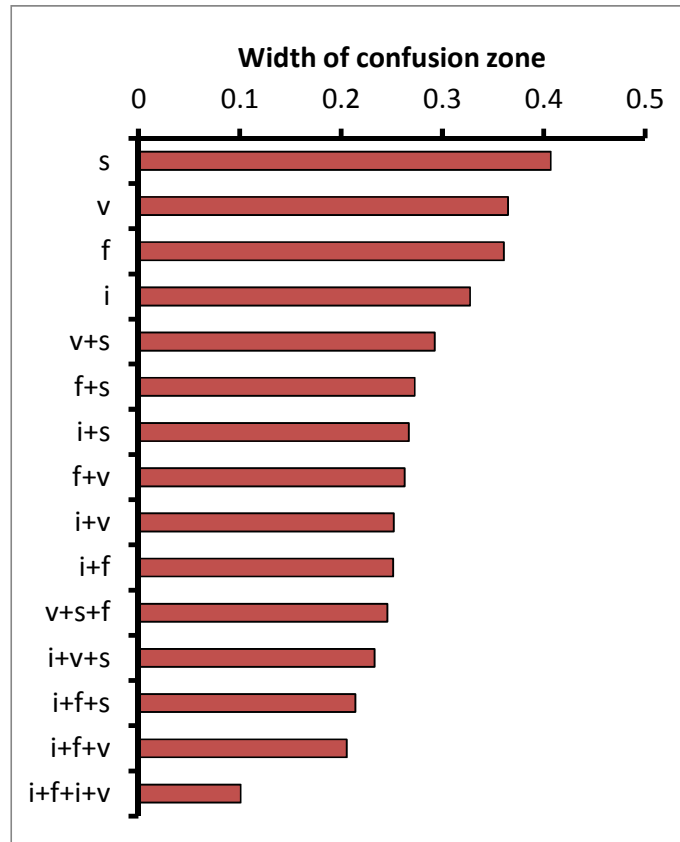


Figure 8.7 Width of the confusion zone.

The success ratio of the trimodal user dependent trait specific weight system is found to be between 92 and 96 percent. The maximum increase in the success ratio of trimodal when compared to the bimodal is found to be 2 percent. Thus, the trimodal system is better than the bimodal system. The success ratio of the tetramodal system is found to be 98 percent. Hence, it is concluded that the tetramodal developed in this study is significantly better than the unimodal, bimodal and trimodal systems.

Table 8.2 Success Ratio (SR) of Biometric identification system

Modal	Trait	Success ratio (percent)
Unimodal	Face(F)	91.32
	Signature (S)	84.45
	Voice (V)	90.48
	Iris (I)	93.26
Bimodal	F+S	88.37
	F+V	91.40
	F+I	94.84
	S+V	88.74
	S+I	91.15
	V+I	93.24
Trimodal	F+S+V	92.75
	F+V+I	96.81
	F+I+S	93.35
	S+V+I	94.50
Tetramodal	F+S+V+I	98.54

It is found that the success ratio of equal weight system is lower than the success ratio of the user dependent trait specific weight fusion method. This is attributed to the fact that the user dependent trait specific weight fusion method accounts for the relative variation of the average performance of the particular trait. The chart of success ratio of the biometric identification system is shown in Figure 8.8.

8.7 Summary

Multimodal biometric systems elegantly address several problems in unimodal systems. By combining multiple sources of information, these systems improve the accuracy, increase population coverage and deter spoofing. Various fusion levels and scenarios are possible in multimodal systems. Fusion at the match score level is the most popular due to the ease in accessing and consolidating matching scores. In this chapter, a

multimodal biometric system, which integrates the decisions made by speaker recognition, face recognition, handwritten signature recognition and iris recognition for personal identification, is developed.

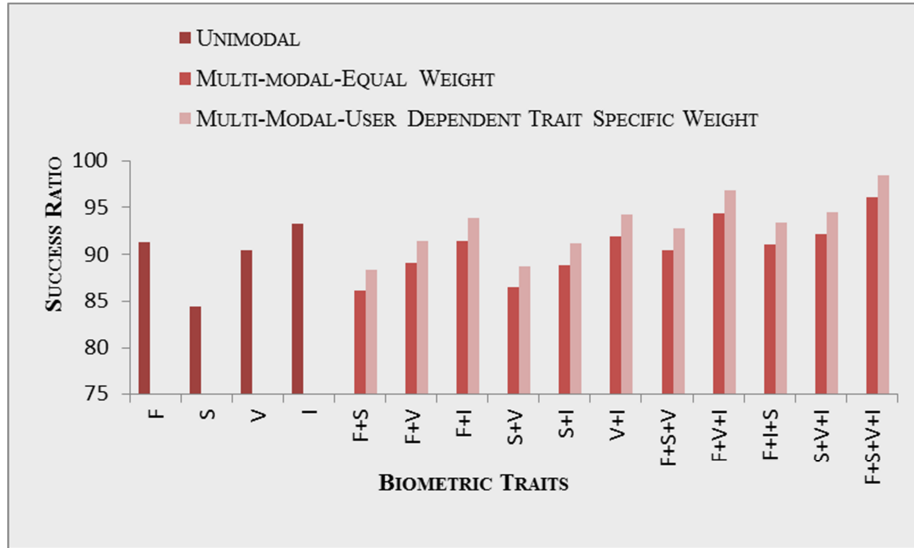


Figure 8.8 Chart of Success Ratio of the biometric identification system

The four biometric traits are fused at the match score level using weighted fusion strategy. These different weights are calculated based on the Equal Error Rate (EER). The system performance is evaluated using the width of the confusion zone and success ratio. Comparing the four scenarios of unimodal, bimodal, trimodal and tetramodal systems, it can be concluded that the fusion of tetra modalities yields a significant improvement in the performance than the other three.

CHAPTER 9

CONCLUSIONS AND FUTURE WORK

The thesis addresses one of the emerging topics in Authentication System, viz., the implementation of Improved Biometric Authentication System using Multimodal Cue Integration, as the operator assisted identification turns out to be tedious, laborious and time consuming. In order to derive the best performance for the authentication system, an appropriate feature selection criteria has been evolved. It has been seen that the selection of too many features lead to the deterioration in the authentication performance and efficiency. In the work reported in this thesis, various judiciously chosen components of the biometric traits and their feature vectors are used for realizing the newly proposed Biometric Authentication System using Multimodal Cue Integration. The feature vectors so generated from the noisy biometric traits is compared with the feature vectors available in the knowledge base and the most matching pattern is identified for the purpose of user authentication. In an attempt to improve the success rate of the Feature Vector based authentication system, the proposed system has been augmented with the user dependent weighted fusion technique.

9.1 *Highlights of the Thesis*

The dependence of society on the usage of information technology for everyday tasks makes the establishment of the identity of a person in a reliable and time-efficient manner a matter of paramount importance.

Biometrics is an efficient technology with great possibilities in the area of security system development for official and commercial applications. The thesis addresses one of the emerging topics in Authentication System, viz., the implementation of *Improved Biometric Authentication System using Multimodal Cue Integration*, as the operator assisted identification turns out to be tedious, laborious and time consuming. In order to derive the best performance for the authentication system, an appropriate feature selection criteria has been evolved. It has been seen that the selection of too many features lead to the deterioration in the authentication performance and efficiency. In the work reported in this thesis, various judiciously chosen components of the biometric traits and their feature vectors are used for realizing the newly developed prototype of the Biometric Authentication System. The feature vectors so generated from the noisy biometric traits is compared with the feature vectors available in the knowledge base and the most matching pattern is identified for the purpose of user authentication. In an attempt to improve the success rate of the developed Feature Vector based Biometric Authentication System, the proposed system has been augmented with the user dependent weighted fusion technique. The following are the salient highlights of this thesis.

9.1.1 Need and Requirement of a Biometric Authentication System

Biometric authentication automatically achieves the establishment of the identity of an individual based on the physiological or behavioural biometric trait rather than knowledge based or possession based method. The introductory chapter of the thesis throws light on the various biometric traits as well as different models of the authentication system. Even though the biometric identification systems out-perform their peer technologies, the unimodal biometric systems have to contend with a variety of problems, namely, noisy data, intra-class variations, restricted degrees of freedom,

non-universality and spoof attacks. Deploying a multimodal biometric system can address the limitations by integrating the evidences presented to it by multiple sources of information. The underlying principle of operation of the proposed multimodal biometric system is also briefly introduced in this chapter.

9.1.2 Preparation of a State-of-the-art Literature

As a part of the work, a state-of-the-art literature survey report has been prepared in the topic covered in the thesis highlighting the characteristic signatures of different biometric traits, principles underlying speaker recognition, face recognition, handwritten signature recognition, iris recognition, multimodal biometrics, normalization methods, different fusion techniques, etc. The consolidated results by various researchers on the success rates of the different methods are discussed. This chapter also covers the recent trends in the fusion technology of various biometric traits.

