

2008-01-14

A Multi-Modal Approach for Face Modeling and Recognition

Mohammad Hossein Mahoor

University of Miami, m.mahoor@umiami.edu

Follow this and additional works at: https://scholarlyrepository.miami.edu/oa_dissertations

Recommended Citation

Mahoor, Mohammad Hossein, "A Multi-Modal Approach for Face Modeling and Recognition" (2008). *Open Access Dissertations*. 32.
https://scholarlyrepository.miami.edu/oa_dissertations/32

This Open access is brought to you for free and open access by the Electronic Theses and Dissertations at Scholarly Repository. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of Scholarly Repository. For more information, please contact repository.library@miami.edu.

UNIVERSITY OF MIAMI

A MULTI-MODAL APPROACH FOR FACE MODELING AND RECOGNITION

By

Mohammad Hossein Mahoor

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

December 2007

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

A MULTI-MODAL APPROACH FOR FACE MODELING AND RECOGNITION

Mohammad Hossein Mahoor

Approved:

Dr. Mohamed Abdel-Mottaleb
Associate Professor of Electrical
and Computer Engineering

Dr. Terri A. Scandura
Dean of Graduate School

Dr. Shahriar Negahdaripour
Professor of Electrical
and Computer Engineering

Dr. Kamal Premaratne
Professor of Electrical and
Computer Engineering

Dr. James W. Modestino
Professor of Electrical and
Computer Engineering

Dr. Ahmad El-gammal
Assistant Professor of
Computer Science
Rutgers University

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Dr. Mohamed Abdel-Mottaleb.
No. of pages in text. (169)

This dissertation describes a new methodology for multi-modal (2-D + 3-D) face modeling and recognition. There are advantages in using each modality for face recognition. For example, the problems of pose variation and illumination condition, which cannot be resolved easily by using the 2-D data, can be handled by using the 3-D data. However, texture, which is provided by 2-D data, is an important cue that cannot be ignored. Therefore, we use both the 2-D and 3-D modalities for face recognition and fuse the results of face recognition by each modality to boost the overall performance of the system.

In this dissertation, we consider two different cases for multi-modal face modeling and recognition. In the first case, the 2-D and 3-D data are registered. In this case we develop a unified graph model called Attributed Relational Graph (ARG) for face modeling and recognition. Based on the ARG model, the 2-D and 3-D data are included in a single model. The developed ARG model consists of nodes, edges, and mutual relations. The nodes of the graph correspond to the landmark points that are extracted by an improved Active Shape Model (ASM) technique. In order to extract the facial landmarks robustly, we improve the Active Shape Model technique by using the color information. Then, at each node of the graph, we calculate the response of a set of log-Gabor filters applied to the facial image texture and shape information

(depth values); these features are used to model the local structure of the face at each node of the graph. The edges of the graph are defined based on Delaunay triangulation and a set of mutual relations between the sides of the triangles are defined. The mutual relations boost the final performance of the system. The results of face matching using the 2-D and 3-D attributes and the mutual relations are fused at the score level.

In the second case, the 2-D and 3-D data are not registered. This lack of registration could be due to different reasons such as time lapse between the data acquisitions. Therefore, the 2-D and 3-D modalities are modeled independently. For the 3-D modality, we developed a fully automated system for 3-D face modeling and recognition based on ridge images. The problem with shape matching approaches such as Iterative Closest Points (ICP) or Hausdorff distance is the computational complexity. We model the face by 3-D binary ridge images and use them for matching. In order to match the ridge points (either using the ICP or the Hausdorff distance), we extract three facial landmark points: namely, the two inner corners of the eyes and the tip of the nose, on the face surface using the Gaussian curvature. These three points are used for initial alignment of the constructed ridge images. As a result of using ridge points, which are just a fraction of the total points on the surface of the face, the computational complexity of the matching is reduced by two orders of magnitude. For the 2-D modality, we model the face using an Attributed Relational Graph. The results of the 2-D and 3-D matching are fused at the score level.

There are various techniques to fuse the 2-D and 3-D modalities. In this dissertation, we fuse the matching results at the score level to enhance the overall perfor-

mance of our face recognition system. We compare the Dempster-Shafer theory of evidence and the weighted sum rule for fusion. We evaluate the performance of the above techniques for multi-modal face recognition on various databases such as Gavab range database, FRGC (Face Recognition Grand Challenge) V2.0, and the University of Miami face database.

*To my beloved mother, deceased father,
and
to my lovely wife.*

ACKNOWLEDGMENTS

First, I would like to express my sincere gratitude to Professor Abdel-Mottaleb, who introduced me to the field of computer vision, and helped me to jump start my research career in this field. He encouraged me to pursue novel ideas and provided me with exceptional experience and knowledge.

My thanks also go to the other members of the committee, Dr. Shahriar Negahdaripour, Dr. James W. Modestino, Dr. Kamal Premaratne, and Dr. Ahmad El-gammal for their valuable comments and suggestions. In particular, I want to thank Dr. Negahdaripour for his valuable support, comments and encouragement during my doctoral studies.

More than everyone, I am indebted to my mother for her enthusiastic encouragement, prayers and unlimited support during all the stage of my life. Also, I must express my appreciation to my beloved wife Eshrat, to my brothers, and to my sisters.

I would like to extend my thanks to the faculty and staff members of the Electrical and Computer Engineering Department at the University of Miami, especially Ms. Clarisa Alvarez, Ms. Rosamund Coutts, and Ms. Marisol Ternas for their role in my education and various resources made available to me to do my research.

Also, I would like to thank my fellow lab members, specially, Dr. Nasser Al-Ansari, Dr. Pezhman Firoozfam, Dr. Omaila Nomir, Dr. Charay Leurdwick, Hamed Pirsivash, Feng Niu, Ali Taatian, Steven Cadavic, Jindan Zhou, Hossein Madjidi,

Behzad M. Dogahe, Dr. Hongsheng Zhang, Muhammad Rushdi, and Dr. Nuno Gracias for their collaboration during the completion of this dissertation.

In addition, I want to thank Dr. Tapia, Dr. Asfour, and Mr. Ali Habashi for their support and help during my study at the University of Miami.

And last but not least, I am indebted to my landlord, Mr. Carlos Fernandez, and Ms. Maria C. Fernandez for their kind and generous hospitality during the past four years that I lived with them. Specially, Maria treated me as her son when I became sick and cooked delicious Mexican food for me.

Contents

• List of Figures	x
• List of Tables	xvi
• List of Acronyms	xviii
1 Introduction	1
1.1 Biometric System	2
1.1.1 History of Biometrics	2
1.1.2 Types of Biometrics	3
1.1.3 Motivation and Overview	4
1.1.4 Biometric Authentication System	5
1.1.5 Biometric Market	7
1.2 Face Recognition	9
1.2.1 Applications of Face Recognition	10
1.2.2 Face Recognition Challenges	15
1.3 Proposed Face Modeling and Recognition System	18
1.3.1 3-D Face Recognition	21
1.3.2 A Unified Approach for 2-D/3-D Face Recognition	24

1.4	Dissertation Contributions	25
1.5	Dissertation Outline	27
2	Related Work	28
2.1	Facial Features Extraction	28
2.1.1	2-D Facial Features Extraction From Frontal Images	29
2.1.2	3-D Facial Features Extraction	36
2.2	Face Recognition	38
2.2.1	2-D face Recognition	39
2.2.2	Holistic-based Approaches	40
2.2.3	Geometric-based Approaches	47
2.2.4	3-D Face Recognition	50
2.2.5	Multi-modal Face Recognition	53
3	Facial Features Extraction	60
3.1	3-D Facial Features Extraction	60
3.1.1	Face Localization Using Template Matching	61
3.1.2	Labeling Feature Points Using Gaussian Curvature	64
3.1.3	Experiments and Results	65
3.2	2-D Facial Features Extraction	69
3.2.1	Active Shape Model and Its Limitations	70
3.2.2	Experiments and Results	73
3.3	Summary	77

4	3-D Face Recognition Based on Ridge Images	79
4.1	3-D Face Matching Based on Ridge Images	80
4.2	Ridge Image	81
4.2.1	Ridge Matching Using Robust Hausdorff Distance	83
4.2.2	Ridge Matching Using Iterative Closest Points	86
4.2.3	Computational Complexity of Ridge Matching	87
4.3	Experiments and Results	88
4.3.1	Experiments on Gavab Database	88
4.3.2	Experiments on FRGC V2.0 Face Database	92
4.3.3	Recognition Using the Ridge Points, Random set of Points, and Entire Face Surface	96
4.3.4	More Results on FRGC V2.0	97
4.4	Summary	99
5	Multi-modal Face Recognition	101
5.1	A Graph Approach For 2-D and 3-D Face Recognition	102
5.2	Graph Modeling	104
5.2.1	Building the Attributed Graph	105
5.2.2	Feature Selection	112
5.3	Recognition; Matching the ARG Models	115
5.3.1	Pose Normalization	116

5.4	Fusing the Information: the 2-D and 3-D	
	Attributes and the Mutual Relation	117
5.4.1	Score Level Fusion	119
5.5	Experiments and Results	123
5.5.1	Experiments on the University of Miami Face Database	123
5.5.2	Accuracy Analysis of the ARG Model For Face Recognition	128
5.5.3	Experiments on the FRGC V2.0 Face Database	129
5.6	Summary	133
6	Conclusion and Future Research Direction	136
6.1	Summary and Contributions	136
6.2	Future Work	139
A	Enhancement of Active Shape Model Using Color Information	141
A.1	Shape Model Initialization and Face Alignment	142
A.2	Improvement of The Local Structure Model	143
A.3	Enhancement of the Location of Features Around the Lips	145
B	The Iterative Closest Point Algorithm	149
C	Sequential Floating Forward Selection Procedure	151
D	Dempster-Shafer Theory of Evidence	153
	Bibliography	156

List of Figures

1.1	Annual biometric industry revenues for the years 2007-2012 [8].	8
1.2	The percent of biometric market by technology in 2006 [8].	9
1.3	The general block diagram of our system for multi-modal face recognition based on 3-D ARG models.	20
1.4	The general block diagram of our system for multi-modal face recognition based on ridge images and 2-D ARG models.	21
1.5	Comparison between the three categories of algorithms for 3-D face recognition, performance of the system versus complexity.	23
1.6	Samples of extracted ridge images.	24
2.1	Labeled facial features in frontal and profile images.	29
2.2	Example of eigenfaces.	42
2.3	The graph representation of faces used in elastic bunch graph matching [163].(a)Elastic graph representation. (b)Bunch graph.	48
3.1	Template matching: (a) Template range image (b) A sample filtered facial range image (c) Detected area of the face in the sample range image.	63

3.2	The relations between surface types and their mean (H) and Gaussian (K) curvatures.	64
3.3	Three feature points labeling: (a) Gaussian curvature on a patch around the nose and eyes (b) Result of thresholding the Gaussian curvature image(c) Final result of feature points labeling	66
3.4	2-D and 3-D range views of an individual: 2-D gray scale images, 3-D range images at 1/4 of the original resolution (both from scanner's point of view) and a rotated version of 3-D range image [115].	67
3.5	Samples of range images in the gallery and the results of preprocessing (a) Original range images, (b) Noise removal and interpolation, (c) Face localization and three feature points labeling.	68
3.6	Extracted facial features sample images from UM database, where our method outperforms the standard method, (o) Improved ASM, (*) Standard ASM. The MSE (pixels) of the enhanced to standard ASM for the images are (a) [28.85, 39.98], (b) [32.74, 33.54], (c) [20.50, 38.85], and (d) [39.29, 110.87].	75
3.7	Extracted facial features sample images from UM database, where standard method outperforms our method, (o) Improved ASM, (*) Standard ASM. The MSE (pixels) of the enhanced to standard ASM for the images are (a) [34.64, 24.87] and (b) [14.02, 9.55].	77
4.1	Block diagram of our system for 3-D face recognition from range data.	80
4.2	Sample of ridge image extracted for different subjects.	82

4.3	Samples of range images in the gallery and the results of preprocessing (a) Original range images, (b) Noise removal and interpolation, (c) Face localization and three feature points labeling, (d) Ridge images .	89
4.4	CMC curves for frontal neutral images in Gavab database based on ridge images matched using the ICP and Hausdorff distance techniques.	91
4.5	ROC curves for frontal neutral images in Gavab database based on ridge images matched using the ICP and Hausdorff distance techniques.	91
4.6	Samples of facial images in the FRGC V2.0 database (texture, range, and extracted ridge images.)	94
4.7	CMC curve for 370 subjects in the FRGC V2.0 database based on ridge images matched with the ICP and Hausdorff techniques.	95
4.8	ROC curve for the neutral faces versus neutral in FRGC2.0 database (total of 2365 facial images for 432 individual subjects); ICP technique is used for matching the ridge images.	98
4.9	ROC I: the data are within the semesters. ROC II: the data are within the year. ROC III: the images are between the semesters.	99
5.1	General block diagram of our system for face modeling and recognition based on Attributed Relational Graph.	103
5.2	A set of 1-D log-Gabor filters seen in logarithmic and linear frequency scales.	108

5.3	Two pairs of log-Gabor wavelets, odd-symmetric and even-symmetric (all tuned to the same frequency, $w_0 = 1$, but having bandwidths of 1 and 3 octaves respectively.)	109
5.4	Angular component at angle = 0 and sigma = 1.5.	110
5.5	Even-symmetric and Odd-Symmetric 2-D log-Gabor filters at wave- length = 1 with 3 octaves bandwidth.	111
5.6	The mutual relation between the nodes.	112
5.7	Extracted landmark points. 57 landmark points (white color) are ex- tracted by the ASM method and the 52 landmark points (blue color) are added by aligning a standard template to the facial points.	113
5.8	Maps that show number of the selected (a) 2-D and (a) 3-D attributes at each node of the ARG model.	114
5.9	Results of feature selection for 3-D and 2-D attributes in the ARG model.	115
5.10	Neutral 3-D model used to estimate and normalize the pose of a de- formed 3-D model. The refernce model is at angles (pitch, yaw, roll) = (0,0,0).	117
5.11	3-D textured model (a) before normalization and (b) after normalization.	118
5.12	Fusion can be accomplished at various levels in a biometric system [140].	118
5.13	Dual threshold scenarios and uncertainty in a verification mode.	123
5.14	Examples of subjects in the UM database: (a) 2-D images and (b) 3-D textured models.	125
5.15	Cumulative Match Characteristic curves using the ARG modeling and two different fusion techniques on the UM database.	126

5.16	Results of rank-one identification using the ARG modeling and two different fusion techniques on the UM database and the results of other matching techniques applied to the UM database.	126
5.17	Receiver Operating Characteristic curves (ROC) using the ARG modeling and two different fusion techniques on the UM database.	127
5.18	Imposing uncertainty and then applying DS rule of combination to fuse the three match scores resulted from ARG modeling (the UM face database), the Equal Error Rate was improved by 1.5%.	128
5.19	Analyzing the robustness of the ARG model in face recognition with respect to added noise in location of the nodes. Results of rank-one identification are represented for 2-D attributes, 3-D attributes, mutual relations, and multi-modal fusion with respect to various values of σ of the Gaussian noise.	129
5.20	CMC curves. (a)for 2-D, 3-D, Mutual relations features, and fusion using DS rule and weighted sum rule for the neutral images versus neutral in FRGC V2.0 data-set; (b) the details in (a).	132
5.21	Results of rank-one identification using the 2-D ARG modeling, 3-D ridge images, and two different fusion techniques on the FRGC database.	132
5.22	ROC curves for 2-D, 3-D, Mutual relations features, and fusion using DS rule and weighted sum rule for the neutral images versus neutral in FRGC V2.0 data-set.	133

5.23	Samples of facial images in FRGC V2.0 database show various factors for poor quality of the dataset: (a) closed eyes, (b) facial hair, (c) poor illumination, (d) occluded facial regions (e) closed eyes and poor illumination and occlusion of facial regions, (f) occluded facial regions.	134
A.1	Detecting the center of the mouth and the eyes as the salient features for initialization.	144
A.2	Searching along a sampled profile to find the best fit.	145
A.3	Lips detection using FDA. (a) Original image.(b) Result of FDA classification. (c) Result of thresholding. (d) Result of applying morphological operators.	148

List of Tables

1.1	Typical applications of face.	10
1.2	Variations in facial appearance Inter-person and intra-person variations.	16
2.1	Facial features extraction techniques.	31
2.2	Classification of 2-D face recognition techniques [179].	41
3.1	Performance Comparison Between the Two Methods and the Manually Labeled Features Based on the Average MSE (pixels) Over All 70 Subjects.	76
4.1	Iterative algorithm to find the optimum pose in Hausdorff distance matching.	85
4.2	Results of first-ranked recognition rate on the Gavab face database using range data.	90
4.3	Comparing the results by Moreno <i>et al.</i> [118, 117, 116] and Ansari [11] on Gavab database and our work.	93

4.4	Comparison between the ridge points, random points selection, and entire surface based on the ICP matching technique; Results are in term of rank-one identification rate (%).	96
4.5	Verification rates (%) at 0.1% FAR for the ROC I, II, and III of the neutral v.s. neutral images.	98
B.1	The iterative closest point algorithm.	150
C.1	Sequential Floating Forward Selection algorithm for feature selection [139].	152

List of Acronyms

2-D: Two Dimensional

3-D: Three Dimensional

ARG: Attributed Relational Graph

ASM: Active Shape Model

CMC: Cumulative Match Characteristic

EBGM: Elastic Bunch Graph Matching

FDA: Fisher Discriminant Analysis

FM: False Match

FMR: False Match Rate

FNM: False Non Match

FNMR: False Non Match Rate

FRGC: Face Recognition Grand Challenge

FRVT: Face Recognition Vendor Tests

HD: Hausdorff distance

ICP: Iterative Closest Points

LTS-HD: Least Trimmed Square-HD

MSE: Mean Square Error

PCA: Principal Component Analysis

ROC: Receiver Operating Characteristic

SFFS: Sequential Floating Forward Selection

Chapter 1

Introduction

Face is a particularly compelling biometric because it is used everyday by everyone as the primary means for recognizing other humans. Face recognition is an important capability of the human perception system. Automated face recognition has become one of the most important applications of image analysis and computer vision in recent years. This has attracted the attention of many groups in research institutes, academia, industries and governmental agencies. This is obvious from the increased number of face recognition contests such as FERET [135], XM2VTS [113], FRVT 2000 [26], FRVT 2002 [134], FRVT 2006 [10], and FRGC [133].