9.1.3 Feature Vector Based Authentication System

The methodology suggested to be adopted for realizing the proposed authentication system in speaker recognition involves the extraction of source features by analysing the speech waveforms and identifying the most matching feature vector using template matching technique leading to the identification of the speaker. Feature vector compiled includes spectral centroid, spectral flux, spectral rolloff and MFCC coefficients retrieved from speech. For making the system fool proof and full-fledged one, the knowledge base has to be updated with the feature vectors comprising of eigen vector from face, combination of static and dynamic features from signature and binary iris code using Haar wavelet generated from iris.

9.1.4 Extraction of Feature Vector Element from Speech

The thesis addresses the speaker enrolment phase, where speech samples that contain the discriminating features are collected from the speakers and feature vectors are generated to form the knowledge base of the model. In the recognition phase, the feature vectors extracted from the unknown person's utterance are compared with the knowledge base of the model to find the similarity score, for the purpose of decision making. Since accuracy of identification is highly dependent on the type and number of features used, feature selection is of great significance. Features are computed from the spectrogram on a frame-by-frame basis and relates directly to some perceptual characteristics of sound, such as loudness, pitch etc. This chapter also touches upon a more systematic approach for computing the cepstral coefficients achieved by estimating the Mel Frequency Cepstral Coefficients (MFCC). The performance analysis of the system was carried out and it was found that along with MFCC, the incorporation of additional spectral features in the feature vector improved the overall performance of the system. Since the success rate stayed at 90% even after including the other spectral features like number of peaks, spectral crest etc., those features were not considered to be a part of the Feature Vector.

9.1.5 Extraction of Feature Vector Element from Face

In the face recognition system, the proposed PCA is found to result in a recognition efficiency of 91.5% when tested for 1000 training images. The computational time efficiency of the present PCA is estimated through covariance matrix with and without dimensionality reduction. The computational time with reduced dimensionality is found to be 34.2 seconds (for 1000 face images) where as it is found to increase by five

times i.e. 171.2 seconds for the case without dimensionality reduction. Hence, the reduction of dimensionality results in reduced computation time.

9.1.6 Extraction of Feature Vector Element from Signature

A signature, widely accepted as a means of identity authentication in legal and commercial transactions, is remarked as a consistent non-invasive authentication procedure by the majority of the users, thereby overcoming some of the privacy issues. The distinct features are extracted from the image of the signature trajectory captured by electronic signature tablet, after pre-processing. In signature recognition, the feature vector, which forms the signature template in the knowledge base, is selected as a combination of static and dynamic features. The success rate of the system with static features alone is 76 percent and by the addition of dynamic features, the success rate is increased to 84 percent. Moreover, the procedures to be adopted for generating static features using 2D Gabor filter as well as the dynamic features under consideration are the x and y stroke as well as the average velocity in x and y directions is described in this chapter. Mahalanobis Distance (MD) which computes the correlation between two signatures used to verify the similarity of images is also described in the thesis.

9.1.7 Extraction of Feature Vector Element from Iris

The thesis also addresses the various steps involved in the extraction of binary code from Iris pattern. The pre-processing stage requires localization of the iris which generally involves the detection of the edge of the iris as well as that of the pupil. Since varying levels of illumination can result in dimensional inconsistencies between eye images due to the stretching of the iris, normalization needs to be performed so that

iris region is transformed to have fixed dimensions. After unwrapping the normalized iris region into a rectangular region, it is encoded using Haar wavelets to generate the iris code. It was also observed that the Canny operator is best suited to extract most of the edges to generate the iris code for comparison. In the recognition stage, Hamming distance is used for the comparison of the iris code, the most discriminating feature of the iris pattern, with the existing iris templates.

9.1.8 Authentication System Based on Multi-modal Approach

In this doctoral thesis, a multimodal biometric system using speech, face, signature and iris biometric identifiers is presented. To combine the information from these four biometric identifiers, user dependent weighted fusion approaches are introduced. The final decision is made by fusion at matching score level architecture, in which the feature vectors of query images are compared with the templates in knowledge base. Based on the proximity of the feature vector and template, each subsystem computes its own matching score. Finally, the individual traits are fused at matching score level using user dependent weighted sum of score technique and is then passed to the decision module. This system is tested on our knowledge base and the overall accuracy of the system is found to be more than 98%.

9.1.9 Towards Improving the Performance of the Prototype

The user dependent weighted fusion approach significantly enhances recognition performance of the multimodal biometric system. The weighted fusion strategy provides more confidence information of the outcomes for the developed multimodal biometric system. The extensive experimentations with multimodal databases indicate that the proposed multimodal system outperforms other commonly used methods and can

help government or public/private sectors to protect valuable property or information and also ensure the overall security of the region or country.

9.2 *Future Scope for Research*

The work presented in this thesis has a significant role to play in view of its practical applications. This work also has substantial scope for further research for improving the overall system performance. A significant progress in the development of fusion strategies schemes that facilitate the design of reliable and secure multibiometric systems is achieved in this work. To begin with, the authentication results presented in this thesis should be validated using other public multimodal real-user databases. Specifically, it would be necessary to measure the performance of the suggested approaches with a larger dataset, containing more individuals. A formal model for cost-benefit analysis of a multibiometric system based on parameters such as performance gain (reduction in FRR/FAR), throughput, physical cost of the system and security needs to be developed in order to enable biometric system developers to rapidly design a multibiometric system that is most appropriate for the application on hand. More research can be conducted to find the optimum matching algorithms for unimodal biometrics to enhance the overall performance of the multimodal system. Dual or tri-level fusion scenarios (different fusion in different levels of the system) can be investigated to make the system faster and significant reduction in the error rate.

9.3 *Summary*

An attempt has been made in this chapter to bring out the salient highlights of the work and the general inferences gathered along with enlisting of the scope and direction for future research in this area. When

the real-time authentication system augmented with a full-fledged backbone knowledge base becomes a reality, the system can outperform the state-of-the-art authentication systems with amazing high success rates.

References

- [1]. Jain A.K., Bolle R., and Pankanti S., *Biometrics: Personal Identification in Networked Society*, Kluwer Academic Publishers, 1999.
- [2]. Jain A.K., Ross A., Pankanti S., *Biometrics: A tool for information security*, IEEE Transaction on Information Forensics and Security, vol.1, No.2, 2006, pp.125-143.
- [3]. Jain A. K., Ross A., and Prabhakar S., *An Introduction to Biometric Recognition*, IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, vol.14, No.1, 2004, pp.4–20.
- [4]. Ross A., Nandakumar K., and Jain A. K., *Handbook of Multibiometrics*, Springer, 2006.
- [5]. Jain A.K. and Ross A., *Multibiometric Systems*, Communications of the ACM, Special Issue on Multimodal Interfaces, 47, No.1, 2004, pp.34–40.
- [6]. Garris M. D., Watson C. I., and Wilson C. L., *Matching Performance for the US-Visit IDENT System Using Flat Fingerprints*. Technical Report 7110, National Institute of Standards and Technology (NIST), 2004.
- [7]. Wilson C., Hicklin A. R., Bone M., Korves H., Grother P., Ulery B., Micheals R., Zoepfl M., Otto S., and Watson C., *Fingerprint Vendor Technology Evaluation 2003: Summary of Results and Analysis Report*. Technical Report NISTIR 7123, NIST, 2004.
- [8]. Matsumoto T., Matsumoto H., Yamada K., and Hoshino S., *Impact of Artificial Gummy Fingers on Fingerprint Systems*. In Optical Security and Counterfeit Deterrence Techniques IV, Proceedings of SPIE, volume 4677, 2002, pp. 275–289.
- [9]. Harrison W. R., *Suspect Documents, their Scientific Examination*, Nelson-Hall Publishers, 1981.
- [10]. Eriksson A. and Wretling P., *How Flexible is the Human Voice? A Case Study of Mimicry*, Proc. of the European Conference on Speech Technology, 1997, pp.1043–1046.

References

- [11]. Hong L., Jain A. K., and Pankanti S., *Can Multibiometrics Improve Performance?*, Proc. of IEEE Workshop on Automatic Identification Advanced Technologies (AutoID), 1999, pp. 59–64.
- [12]. Ho T.K., *Data complexity analysis for classifier combination*, Proc. of the 2nd International Workshop of Multiple Classifier System, Lecture Notes in Computer Science, Springer-Verlag, Cambridge, UK, 2001, pp.53–67.
- [13]. Zanuy M. F., *Data fusion in biometrics*, IEEE Aerospace and Electronic Systems Magazine, vol. 20, 2005, pp.34-38.
- [14]. Fierrez-Aguilar J., *Adapted Fusion Schemes for Multimodal Biometric Authentication*, PhD thesis, Universidad Politecnica de Madrid, 2006.
- [15]. Maltoni D., Maio D., Jain A. K., Prabhakar S., *Handbook of Fingerprint Recognition*, Springer-Verlag, 2003.
- [16]. Nandakumar K., *Integration of Multiple Cues in Biometric Systems*, Master thesis, Department of Computer Science and Engineering, Michigan State University, 2005
- [17]. Iyengar S. S., Prasad L., and Min H., *Advances in Distributed Sensor Technology*, Prentice Hall, 1995.
- [18]. Ross A. and Govindarajan R., *Feature Level Fusion Using Hand and Face Biometrics*, Proc. of SPIE Conference on Biometric Technology for Human Identification II, volume 5779, Orlando, USA, March 2005, pp. 196–204.
- [19]. Jain A. K. and Chandrasekaran. B., *Dimensionality and Sample Size Considerations in Pattern Recognition Practice*, Handbook of Statistics, P.R. Krishnaiah and L. N. Kanal editors, vol 2, . North-Holland, Amsterdam, 1982, pp. 835–855.
- [20]. Ho T. K., Hull J. J., and Srihari S. N., *Decision Combination in Multiple Classifier Systems*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.16, No.1, 1994, pp. 66–75.
- [21]. Daugman J., *Combining Multiple Biometrics*. Available at <http://www.cl.cam.ac.uk/users/jgd1000/combine/combine.html>, 2000.
- [22]. Lam L. and Suen C. Y., *Application of Majority Voting to Pattern Recognition: An Analysis of its Behavior and Performance*, IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans, vol 27, No.5, 1997, pp. 553–568.

- [23]. Kuncheva L. I., *Combining Pattern Classifiers - Methods and Algorithms*, Wiley, 2004.
- [24]. Xu L., Krzyzak A. and Suen C. Y., *Methods for Combining Multiple Classifiers and their Applications to Handwriting Recognition*, IEEE Transactions on Systems, Man, and Cybernetics, vol. 22, No.3, 1992, pp. 418–435.
- [25]. Shaneh M. and Taheri A., *Voice Command Recognition System based on MFCC and VQ Algorithms*, World Academy of Science, Engineering and Technology vol. 33, 2009, pp. 534-538.
- [26]. Zhou. G. and Mikhael W.B, *Speaker Identification based on adaptive discriminative vector quantisation*, IEEE Proc. Vis. Image Signal Process vol.153. No.6, 2006, pp. 754-760.
- [27]. Shi Huang Chen and Yu-Ren Luo, *Speaker verification using MFCC and support vector machine*, Proc. of IMEC, vol 1, 2009, pp. 121-124.
- [28]. Jabloun F., Cetin A. E and Erzin E., *Teager Energy Based Feature Parameters for Speech Recognition in Car Noise*, IEEE Signal Processing Letters vol 6. No.10, 1999, pp. 259-261.
- [29]. Revathi A, Ganapathy R, and Venkataramani, *Text Independent Speaker Recognition and Speaker Independent Speech Recognition using Iterative clustering approach*, International Journal of Computer Science and Information Technology, vol. 1, No 2, 2009, pp. 30-41.
- [30]. Shafran I., Riley M and Mohri M., *Voice signature*, IEEE Automatic Speech Recognition and Understanding Workshop (ASRU), 2003, pp. 31-36.
- [31]. Ming J, Hazen T. J, Glass J. R and Reynolds D.A, *Robust Speaker Recognition in noisy condition*, IEEE Trans. on Audio Speech and language Processing, vol. 15, No.5, 2007, pp. 1711-1723.
- [32]. Togneri R and Pallela D, *An overview of Speaker Identification: Accuracy and Robustness issues*, IEEE circuits and systems magazine second quarter, 2011, pp. 23-61.
- [33]. Campbell J.P, *Speaker Recognition: A Tutorial*, Proc. of IEEE, vol. 85, 1997, pp. 1437-1462.
- [34]. Cui X and Alwan A, *Noise Robust Speech Recognition using Feature compensation based on polynomial Regression of utterance SNR*, IEEE Trans. on Speech and Audio Processing, vol.13, No.6, 2005, pp. 1161-1172.

References

- [35]. Lee K.F, Hon H. W and Reddy R, *An overview of the SPHINX Speech Recognition System*, IEEE Trans. on Acoustics speech and signal processing, vol. 38, No.1, 1990, pp. 35-45.
- [36]. Chulhee Lee, Donghon Hyun, Euism Choi and Jinwook Go, *Optimizing feature extraction for Speech Recognition*, IEEE Transactions on Speech and Audio Processing, vol. 11. No.1, 2003, pp. 80-87.
- [37]. Reynolds D.A, Quatieri T.F and Dunn R.B, *Speaker verification using adapted Gaussian mixture models*, Digital Signal Processing, vol. 10, 2000, pp. 19-41.
- [38]. Furui. S, *Cepstral Analysis Technique for Automatic Speaker Verification*, IEEE Acoustics, Speech and Signal Processing(ASSP), vol.29, No.2, 1981, pp. 254-272.
- [39]. Markel D J and Davis B.S, *Text Independent Speaker Recognition from a Large Linguistically unconstrained Time Speed Data Base*, IEEE Transaction on Acoustics, Speech and Signal Processing (ASSP), vol. 27, No.1, 1979, pp. 74-82.
- [40]. Soong F.K., Rosenberg A.E, Rabiner L.R,Juang B.H , *A Vector Quantization Approach to Speaker Recognition*, IEEE Int. Conf. on Acoustics speech and signal Processing, 1985, pp. 387 - 390.
- [41]. Lin. Q., Jan E. A. and Flanagan J., *Microphone Arrays and Speaker Identification*, IEEE Trans. on Speech and Audio Processing (SAP), vol. 2, No. 4, 1994, pp. 622-629.
- [42]. Reynolds D. A. and Rose R. C., *Robust Text- Independent Speaker Identification Using Gaussian Mixture Speaker Models*, IEEE Trans. on Speech and Audio Processing, vol. 3, No. 1, 1995, pp. 72-83.
- [43]. Maleh K. E, Klein M., Petrucci G and Kabal P, *Speech/Music discrimination for multimedia applications*, Proc. IEEE Int. Conf. on acoustics, Speech, Signal Processing, pp. 2445-2448, 2000.
- [44]. Meindo H and Neto J, Audio Segmentaion, *Classification and Clustering in a Broadcast News Task*, IEEE Int. Conf. on Acoustics speech and signal Processing, 2003, pp. 115-118.
- [45]. Panagiotakis. C and Tziritas G, *A Speech/Music Discriminator Based on RMS and Zero-Crossings*, IEEE Trans. on multimedia, 2004, pp. 155-166,

- [46]. Zhang T and Kuo J. C.C., *Heuristic Approach for Generic Audio Data Segmentation and Annotation*, ACM Multimedia, vol. 1, 1999, pp. 67-76.
- [47]. Sourjya S, Sreenivasa R. K, *Stochastic feature compensation methods for speaker verification in noisy environment*, Elsevier Applied Soft Computing, , vol. 19 , 2014, pp. 198–214.
- [48]. Mohan A, Rose R, Ghalehjegh S. H and Umesh S, *Acoustic modelling for speech recognition in Indian languages in an agricultural commodities task domain*, Elsevier Speech Communication, vol. 56, 2014, pp. 167–180.
- [49]. Dufour R., Esteve Y., Deleglise P., *Characterizing and detecting spontaneous speech: Application to speaker role recognition*, Elsevier Speech Communication, vol. 56, 2014, pp. 1–18.
- [50]. Sahidullah M. and Saha G, *Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition*, Elsevier Speech Communication, vol. 54, 2012, pp. 543–565.
- [51]. Phillips. P. J, Flynn P L, Seruggo T, Bowyer K. W, Chang J, Hoffman K, Marques J, Min J and Worek W., *Overview of the face recognition Grand Challenge*, Proc. of the IEEE computer society conference on Computer Vision and Pattern Recognition(CVPR), 2005, pp. 1063-6919.
- [52]. Prakash J and Rajesh K, *Human Face Detection and Segmentation using Eigen values of covariance matrix, Hough Transform and Raster Scan Algorithm*, World Academy of Science, Engineering and Technology, vol. 39, 2008, pp. 372-380.
- [53]. Kim K I, Jung K, and Kim H J, *Face Recognition Using Kernel Principal Component Analysis*, IEEE Signal Processing Letters, vol. 9, No. 2, 2002, pp. 40-42.
- [54]. Demirel. H and Anbarjafari. G, *Pose Invariant Face Recognition Using Probability Distribution Functions in Different Color Channels*, IEEE Signal Processing Letters, vol.15, 2008, pp. 537-540.
- [55]. Spies H., and Ricketts I., *Face Recognition in Fourier Space*, Vision Interface, 2000, pp. 38-44.
- [56]. Temdee P., Khawparisuth D., and Chamnongthai K., *Face Recognition by Using Fractal Encoding and Backpropagation Neural Network*, Fifth Int. Symp. on Signal Processing and its