There two major reasons for this enormous interest on face recognition technology. The first reason is the existence of a large number of forensic, security, and commercial applications. These applications include video camera surveillance, access control, mug shot identification (e.g., for issuing passport), design of human computer interactions (HCI), multimedia communication, and content-based image retrieval. The second reason is the availability of technologies developed by researchers in the field of

computer vision, pattern recognition, image processing, and computer graphics that has motivated the use of the technology for this purpose.

1.1 Biometric System

Recognizing the identity and authenticity of the people is the fundamental process for many activities and applications in our life. Biometric identification or biometrics refers to identifying an individual based on his/her distinguishing characteristics. In other words, biometrics is the science of identifying, or verifying the identity of a person based on the behavioral or physiological characteristics.

1.1.1 History of Biometrics

Possibly the first known example of the use of biometrics in practice was the use of a form of finger printing in China in the 14th century, as reported by explorer Joao de Barros. He wrote that the Chinese merchants were stamping children's palm prints and footprints on paper with ink to distinguish the young children from one another. This is one of the earliest known cases of biometrics in use and is still being used today [55].

Elsewhere in the world up until the late 1800s, identification largely relied upon "photographic memory." In the 1890s, an anthropologist and police desk clerk in Paris named Alphonse Bertillon sought to fix the problem of identifying convicted criminals and turned biometrics into a distinct field of study. He developed a method of multiple body measurements which got named after him (Bertillonage). His system

was used by police authorities throughout the world, until it quickly faded when it was discovered that some people shared the same measurements and based on the measurements alone, two people could get treated as one. After the failure of Bertillonage, the police started using finger printing, which was developed by Richard Edward Henry of Scotland Yard, essentially reverting to the same methods used by the Chinese for years [55].

1.1.2 Types of Biometrics

In recent years biometrics moved from simple fingerprinting to many different methods that use various physical and behavioral measures. The characteristics used in each category are as follows:

- Physiological
 - Iris
 - Fingerprint (including nail)
 - Hand (including knuckle, palm, vascular)
 - Face
 - Retina
 - DNA
 - Vein
 - Ear
 - Even Odor, Sweat pore, Lips

- Behavioral
 - Signature
 - Keystroke
 - Voice
 - Gait

Among the biometric features/identifiers, Fingerprint, Face, Hand geometry, Iris, Signature, and Voice are the most commonly used in today's automated authentication systems. These biometrics received more attentions than the others by researches in the filed of computer science.

1.1.3 Motivation and Overview

Biometric technologies are becoming the foundation of an extensive array of highly secure identification and personal verification solutions. As the level of security breaches and transaction fraud increases, the need for highly secure identification and personal verification technologies is becoming apparent. The uses of biometrics have also increased from just identification to verification as used in security systems and more. Biometric-based solutions are able to provide for confidential financial transactions and personal data privacy. The need for biometrics can be found in federal, state and local governments, in the military, and in commercial applications. Enterprise-wide network security infrastructures, government IDs, secure electronic banking, investing and other financial transactions, retail sales, law enforcement, and health and social services are already benefiting from these technologies.

1.1.4 Biometric Authentication System

A biometric authentication system can be considered as a pattern recognition system with three modules consisting of: biometric sensor(s) to collect the data from the biometric identifier, feature extractor, and matcher. For example, in case of face recognition, the sensor is a camera and the biometric identifier is a facial image. The task of biometric authentication mainly is divided into two categories:

- **Identification** is a closed-universe (one-to-many) comparing process for a biometric sample from a given probe against all the known biometric reference templates in the database. In other words, this is the answer to the question “Who am I?” If the acquired sample matches a stored template within an acceptable margin of error, then the identity of the probe is matched to that of the previously stored reference. During the matching process, a set of similarity matching scores are obtained for the probe sample (i.e., one-to-many comparison process). These similarity scores are numerically ranked such that the highest similarity score is first and the smallest similarity score is ranked n , where n is the number of the subjects enrolled in the database. In an ideal case, the highest similarity score is the comparison of the claimed person’s biometric with the same person’s biometric that was previously stored in the database. The percentage of time that the highest similarity score is the correct match for all individuals, is referred to as the identification rate.

In order to evaluate the performance of identification, the percentage of time when one of the top- r matches is correct is considered and called as “Cumula-

tive Match score”. In other words, the “Cumulative Match Score” curve is the percentage of correct identification versus the rank r . Usually the percentage of the correct identification for the rank-one is reported as the performance of a biometric system.

- **Verification** is an open-universe (one-to-one) process of comparing a submitted biometric sample against single biometric reference of a single enrollee whose identity or role is being claimed. In other words, this is the answer to the question, “Am I who I claim I am?” The result of the verification is to confirm that the identity is matched or not matched. During the process of matching, a similarity score is computed by the biometric matcher; if the similarity score is higher than a preset threshold T , then the submitted biometric sample is approved to be the same as the biometric reference claimed. If the similarity match score is less than the preset threshold T , then the claimed identity for the submitted biometric is rejected.

In order to evaluate the verification performance, two kinds of errors can be made by the system: False Match (FM) and False Non-Match (FNM). FM is the error made by deciding that a (claimed) identity is a legitimate one while in reality it is an imposter and FNM is the error made by deciding that a (claimed) identity is not a legitimate while in reality the person is genuine. The frequency rate at which FM occurs is called False Match Rate (FMR), and the frequency rate at which FNM occurs is called False Non-Match Rate (FNMR). The error rates can be evaluated for any threshold T . Therefore, the functions

$FMR(T)$ and $FNMR(T)$ give the error rates when the match decision is made at threshold T . The error rates can be plotted against each other as a two-dimensional curve, $(FMR(T), FNMR(T))$.

This two-dimensional curve is called Receiver Operating Characteristic (ROC) curve. The ROC curve precisely defines the complete specification of a biometric matcher and shows the trade-off between the FMR and FNMR errors over a wide range of threshold. The biometric matcher can operate using any threshold T which defines a point on the ROC curve. In addition, the ROC can be used to compare the performance of two biometric matchers against each other.

1.1.5 Biometric Market

The research service from the Auto ID & Security business and financial services group highlights growth sectors of notable interest and also provides a comprehensive financial analysis of the biometrics market. The spotlight on security has intensified considerably in the wake of global terror attacks and increasing threats to safety, driving governments across the world to tighten security measures. The demand for sophisticated security solutions is greater now than ever before. Figure 1.1 shows the annual biometric industry revenues for the years 2007-2012 in \$m US. As the Figure shows, the annual revenues in the biometric market are growing up with a rate of more than 15% every year. This is due to the huge demand for the applications of biometric technology in different fields. Figure 1.2 shows the percentage share of the different biometrics in the market in 2006. As the Figure shows, after Fingerprint

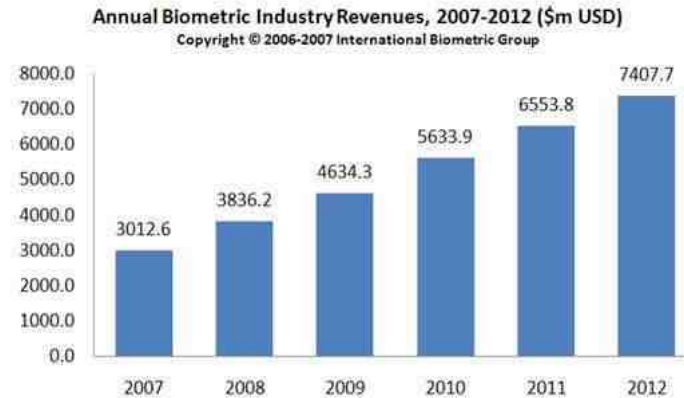


Figure 1.1: Annual biometric industry revenues for the years 2007-2012 [8].

(43.6%), great attention is paid to face recognition (19.0%) in the biometric market. Advanced face recognition biometrics are ideally positioned to address the demand for security solutions and are set to witness a compound annual growth rate (CAGR) of 27.5 percent from \$186 million in 2005 to \$1021.1 million in 2012 [53].

Enhanced credibility of this technology combined with its rapidly growing awareness is also likely to provide a strong impetus to growth of the face recognition biometrics market throughout the forecast period. Concrete evidence in the form of successful deployments has also helped contribute to continued market growth.

North America is clearly leading the way in the uptake of face recognition biometrics, and this trend is likely to continue throughout the forecast period. “However, Europe, the Middle East, and Africa (EMEA) are likely to catch up very soon, with Western Europe expected to significantly contribute to revenue growth in this region,” says the analyst of this research service. Asia Pacific is also set to emerge as a key region for the implementation of face recognition biometrics technologies.

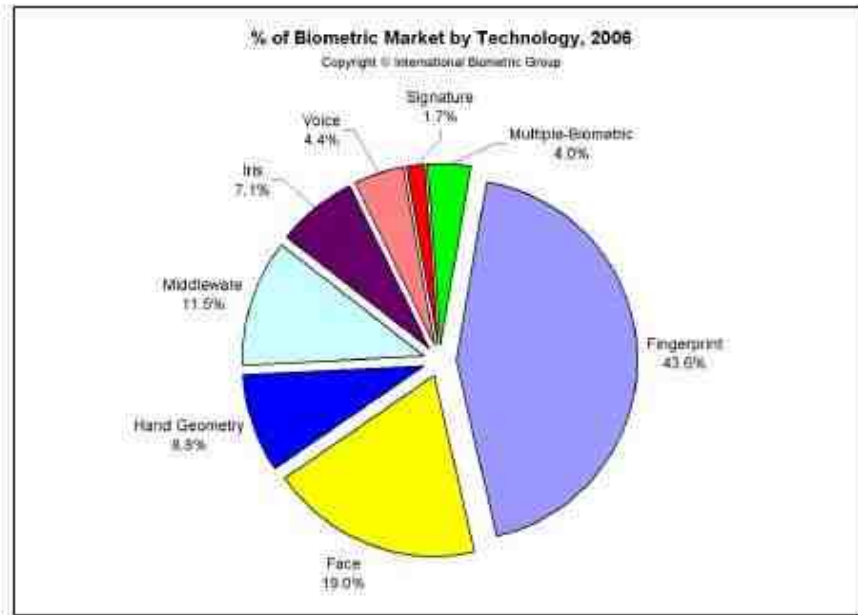


Figure 1.2: The percent of biometric market by technology in 2006 [8].

1.2 Face Recognition

Face recognition has a number of advantages over some of the other biometrics. Firstly, it is non-intrusive, whereas many biometrics require the subject's cooperation and awareness in order to perform an identification or verification, such as looking into an eye scanner or placing a finger on a fingerprint reader, while face recognition could be performed even without the subject's knowledge. Secondly, the biometric data used to perform face recognition is in a format that is readable and understood by humans. This means that a potential face recognition system can always be backed up and verified by a human instantly, unlike iris scan or finger prints. Besides the biometric application of face recognition, there are other applications for this technology. In the following subsection we review the applications of face recognition.

Category Area	Applications
Face ID	Voter registration, Driver licenses, national ID, immigration
Access Control	Building/room access, computer access
Security	Terrorist alert, secure flight boarding system
Surveillance	Advanced video surveillance, nuclear plants surveillance, neighborhood watch, power grid surveillance, portal control
Smart	Cards stored valued security, user authentication
Law Enforcement	Crime stopping and suspect alert, shoplifter recognition, suspect background check, post event analysis
Face-based database	Face-based search and retrieval
Multimedia management	Indexing, segmentation, classification, or event detection
Human computer interaction	Interactive gaming, animation

Table 1.1: Typical applications of face.

1.2.1 Applications of Face Recognition

Table 1.1 lists typical applications for face recognition in nine categories. These categories are neither exclusive nor exhaustive. For each category, some example applications are also listed and briefly discussed in this section. In all categories, the input to the system is a facial image either from still camera, video camera, or 3-D scanner.

Face ID

Face recognition systems identify people by their faces. This process is approached in one of two ways: Face recognition (identification) and Face verification (authentication). In general, there are three ways to identify an individual: The person knows something (a PIN, a password, etc.), the person possess something (an ID card, a drivers license, etc.), or by measuring something about the person's body or activity. The later encompasses biometric identification, and face recognition falls in this category. The system establishes the presence of an authorized person rather than

checking for a valid ID card or password PIN number. The security advantage of using such a system eliminates the misuse of lost or stolen cards. In 2000, the commercial product, FaceIt [3], was used to eliminate duplicates in a nationwide voter registration when the same person was assigned more than one identification number. The face recognition system compares the face image of the voters to differentiate from one from the other. When the top two matched faces are very similar to the query face image, manual inspection is required to make sure that they are for different persons in order to eliminate the duplicates. In the future, targeted face ID applications will include large scale applications such as e-commerce, student ID, and national ID.

Access Control

Face recognition is being used in access controls such as in accessing computers or buildings. A commercial system by FaceGate [6] was used along with an entry code control that acts as a specific label for a stored face of an individual in the database. An entry code and a face image are captured by the system at the door. The system simultaneously checks the person's entry code and verifies if the face image matches that corresponding to the entered key. Access is denied to anyone whose face does not match the stored key. Recently, multi-modal systems are available [2], which integrate more than one biometrics, such as finger prints, speech, and face to reduce the recognition error rate.

Security

Recently, security has become a primary concern at airports, stadiums, and big cities. Face recognition technology have been implemented at many airports around the globe. A Viisage system [7] was installed at Fresno Yosemite International airport in California to alert public safety officers whenever an individual matching the appearance of a known terrorist suspect enters the airport's security check point. Anyone recognized by the system would undergo further investigative procedures by the safety officer. Viisage's faceFinder was also used to scan the stadium audience at games events in Tampa Florida in search of criminals. Everyone entering the stadium was scanned by video cameras installed at the entrance. The cameras were tied to security command center that Compare the face images against a list of images of known criminals.

Surveillance

Similar to security applications, surveillance by face recognition was first used in 1998 with 300 cameras in areas of London. The city council claims that the technology has helped achieve 34% reduction in crime rate. Virginia Beach, Virginia is the second U.S city to install FaceIt system on its public streets to scan pedestrian's faces to look for 2500 missing persons or runaways.

Smart Card

Smart cards are used mainly in face verification scenarios. In this case the characteristics of a face are stored in the card. The user to be verified first scans his card

and has his image captured by the system. Comparison is achieved by measuring the similarity between the live captured image and the facial characteristics stored in the card. This technology was also integrated with fingerprint recognition as in the system by Maximus [5].