References

- Applications, ISSPA '99, Brisbane, Australia, August 1999, pp. 159-161.
- [57]. Wiskott L., Fellous J. M., Krüger N., and von der Malsburg C., *Face Recognition by Elastic Bunch Graph Matching*, Proc. IEEE Int. Conf. on Image Processing (ICIP'97), vol.1, 1997, pp. 129-132.
- [58]. Chung K. C., Kee S. C., and Kim S. R., *Face Recognition using Principal Component Analysis of Gabor Filter Responses*, IEEE, 1999, pp. 53-57.
- [59]. Joo M., Shiqian W., Lu J., *Face recognition using radial basis function (RBF) neural networks*, Proc. of the 38th IEEE Conference on Decision and Control ,vol.3, 1999, pp. 2162 - 2167.
- [60]. Pan G, Shi Han, Zhaohui W and Wang Y., *3D Face Recognition using mapped depth images*, Proc. of the IEEE Computer Society Conf. on Computer Vision and pattern Recognition(CVPR), 2005, pp. 111-117.
- [61]. Belhumeur P.N, Hespanha J. P and Kriegman D.J, *Eigen faces vs Fisherfaces: Recognition using class specific Linear Projection*, IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 1997, pp. 711-720.
- [62]. Chen L., Liao H., Ko M., Liin J., and Yu G, *A New LDA based Face Recognition System which can solve the small sample size problem*, Elsevier Pattern Recognition, vol. 33, No.10, 2000, pp 1713-1726.
- [63]. Chen L., Man H, Nefian A.V, *Face recognition based on multi-class mapping of Fisher Scores*, Elsevier Pattern Recognition, vol.38, 2005, pp. 799-811.
- [64]. Graf H. P, Chen T., Petajan E., and Cosatto E., *Locating Faces and Facial Parts*, Proc. First Int'l Workshop on Automatic Face and Gesture Recognition, 1995, pp. 41-46.
- [65]. Han C.C, Liao H. M, Yu K.C, and Chen L.H, *Fast Face Detection via Morphology-Based Pre-Processing*, Proc. Ninth Int'l Conf. Image Analysis and Processing, 1998, pp. 469-476.
- [66]. Amit Y, Geman D., and Jedynak B., *Efficient Focusing and Face Detection*, Face Recognition: From Theory to Applications, vol. 163, 1998,pp. 124-156.

- [67]. Breiman. L, Friedman. J, Olshen. R, and Stone. C, *Classification and Regression Trees*. Wadsworth, 1984, pp. 344-351.
- [68]. Samaria F.S, *Face Recognition Using Hidden Markov Models*, Ph.D thesis, Univ. of Cambridge, 1994
- [69]. Augusteijn. M.F and Skujca. T.L, *Identification of Human Faces through Texture-Based Feature Recognition and Neural Network Technology*, Proc. IEEE Conf. Neural Networks, 1993, pp. 392-398.
- [70]. Haralick R.M, Shanmugam K, and Dinstein I, *Texture Features for Image Classification*, IEEE Trans. Systems, Man, and Cybernetics, vol. 3, No. 6, 1973, pp. 610-621.
- [71]. Fahlman. S and Lebiere. C, *The Cascade-Correlation Learning Architecture*, Advances in Neural Information Processing Systems, Touretsky D.S. ed., 1990, pp. 524-532.
- [72]. Kohonen T, *Self-Organization and Associative Memory*, Springer, 3rd edn, 1989.
- [73]. Dai. Y and Nakano. Y, *Face-Texture Model Based on SGLD and its Application in Face Detection in a Color Scene*, Elsevier Pattern Recognition, vol. 29, no. 6, 1996, pp. 1007-1017.
- [74]. Adini. Y, Moses. Y, and Ullman S., *Face Recognition: the Problem of Compensating for Changes in Illumination Direction*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, 1997, pp. 721-732.
- [75]. Zhao W, Chellappa R. *3D Model Enhanced Face Recognition*, Proc. on Int. Conf. Image Processing, Vancouver, 2000, pp 50-53.
- [76]. Cutler R., *Face Recognition Using Infrared Images and Eigen faces*, citeseer.nj.nec.com/456378.html, 1996.
- [77]. Pentland A, Moghaddom B., Starner T., *View-Based and Modular Eigen faces for Face Recognition*, Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'94), 1994, pp. 133-139.
- [78]. Liu H. D, Yang M, Gao Y., Cui C., *Local histogram specification for face recognition under varying lighting conditions*, Elsevier Image and Vision Computing, vol. 32, 2014, pp. 335–347.
- [79]. Gaidhane V. H., Hote Y.V, Vijander S., *An efficient approach for face recognition based on common eigen values*, Elsevier Pattern Recognition, vol 47, 2014, pp. 1869–1879.

References

- [80]. Zhou C., Wang L., Zhang Q, Wei X., Face recognition based on PCA image reconstruction and LDA, Elsevier Optik, vol. 124, 2013, pp 5599– 5603.
- [81]. Guo J. K., Doermann D., Rosenfeld A., *Local correspondence for detecting random forgeries*, Proc. of 4th Int. Conf. on Document Analysis and Recognition, 1997, pp. 319-323.
- [82]. Justino E. J., Bortolozzi F., and Sabourin R, *Off-line signature verification using HMM for random simple and skilled forgeries*, Proc. of 6th Int. Conf. on Document Analysis and Recognition, 2001, pp. 450-453.
- [83]. Mizukami Y., Miike H., Yoshimura M., Yoshimura I., *An off-line signature verification system using an extracted displacement function*, Proc. of the 5th Int. Conf. Document Analysis and Recognition (ICDAR 99), 1999, pp. 757-760.
- [84]. Fang. B, Leung.C.H, Tang Y. Y, Tse K. W, Kwok P. C. K, Wong Y. K, *Off-line Signature Verification by the Tracking of Feature and Stroke Positions*, Elsevier Pattern Recognition, vol. 36, 2003, pp. 91-101.
- [85]. Sato Y. and Kogure K., *On-Line Signature Verification Based on Shape, Motion and Writing Pressure*, Proc. Sixth Int. Conf. Pattern Recognition, 1982, pp. 823-826.
- [86]. Parizeau M. and Plamondon R., *A Comparative Analysis of Regional Correlation, Dynamical Time Warping and Skeletal Tree Matching for Signature Verification*, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 7, 1990, pp. 710-717.
- [87]. Hastie T., Kishon E., Clark M., and Fan J., *A Model for Signature Verification*, Proc. IEEE Conf. Systems, Man and Cybernetics, 1991, pp. 191-196.
- [88]. Nalwa V.S., *Automatic On-Line Signature Verification*, Proc. IEEE, vol. 85, no. 2, 1997, pp. 215-239.
- [89]. Hairong L , Wang W, Chong W., Zhuo Q., *Off-line Chinese signature verification based on support vector machines*, Elsevier Pattern Recognition Letters, vol. 26, 2005, pp. 2390–2399.
- [90]. Kalera. M. K, Srihari S and Xu A, *Offline Signature Verification and Identification using Distance Statistics*, Int. Journal of Pattern Recognition and Artificial Intelligence, vol. 18, No.7, 2004, pp.1339-1360.

- [91]. Munich M. E and Perona P, *Visual Identification by signature tracking*, IEEE Trans. on Pattern Analysis and Machine Intelligence vol. 25, No.2, 2003, pp. 200- 217.
- [92]. Jayasekara B., Jayasiri A, Udawatta L. , *An Evolving Signature recognition System*, IEEE First Int. Conf. on Industrial and Information Systems (ICIIS), 2006, pp. 529-534.
- [93]. Vargas J.F, Ferrer M.A., Travieso C.M, Alonso J.B , *Offline Signature Verification Based on Pseudo-Cepstral Coefficients*, Proc. of the 10th Int. Conf. on Document Analysis and Recognition(ICDAR '09), 2009, pp. 126 – 130.
- [94]. Mohamadi S. Z., *Persian Static Signature Recognition*, Thesis of Bachelor of Electrical Engineering, Ferdowsi University of Mashad, Iran, 2006.
- [95]. Sigari M. H., Pourshahabi M. R., *Static Handwritten Signature Identification and Verification*, Thesis of Bachelor of Computer Engineering, Ferdowsi University of Mashad, Iran, 2006.
- [96]. Ozgunduz E., Mentürk T., Karslıngil M. E., *Off-Line Signature Verification and Identification by Support Vector Machine*, 11th Int. Conf. on Computer Analysis of Images and Patterns, 2005, pp. 799-805.
- [97]. Martinez E. F., Sanchez A., Velez J., *Support Vector Machines versus Multi-Layer Perceptrons for Efficient Off-Line Signature Identification*, Elsevier Engineering Applications of Artificial Intelligence, vol. 19, Issue 6, 2006, pp. 693-704.
- [98]. Coetzer J., Herbst B. M. and Preez J. A, *Offline Signature Verification Using the Discrete Radon Transform and a Hidden Markov Model*, EURASIP Journal on Applied Signal Processing , vol. 4, 2004, pp. 559–571.
- [99]. Fakhilai M, Pourreza H. R, Moarefdost R and Shadroo. S., *Off line signature recognition based on contourlet transform*, Int. Conf. on Machine Learning and Computing, vol.3 , 2011, pp 670-673.
- [100]. Yang L., Widjaja B., Prasad R., *Application of hidden Markov models for signature verification*, Elsevier Pattern Recognition, vol.28, No.2, 1995, pp.161–170.
- [101]. Shafiei M.M., Rabiee H.R., *A new online signature verification algorithm using variable length segmentation and hidden Markov models*. Proceedings of the 7th Int. Conf. on Document Analysis and Recognition, vol.1, 2003, pp. 443–446.

References

- [102]. Lee L. L., *Neural approaches for human signature verification*, Proceedings of the Third Int. Conf. on Signal Processing, 1996, pp. 1346–1349.
- [103]. Lejtman D. Z. and George S. E., *On-line handwritten signature verification using wavelets and back-propagation neural networks*, Proceedings of the 6th Int. Conf. on Document Analysis and Recognition, 2001, pp. 596-598.
- [104]. Zanuy M.F., *On-line signature recognition based on VQ-DTW*, Elsevier Pattern Recognition, vol. 40, No.3, 2007, pp. 981–992.
- [105]. Nanni L. and Lumini A., *A novel local on-line signature verification system*, Elsevier Pattern Recognition Letters, vol.29, No.5, 2008, pp.559–568.
- [106]. Roy S, Maheshkar S., *Offline Signature Verification using Grid based and Centroid based Approach*, Int. Journal of Computer Applications, vol 86, No. 8, 2014, pp. 35-39.
- [107]. Kumar R., Sharma J. D. and Chanda B., *Writer-independent off-line signature verification using surroundedness feature*, Elsevier Pattern Recognition Letters, vol. 33, 2012, pp. 301–308.
- [108]. Vargas J F, Ferrer M. A, Travieso C.M and Alonso J.B, *Off-line signature verification based on grey level information using texture features*, Elsevier Pattern Recognition, vol. 44, 2011, pp. 375–385.
- [109]. Shukla A.K , Mohan P., Ojha G. and Wariya M., *Offline Signature Verification System Using Grid and Tree Based Feature Extraction*, Int. IEEE Conf. on Issues and Challenges in Intelligent Computing Techniques, 2014, pp. 784-789.
- [110]. Daugman J, *How iris recognition works*, IEEE Trans. on Circuits and Systems for Video Technology, vol.14, No.1, 2004, pp. 21-30.
- [111]. Daugman J., *High Confidence Visual Recognition of Persons by a Test of Statistical Independence*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.15, No.11, 1993, pp.1148-1161.
- [112]. Wildes R. P., *Iris recognition: An Emerging Biometric Technology*, Proceedings of the IEEE, vol.85, No.9, 1997, pp.1348–1363.
- [113]. Boles W.W. and Boashash B., *A Human Identification Technique Using Images of the Iris and Wavelet Transform*, IEEE

- Transactions on Signal Processing, vol. 46, No.4, 1998, pp.1185-1188.
- [114]. Ma L., Tan T., Wang Y. and Zhang D., *Efficient Iris Recognition by Characterizing Key Local Variations*, IEEE Transactions on Image Processing, vol.13, No.6, 2004, pp. 739-750.
- [115]. Tan T., Sun Z., *Ordinal Measures for Iris Recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.31, No.12, 2009, pp. 2211-2226.
- [116]. Proenca H, Filipe S, Santos R, Olivera J and Alexandre L A, *The UBIRIS v2: A Database of visible wavelength Iris Images Captured on-the-move and at a distance*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, No.8, 2010, pp.1529-1535.
- [117]. Pillai J. K, Patel V.M., Chellappa, R. and Ratha N.K., *Secure and Robust Iris Recognition Using Random Projections and Sparse Representations*, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 33, No. 9 2011, pp: 1877 - 1893.
- [118]. Lim S., Lee K., Byeon O., and Kim T., *Efficient Iris Recognition through Improvement of Feature Vector and Classifier*, ETRI Journal, vol. 23, No. 2, 2001, pp. 61-70.
- [119]. Donald M. M, Rakshit S. and Zhang D, *DCT-Based Iris Recognition*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, No. 4, 2007, pp. 586-595.
- [120]. Kang B. J and Park K. R, *Real-Time Image Restoration for Iris Recognition Systems*, IEEE Transactions on Systems, Man, and Cybernetics, vol. 37, No. 6, 2007, pp. 1555-1566.
- [121]. Zhou. Z, Du Y., and Belcher C., *Transforming Traditional Iris Recognition Systems to Work in Non-ideal Situations*, IEEE Transactions on Industrial Electronics, vol. 56, No. 8, 2009, pp. 3203-3213.
- [122]. Rakvic R. N., Ullis B.J, Broussard R. P, Ives R.W., and Steiner N., *Parallelizing Iris Recognition*, IEEE Transactions on Information Forensics and Security, vol. 4, No. 4, 2009, pp. 812-823.
- [123]. Chou C.T, Shih S.W, Chen W.S, Cheng V. W, and Chen D.Y, *Non-Orthogonal View Iris Recognition System*, IEEE Transactions on Circuits and Systems for Video Technology, vol. 20, No. 3, 2010, pp. 417-430.

References

- [124]. Kekre. H. R, Sarode. T. K, Bharadi V.A, Agarwal A.A, Arora R.J, and Nair M.C, *Iris Recognition Using Vector Quantization*, IEEE Int. Conf. on Signal Acquisition and Processing, 2010, pp. 58-62.
- [125]. Dong W, Sun Z, and Tan T, *Iris Matching Based on Personalized Weight Map*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, No. 9, 2011, pp. 1744-1757.
- [126]. Tan T, Zhang X., Sun Z. and Zhang H., *Noisy iris image matching by using multiple cues*, Elsevier Pattern Recognition Letters, vol. 33, 2012, pp. 970–977.
- [127]. Hong. L and Jain A. K., *Integrating faces and fingerprints for personal identification*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.20, No.12, 1998, pp.1295-1307.
- [128]. Turk. M. and Pentland. A., *Eigen faces for recognition*, Journal of Cognitive Neuroscience, vol.3, No.1, 1991, pp. 71-86.
- [129]. Jain A. K, Hong L., and Bolle R., *On-line fingerprint verification*, IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 19, No.4, 1997, pp. 302-314.
- [130]. Jain A. K., Hong. L., and Kulkarni. Y., *A multimodal biometric system using fingerprint, face and speech*, Proceedings of the Second Int. Conf. on Audio and Video based Biometric Person Authentication(AVBPA), 1999, pp. 182-187.
- [131]. Jain. A. K. and Ross. A., *Learning user-specific parameters in a multi-biometric system*, Proceedings of the IEEE Int. Conf. on Image Processing, 2002, pp. 57-60.
- [132]. Ross A. and Jain A. K., *Information fusion in biometrics*, Elsevier Pattern Recognition Letters, vol. 24, No.13, 2003, pp. 2115-2125.
- [133]. Snelick. R., Indovina M., Yen J., and Mink A., *Multimodal Biometrics: Issues in design and testing*, Proceedings of Fifth Int. Conf. on Multimodal Interfaces, 2003, pp. 68-72.
- [134]. Phillips P.J, Moon H, Rizvi. S. A and Rauss P.J, *The FERET Evaluation Methodology for Face Recognition Algorithm*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, No.10, 2000, pp. 504-524.
- [135]. Snelick. R, Uludag U, Mink. A, Indovina. M, and Jain A. K, *Large scale evaluation of multimodal biometric authentication using state-of-the-art systems*, IEEE Transactions on Pattern

- Analysis and Machine Intelligence, vol. 27, No.3, 2005, pp. 450-455.
- [136]. Doddington G., Liggett W., Martin A., Przybocki M., and Reynolds D., *Sheeps, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 Speaker Recognition Evaluation*, Proceedings of International Conference on Spoken Language Processing, 1998, pp. 1351-1354.
- [137]. Brunelli R. and Falavigna D., *Person Identification using multiple cues*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.12, No.10, 1995, pp. 955-966.
- [138]. Bigun E.S., Bigun J., Duc. B, and Fischer S., *Expert conciliation for multimodal person authentication systems using Bayesian statistics*, Proceedings of First Int. Conf. on Audio and Video based Biometric Person Authentication(AVBPA), 1997, pp. 291-300.
- [139]. Pigeon S. and Vandendrope L., *M2VTS Multimodal Face Database Release 1.00*, Available at:<http://www.tele.ucl.ac.be/PROJECTS/M2VTS/m2fdb.html>, 1996.
- [140]. Verlinde. P and Cholet G., *Comparing decision fusion paradigms using k-NN based classifiers, decision trees and logistic regression in a multi-modal identity verification application*, Proceedings of the Second International Conference on Audio and Video based Biometric Person Authentication (AVBPA), 1999, pp.188-193.
- [141]. Chatzis V., Bors A. G., and Pitas I., *Multimodal decision-level fusion for person authentication*, IEEE Transactions on Systems, Man, and Cybernetics, vol. 29, No. 6, 1999, pp. 674-681.
- [142]. Ben-Yacoub S., Abdeljaoued Y., and Mayoraz E., *Fusion of face and speech data for person identity verification*, IEEE Trans. on Neural Networks, vol. 10, No.5, 1999, pp. 1065-1075.
- [143]. Messer K., Matas J., Kittler J., Luettin J., and Maitre G., *XM2VTSDB: The Extended M2VTS Database*, Proceedings of the Second Int. Conf. on AVBPA, 1999, pp. 72-77.
- [144]. Lades M., Vorbruggen J., Buhmann J., Lange J., Malburg C. V. D, and Wurtz. R., *Distortion invariant object recognition in the dynamic link architecture*, IEEE Transactions on Computers, vol. 42, No.3, 1993, pp. 300-311.