Law Enforcement

Face recognition empowers the law enforcement agencies with the ability to search and identify suspects using face recognition and retrieval programs. The system by Imagis [4] provides Huntington Beach, California's police officers and detectives with current arrest information and photographs, readily available via internet protocol and secured wireless laptops. The Imagis system includes biometric facial recognition and image database management, which provides investigators with invaluable tools to accomplish their work.

Multimedia Management of Face Databases

Because of the emergence of large image databases, content-based image retrieval was developed to index and retrieve images by their own visual contents such as texture, color, and/or edges [12, 13]. However these general techniques have their own limitations. Recently, researchers have combined traditional retrieval methods with face detection and recognition in order to improve the retrieval accuracy. Applications of such methods have been used not only in a face-only database but also in a face and non-faces databases such as photo albums [86].

Multimedia Management

Human faces are frequently seen in sports, news, videos, and other multimedia contents. Indexing this multimedia content by face detection, recognition, and face change detection is important to generate segments of video contents for video browsing, skimming, or summarization. Together with other image or speech processing techniques, face recognition becomes a powerful tool for indexing and retrieval of growing multimedia contents. One difficulty of directly using face recognition in multimedia applications is that usually the gallery set is not available. Houghton [62] developed a method to index and retrieve faces from the web. A gallery for his method is created by searching the web pages associated with broadcast television stations using three text processing methods and names in images using optical character recognition (OCR). The names are then linked to faces, detected using FaceIt, in the images. The “face-naming” method compares unknown faces with the gallery and returns the identities.

Human Computer Interaction

Human-computer interaction (HCI) is the study of interaction between people (users) and computers. To achieve efficient and user-friendly interaction, the human body part (e.g., the face) could be considered as a natural input device. For instance the movements of the face can be used in human tracking system. We recently developed an efficient tracking system of people based on their facial skin and body (cloth) colors using a single video camera [100]. Also, the tracked faces can be used as first

step to localize the location of faces, in video images, for face recognition. Facial expression recognition is the ability of computers to understand human emotions. Cohen *et al.* [43] reported on several Advances they have made in building a system for classifying facial expression from continuous video input. They used Bayesian network classifiers for classifying expressions from video. Another application of HCI is realistic synthesis and animation of faces which are widely used in the video and motion picture industries as well as the video game industry. Hong *et al.* [125] designed a system that provides functionalities for 3-D face modeling and animation with the help of user interactions. Text and speech streams can be used to drive the face animation which is used in computer aided education.

1.2.2 Face Recognition Challenges

In spite of the large amount of work on automatic face recognition, it still remains a very challenging task and it is not robust for large scale applications. This is not only because the techniques used for face recognition need to be improved, but also because the presence of many conflicting factors which alter the facial appearance and make the task difficult. The variations in facial appearance can be categorized into two types: intrinsic and extrinsic sources of variations.

- Intrinsic variations are independent of any external sources and are due to the physical nature of the human face.
- Extrinsic variations are caused by the sources that do not depend on the human face and are due to factors such as illumination and viewing geometry.

Variation in appearance	Source	Effect/possible task
Extrinsic	Viewing geometry, Illumination, Imaging process Other objects	Head Pose light variations, shadow, self shadow Resolution, scale, focus, sampling Occlusion, shadowing, indirect illumination, hair, make-up, surgery
Intrinsic	Identity, Facial expression, Age, Sex, Speech	Identification, known-unknown Inference of emotion or intension Estimating age Decide if male or female Lip reading

Table 1.2: Variations in facial appearance Inter-person and intra-person variations.

Table 1.2 summarizes these two types of variations and their effects on face recognition. Among these effects, illumination, variations in pose, aging, and facial expressions are the most challenging for face recognition.

- Illumination:** Changes in lighting conditions, e.g., indoor or outdoor, under which the facial images are captured, affect the accuracy of face recognition. Variations in illumination could be caused either by variations in the light source or by variations in physical parameters of the cameras and the capturing devices. A solution for this problem is by utilizing the 3-D surface information of the face. So, by having the 3-D model of the face surface, the problem reduces to matching the surface geometry of two faces which are invariant under the effect of illumination.
- Head Pose:** Pose variation is another challenging problem in face recognition. The variations in pose could be because of the changes in viewing angle of the camera which causes pose variation in the 2-D or 3-D captured face image.

Because face is a 3-D object, 2-D face recognition under the effect of pose variations is difficult, while having the 3-D face data, the problem of pose variation can be handled either in 3-D versus 3-D face recognition or 2-D versus 3-D.

- **Facial Expressions:** The development of robust face recognition algorithm insensitive to facial expression is one of the biggest challenges of current research in this field. The change in the face appearance due to its non-rigid structure makes modeling and analyzing the facial expressions difficult. In addition, facial expressions vary from person to person, which makes the task of modeling the facial expressions more difficult.
- **Aging Effect:** Aging is the inherent problem of face recognition because face is an identifier that changes with age and the aging effect cannot be controlled or ignored. The facial aging effects are manifested in different forms in different ages. It is manifested as changes in the shape of the cranium from infancy to teenage while during the adulthood it is demonstrated as changes in the skin texture. Thus, because facial aging has different sources, having a unified solution for this problem is difficult.

Another challenge for face recognition is the need for an evaluation standard for measuring recognition performance under different environments and conditions. As a result of this necessity, an independent government evaluation standard was born, which is called, Face Recognition Vendor Tests (FRVT). FRVT was developed to provide evaluations of commercially available and prototype face recognition technologies. These evaluations are designed to provide U.S. government and law enforcement

agencies with information to assist them in determining where and how facial recognition technology can best be deployed. In addition, FRVT results help identify future research directions for the face recognition community. In the past, many factors have been evaluated in FRVT 2002 [134]. For example, in a verification test with reasonably controlled lighting, when the gallery consisted of 37,437 individuals with one image per person and the probe set of 74,854 probes with two images per person, the best three systems averaged a verification rate of 90% at false accept rate of 1%, 80% at false accept rate of 0.1%, and 70% at false accept rate of 0.01%. This level of accuracy may be suitable for access control with a small database of hundreds of people but not for a security system at airports where the number of passengers is much larger. When evaluating the performance with respect to pose changes with a database of 87 individuals, the best system achieved an identification rate of 42% for faces within ± 45 degrees of panning and 53% within ± 45 degrees of tilting. Lighting changes between outdoor probe images and indoor gallery images degrade the best systems from a verification rate of 90% to 60% at a false accept rate of 1%.

1.3 Proposed Face Modeling and Recognition System

In Face Recognition Grand Challenge (FRGC) contest, three contenders for improving face recognition algorithms were considered: high resolution images, three-dimensional (3-D) face recognition, and multiple still images. With the 3-D data, the two main challenges of face recognition, pose variation and illumination, are handled more readily. This is due to the fact that the 3-D shape of a person's face is not

affected by changes in head orientation and lighting. Hence, 3-D face recognition has the potential of improving the recognition performance under these conditions [31]. Nevertheless, a pure 3-D face recognition system has its own following limitations.

- Capturing the 3-D face data either by a range scanner or by a stereo-based system is slow and expensive with the current technologies.
- Capturing the 3-D data is intrusive.
- Extraction of facial landmarks in 3-D is a very challenging task.
- Shape matching techniques are complex and time consuming.
- 3-D range data lacks texture cue.

Based on the above discussion, a multi-modal system would benefit from both modalities. For example, the pose variations and changing in illumination can be handled by 3-D data while extracting facial features are much easier in 2-D (texture) data. In particular, texture provides more discriminative information for face recognition. In a multi-modal scheme, the 2-d and 3-D face recognition can be fused at different levels; e.g., feature level, decision level, score level fusion. Beside resolving the above issues, the overall performance of the system would be increased by fusing the 2-D and 3-D modalities.

In this dissertation, our aim is to develop a multi-modal system for face modeling and recognition. We consider two different scenarios for the multi-modal face recognition. In the first scenario, where the 3-D (shape) and 2-D (texture) data are registered, i.e., each point in 2-D has a corresponding point in 3-D, we develop a

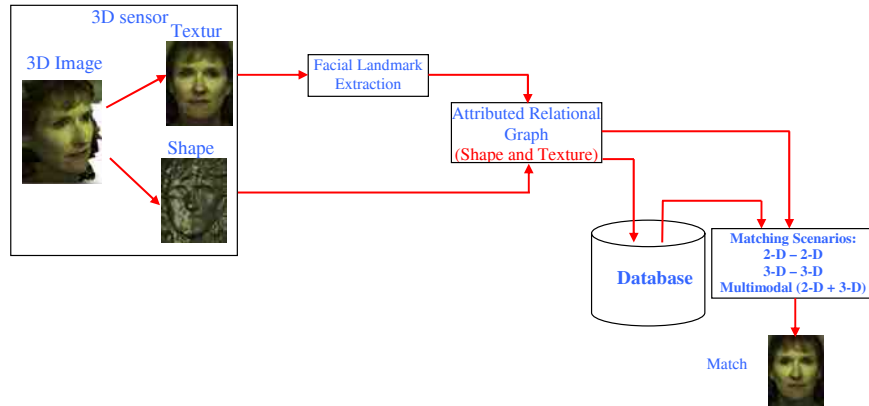


Figure 1.3: The general block diagram of our system for multi-modal face recognition based on 3-D ARG models.

technique for multi-modal face modeling and recognition based on 3-D ARG models. In fact, by using the 3-D ARG framework, the modeling of the 2-D and 3-D data is approached in a unified manner. More precisely, the 2-D and 3-D features are represented in one integrated geometric graph model. Figure 1.3 shows the general block diagram of our system for multi-modal face recognition based on 3-D ARG models.

In the second case, the 2-D and the 3-D data are not registered. The lack of registration could be due to the time laps in capturing the 2-D and 3-D data or the nature of the imaging system (e.g., scanning a face with a laser scanner with current technologies takes few seconds.) Therefore, we cannot directly establish correspondence between the facial landmark points in the 2-D and 3-D images. In this case, the 2-D and 3-D modeling and recognition are carried independently and then the results are fused at the score level. For the 3-D modeling, we develop an approach based on ridge images. This approach is discussed in Chapter 3. For the 2-D face recognition, we develop a technique based on Attributed Relational Graphs (ARG). Figure 1.4 shows

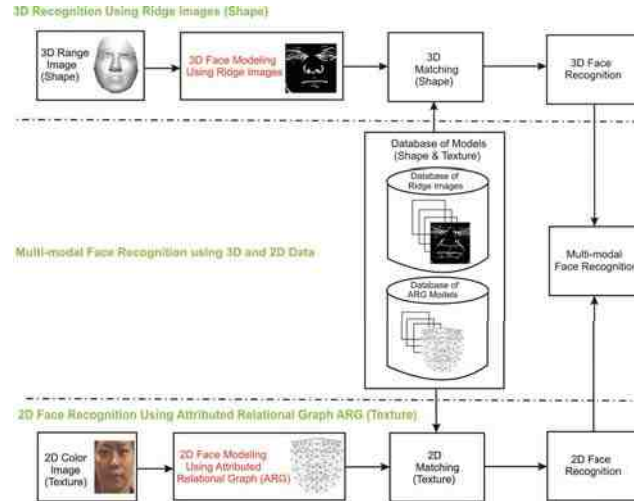


Figure 1.4: The general block diagram of our system for multi-modal face recognition based on ridge images and 2-D ARG models.

the general block diagram of our system for multi-modal face recognition based on ridge images (3-D) and ARG models (2-D). As shown in the figure, the modeling of the 2-D and 3-D are independent of each other and the results of the 2-D and 3-D are fused to obtain the final result (i.e., multi-modal face recognition.)

1.3.1 3-D Face Recognition

There are mainly three categories of approaches for 3-D face recognition: 1) Principal Component Analysis (PCA) based approaches [133, 126], 2) feature based approaches [130, 44, 69], and 3) surface matching approaches [145, 37, 109]. In the first category, similar to the 2-D Eigenface recognition algorithm, PCA analysis is applied to range data to reduce the dimension and then the recognition is performed by matching a probe image with gallery images in a lower dimension space. These approaches are simple, fast and straightforward, but they have low performance rate

compared to approaches in the other two categories. In the second category, the response of the range image or its representation at certain landmark points to a set of filters is calculated and considered as a set of features. Then, recognition is done based on the similarity of these features. Generally, these approaches are fast and have high performance rate compared to approaches in the other two categories, but localization of the landmarks is very important. For example in [69] the 2-D texture images were used for landmark localization which means that the approach is not a pure 3-D algorithm. In the third category, the researchers mainly utilize the Iterative Closest Points (ICP) or Hausdorff distance to match the 3-D surface points of a probe face to those of the face images in the gallery and then perform the recognition based on the Mean Square Error (MSE) distance between the matched points of the two faces. As mentioned in [31], the main problem with the approaches that rely on ICP or Hausdorff distance for matching is speed and computational complexity, but these approaches have high performance rate. Figure 1.5 illustrates the above comparison (i.e., performance versus computational complexity.)

3-D Face Recognition Using 3-D Ridge Images

In this dissertation, we present a novel method for 3-D face recognition (shape matching) based on ridge lines extracted from the 3-D range facial images. Compared to other shape matching based approaches for 3-D face recognition, such as [108, 145, 37, 109], our approach is faster and requires less computations. This reduction in computations is due to the fact that we only use the points around the important facial regions on the face (i.e., the eyes, the nose, and the mouth) and ignore other

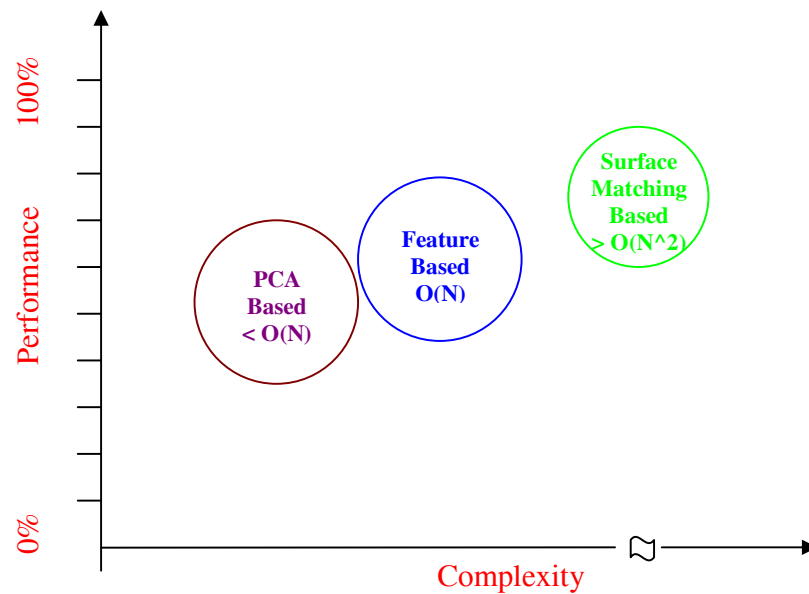


Figure 1.5: Comparison between the three categories of algorithms for 3-D face recognition, performance of the system versus complexity.

surface patches on the face during the matching process. These points correspond to the extreme ridge points on the considered surface. An extreme ridge point is a point where the principal curvature k_{max} , has large positive value. There are different approaches to locate the ridges, here we threshold the k_{max} values to find these points. Figure 1.6 shows few examples of the ridge images obtained by thresholding the k_{max} values. These are 3-D binary images that show the location of the ridge lines on the surface of the face. In this work, the number of the points in a ridge image of the face is $12\% \pm 2\%$ of the total number of points that cover the face. For matching the ridge images (probe image versus gallery image), either the Hausdorff Distance or the ICP method can be used.

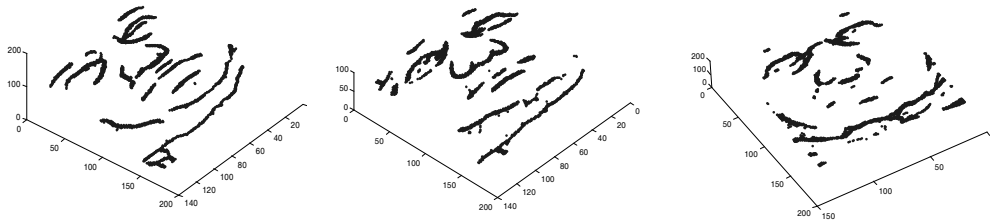


Figure 1.6: Samples of extracted ridge images.