References

- [145]. Bimbot F., Magrin-Chagnolleau I., and Mathan L., *Second-order statistical measure for text-independent speaker identification*, Speech Communication, vol.17, 1995, pp.177-192.
- [146]. Fierrez-Aguilar J., Ortega-Garcia J., Garcia-Romero D. and Gonzalez-Rodriguez. J., *A comparative evaluation of fusion strategies for multimodal biometric verification*, Int. Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA), Springer LNCS-2688, 2003, pp.830–837.
- [147]. Kumar A., Wong D. C. M., Shen H. C. and Jain A. K., *Personal verification using palmprint and hand geometry biometric*, Proceedings of the 4th International Conference on Audio- Video-Based Biometric Person Authentication, J. Kittler and M. Nixon, Eds., vol. LNCS 2668, 2003, pp. 668–678.
- [148]. Frischholz. R and Dieckmann. U., *BioID: A multimodal biometric identification system*, IEEE Transactions on Computers, vol.33, No.2, 2000, pp. 64–68.
- [149]. Frischholz R.W, Boebel F.G, and Spinnler K.P., *Face recognition with the synergetic computer*, Proceedings of the Int. Conf. on Applied Synergetics and Synergetic Engineering, 1994, pp. 107–110.
- [150]. Wang Y., Tan T., and Jain A. K., *Combining face and iris biometrics for identity verification*, Proceedings of 4th Int. Conf. on AVBPA, 2003, pp. 805–813.
- [151]. BBC News. *Long lashes thwart ID scan trial*. Available at [urlhttp://news.bbc.co.uk/1/hi/uk_news/politics/3693375.stm](http://news.bbc.co.uk/1/hi/uk_news/politics/3693375.stm), May 2004.
- [152]. Ma L., Tan T., Wang Y., and Zhang D., *Personal identification based on iris texture analysis*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, No.12, 2003, pp.1519–1533.
- [153]. Metallinou A., Lee S. and Narayanan S., *Audio Visual Emotion-Recognition using Gaussian Mixture Model for Face and Voice*, Tenth IEEE Int. Symposium on Multimedia, 2008, pp 250-257.
- [154]. Kumatani K and Stiefelhagen R., *State Synchronous Modelling on Phone Boundary for Audio Visual Speech Recognition and Application to Multi-view Face Images*, IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2007, pp. 417-420.

- [155]. Kuar M, Giridhar A and Kaur M, *Multimodal Biometric Systems using speech and signature modalities*, Int. Journal of Computer Applications, vol. 5, No.12, 2010, pp. 13-16.
- [156]. Kale A., Chowdhury A. K. R, and Chellappa R., *Towards a view-invariant gait recognition algorithm*, Proceedings of IEEE Int. Conf. on Advanced Video and Signal based Surveillance, 2003, pp. 143–150.
- [157]. Zhou S., Krueger V., and Chellappa R., *Probabilistic recognition of human faces from video*, Computer Vision and Image Understanding, vol. 91, 2003, pp. 214–245.
- [158]. Kale A., RoyChowdhury A.K., and Chellappa R., *Fusion of gait and face for human identification*, Proceedings of the IEEE Int. Conf. on Speech, Acoustics, and Signal Processing, 2005, pp. 901-905.
- [159]. Beumier C., and Acheroy M., *Automatic face verification from 3D and grey level clues*, Eleventh Portuguese Conference on Pattern Recognition, 2000, pp. 315-321.
- [160]. Wang Y., Chua C., and Ho Y., *Facial feature detection and face recognition from 2D and 3D images*, Elsevier Pattern Recognition Letters, vol. 23, No. 10, 2002, pp.1191–1202.
- [161]. Chang K.I., Bowyer K.W., and Flynn P.J., *Multimodal 2D and 3D biometrics for face recognition*, Proceedings of the IEEE Int. Workshop on Analysis and Modeling of Faces and Gestures, 2003, pp. 187–194.
- [162]. Lu X., and Jain A., *Integrating range and texture information for 3D face recognition*, Proceedings of IEEE Computer Society Workshop on Application of Computer Vision, 2005, pp. 156-163.
- [163]. Xiuqin P., Xiaona X., Yong L., Youngcun C, *Feature fusion of multimodal recognition based on ear and profile face*, Proceedings SPIE on Biometric Technology for Human Identification, 2008, pp. 253-261.
- [164]. Rattani A, Tistarelli M., *Robust multimodal and multiunit feature level fusion of face and iris biometrics*, Int. Conf. of Biometrics, Springer, 2009, pp. 960–969.
- [165]. Ross A., Govindarajan R., *Feature level fusion using hand and face biometrics*, Proceedings of SPIE Conference on Biometric Technology for Human Identification, 2004, pp. 196–204,

References

- [166]. Toh K. A, Jiang X. D, Yau W.Y, *Exploiting global and local decisions for multi-modal biometrics verification*, IEEE Transactions on Signal Processing, vol.52, No.10, 2004, pp. 3059–3072.
- [167]. Veeramachaneni K., Osadciw L. A., , and Varshney P. K., *An Adaptive Multimodal Biometric Management Algorithm*, IEEE Transactions on Systems, Man, And Cybernetics, vol. 35, No. 3, 2005, pp. 535-543.
- [168]. Lupu C. and Lupu V., *Multimodal Biometrics for Access Control in an Intelligent Car*, 3rd Int. Symposium on Computational Intelligence and Intelligent Informatics (ISCIII), 2007, pp. 215-221.
- [169]. Yuan S., Zhang T., Zhou X, Liu X and Liu M, *An optical authentication system based on encryption technique and multimodal biometrics*, Elsevier Optics & Laser Technology, vol 54, 2013, pp 120–127.
- [170]. Huang Z , Liu Y, Li C, Yang M, Chen L, *A robust face and ear based multimodal biometric system using sparse representation*, Elsevier Pattern Recognition, vol 46, 2013, pp 2156–2168.
- [171]. Emerich S, Lupu E, Rusu C, *A new set of features for a bimodal system based on on-line signature and speech*, Elsevier Digital Signal Processing, vol 23, 2013, pp. 928–940.
- [172]. Poh N., Ross A., Lee W., Kittler J., *A user-specific and selective multimodal biometric fusion strategy by ranking subjects*, Elsevier Pattern Recognition, vol 46, 2013, pp 3341–3357.
- [173]. Rabiner L and Juang B.H., *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Clis, NJ, 1993.
- [174]. Atal B. S. and Hanauer S. L., *Speech Analysis and Synthesis by Linear Prediction of the Speech Wave*, Journal of the Acoustical Society of America ,vol. 50 No. 2 (Part 2), 1971, pp. 637 – 655.
- [175]. Laitinen N., Puhujantunnistus T., *Licentiate's thesis*, University of Helsinki, Department of Phonetics, Helsinki, Finland, 1999.
- [176]. Genoud D., and Chollet, G. *Speech pre-processing against intentional imposture in speaker recognition*, Proc. Int. Conf. on Spoken Language Processing (ICSLP) 1998, pp. 111-115.
- [177]. Huang, X., Acero, A., and Hon H. W., *Spoken Language Processing: a Guide to Theory, Algorithm, and System Development*, Prentice-Hall, New Jersey, 2001.

- [178]. Quatieri, T., Reynolds, D., and O'Leary G., *Estimation of handset nonlinearity with application to speaker recognition*, IEEE Trans. on Speech and Audio Processing, vol.8, No. 5, 2000, pp. 567–584.
- [179]. Phythian M., Ingram J., and Sridharan S., *Effects of speech coding on text-dependent speaker recognition*, Proc. of IEEE Conf. on Speech and Image Technologies for Computing and Telecommunications, 1997, pp. 137–140.
- [180]. Besacier L., Grassi S., Dufaux A., Ansorge M., and Pellandini F., *GSM speech coding and speaker recognition*, Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2000, pp. 1085–1088.
- [181]. Mammone R., Zhang X., and Ramachandran R., *Robust speaker recognition: A feature based approach*, IEEE Signal Processing Magazine, vol. 13, No.5, 1996, pp.58–71.
- [182]. Reynolds D., *An overview of automatic speaker recognition technology*, Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), 2002, pp. 4072–4075
- [183]. Tzanetakis G. and Perry Cook., *Musical Genre Classification of Audio Signals*, IEEE Transactions On Speech And Audio processing, vol. 10, No. 5, 2002, pp. 293-302.
- [184]. Phadke S, Limaye R., Verma S. and Kavitha S., *On Design and Implementation of an Embedded Automatic Speech Recognition System*, Proc. of the 17th IEEE International Conference on VLSI Design, 2004, pp. 127- 132.
- [185]. Kemerait R.C. and Childers D.G., *Signal Detection and Extraction by Cepstrum Techniques*, IEEE Trans. on Information Theory, vol.18, No.6, 1972, pp. 745-757.
- [186]. Tan Li and Karnjanadecha M., *Modified Mel-Frequency Cepstrum Coefficient*, Proc. of the Information Engineering Postgraduate Workshop, 2003, pp. 127-130.
- [187]. Davis S. and Mermelstein, P., *Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences*, IEEE Transactions on Signal Processing, vol.28, No.4, 1980, pp. 357 – 366.
- [188]. Molau S, Michael Pitz, Ralf Schluter and Hermann Ney, *Computing Mel-Frequency Cepstral Coefficients on the Power*

References

- Spectrum*, Proc. of the IEEE Int. Conf. on Acoustics, Speech and Signal Processing, vol.1, 2001, pp. 73-76.
- [189]. Skowronski M. D. and Harris J. G., *Improving the Filter Bank of a Classic Speech Feature Extraction Algorithm*, Proc. of the IEEE Int. Symposium on Circuits and Systems, vol.4, 2003, pp. 281-284
- [190]. Siniscalchi S.M, Fulvio Gennaro, Salvatore Andolina, Salvatore Vitabile, Antonio Gentile, and Filippo Sorbello, *Embedded Knowledge-based Speech Detectors for Real-Time Recognition Tasks*, Proc. of Int. Conf. on Parallel Processing Workshops, 2006, pp. 66-72.
- [191]. Bertocci G., Schoenherr B. W. and Messerschmitt D. G., *An approach to the Implementation of a Discrete Cosine Transform*, IEEE Trans. on Communications, vol.30, No.4, 1982, pp. 635 – 641.
- [192]. Linde Y, Buzo A, and Gray R., *An Algorithm for Vector Quantizer Design*, IEEE Transactions On Communications, vol.28, No.1, 1980, pp. 84 – 95
- [193]. Lin Y. and Shen-Chuan Tai, *A fast Linde-Buzo-Gray algorithm in image vector quantization*, IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing, vol.45, No.3, 1998, pp. 432-435.
- [194]. Matsui T., and Furui S., *Comparison of text-independent speaker recognition methods using VQ-distortion and Discrete/continuous HMMs*, IEEE Trans. Speech Audio Process., vol. 2, No.3, 1994, pp. 456-459.
- [195]. Gray R., *Vector quantization*, IEEE Acoustics, Speech, Signal Process. Mag., vol. 1, 1984, pp. 4-29.
- [196]. Gose E., Richard J. and Jost S., *Pattern Recognition and Image Analysis*, Prentice-Hall, 2002.
- [197]. Hietmeyer R., *Biometric identification promises fast and secure processing of airline passengers*, Int. Civil Aviation Organization Journal, vol. 55, no. 9, 2000, pp. 10–11.
- [198]. Bartlett M. S., Lades H.M., and Sejnowski T.J., *Independent component representations for face recognition*, Proc. of SPIE, vol. 3299, 1998, pp. 528–539.

- [199]. Swets D. L. and Weng J., *Using discriminant eigenfeatures for image retrieval*, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 8, 1996, pp. 831–836.
- [200]. Cootes T. F, Taylor, C. J. and Edwards, G. J, *Face recognition using active appearance models*, Proceedings ECCV, vol. 2, 1998, pp. 581–695.
- [201]. Blanz V., Huang J., and Heisele B., *Component-based Face Recognition with 3D Morphable Models*, Proceedings of the Fourth International Conference on Audio- and Video-based Biometric Person Authentication, Survey, UK, 2003.
- [202]. Blanz V. S. R. and Vetter, T. *Face identification across different poses and illuminations with a 3D morphable model*, IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2002, pp. 202–207.
- [203]. Hyvarinen A., *Survey on independent component analysis*, Neural Computing Surveys, vol.2, 1999, pp. 94–128.
- [204]. Kirby M. and Sirovich L., *Application of the Karhunen-Lo´eve procedure for the characterization of human faces*, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, No. 1, 1990, pp. 103–108.
- [205]. Sirovich L. and Kirby, *Low-dimensional procedure for the characterization of human faces*, Journal of Optical Society of America, vol. 4, No.3, 1987, pp. 519-524.
- [206]. Tat Jun Chin and David Suter, *A Study of the Eigen face Approach for Face Recognition*, Technical Report MECSE, 2004, pp. 302-309.
- [207]. Plamondon R. and Srihari S. N, *On-line and Offline Handwriting Recognition: A Comprehensive Survey*, IEEE Trans. Pattern Recognition and Machine Intelligence, vol. 22, No.1, 2000, pp.63-84.
- [208]. Leclerc F. and Plamondon R., *Automatic Signature Verification*, Int. Journal Pattern Recognition and Artificial Intelligence, vol. 8 no. 3, 1994, pp. 643-660.
- [209]. Raafat H., Tolba A.S., Almayaan W., *A glove-based system for automated signature identification using neural networks*, Int. Conf. on Systems Analysis, and Synthesis (ISAS '99), 1999, pp. 328-333.

References

- [210]. Jain A., Uludag U., *Hiding fingerprint minutiae in images*, Proc. of Third Workshop on Automatic Identification Advanced Technologies, 2002, pp. 97-102.
- [211]. Jain A. K., Ross A., Pankanti S., *Biometrics: A tool for information security*, IEEE Trans. on Information Forensics and Security, vol.1, No.2, 2006, pp.125-143.
- [212]. Plamondon R., Lorette G., *Automatic signature verification and writer identification- the state of the art*, Elsevier Pattern Recognition, vol. 22, 1989, pp. 107–131.
- [213]. Jiang W., Lam K. and Shen T. Z., *Edge Detection using simplified Gabor wavelet*, IEEE Int. conference Neural Networks and Signal Processing, 2008, pp. 586-591.
- [214]. Lei H., Govindaraju V., *A study on the consistency of features for on-line signature verification*, Structural, Syntactic, and Statistical Pattern Recognition, Springer- Heidelberg, 2004, pp.444-451.
- [215]. Flom L., Safir A., *Iris recognition system*, US Patent 4641394, 1987.
- [216]. Daugman J., *Iris recognition*, American Scientist, vol.89, No.4, 2001, pp.326–333.
- [217]. Oyster C., *The Human Eye Structure and Function*, Sinauer Associates, 1999.
- [218]. Marieb E. N., *Human Anatomy and Physiology*, Benjamin-Cummings Publishing Company, 7th edition, 2007.
- [219]. Pierscionek B., Crawford S., Scotney B., *Iris recognition and ocular biometrics- the salient features*. Proc. of Machine Vision and Image Processing Conference IMVIP '08. International, 2008, pp. 170-175.
- [220]. Barros J., French J., Martin W., *Indexing Multi-Spectral Images for Content-Based Retrieval*, University of Virginia Technical Report, CS-94-40, 1994.
- [221]. Anderson R. J.. *Security Engineering: A Guide to Building Dependable Distributed Systems*. New York: Wiley, 2001.
- [222]. Barros J., French J., Martin W., *System for indexing multi-spectral satellite images for efficient content-based retrieval*, Proc. of the SPIE, vol. 2420, 1995, pp. 228–237.

- [223]. Giacinto G., Roli F., *Dynamic classifier selection based on multiple classifier behavior*, Elsevier Pattern Recognition, vol.34, 200, pp.1879–1881.
- [224]. Kumar A., Wong D. C., Shen H. C., Jain A.K., *Personal authentication using hand images*, Elsevier Pattern Recognition Letters, vol.27, No.13, 2006,pp.1478–1486.
- [225]. Proença H., Alexandre L.A., *Iris Recognition: Measuring Feature's Quality for the Feature Selection in Unconstrained Image Capture Environments*, IEEE Proc. of Int. Conf. on Computational Intelligence for Homeland Security and Personal Safety (CIHSPS), vol. 1, 2006, pp.35-40.
- [226]. Bowyer K., Hollingsworth K., Flynn P., *Image Understanding for Iris Biometrics: A Survey*, Computer Vision and Image Understanding, vol. 110, No.2, 2008, pp. 281–307.
- [227]. Daugman J., Downing C., *Epigenetic randomness, complexity and singularity of human iris patterns*, Proceedings of the Royal Society of London - B, vol.268, 2001, pp. 1737–1740.
- [228]. Zhang D., *Detecting eyelash and reflection for accurate iris segmentation*, Int. Journal of Pattern Recognition and Artificial Intelligence, vol. 1, No.6, 2003, pp.1025-1034.
- [229]. Gonzalez R. and Woods R., *Digital Image Processing*, Addison Wesley, 1992, pp. 414 - 428.
- [230]. Canny J., *A computational approach to edge detection*, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 8, 1986, pp. 679-698.
- [231]. Falkowski B. J., *Forward and inverse transformations between Haar wavelets and arithmetic functions*, Electronics Letters, vol. 34, no. 11, 1998, pp. 1084-1085.
- [232]. Grochenig K. and Madych W. R., *Multiresolution Analysis, Haar Bases, and Self-similar Tilings of R*, IEEE Transactions on Information Theory, vol. 38, No. 2, 1992, pp. 556-568.
- [233]. Jain A.K, Nandakumar K., Ross A., *Score Normalization in multimodal biometric systems*, Pattern Recognition, vol. 38, 2005, pp. 2270-2285.
- [234]. Kittler J., Hatef M., Duin R. and Matas J., *On Combining Classifiers*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, No.3, 1998, pp. 226-236.