1.3.2 A Unified Approach for 2-D/3-D Face Recognition

Graph representation has shown to be successful [163, 29] in 2-D face modeling and recognition. The idea is to use a graph to model the face such that the nodes of the graph represent the facial landmarks and a set of features are extracted and assigned to each node of the graph. However, the graph models that are in the literature have some limitations. For example, there is no justification for defining the edges of the graph. Also, no relations are defined between the nodes/edges of the graph. In this dissertation, we develop a technique for face modeling and recognition based on Attributed Relational Graphs (ARG).

The ARG is used to model the 2-D or 2-D/3-D face data. In the case where the 2-D and 3-D data are registered, a 3-D ARG model is built to represent both the texture and shape by a single graph model. If the 3-D data is available and is not registered with the 2-D data, the 3-D faces are modeled using ridge images. The results of recognition of the 2-D and 3-D modalities are fused at the score level.

In this dissertation, we use the ARG to represent the local as well as the global geometric structures of the face. The ARG consists of nodes and edges such that the nodes of the graph correspond to the landmark points that are extracted by an

improved Active Shape Model (ASM) technique. The edges of the graph represent the mutual relations between the nodes. For each node of the graph, we calculate the response of both the shape and the texture of the face, at the corresponding landmark, using log-Gabor filters. These are feature vectors that model the local structure of the face at each node. Moreover, we define a set of mutual relations features between the nodes of the Graph (i.e., edges of the graph). As our experiments indicate, these mutual relations increase the performance of the face recognition using the graph.

Unlike the work presented by Park *et al.* [129], where the nodes of the ARG do not have correspondences and rely on stochastic analysis to find the feature correspondences, in our work, the nodes of the ARG have direct one-to-one correspondences with facial features. In addition, based on the availability of the shape information, the constructed ARG is a graph that combines both the 2-D and the 3-D information in a single graph. Compared to [163], the nodes of the graph in our work correspond to landmarks that are extracted by ASM, and our graph contains the mutual relation between the edges of the graph and these increase the accuracy of the face recognition.

1.4 Dissertation Contributions

The major contributions of this dissertation are as follows:

- Improving the Active Shape Model for 2-D facial features extraction from color image. We present solutions for some of the limitations of Active Shape Model (ASM) to extract facial feature extraction in color images.

- Developing an algorithm for 3-D facial feature extraction from range data. Extracting 3-D facial features from 3-D range images is more difficult compared to 2-D facial feature extraction, because of the lack of texture in range images. In this dissertation, we develop an algorithm for extracting three facial feature points (i.e., the inner corners of the two eyes and the tip of the nose) from facial range images. These points are used to initially align the ridge images during the matching process.
- Developing an algorithm for 3-D face modeling and recognition based on ridge images. The ridge lines in the range image carry the most important distinguishing information of the 3-D face and have high potential for face recognition. We develop a system for 3-D face recognition based on ridge lines. For matching the ridge images of two faces (probe and gallery), the Hausdorff and Iterative Closest Points are utilized.
- Developing a novel algorithm for 3-D face recognition based on Attributed Relational Graphs (ARG). The nodes of the graph represent the facial landmark points. A set of attributes are extracted using Gabor filters and assigned to each node of the graph. Also, a set of features that defines the mutual relations between the edges of the graph are extracted and used to increase the performance of the graph model for face recognition.
- Developing a multi-modal technique based on the Dempster-Shafer theory of evidence and the weighted sum rule for fusion at the score level.

1.5 Dissertation Outline

This dissertation is organized as follows: In Chapter two, we present related work for facial features extraction, two dimensional (2-D), three dimensional (3-D), and multi-modal (2-D + 3-D) face recognition. Chapter three explains our algorithm for 2-D facial feature extraction from frontal face images (i.e., Improved ASM) and our algorithm for 3-D facial feature extraction (i.e., the extraction of the three feature points) along with the experimental results. Chapter four presents our approach for 3-D face modeling and recognition based on ridge images. Chapter five describes our multi-modal face modeling and recognition (2-D/3-D) based on attributed relational graphs along with the experiments. In addition, we present two fusion techniques for combining the 2-D and 3-D modalities in this chapter. Finally in Chapter six, we present the conclusion and the future research directions.

Chapter 2

Related Work

In this chapter we present an overview of the previous related works. We start by reviewing the literature for facial features extraction from frontal 2-D images and 3-D range images. Afterwards, we review the literature for 2-D and 3-D face recognition. Finally we review the algorithms for multi-modal (2-D + 3-D) face recognition.

2.1 Facial Features Extraction

In both 2-D and 3-D face recognition systems, alignment (registration) between the query and the template images or models is necessary [146]. This is the main step before recognition in a typical face recognition system. This step is usually based on extracted facial features (i.e., fiducial points). Figure 2.1 shows a sample of labeled facial features on both frontal and profile face images. Beside the applications of facial features extraction for face recognition, other applications such as tracking, expression analysis, and animation rely on facial features extraction. Automatic

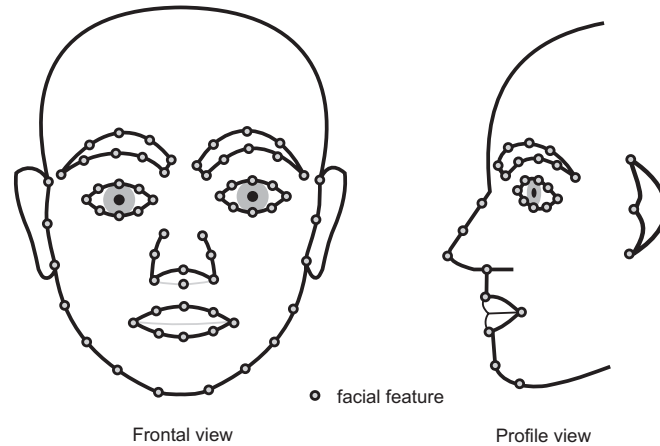


Figure 2.1: Labeled facial features in frontal and profile images.

extraction of facial features has been one of the important and challenging tasks and many approaches and algorithms have been developed and presented for facial features extraction. Many of these methods are cited and reviewed in face detection and face recognition surveys [179, 38, 171]. In this Section, we review the most important algorithms and methods for 2-D and 3-D facial features extraction.

2.1.1 2-D Facial Features Extraction From Frontal Images

Facial features extraction is defined as the process of locating specific region, points, landmarks, or curves/contours in a given 2-D image or a 3-D range image [84, 120, 121, 94, 172, 149, 166, 82]. Although many facial feature extraction algorithms have been proposed so far, facial feature extraction is difficult to the applications due to its high complexity. The computational cost of facial features extraction is dominated by the localization of the face region and the searching for the feature points. Generic methods extract features from images without relying on extensive knowledge about

the object of interest. They have the advantage of being typically fast and simple. However, these approaches can become unreliable when the quality of the image is poor or the face in the image has a cluttered background. The algorithms for facial feature extraction can be divided into four types:

1. Generic methods based on edges, lines, and curves.
2. Template-based methods that are used to detect facial features such as the eyes.
3. Structural matching methods that take into consideration the geometrical constraints on the features (i.e. everyone has two eyes above the mouth).
4. Hybrid methods which combine some of the above previous methods.

Table 2.1 reviews some of the important published methods for 2-D facial features extraction from frontal images.

Kobayashi *et al.* [84] described an automated algorithm for face detection and facial features extraction from video images. The extracted features are points around the eyes, mouth, nose and facial contours. The authors used spatiotemporal difference images to extract these feature points. For example blinking is used for detecting the eyes. The method proposed by De Natalie *et al.* [120] aimed at identifying the position of characteristic facial elements (eyes, nose and mouth). The proposed detection strategy is based on the identification of the face symmetry axis, and the successive detection of eyes, mouth and other relevant facial features using correlation.

Nikolaidis *et al.* [121] described a method for extracting facial features with the goal of using them in defining a sufficient set of distances between them so that a

Author	Face detection	Approach	Number of feature points	Video/Still
S. Kobayashi <i>et al.</i> 1995 [84]	Included	Using spatiotemporal difference image to extract the feature points	Representing eyes, eyebrows, boundary, mouth and face boundary by contours	Video - Frontal
T.F. Cootes <i>et al.</i> 1996 [45, 92]	-	Using active shape models	Flexible in term of number of features	Still - Frontal
F.G.B. De Natalie 1997 [120]	Included	Identification of the face symmetry axis and then feature points using correlation	Location of eyes, and mouth	Still - Frontal
A. Nikoladis 1997 [121]	Included	Using adaptive Hough transform, template matching and active contour models	Eyes, eyebrows, mouth, nostrils, cheeks and chin	Still - Frontal
K. Lam and Y. Li <i>et al.</i> 1998 [90]	-	Eye corner detection and template matching	Eyes	Still - frontal
Y. Yagi 2000 [168]	Included	-	Pupil, and contours	Still - Frontal
C.M. Lau <i>et al.</i> 2001 [94]	Included	Using energy function to extract facial features	Face region, iris, eyebrows, nose, mouth, eye corners, face-hair	Still - Frontal
G. Yen and N. Nithianandan 2002 [172]	Included	Using the edge density distribution of the image and genetic algorithm	Face region, eyes, nose, mouth	Still - Frontal
K. Seo <i>et al.</i> 2002 [149]	Included	Using active contour model and color information	Face region, eyes, nose, mouth	Still - Frontal
D. Xi and S.W. Lee <i>et al.</i> 2003 [166]	Included	Using Support Vector Machines and Multi wavelet decomposition	Face region, eyes, nose, mouth	Still - Frontal
Z. Xue <i>et al.</i> 2002 [167]	-	Using Bayesian shape model	A mesh model of facial features	Still - Frontal
R. Hsu <i>et al.</i> 2002 [63]	Included	Using color information for face detection	Face boundary, eyes, mouth	Still - Frontal with pose
Y. Hu <i>et al.</i> 2003 [66]	-	Using linear combination model	Eyes, eyebrows, nose, mouth	Still - Frontal
A. Gundaz and H. Krim 2003 [60]	-	Using topological operators	Eye corners and mouth center	Still - Frontal
K. Kim <i>et al.</i> 2004 [82]	-	Using PCA and Wavelet multi resolution images	Eyes, eyebrows, nose, mouth	Still - Frontal
K. Nagao 2004 [119]	-	Using Bayesian approach with nonlinear kernels	Eye centers	Still - Frontal with pose

Table 2.1: Facial features extraction techniques.

unique description of the structure of a face is obtained. Eyebrows, eyes, nostrils, mouth, cheeks and chin are considered as interesting features. Candidates for eyes, nostrils and mouth are determined by searching for minima and maxima in the x and y projections of the gray level pixels in the image. Candidates for cheeks and chin are determined by performing adaptive hough transform on a sub-image defined

according to the position of the eyes, mouth, and the ellipse containing the main connected component of the image. In order to acquire a more accurate model of the face, a deforming technique is also applied to the ellipse representing the main face region. Candidates for eyebrows are determined by adapting a proper gray level template to an area restricted by the position of the eyes.

Lam *et al.* [90] devised an efficient approach for detecting and locating the eyes in frontal images. Possible eye candidates in an image are identified by means of the valley features and corners of the eyes. Two possible eye candidates are considered to belong to the eyes of a human face if their respective local properties are similar; an eye window is then formed. Each eye region candidate is then further verified by comparison with a standard eye template, and by measuring its symmetry. Yagi [168] presented a system that integrates a library of 32 functions for automatic facial contour extraction. The 32 functions can be classified into five groups such as face detection, pupil detection, facial parts detection, facial parts contour extraction, and face contour extraction. The system is not only useful for automatic 3-D facial model fitting, but also for range facial image processing applications such as personal authentication and facial expression analysis. Lau *et al.* [94] proposed an energy function which is the sum of seven weighted terms for facial features extraction. By allocating different values for the weighting factors, the function can extract different fiducial points.

Yen *et al.* [172] presented a method for facial features extraction that uses the edge density distribution of the image. In the preprocessing stage, a face is approximated by an ellipse, and a genetic algorithm is applied to search for the best matching region.

In the feature extraction stage, a genetic algorithm is applied to extract the facial features, such as the eyes, nose and mouth, in the predefined sub regions. The authors validated their method by experimenting on various video images under natural lighting environments and in the presence of noise and different face orientations.

Seo *et al.* [149] presented an active contour model based upon color information for extracting facial features. Their algorithm is composed of three main parts: the face region estimation part, the detection part and the facial features extraction part. In the face region estimation part, images are segmented based on human skin color. In the face detection part, a template matching method is used, and in the facial features extraction part, an algorithm called “color snake” is applied to extract facial feature points within the estimated face region.

Xi *et al.* [166] developed an algorithm for detecting human face and extracting facial features based on several support vector machines. A face model for both detection and extraction was designed based on multi-resolution wavelet decomposition (MWD). The MWD and a small number of feature points were applied to roughly detect the face. More accurate results were achieved by a series of support vector machines. Xue *et al.* [167] presented a novel application of the Bayesian Shape Model (BSM) for facial features extraction. First, a full-face model is designed to describe the shape of a face, and the PCA is used to estimate the shape variance of the face model. Then, the BSM is applied to match and extract the face patch from input face images. Finally, using the face model, the extracted face patches are easily warped or normalized to a standard view.

Hsu *et al.* [63] proposed a face detection algorithm from color images in the presence of varying lighting conditions as well as complex backgrounds. Based on a novel lighting compensation technique and a nonlinear color transformation, this method detects skin regions over the entire image and then generates face candidates based on the spatial arrangement of these skin patches. The algorithm constructs eye, mouth, and boundary maps for verifying each face candidate. Hu *et al.* [66] proposed a facial feature extraction method based on a linear combination model. The model uses the knowledge of prototype faces, which are manually labeled, to interpret novel faces. Generally, the construction of the linear combination model depends on pixel-wise alignments of prototypes, and the alignments are computed by an optical flow algorithm or bootstrapping algorithm which is a full-scale optimization and without including local information such as facial feature points. To combine local facial features with the linear combination model, an optical flow algorithm is proposed to compute the pixel-wise alignments.

Gundaz *et al.* [60] presented a method for facial features extraction by considering the face image as a surface. Topological properties of the facial surface, such as principal curvatures are used to extract the eyes and mouth, which form deep valleys on the surface. The basic idea of the proposed method is to model the facial features as ravines on the facial surface. Ravines are points on the surface where the maximum curvature is a local maximum in the corresponding principal direction. Kim *et al.* [82] proposed an algorithm for extracting facial feature fields (eyebrow, eye, nose, and mouth) from gray scale face images. The foundation of this method is that eigenfeatures, derived from the eigenvalues and eigenvectors of the gray scale data set

constructed from the feature fields, are very useful to locate these fields efficiently. In addition, multi-resolution images, derived from a 2-D DWT (Discrete Wavelet Transform), are used to reduce the search time for the facial features. Nagao [119] described a method for finding the positions of features in facial images. A large class of image variations, including those resulting from object rotation in 3-D space and scaling (i.e., translation in depth), are handled. A MAP (Maximum a Posteriori) estimation technique using Gaussian distribution is exploited to model the relationship between images and feature positions.

Deformable models used for non-rigid object segmentation, received attention in recent years. These models have proven to be efficient in many applications such as object segmentation, appearance interpretation, motion tracking etc. A deformable model can be characterized as a model, which under an implicit or explicit optimization criterion deforms to match the shape of a known object in a given image. For a general review of the most commonly used models, we refer the readers to [111, 73].

In this dissertation, we improve the Active Shape Model (ASM) for facial features extraction. The original ASM developed by Cootes *et al.* [45] suffers from factors such as, poor model initialization, intensity modeling of the local structure of the facial features, and alignment of the shape model to a new instant of the object in a given image using simple similarity transformation. The core of our enhancement relies on three improvements (a) initializing the ASM model using a given set of points (e.g., the centers of the mouth and eyes, which are located using color information), (b) incorporating color information to represent the local structure of the feature points, and (c) applying 2-D affine transformation in aligning the facial features that are

perturbed by head pose variations, which effectively aligns the matched facial features to the shape model and compensates for the effect of the head pose variations. The details of our improved Active Shape Model (ASM) for facial features extraction are presented in Chapter 3.

2.1.2 3-D Facial Features Extraction

The previous works that utilized 3-D features for face recognition can be categorized into four groups:

1. Curvature-based methods.
2. Direct spatial surface matching.
3. Shape representation-based methods.
4. Multi-modal based methods that combines information from 2-D intensity images with 3-D range images.