References

- [235]. Wang Y. and Tan T., *Combining fingerprint and voice biometrics for identity verifications-An Experimental Comparison*, Proc. of Int. Conf. on Biometric Authentication (ICBA), Springer LNCS 3072, 2004, pp. 663-670.
- [236]. Bigun J, Fierrez-Aguilar, Ortega-Garcia and Gonzalez R, *Multimodal Biometric Authentication using quality signals in mobile communications*, Proc. of Int. Conf. on Image Analysis and Processing, IEEE CS Press, 2003, pp. 2-13.

List of Publications

1. **Binsu C. Kovoov**, Supriya M.H and K. Poullose Jacob, *Speaker Identification Biometric System using Formant Analysis*, Proceedings of National Conference in Software Engineering (NCSOFT),CUSAT, Kochi, India, 2008.
2. **Binsu C. Kovoov**, Supriya M.H and K. Poullose Jacob, *Parametric study on speaker identification biometric system using formant analysis*, Journal of Acoustical Society of America, vol. 125, No.4, pp. 2530, 2009.
3. **Binsu C. Kovoov**, Supriya M.H and K. Poullose Jacob, *Streaming audio in on-line biometric security system*, Proceedings of All India Seminar on Innovations in Electronics Technology for Futuristic Communications, Division of Institution of Engineers, India, pp. 79-84, 2009.
4. **Binsu C. Kovoov**, Supriya M.H and K. Poullose Jacob, *Implementation of Multimodal Biometric Security System based on Audio and Video Streaming*, Proceedings of National Symposium on Acoustics(NSA),RCI, Hyderabad, India, pp. 26, 2009
5. **Binsu C. Kovoov**, Supriya M.H and K. Poullose Jacob, *A Prototype For a Multimodal Biometric Security System Based on Face and Audio Signatures*, International Journal of Computer Science (IJCS), vol 2, No.1, pp. 143-147, Jan-June 2011.
6. **Binsu C. Kovoov**, Supriya M.H and K. Poullose Jacob, *Iris Biometric Recognition System Employing Canny Operator*, Proceedings of International Conference on Computational Science and Engineering (CSE 2013), Dubai, UAE, pp. 65-74, 2013.
7. **Binsu C. Kovoov**, Supriya M.H and K. Poullose Jacob., *Effectiveness of Feature Detection operators on the performance of iris biometric recognition system*, International Journal of Network Security & Its Applications(IJNSA), vol 5, No.5, pp. 73-81, Sept 2013.
8. **Binsu C. Kovoov**, Supriya M.H and K. Poullose Jacob., *Multimodal Behavioural Biometric Authentication based on handwritten signature and speech*, International Journal of Advanced Research in Computer Science and Software Engineering vol.4, No.9, pp. 530-535, September 2014.

Subject Index

A

Acceptability 35
Accuracy..... 26, 85, 245
ambient noise 122, 236
Appearance-based Face Recognition..... **146**

B

Biometrics ii, 21, 22, 26, 42, 117, 243, 244,
254, 257, 258, 262, 263

C

Canny **184, 197, 198, 203, 207, 208, 234, 263**
Cepstral..... 122, 123, 259
cepstral analysis..... 122, 237
Cepstral Coefficients 123, 259
Circumvention..... 35
Classification **145, 175, 246, 249, 259**
Collectability 35

D

DCT..... **72, 75, 81, 124, 127, 136, 199, 253**
Decimal Scaling **213**
Decision logic **132**
DFT..... **125, 127**
discrete wavelet transform.....*See* DWT
Distinctiveness 35
DWT **199, 200, 201**
dynamic features ...**44, 46, 70, 104, 106, 111,**
166, 174, 175, 181, 182, 183, 233

E

Edge detection **197**
EER24, 25, 46, 57, 58, 76, 78, 81, 93, 97, 106,
107, 114, 179, 183, 203, 211, 218, 220,
230
Elastic Bunch Graph Matching ... **61, 147, 248**
enrolment .. 22, 23, 27, 28, 29, 30, 32, 47, 48,
78, 105, 106, 116, 117, 168, 176, 209
Equal Error Rate*See* EER
Euclidean distance 238

F

face **22, 26, 29, 30, 32, 37, 41, 43, 45, 46, 48,**
58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 82,
87, 89, 90, 92, 93, 94, 95, 96, 97, 98, 99,
100, 102, 104, 105, 106, 107, 109, 110,
113, 115, 139, 142, 143, 144, 145, 146,
147, 148, 149, 150, 151, 152, 153, 155,
157, 158, 159, 160, 161, 162, 164, 185,
209, 219, 226, 227, 230, 232, 233, 234,
235, 247, 249, 254, 255, 256, 257, 258,
260
Failure to Capture Rate*See* FTCT
Failure to Enroll Rate*See* FTER
False Match Rate 24, *See* FMR
False Non Match Rate.....*See* FNMR
FAR. 24, 25, 36, 76, 77, 85, 86, 88, 89, 91, 92,
93, 95, 96, 97, 101, 102, 106, 114, 138,
139, 179, 211, 220, 235
Feature Extraction ... **172, 195, 196, 252, 260**
Feature Level fusion..... 38
feature vector 237, 238, 239, 241

Subject Index

Features 122
FER 25
FMR 24, 25, 27
Fourier transform **123, 199**
FRR. 25, 36, 76, 85, 86, 88, 97, 101, 102, 106,
114, 138, 139, 179, 211, 220, 235
FTC 26
FTCR 24, 27, 29
FTE 25, 98
FTER 24, 27, 29

G

Gabor... **44, 46, 61, 66, 72, 77, 78, 80, 95, 96,**
111, 148, 166, 172, 173, 174, 177, 178,
183, 248, 262
GAR 26, 27, 89, 91, 92, 93, 96, 114
Gaussian..... **50, 56, 68, 78, 83, 96, 173, 174,**
193, 197, 198, 238, 246, 256

H

Hamming Distance..... **77, 201**
harmonic structure 123
Hierarchical mode..... 34
Hierarchical Target Trimming Approach.. 238

I

identification..... **22, 23, 24, 27, 28, 29, 30, 43,**
44, 46, 47, 49, 51, 52, 54, 55, 60, 64, 66,
67, 68, 70, 72, 87, 93, 97, 105, 106, 111,
116, 117, 118, 122, 142, 146, 150, 151,
152, 160, 161, 169, 172, 176, 183, 185,
201, 203, 215, 216, 226, 228, 229, 230,
231, 254, 256, 257, 260, 261, 262

Iris **77, 87, 98, 108, 111, 184, 185, 186, 187,**
188, 189, 190, 191, 194, 202, 221, 228,
252, 253, 254, 262, 263

K

knowledgebase ... **26, 45, 106, 111, 113, 115,**
166, 169, 172, 190, 218

L

LBG..... **83, 109, 129, 137**
Linear Subspace **149**
localization. **79, 111, 144, 184, 188, 190, 194,**
197, 199

M

Machine Readable Travel Documents **144**
MAD **85, 92, 214**
Mahalanobis Distance..... **166, 175, 179**
Median Absolute Deviation See MAD
Median-MAD **212, 214**
Mel Frequency Cepstral Coefficients 123, See
MFCC
MFCC **45, 49, 50, 58, 104, 105, 109, 116, 123,**
124, 126, 129, 130, 131, 133, 136, 140,
141, 233, 245, 247
Min-max Normalization **213**
Model-based Face Recognition..... **147**
multimodal. 21, 23, 29, 30, 31, 32, 34, 36, 44,
45, 46, 48, 49, 84, 85, 88, 89, 90, 91, 92,
93, 94, 97, 98, 99, 100, 101, 102, 103,
104, 105, 106, 112, 113, 115, 209, 215,
216, 219, 226, 228, 230, 231, 232, 233,
235, 254, 255, 256, 257, 258, 263

Subject Index

209, 226, 232, 233, 235, 246, 247, 254,
255, 257, 258, 259
Static Features **172**
success rates 241

T

Tanh transformation **214**
target
 identification, classification 236, 237, 238,
 240, 241
target feature record 238
Template-based Face Recognition **145**

U

UBIRIS **79, 85, 108, 202, 203, 253**
Universality 35
User-Dependent Fusion Strategy **211**

V

Vector Quantization. **45, 49, 68, 80, 129, 246,**
254
verification . 22, 23, 27, 28, 29, 31, 32, 46, 47,
48, 50, 51, 53, 54, 56, 67, 68, 69, 70, 71,
72, 73, 74, 75, 76, 88, 89, 90, 92, 94, 96,
97, 98, 102, 105, 117, 168, 180, 181, 183,
245, 246, 247, 250, 251, 252, 254, 255,
256, 257, 258, 262
voice..... 22, 28, 30, 45, 46, 48, 49, 94, 95, 97,
100, 102, 104, 105, 109, 116, 117, 119,
132, 219, 264

W

Window function **127**

Z

zero-normalisation..... *See* Z-norm
Z-norm **212, 213**