Most of the early studies concentrate on curvature analysis. In [59], the surface regions from range images are classified as convex, concave, and saddle by calculating the minimum and maximum principal curvature. Then locations of facial features are determined, which are used for template comparison. Lee *et al.* [95] detects corresponding regions in two range images by graph matching based on Extended Gaussian Image (EGI). Tanaka *et al.* [156] also use EGI. For each face, two EGIs are constructed from maximum principal direction and minimum principal direction.

The EGI similarity is measured by Fisher’s spherical correlation. In spatial matching approaches, recognition is achieved via matching facial surfaces directly in 3-D Euclidean space [23, 127]. In shape representation methods, the 3-D facial shape or surface is converted to another shape representation such as point signature (PS) [42], spin image (SP) [76], shape distribution [124], or Local Shape Map (LSM) [164]. LSM is based on point signature approach and spin images. Thus the recognition task can be achieved in the representation domain.

Johnson *et al.* [76] compute a spin image which describes the shape of the surface. It is created by projecting a surface 3-D point P onto 2-D coordinates via a spin map function S using oriented point on the surface. The oriented point is the coordinate location on the surface with the normal, i.e., (x, y, n) . An image is created by applying S to all points on the surface. This image can be used in object recognition or 3-D surface registration. Point signature encodes a surface point p in a predefined periphery to a tangential plane, P passing through p [42]. Neighbors of point p are found by intersecting the original surface with a sphere. P is formed by fitting a plane to the neighboring points, and by translating it to the original point p . Signed distances are sampled by $\Delta\theta$ degree intervals, thus forming a 1D parametric curve, $d(\theta)$. In the recognition phase, each point signature extracted from the test image is compared with each image’s point signatures, for the database images and a total similarity between the database and the probe image is computed according to the sum of individual point signature distances.

The facial structure can also be described with the aid of 2-D or 3-D data sources. Assisted by a statistical feature location model, Lu *et al.* [107] automatically combine

3-D features shape index response, derived from the range map, with 2-D intensity corner-ness response to determine the correct positions of the corners of the eyes and the mouth. A feature extractor based on the directional maximum is presented to estimate the tip of the nose and the pose angle simultaneously. In other words, the nose tip has the largest depth value if projected onto the corrected pose direction. The limitation of the nose extraction approach is due to not making full use of the entire 3-D directional rotations (rotation about the z and x) and only yaw pose variation (rotation about the y -axis) is assumed. Similarly, Wang *et al.* [162] showed that by combining 2-D Gabor wavelet-based image intensity features with point signature-based 3-D shape features they obtained a superior performance than using each modality alone.

In Chapter 3, we present an algorithm based on Gaussian curvature to extract three feature points which are utilized in face alignment during the recognition process. The extracted feature points are the two inner corners of the eyes and the tip of the nose. Before extracting the facial features from 3-D range data, we locate the face in the range image and only keep the face areas and exclude the background, hair, and neck in the image. Therefore, we developed a method for localizing faces in range data using template matching.

2.2 Face Recognition

A man-machine facial recognition system dates way back to 1965 [35]. The authors showed that a computer program provided with facial features extracted

manually could perform recognition with satisfactory performance. In the past few years, face recognition has received great attention. A recent literature survey of face recognition is given in [179], where most of the paper surveys 2-D algorithms. In addition, the work and survey by Bowyer *et al.* [30] in 2004 gives a comparison among face recognition techniques based on 2-D data, 3-D data, and 2-D + 3-D data fusion. They reported that 3-D face recognition approaches outperform 2-D approaches and the fusion of 2-D + 3-D data produces slightly better results than 3-D alone. A very recent survey by Bowyer *et al.* [31] in 2006 cited some approaches that some 2-D recognition approaches outperform 3-D approaches. There is a belief that it is still premature to make this judgment at this time because current approaches did not yet make full use of 3-D data either in the recognition algorithms or the rigorous experimental methodology.

2.2.1 2-D face Recognition

Many algorithms for face recognition have been proposed during the past three decades. The literature on face recognition is vast and diverse. Zhao *et al.* [179] presented a literature survey of 2-D face recognition. We refer the readers to this paper for a complete survey of the state of the art in the area of 2-D face recognition. In this Section, we review the important approaches for 2-D face recognition.

The algorithms for 2-D face recognition are divided into three categories in [179]. This is a clear and high-level categorization based on a guideline suggested by the psychological study of how humans use holistic and local features [179].

1. *Holistic matching methods*: These methods use the whole face region as the raw input to a recognition system. One of the most widely used representations of the face region is eigenfaces [159], which are based on principal component analysis.
2. *Feature-based (structural) matching methods*: Typically, in these methods, local features such as the eyes, nose, and mouth are first extracted and their locations and local statistics (geometric and/or appearance) are fed into a structural classifier.
3. *Hybrid methods*: Just as the human perception system uses both local features and the whole face region to recognize a face, a machine recognition system should use both. One can argue that these methods could potentially offer the best of the two types of methods.

Table 2.2 classifies each category into sub-classes. In the following subsections, we review in details the most important methods in each class.

2.2.2 Holistic-based Approaches

Principal Component Analysis

Successful low-dimensional representation of faces using Kullback-Leibler (KL) or Principal Component Analysis (PCA) projections starts from the works by Kirby and Sirovich in 1987 [153] and 1990 [83]. Eigenpictures have been one of the major driving forces behind face representation, detection, and recognition. It is well known

Approach	Works
Holistic methods	
<i>Principal component analysis (PCA)</i>	
Eigenfaces	Direct application of PCA [83, 159]
Bayesian	Two-class problem with probability measure [16]
Fisherfaces/LDA	FLD on eigenspace [19, 177, 155]
ICA	ICA-based feature analysis [18]
<i>Other representations</i>	Neural Network based method [102]
Feature-based methods	
Pure geometry methods	Earlier works [80, 81]
Dynamic link architecture	Elastic Graph Matching [122, 163, 29]
Hidden Markov model	HMM methods[147, 148]
Hybrid methods	
Modular eigenfaces	Eigenfaces and eigenmodules [132]
Hybrid LFA	Local feature method [131]
Shape Normalized	Flexible appearance models [91]

Table 2.2: Classification of 2-D face recognition techniques [179].

that there exist significant statistical redundancies in natural images [142]. For a limited class of objects such as face images that are normalized with respect to scale, translation, and rotation, the redundancy is even greater [131, 176]. One of the best global compact representations is KL/PCA, which decorrelates the outputs. More specifically, sample vectors x can be expressed as linear combinations of the orthogonal basis Φ_i :

$$x = \sum_{i=1}^n a_i \Phi_i \approx \sum_{i=1}^m a_i \Phi_i \quad (2.1)$$

where typically $m \ll n$. The basis Φ_i are calculated by solving the eigenproblem:

$$C\Phi = \Phi\Lambda \quad (2.2)$$

where C is the covariance matrix for input x .

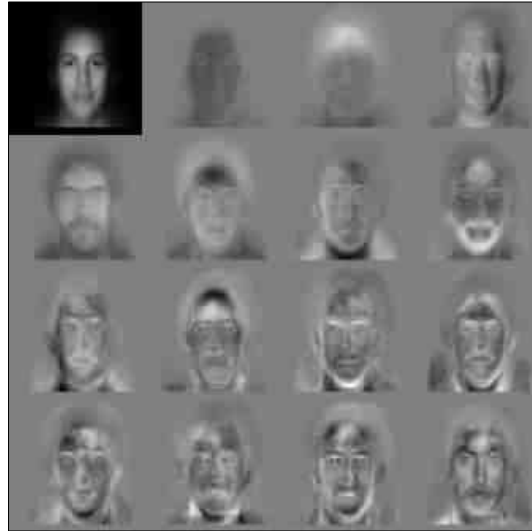


Figure 2.2: Example of eigenfaces.

Such representation is less sensitive to noise that may be due to small occlusions, as long as the topological structure does not change.

Turk and Pentland [159] made a very successful demonstration of machine recognition on faces using eigenpictures (known as eigenfaces) for face detection and identification. Figure 2.2 shows an example of eigenfaces. Given the eigenfaces, every face in the database can be represented as a vector of weights; the weights are obtained by projecting the image into eigenface components by a simple inner product operation. When a new test image whose identification is required is given, the new image is also represented by its vector of weights. The identification of a face image is done by locating the image in the database whose weights are the closest to the weights of the test image.

Moghaddam and Pentland [16] extended the standard eigenface approach to a Bayesian approach. They used a probabilistic measure of similarity, instead of the

simple Euclidean distance used with eigenfaces in [159]. The major drawback of a Bayesian method is the need to estimate probability distributions in a high-dimensional space from a very limited number of training samples per class. To avoid this problem, a much simpler two-class problem was created from the multi-class problem. Two mutually exclusive classes were defined: Ω_I , representing intra-personal variations between multiple images of the same individual, and Ω_E , representing extra-personal variations due to differences in identity. Assuming that both classes are Gaussian-distributed, likelihood functions $P(\Delta|\Omega_I)$ and $P(\Delta|\Omega_E)$ were estimated for a given intensity difference $\Delta = I_1 - I_2$. Given these likelihood functions and using the MAP rule, two face images are determined to belong to the same individual if $P(\Delta|I) > P(\Delta|E)$. A large performance improvement of this probabilistic matching technique over standard nearest-neighbor eigenspace matching was reported using large face data sets including the FERET database [135].

Successful face recognition systems using Linear Discriminant Analysis/Fisher Linear Discriminant (LDA/FLD) were reported by [19, 50, 155, 177, 178]. LDA training is carried out via scatter matrix analysis [54]. For an M -class problem, the within- and between-class scatter matrices S_w , S_b are computed as follows:

$$\begin{aligned} S_w &= \sum_{i=1}^M Pr(\omega_i) C_i, \\ S_b &= \sum_{i=1}^M Pr(\omega_i) (m_i - m_0)(m_i - m_0)^t, \end{aligned} \tag{2.3}$$

where $Pr(\omega_i)$ is the prior class probability, and is usually replaced by $1/M$ in practice with the assumption of equal priors. Here S_w is the within-class scatter matrix, showing the average scatter C_i of the sample vectors x of different classes ω_i around

their respective means m_i :

$$C_i = E[(x(\omega) - m_i)(x(\omega) - m_i)^t | \omega = \omega_i]. \quad (2.4)$$

Similarly, S_b is the Between-class Scatter Matrix, representing the scatter of the conditional mean vectors m_i around the overall mean vector m_0 . A commonly used measure for quantifying discriminatory power is the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix:

$$\mathcal{J}(T) = |T^t S_b T| / |T^t S_w T|. \quad (2.5)$$

The optimal projection matrix W which maximizes $\mathcal{J}(T)$ can be obtained by solving a generalized eigenvalue problem:

$$S_b W = S_w W \Lambda_W. \quad (2.6)$$

It is helpful to make comparisons among the so-called (linear) projection algorithms. Here we illustrate the comparison between eigenfaces and Fisherfaces. Similar comparisons can be made for other methods, for example, ICA projection methods. In all these projection algorithms, classification is performed by (1) projecting the input x into a subspace via a projection/basis matrix P_{roj} :

$$z = P_{roj} x \quad (2.7)$$

(2) comparing the projection coefficient vector z of the input to all the pre-stored projection vectors of labeled classes to determine the input class label. The vector comparison varies in different implementations and can influence the system's performance dramatically [114]. For example, PCA algorithms can use either the angle or the Euclidean distance (weighted or unweighted) between two projection vectors. For LDA algorithms, the distance can be unweighted or weighted.

A comparative performance analysis was carried out by Belhumeur *et al.* [19]. They compared four methods: (1) a correlation-based method, (2) a variant of the linear subspace method suggested in [151], (3) an eigenface method by Turk and Pentland [159], and (4) a Fisherface method which uses subspace projection prior to LDA projection to avoid the possible singularity in S_w as in [155]. Experiments were performed on a database of 500 images created by [61] and a database of 176 images created at Yale [1]. The results of the experiments showed that the Fisherface method performed significantly better than the other three methods. However, no claim was made about the relative performance of these algorithms on larger databases.

Bartlett *et al.* [18] presented an argument that for tasks such as face recognition, much of the important information is contained in high-order statistics. So, they proposed to use independent component analysis (ICA) to extract features for face recognition. ICA is a generalization of principal component analysis, which decorrelates the high-order moments of the input in addition to the second-order moments.

Neural Network-based Approaches

A fully automatic face detection/recognition system based on a neural network is reported in [102]. The proposed system is based on a probabilistic decision-based neural network (PDBNN, an extended (DBNN) [87]) which consists of three modules: a face detector, an eye localizer, and a face recognizer. Unlike most methods, the facial regions contain the eyebrows, eyes, and nose, but not the mouth. The rationale of using only the upper face is to build a robust system that excludes the influence of facial variations due to expressions that cause motion around the mouth. To improve robustness, the segmented facial region images are first processed to produce two features at a reduced resolution of 1410: normalized intensity features and edge features, both in the range $(0, 1)$. These features are fed into two PDBNNs and the final recognition result is the fusion of the outputs of these two PDBNNs.

Compared to most multi-class recognition systems that use a discrimination function between any two classes, PDBNN has a lower false acceptance/rejection rate because it uses the full density description for each class. In addition, this architecture is suitable for hardware implementation such as distributed computing. However, it is not clear how to accurately estimate the full density functions for the classes when there are only limited numbers of samples. Further, the system could have problems when the number of classes grows exponentially.

2.2.3 Geometric-based Approaches

Many methods in the geometric matching category have been proposed, including early methods based on geometry of local features [80, 81] as well as 1D [147] and pseudo-2-D [148] HMM methods. One of the most successful of these systems is the Elastic Bunch Graph Matching (EBGM) system [122, 163], which is based on Dynamic Link Architecture (DLA) [34, 88]. Gabor wavelets play a building block role for facial representation in these graph matching methods. A typical local feature representation consists of wavelet coefficients for different scales and rotations based on fixed wavelet bases (called jets in [122]). These locally estimated wavelet coefficients are robust to illumination change, translation, distortion, rotation, and scaling. The basic 2-D Gabor function and its Fourier transform are

$$\begin{aligned} g(x, y : u_0, v_0) &= \exp(-[x^2/2\sigma_x^2 + y^2/2\sigma_y^2] + 2\pi i[u_0x + v_0y]), \\ G(u, v) &= \exp(-2\pi^2(\sigma_x^2(u - u_0)^2 + \sigma_y^2(v - v_0)^2)), \end{aligned} \tag{2.8}$$

where σ_x and σ_y represent the spatial widths of the Gaussian and (u_0, v_0) is the frequency of the complex sinusoid.

The DLA architecture was recently extended to Elastic Bunch Graph Matching [163] (Figure 2.3.) This is similar to the EGBM method described above, but instead of attaching only a single jet to each node, the authors attached a set of jets (called the bunch graph representation, Figure 2.3(b)), each derived from a different face image. To handle the pose variation problem, the pose of the face is first determined [85], and the “jet” transformations under pose variations are learned [110]. Systems based

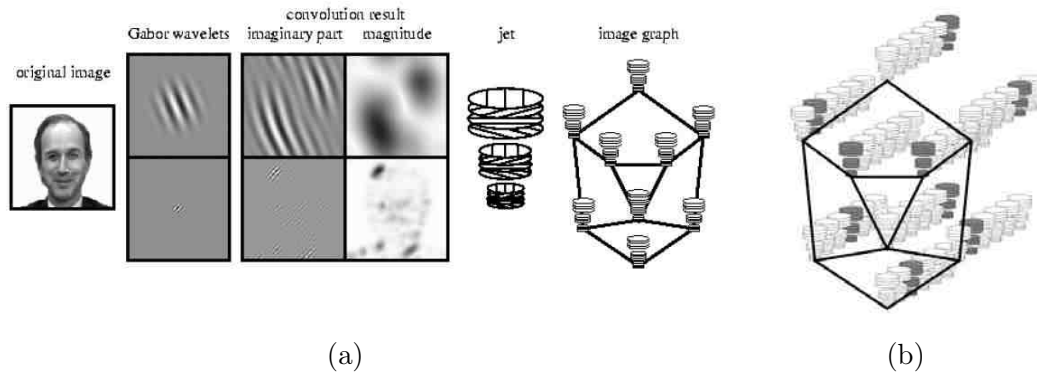


Figure 2.3: The graph representation of faces used in elastic bunch graph matching [163].(a)Elastic graph representation. (b)Bunch graph.

on the EBGm approach have been applied to face detection and extraction, pose estimation, gender classification, sketch-image-based recognition, and general object recognition. The success of the EBGm approach may be due to its resemblance to the human visual system [25].

Hybrid Approaches

Hybrid approaches use both holistic and local features. For example, the presented modular eigenfaces approach by Pentland *et al.* in [132] uses both global eigenfaces and local eigenfeatures. They extended the capabilities of the earlier system [159] in several directions. In mugshot applications, usually a frontal and a side view of a person are available; in some other applications, more than two views may be appropriate. The first approach pools all the images and constructs a set of eigenfaces that represent all the images from all the views. The other approach uses separate eigenspaces for different views, so that images taken from each view have their own eigenspace. The second approach, known as view-based eigenspaces, performs better.

It has been argued that practical systems should use a hybrid of PCA and LFA. Such view has been long held in the psychology community [33]. It seems to be better to estimate eigenmodes/eigenfaces that have large eigenvalues (and so are more robust against noise), while for estimating higher-order eigenmodes it is better to use LFA. To support this point, it was argued in [131] that the leading eigenpictures are global, integrating, or smoothing filters that are efficient in suppressing noise, while the higher-order modes are ripply or differentiating filters that are likely to amplify noise.

A flexible appearance model-based method for automatic face recognition was presented in [91]. To identify a face, both shape and gray-level information are modeled and used. The shape model is an ASM; these are statistical models of the shapes of objects which iteratively deform to fit to an example of the shape in a new image. The statistical shape model is trained on example images using PCA, where the variables are the coordinates of the shape model points. For the purpose of classification, the shape variations due to interclass variations are separated from those due to within-class variations (such as small variations in 3-D orientation and facial expression) using discriminant analysis. Based on the average shape of the shape model, a global shape-free gray level model can be constructed, again using PCA. To further enhance the robustness of the system against changes in local appearance such as occlusions, local gray-level models are also built on the shape model points. Simple local profiles perpendicular to the shape boundary are used. Finally, for an input image, all three types of information, including extracted shape parameters, shape-free image parameters, and local profiles, are used to compute a Mahalanobis

distance for classification. Based on 10 images for training and 13 images for testing for each of 30 individuals, the classification rate was 92% for the 10 normal testing images and 48% for the three difficult images with considerable changes in facial expressions, head pose or lighting condition.

2.2.4 3-D Face Recognition

3-D-based approaches provide a better solution to deal with variations in pose and illumination [31]. Work on 3-D face recognition based on range image started in late 80's and has grown significantly in the last few years. It is difficult to describe all the approaches reported in the literature [15, 37, 126, 130, 145, 105, 72, 109, 56, 42, 32, 64, 96, 180, 98, 171, 79], so we focus only on the most important works.

Chang *et al.* [37] describe a “multi-region” approach to 3-D face recognition from range data. It is a type of classifier ensemble approach in which multiple overlapping sub-regions around the nose are independently matched using (Iterative Closest Points) ICP, and the results of the multiple 3-D matches are fused. Their experimental evaluation is based on the Face Recognition Grand Challenge (FRGC) version 2 data set, representing over 4,000 images from over 400 persons. With one neutral-expression image enrolled as the gallery for each person and all subsequent images (of varied facial expressions) used as probes, performance of 92% rank-one recognition is reported.

Lee *et al.* [97] propose an approach to 3-D face recognition based on the curvature values at eight feature points of the face. Using a support vector machine for classifi-

cation, they report a rank-one recognition rate of 96% for a data set representing 100 persons. They use a Cyberware sensor to acquire the enrollment images and a Genex sensor to acquire the probe images. The recognition results are called “simulation” results, apparently because the feature points are manually located.

Russ *et al.* [145] developed an approach using Hausdorff distance matching on the range image representation of the 3-D face. An iterative registration procedure similar to that in ICP is used to align probe data to gallery data. Various means of reducing space and time complexity of the matching process are explored. Experimental results are presented on a part of the FRGC version 1 data set, using one probe per person rather than all available probes. Performance as high as 98.5% rank-one recognition, or 93.5% verification at a false accept rate of 0.1%, is achieved.

Medioni and Waupotitsch [112] performed 3-D face recognition using an iterative closest point (ICP) approach to match face surfaces. They used 3-D shapes acquired by a passive stereo sensor. Experiments with seven images each from a set of 100 subjects were reported, with the seven images sampling different poses. An EER of “better than“ 2% was reported.

Pan *et al.* [126] mapped the range data to a circular range image and then applied the PCA technique for matching. Mapping of the range data was achieved by finding the tip of the nose as a center point and an axis of symmetry for alignment. Experimental results are reported using the FRGC version 1 data set with 95% rank-one recognition rate or 2.8% EER in a verification scenario.

Chowdhury and Chellappa [41] obtain 3-D range data of the face from a 3-D acquisition system which contains a digital camera and a light projector that projects

parallel stripes of light on the face, i.e., structured light method. The camera captures the image with light stripes deformed on the surface of the face. They use the deformation of the stripes along with their location on the image to estimate depth. As reported by the authors, this method can lead to inaccurate results near the mouth and the eyes regions. Once the 3-D geometry and the facial texture are available, images can be generated for that face from almost any pose and under any illumination using computer-graphics methods [27]. For example, in [27], morphable models are utilized to map one view of the face to another. Given a face view, 3-D shape and texture from the 3-D morphable model are generated to match that image and then using graphic approaches, it generates the other views for recognition.

Lengagne *et al.* [99] proposed a 3-D face reconstruction scheme using a pair of stereo images for modeling and recognition. However, they did not implement the recognition module. Atick *et al.* [72] proposed a reconstruction method of 3-D face surfaces based on the Karhonen-Loeve (KL) transform and shape from shading. They discussed the possibility of using eigenhead surfaces in face recognition applications.

Yan *et al.* [170] proposed a 3-D reconstruction method to improve the performance of face recognition by making Atick *et al.*'s reconstruction method rotation-invariant. BenjArie *et al.* [21] proposed a volumetric frequency representation (VFR) for pose invariant face recognition.

Zhao *et al.* [180] proposed a method to adapt a 3-D model from a generic range map to the shape obtained from shading for enhancing face recognition performance in different lighting and viewing conditions. Georghiadis *et al.* [56] modeled variations in illumination and pose using illumination cone models. Zhang *et al.* [175, 174]

utilized a 3-D generic head model to estimate the head pose in face images, and then made use of Euclidean distance transform to compute feature curves extracted in the image with those in the projected head model. Their algorithm depends on manual selection of certain feature points in the face image.

Chowdhury and Chellappa [41] built the 3-D model of the face from monocular video sequences using structure from motion. The accuracy of this method depends on the quality of the video sequences. Analysis-by-synthesis was used for building 3-D face models from a single image of the face [27, 28], with excellent results. The method in [28] uses 200 pre-scanned face models to build an eigenvector space so that an image of a face can be represented as a projection of a linear combination of the eigenmodels. This method requires labeling seven to eight facial features points manually for images used in building the database and for probe images. This method takes a personal computer several minutes to obtain the morphing parameters of the face from an image. In [175] a method for building 3-D models by morphing a generic model based on different views of a face was presented. In this method manual intervention is needed to select a point on each facial feature component such as the eye, brow, nose, and chin. Also, their use of texture synthesis would not work for different lighting conditions.

2.2.5 Multi-modal Face Recognition

In biometrics, the classical definition of “multi-modal” refers to the use more than one modality or multiple sensors, in order to increase the accuracy and robustness

of the biometric recognition system. The goal is to counterbalance the imperfection of one modality by the remaining ones. The fusion of modalities can take place at different levels, namely the input sensor level, the feature level, and/or the decision level. An example of fusion at the feature level is when features such as those of the face and hands are combined and represented in one feature vector. One of the most applicable strategies is the fusion at the decision level. Each algorithm for a single biometric modality is regarded as a single classifier that represents the decision or matching score of the algorithm. The fusion is primarily based on the distributions of the classifiers' outputs. As long as the features from the different modalities are statistically independent, the fusion of the classifiers' outputs seems to be a promising approach. An example for independent biometric modalities might be face and fingerprint, where it cannot be assumed that the structure of the fingerprint contains information about the face and vice versa. However, in the fusion of 2-D and 3-D face recognition, the situation is different as both modalities are likely to be dependent to a certain degree. For instance, the relative positions of the eyes, nose, and mouth will be the same in both the 2-D and 3-D representation of the face. Even more, there may occur effects during acquisition like pose or occlusions that will affect both modalities. The disadvantage of this dependency is that the fusion at the level of the classifiers outputs will be less beneficial than in the case of completely independent classifiers' outputs. However, the great benefit of the dependency is that it enables additional levels of fusion during the enrollment, where the information from one modality might support the other. For example, if both texture and shape information of each face are acquired using 3-D scanners, the image from the texture

modality is automatically registered to the shape modality. Thus it is sufficient to find landmarks in one of the two modalities where certain landmarks might be easier to find with higher accuracy than from the other modality.

Lao *et al.* [93] proposed a framework for 3-D face recognition. A sparse depth map is constructed from three stereo images using Iso-luminance lines for the stereo matching. A 3-D model is constructed from 3-D geometrical features representing arcs and line edges which are extracted from the 3-D data. By searching for arcs whose radii are of certain ranges, they first locate the irises and the mouth and then use this information to estimate the pose. 3-D Recognition is performed by calculating the mean differences in depth between corresponding data points in the test 3-D model and all the models in the database. Using a dataset of 10 persons, they reported 87-96% recognition rate. One drawback of this approach is that only depth is used for recognition. This approach is multi-modal in a specific limited sense (using 2-D stereo images for building the 3-D model.)

Bueumier and Acheroy [24] used the fusion of 3-D facial surfaces obtained from structured light method and grey level in 2-D images. For detail explanations on structured light method, we refer the readers to [41]. The 2-D and 3-D data of the face are each represented with a central profile and a lateral profile. Therefore, they have a total of four classifiers, and an overall decision is made using a weighted sum of 3-D and 2-D similarity measures. For experiments on a subset of the data of 100 persons, using a 27-person gallery and a 29-person probe set, they reported an EER as low as 1.4% for recognition that merges multiple probe images per subject.

Wang *et al.* [162] presented a feature-based face recognition system based on both 3-D range data as well as 2-D gray scale facial images. Feature points are extracted and described by Gabor filter responses in the 2-D domain and Point Signatures in the 3-D domain. The corresponding normalized shape and texture weight vectors are then integrated to form an augmented vector which is used to represent each facial image. For a given test facial image, the best match is identified according to a similarity function. Experimental results involving 50 persons, with six images per person, demonstrated a recognition rate of 90%.

Bronstein *et al.* [32] used range images with texture and presented a 3-D face recognition approach based on geometric invariants, which maps 2-D facial texture images into special images that incorporate the 3-D geometry of the face. Although they never reported any quantitative performance, they claimed that their system provides high recognition and can cope with variations caused by facial expressions.

Tsalakanidou *et al.* [158] evaluated three different approaches, color, depth, and fusion of color and depth for face recognition. Experimental results, on a dataset of 40 persons, are reported for color images alone, 3-D alone, and 3-D plus color. Using the Principal Component Analysis approach (PCA), recognition rate is as high as 99% for the multi-modal part which is found to be higher than either 2-D or 3-D data alone. Similarly, using a PCA-based recognition approach, Chang *et al.* [36] performed experiments with 3-D and 2-D images for 200 subjects and reported approximately 99% rank-one recognition for multi-modal fusion of 3-D with 2-D, 94% for 3-D alone, and 89% for 2-D alone. The multi-modal results are obtained using a weighted sum of the distances from individual 3-D and 2-D faces.

Godil *et al.* [58] considered fusion at image level and score level using shape from range images and texture map information of 200 subjects from CAESAR anthropometric database. The image level fusion is created by concatenating of 3-D shape and color map information. The score level fusion combines scores using min, max, mean, and product rule. They also used PCA for matching both the 2-D and the 3-D data. Using the multiple score level, the reported performance is as high as 82%.

Papatheodorou *et al.* [128] used a commercial stereo camera system for 3-D data acquisition. The stereo system is made up of three video cameras and a speckled pattern projector. The projector projects a random light pattern of dots on the surface of the face, used to establish correspondences between two of the three cameras, allowing the reconstruction of depth information. The third camera captures the texture information and uses a filter to eliminate the speckled pattern projected onto the face. Their approach combines modality at the sensor stage in 4D space as in $(x, y, z, \text{intensity})$. Recognition experiments, for 62 persons, based on Iterative Closest Point (ICP), show 98-100% recognition rate from the frontal view probes, 73-94% for probes with varying poses, and 69-89% for probes with smiling expressions.

In a multi-modal recognition approach integrating intensity and 3-D range data, Tsalakanidou *et al.* [157] used embedded hidden Markov model technique applied to depth map. Their experimental data set represents a small number of different persons, but each has 12 images representing various facial poses and expressions. They reported a higher EER for 3-D than for 2-D in matching frontal neutral-expression probes to frontal neutral-expression gallery. They also report that depth data mainly suffers from pose variations.

Lu and Jain [106, 104] reported better performance with 3-D matching alone than with 2-D matching alone. They also reported 98% rank-one recognition for 3-D + 2-D recognition on neutral face expressions and 91% on the larger set of neutral and smiling expressions. In their algorithm, they use ICP style matching of 3-D shape [165] and a linear discriminant analysis approach for the 2-D matching component. Their experimental data set consists of multiple scans of each of 100 persons. Five scans with a Minolta Vivid 910 system are taken in order to create a more accurate 3-D face model for enrolling a person. Enrollment is done with neutral expression. An individual probe for testing uses six scans for each person; three with neutral expression and three with smiling expression.

Maurer *et al.* [109] described the approach used by Geometrix, which is based on multi-modal (3-D + 2-D) face recognition. The 3-D matching builds on the approach described by Medioni and Waupotitsch [112], whereas the 2-D matching uses the approach of Neven Vision [9]. A weighted sum rule is used to fuse the two results, with the exception that, when the shape score (3-D) is very high, the texture score (2-D) is ignored [109]. Experimental results on the FRGC V2.0 data set (all versus all matching of the 4,007 images) show 87% verification at 0.01 FAR. They also reported that the performance of 3-D+2-D outperforms that of 3-D alone by a noticeable amount, and that the verification rates for 2-D alone are below those for 3-D alone.

Husken *et al.* [69] described the Viisage approach for multi-modal recognition. The 3-D matching follows the style of hierarchical graph matching already used in Viisage's 2-D face recognition technology. This approach is fast in matching compared with techniques based on ICP or similar iterative techniques. Fusion of the results

from the two modalities is done at the score level. Multi-modal performance on the FRGC version 2 data is reported as 93% verification at 0.01 FAR. In addition, it is reported that the performance of 2-D alone is only slightly less than that for multi-modal performance, and that performance of 3-D alone is substantially less than that of 2-D alone. In this context, it is interesting to note that results from the Geometrix group, that originally focused on 3-D face recognition, show that 3-D alone outperforms 2-D alone, whereas results from the Viisage group that originally focused on 2-D alone, show that 2-D alone outperforms 3-D alone.

A-Ansari [11] provided a methodology for building 3-D mesh models of faces, for 3-D to 3-D and 2-D to 2-D face recognition applications. In particular, he developed two algorithms for face modeling from either stereo or range images and one algorithm based on 3-D model face image synthesis. The first algorithm produces a deformed 3-D model using stereo images from two parallel frontal views and one orthogonal profile view of the face, while the second algorithm produces a deformed 3-D model from one range image of the face. Both algorithms obtain 3-D models which are applied in 3-D face recognition. Also, the stereo-image-based algorithm is capable of synthesizing multiple views 2-D facial images which are applied in 2-D face recognition. He combines 3-D information of depth from stereo with 2-D extracted labels for facial feature points. He then used a generic mesh model which establishes direct feature vertices correspondences between all subjects, resulting in a faster face recognition comparison. He tested the performance of his system for 3-D face recognition on a face database of 112 subjects captured at the University of Miami. He achieved 96.4% rank-one identification by using a mesh model with 6080 vertices.

Chapter 3

Facial Features Extraction

In this chapter we present our techniques for 3-D and 2-D facial features extraction. For 3-D range images, we develop an algorithm to extract three feature points (i.e., the two inner corners of the eyes and the tip of the nose) based on Gaussian curvature. In addition, we improve the ASM technique for 2-D facial features extraction which is a statistical based shape modeling approach. The 2-D extracted facial features will be used in building the ARG model in Chapter five.

3.1 3-D Facial Features Extraction

As we reviewed in Chapter two, the location of the facial features are determined by calculating the minimum and maximum principal curvature. The extracted location of the features (e.g., landmark points) may be utilized for initial alignment of a probe facial image to the gallery image. Moreover, the extracted feature points can be used as signature for shape matching. In this Section, we present an algorithm for

extraction of three feature points in range images. The feature points (i.e., landmark points), are the two inner corners of the eyes and the tip of the nose. These features are utilized for initial alignment of the probe ridge image to the gallery ridge image.

Before extracting the facial features, the range images needs to be preprocessed and filtered to remove the noise and artifacts of range images. In order to remove sharp spikes that occur during scanning of the face, we apply median filtering with a window size of 3×3 . Afterwards, we use interpolation (nearest neighbor points) to fill the gaps on the face region and finally we use a low pass filter (disk with radius 3 pixels) to slightly smooth the surface of the face that suffers from rapid changes due to facial hair or any other artifacts.

After preprocessing, the area of the face in the range data is localized and the neck, hair and the background areas of the range image are discarded. Then, the inner corners of the eyes and the tip of the nose are detected. In the following subsections, we explain in detail the process of face localization using template matching and the labeling of the three feature points using Gaussian curvature.

3.1.1 Face Localization Using Template Matching

Generally, face detection is one of the first preliminary steps for any face identification or processing system. In the case of 2-D textured facial images, there are different methods for face detection [160, 141, 63]. For range data, we need to find the location of the face in the range image, keep the face area, and exclude the background, hair, and neck in the image. Therefore, we developed a method for localizing

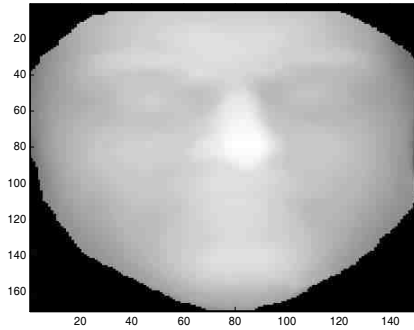
faces in range data using template matching. We adopt an algorithm similar to the normalized cross correlation (Equation 3.1) to match a facial template range image to the range images:

$$C(s, \theta) = \frac{\sum_{x,y} [f(x,y) - \bar{f}] [H_{s,\theta}(\tau(x,y)) - \bar{\tau}]}{\{\sum_{x,y} [f(x,y) - \bar{f}]^2 \sum_{x,y} [H_{s,\theta}(\tau(x,y)) - \bar{\tau}]^2\}^{0.5}} \quad (3.1)$$

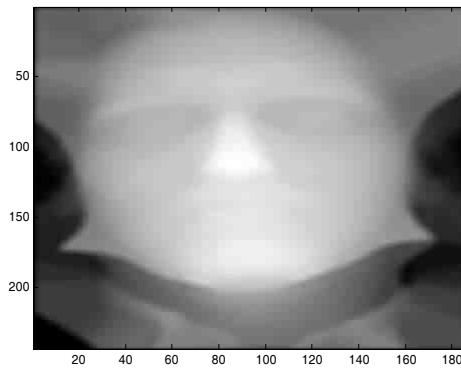
where $f(x, y)$ and $\tau(x, y)$ are the image and the template respectively. \bar{f} is the mean value of the portion of the image underneath the template and $\bar{\tau}$ is the mean value of the template. The $H_{s,\theta}$ is similarity transformation matrix with parameters s and θ for scale and rotation, respectively.

For template matching, at first we roughly detect the location of the nose tip. Then, we translate the template face such that the detected tip of the nose is placed on the location of the nose tip of the range image under test. Afterwards, we iteratively apply a 3-D similarity transformation (only scale and rotation is considered) to the template image and calculate the normalized cross correlation, i.e. $C(s, \theta)$, to obtain the optimum scale and pose orientation of the template that results in the maximum correlation between the template and the range image.

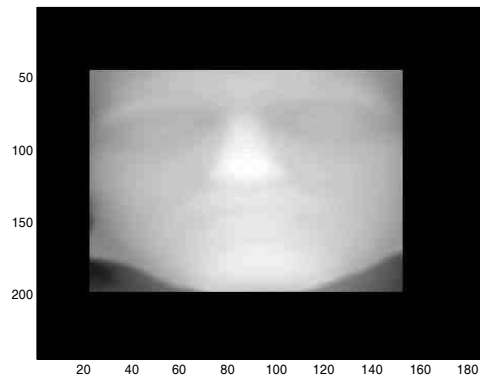
Figure 3.1 shows the template image as well as the result of the template matching to one of the facial range images in the database. This approach is robust for cropping the face region in the range data.



(a)



(b)



(c)

Figure 3.1: Template matching: (a) Template range image (b) A sample filtered facial range image (c) Detected area of the face in the sample range image.










	$K > 0$	$K = 0$	$K < 0$
$H < 0$	peak 	ridge 	saddle ridge 
$H = 0$	none 	flat 	minimal 
$H > 0$	pit 	valley 	saddle valley 

Figure 3.2: The relations between surface types and their mean (H) and Gaussian (K) curvatures.

3.1.2 Labeling Feature Points Using Gaussian Curvature

In order to extract three feature points that is utilized in face alignment during the recognition process, we use Gaussian curvature. The extracted feature points are the two inner corners of the eyes and the tip of the nose. For a given surface $z = f(x, y)$, the mean curvature H , the Gaussian curvature K , and the principal curvatures k_{max} , k_{min} , are defined as:

$$\begin{aligned}
 K &= \frac{f_{xx}f_{yy} - f_{xy}^2}{(1 + f_x^2 + f_y^2)^2} \\
 H &= \frac{f_{xx} + f_{yy} + f_{xx}f_y^2 + f_{yy}f_x^2 - 2f_xf_yf_{xy}}{2(1 + f_x^2 + f_y^2)^{1.5}} \\
 k_{max} &= H + \sqrt{(H^2 - K)} \\
 k_{min} &= H - \sqrt{(H^2 - K)}
 \end{aligned} \tag{3.2}$$

Because, the calculation of Gaussian curvature involves the second derivative of the surface function, the noise and the artifacts affect the final result and applying a low-pass filter to smooth the data is required. Figure 3.2 shows the representation of different shapes and their corresponding mean and Gaussian curvatures. As the

Figure shows, the surface that either has a peak or a pit shape has a positive Gaussian curvature value ($K > 0$). Each of the two inner corners of the eyes has a pit surface type, and the tip of the nose has a peak surface type that are detectable based on the Gaussian curvature. These points have the highest positive Gaussian curvature values locally, among the points on the face surface. Figure 3.3 shows the result of calculating Gaussian curvature for one of the sample range images in the gallery. The highest points in Figure 3.3.(a) correspond to the points with pit/peak shape. We threshold the Gaussian curvature to find the areas that have positive Gaussian curvature values greater than a threshold, producing a binary image (Fig. 3.3.b). This threshold is calculated based on a small training data set different from the images used in the recognition experiments. The three regions with the largest average value of Gaussian curvature are the candidate regions that include the feature points. The locations of the points with maximum Gaussian curvature in these regions are labeled as feature points. Figure 3.3.(c) shows a final result of feature points labeling.

3.1.3 Experiments and Results

We evaluated our algorithm for locating the face and 3-D features extraction using the 3-D Gavab database [115]. The Gavab database contains 549 range images for 61 individuals (45 males and 16 females). For each person, there are nine different images, two neutral frontal images, two neutral images with pose (looking down and up), two profile images, and three frontal images in which the subject presents different and accentuated facial expressions. The digitizer is a Minolta VI-700 digitizer, a

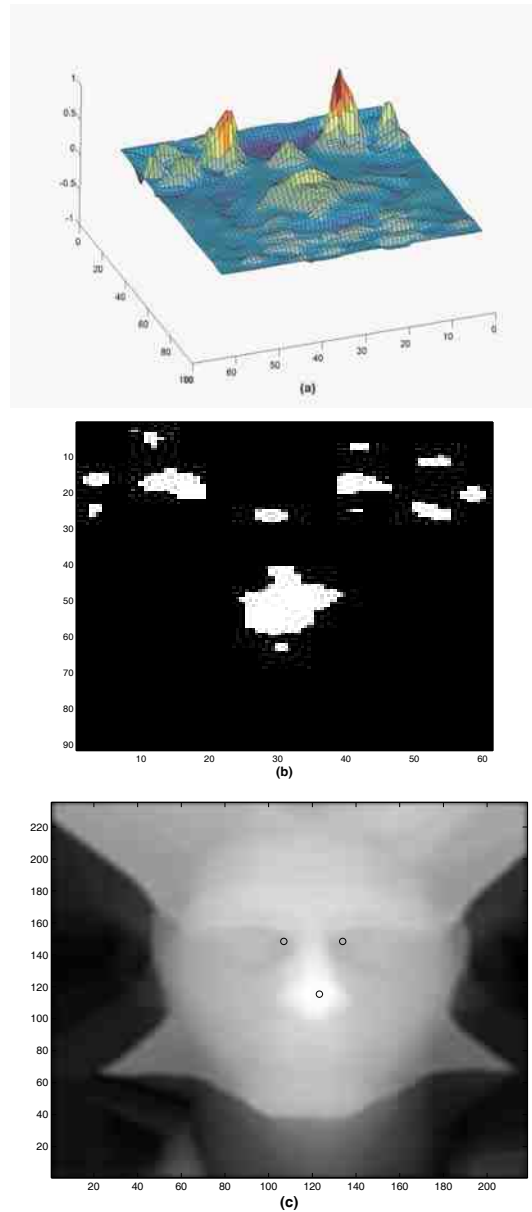


Figure 3.3: Three feature points labeling: (a) Gaussian curvature on a patch around the nose and eyes (b) Result of thresholding the Gaussian curvature image(c) Final result of feature points labeling

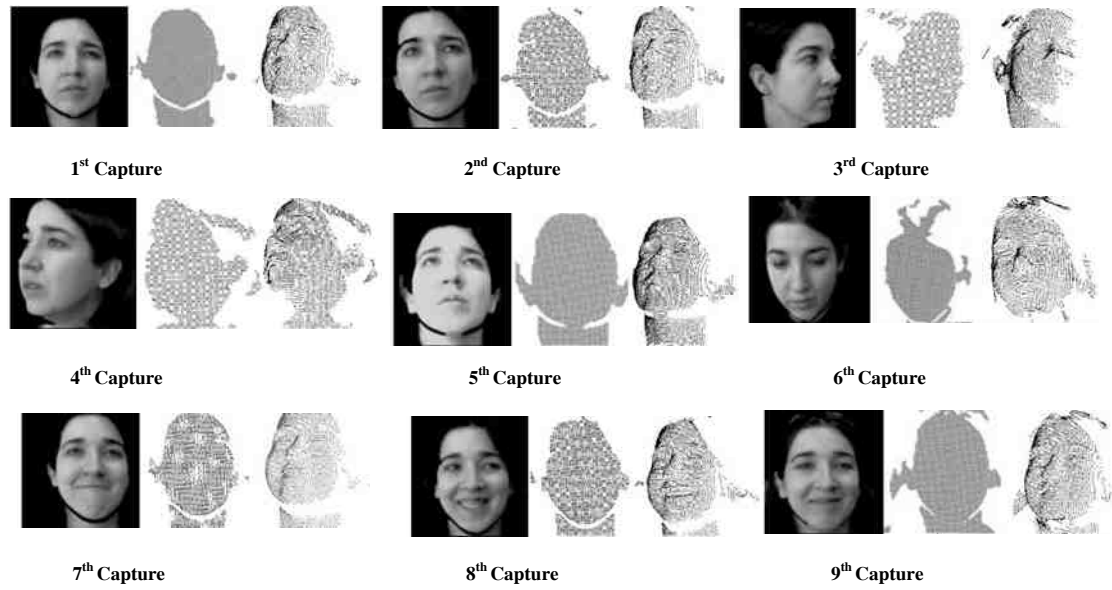


Figure 3.4: 2-D and 3-D range views of an individual: 2-D gray scale images, 3-D range images at 1/4 of the original resolution (both from scanner’s point of view) and a rotated version of 3-D range image [115].

laser sensor which captures in less than a second a range image of the scene as well as a color image. The faces were at a distance of 1.5 ± 0.5 m from the scanner. Figure 3.4 shows the range images for one of the subjects in the database along with the textured images [115]. The texture images for each person are not released and only the range images are available for public access.

In our experiments, we used the two neutral frontal images (the 1st and the 2nd captures), the two neutral looking up and down images (the 5th and the 6th captures), the frontal images with smile expression (the 7th capture), the frontal images with laughing expression (the 8th capture), and a frontal image with random gesture (the 9th capture) (In this case, occlusions of the face by the hand or by the tongue are permitted.)

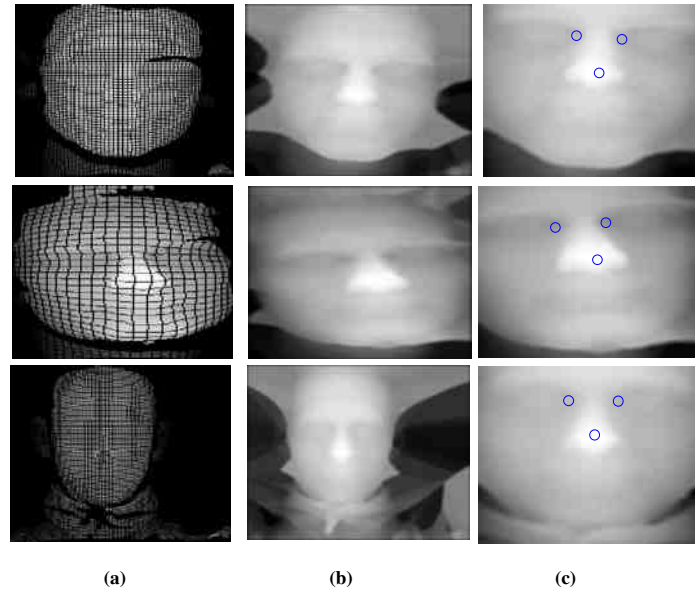


Figure 3.5: Samples of range images in the gallery and the results of preprocessing (a) Original range images, (b) Noise removal and interpolation, (c) Face localization and three feature points labeling.

Figure 3.5 shows three samples of the original images in the Gavab database, results of noise removal and interpolation, face localization, and feature points labeling. The process of labeling three feature points is successful but fails for few subjects (15% of the images in the database), where the noise and the disturbance in the image around the eyes, (i.e., eyelash) are high. Also, for few cases (10% of the images in the database), where the face has pose (looking up or down), the initial detection of the nose is difficult and the nose is mistaken with the chin or forehead. For images where our algorithm for labeling the three feature points failed, we manually labeled the locations of the feature points to be utilized in the matching process.

3.2 2-D Facial Features Extraction

In Chapter two we reviewed the techniques for facial features extraction. We mentioned that extracted facial features can be used for either alignment (registration) between the query and the template images or the models. In addition, extracted facial features can be used for constructing a face model where the labeled feature points play important role in building the model. In this dissertation, the aim is to label fiducial points in a given 2-D color facial image and use them for representation of the given face by an Attributed Relational Graph (ARG), where the nodes of the graph correspond to the landmark points that are extracted by the presented algorithm in this chapter.

In order to extract facial feature points, we improve the Active Shape Model (ASM) [45] with respect to three factors: model initialization, modeling of the local structure of the facial features, and alignment of the shape model to a new instant of the object in a given image. The core of our contributions relies on three improvements: (a) initializing the ASM model using the centers of the mouth and eyes, which are located using color information; (b) incorporating color information to represent the local structure of the feature points; (c) applying 2-D affine transformation in aligning the facial features that are perturbed by head pose variations, thus effectively aligning the matched facial features to the shape model and compensating for the effect of the head pose variations.

3.2.1 Active Shape Model and Its Limitations

Active Shape Model (ASM) is a statistical approach for shape modeling and feature extraction. It represents a target structure by a parameterized statistical shape model obtained from training. This method was introduced by Cootes *et al.* [45, 92] and improved by other researchers over the past few years. In the original version of the ASM, the initial locations of the feature points are obtained from the mean shape, which is derived from training data, and its accuracy depends on the size of the training set. In addition, the local structure of the feature points is represented by the change in intensity values of the pixels along a profile line (i.e., edge location) that goes through the feature points. This is based on the assumption that usually the facial features are located on strong edges. But, finding the correct locations of the feature points on the edges is not always possible and this affects the robustness of the ASM technique for feature extraction.

Ginneken *et al.* [57] proposed to use a non-linear gray level appearance instead of the first derivative profile to model the local structure of the features. In [161], Wei Wang *et al.* had some improvements on the ASM for face alignment. Other authors [173, 74, 65] used the wavelet transform to model the local structure of features and improve the face alignment. Unfortunately, their approaches using wavelet transform are computationally expensive.

The most important limitations of the ASM for facial feature extraction can be summarized as follows:

- Representation of complex multi-part objects by a single shape model. Although a single shape model may preserve the general shape of the whole object, its constraint may fail in extracting some of its parts.
- Representing the local structure of each point by independent models. This drawback may lead to a final shape far from the actual shape model.
- The need for a large training set to cover shape variations. The shape model may fail in characterizing the shape variations if instances of the shape are not incorporated in the training set.
- The initialization of the shape model. This is a major drawback of ASM. If the shape is initialized far from the object of interest, the searching process may either fail or become very slow.
- The choice of modeling the local structure of the points. Many variations of ASM model the local structure by edges or statistical models of gray level variations.
- Alignment of the shape model to a new instant of the object. Alignment has the same effect as initialization. Successful alignment leads to faster and accurate model fitting to the object of interest in the image.
- Search for the best candidate feature points. Most existing algorithms rely on Euclidean or Mahalanobis distance between the candidate feature points and the trained model of the local structure of the feature points.

Some of the above are inherent limitations in ASM and some of them depend on the method used to handle them. For example, the representation of complex multi-part objects by a single shape model is an inherent problem of ASM; while modeling the local structure of the feature points and alignment of the shape model to a new instant of the object are not. In this dissertation, we deal with model initialization, modeling the local structure of feature points, and shape model alignment to a new instant of the object.

For initial alignment we use the color information to initially detect few salient facial features such as the centers of the mouth, and the eyes. These points are employed to initialize the ASM. However, these three feature points are not the only feature points that can be used for initial alignment. Sometimes, an expert user manually labels a set of feature points for the purpose of initial alignment. For example, four labeled feature points (two outer corner of the eyes, the tip of the nose, and a point on cheek) were extracted manually and released with the FRGC face database which is used in this work.

We also use the color information to improve the model that characterizes the local structure of the feature points. A weighted sum of three multivariate Gaussian models for the three components (i.e. Hue, Saturation, and Value) is used to represent the normalized first derivative of pixel values along a profile line. Furthermore, for the lips, we use the color information to detect their boundary. This enhances the localization of geometric features that represent the external boundary of the lips in face images. In addition, we use 2-D affine transformation to align the extracted facial features to the shape model. The 2-D affine transformation compensates the effect

of head pose variations and the projection of 3-D data to 2-D. In fact, ASM needs a large training set to cover the variations caused by head pose. For frontal images, the similarity transformation is a suitable transformation to align the extracted facial features to the shape model, but for large variations in the head pose, it is not suitable.

The use of the color information for modeling the local structure of the feature points and the use of the 2-D affine transformation are general improvements to the ASM approach. On the other hand, the initialization of the ASM using the centers of the mouth and the eyes, and localizing the feature points around the lips are specific for facial feature extraction. Our experimental results show that the proposed approach outperforms the standard ASM technique for facial feature extraction. The details of our improvement for Active Shape Model is presented in Appendix A.

3.2.2 Experiments and Results

In this Section, we validate our algorithm for 2-D facial features extraction. We compare between performance of the standard ASM algorithm and our improved ASM approach. In our comparison, we use a subset of the University of Miami (UM) face database which includes 70 different subjects with a total of 140 near frontal color images [11]. For each subject, we captured three near frontal images. One image is used as a probe and another image is used for database storage. The third image is used for manual feature extraction, labeled by an expert person, for ground truth comparison. We use a trained shape model for 75 facial feature points which is provided from a completely different source of images and other public databases.

Figure 3.6 shows few samples of the facial images in UM database and the extracted facial features by both our method and the original ASM method. The visual inspection shows that our approach is more successful in extracting the locations of the facial feature points, especially around the lips and the corner of the eyes and the eyebrows. From these few examples, it is clear that our method leads to more accurate localization of the facial features as we further show next.

Performance Evaluation

To evaluate the performance of our method, we manually labeled 75 feature points in each of the images in the database. For each subject image, the 75 facial feature points were extracted using the original ASM and our improved approach. The performance was evaluated using the average mean square error (MSE) in pixel unit over all the images. The error is defined as the distance between the manually labeled feature points and the corresponding feature points obtained from both the original and our improved version of the ASM approach. This is defined as follows:

$$AverageMSE = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{n} \sum_{j=1}^n \|P_{ij} - P'_{ij}\|^2 \right) \quad (3.3)$$

where N is the total number of probe images, n is the number of the landmark points in the shape model, P_{ij} is the j^{th} landmark point in the manually labeled shape of the i^{th} test image, P'_{ij} is the j^{th} landmark point in the resulting shape of ASM for the i^{th} test image, and $\|\cdot\|$ denotes the Euclidean distance.

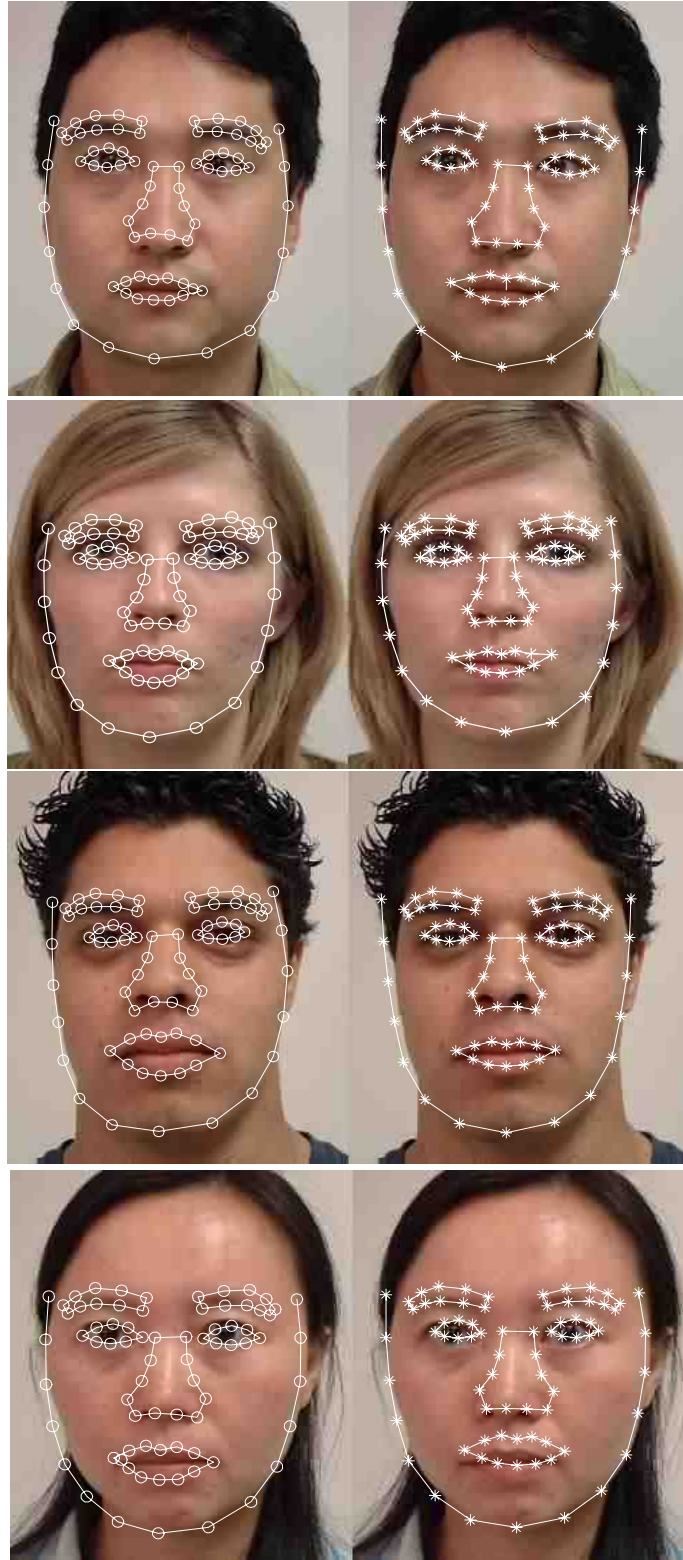


Figure 3.6: Extracted facial features sample images from UM database, where our method outperforms the standard method, (o) Improved ASM, (*) Standard ASM. The MSE (pixels) of the enhanced to standard ASM for the images are (a) [28.85, 39.98], (b) [32.74, 33.54], (c) [20.50, 38.85], and (d) [39.29, 110.87].

Method	Ave.MSE for 49 subjects out of 70	Ave. MSE for 21 subjects out of 70	% of cases which have minimum error
Standard ASM	70.3 (pixels)	23.3 (pixels)	30.0%
Enhanced ASM	40.0 (pixels)	31.9 (pixels)	70.0%

Table 3.1: Performance Comparison Between the Two Methods and the Manually Labeled Features Based on the Average MSE (pixels) Over All 70 Subjects.

As given in Table 3.1, we categorize our results in two sets based on the average MSE. In one set, 49 subjects out of the 70 subjects, the average MSE between the manually labeled feature points and the feature points extracted by our improved ASM are lower than the average MSE between the manually labeled feature points and the feature points extracted by the original ASM method. For the second set, the remaining 21 subjects, the original ASM has lower average MSE. Our method has a 70% improvement (minimum error) over the entire database compared to 30% given by the original ASM.

Figure 3.6 shows sample images along with the MSE errors in which our method outperforms the standard ASM approach with minimum error. Similarly, Figure 3.7 gives sample cases where the standard ASM performs better than our method. In these sample examples, the figure visually shows comparable performance between the two methods. Based on the MSE, the standard ASM attains lower error in 30% of the cases.

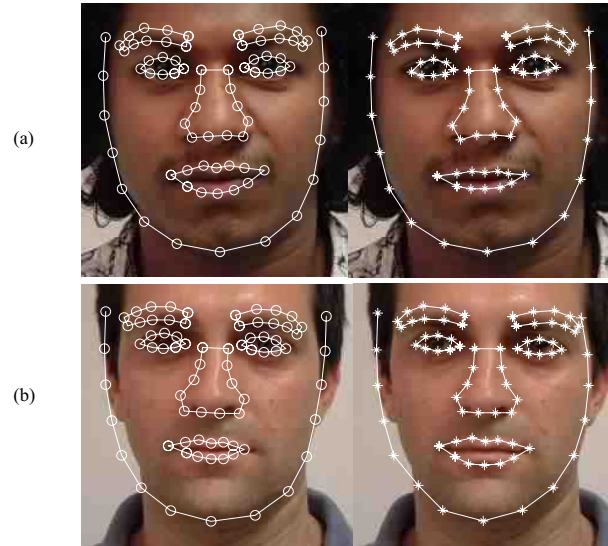


Figure 3.7: Extracted facial features sample images from UM database, where standard method outperforms our method, (o) Improved ASM, (*) Standard ASM. The MSE (pixels) of the enhanced to standard ASM for the images are (a) [34.64, 24.87] and (b) [14.02, 9.55].

3.3 Summary

For 3-D range data, we have developed a method based on template matching to find the area of the face in the range data. The goal is to localize the face area and discard the neck, hair and the background areas of the range image. Moreover, we have presented an algorithm based on Gaussian curvature for detecting three feature points, the inner corners of the two eyes and the tip of the nose. For 2-D images, we have improved the Active Shape Model approach for facial features extraction. We have used the color information to localize the centers of the mouth and the eyes to improve the initialization step in the standard ASM. In addition, we have modeled the local structure of the feature points in the HSV color space and we used a 2-D affine transformation to align the facial features that are perturbed by head pose.

Experimental results have showed that our improved version of the ASM is accurate and outperforms the standard ASM. The extracted feature points either in 2-D or 3-D are utilized in the next chapters.

Chapter 4

3-D Face Recognition Based on Ridge Images

In this chapter, we present a method for 3-D face recognition from frontal/near-frontal range images based on the ridge lines on the surface of the face. As we discussed in chapter one, the surface matching techniques suffer from computational complexity. In our approach, a subset of points on the surface of the face are selected using the principal curvature, k_{max} . These points show the locations of the ridge points around the important facial regions on the face, i.e., the eyes, the nose, and the mouth. Instead of matching all the points on the surface of the face, the ridge points are used for matching which leads to huge reduction of the computational complexity while keeping the performance of the system for face recognition promising. We compare the robust Hausdorff distance versus the Iterative Closest Points (ICP) for matching the ridge image of a given probe image to the ridge images of the facial images in the gallery.

4.1 3-D Face Matching Based on Ridge Images

Figure 4.1 shows the block diagram of our method. In the first step, because of noise and artifacts in the range images, we use median filtering and low-pass filtering to remove sharp spikes and smooth the images and then we use interpolation to fill the gaps in the image. In the next step, we roughly find the tip of the nose which is the closest point to the scanner. Because the facial range images in the databases that we work on are frontal/near-frontal, the claim that the tip of the nose is the closest point to the scanner is valid. Then, we apply template matching to localize the face region in the filtered range data. Afterward, we use Gaussian curvature to label three feature points, i.e., the inner corners of the two eyes and the accurate position of the nose tip. We represent the range images by the points on the 3-D surface of the face which have maximum principal curvature, k_{max} , greater than a threshold. Therefore, each range image is represented by ridge lines on the 3-D surface of the face using a 3-D binary image, called ridge image. The details of the preprocessing techniques and 3-D facial features labeling were presented in Chapter three.

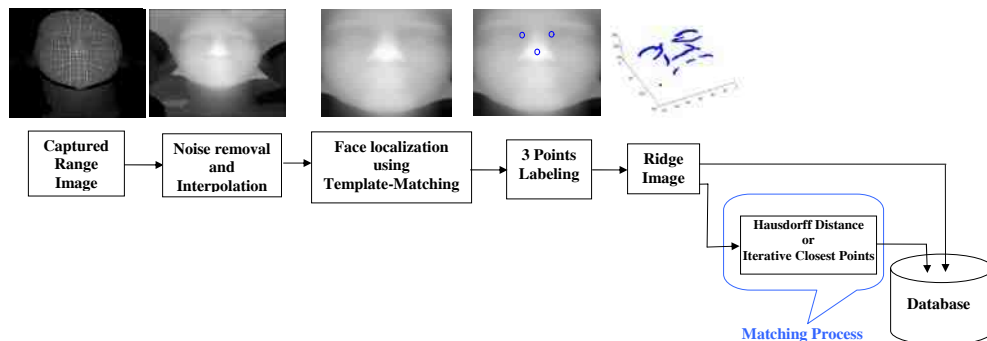


Figure 4.1: Block diagram of our system for 3-D face recognition from range data.

For recognition, we use two different techniques for ridge image matching: the robust Hausdorff distance and the Iterative Closest Points (ICP). First, we apply similarity transformation (scale, rotation and translation) to find the best pose that matches the probe image with the gallery image. The three labeled feature points plus an auxiliary point in the middle of the triangle formed by the three labeled points are used to find the initial transformation that aligns the probe ridge image to the test ridge image. The x and y values of the auxiliary point are the average value of the x and y values of the other three points and z value comes from the range image. After initial alignment, for matching based on robust Hausdorff distance, an iterative algorithm is applied to find the optimum pose that results in the minimum Hausdorff distance. For matching based on ICP, we utilized a fast variation of ICP to find the best geometric alignment between a 3-D ridge probe image and a given 3-D ridge gallery image and compute the Mean Square Error (MSE) distance between the ridge points.

4.2 Ridge Image

Our goal is to extract and use the points lying on ridge lines as the feature points. These points correspond to the extreme ridge points of the surface. In the literature [14], the ridges are defined as the umbilic points at which the k_{max} attains a local positive maximum. An umbilic point is a point on a surface where the principal curvatures are equal and are non-zero (in the case of zero curvature, the point is called a flat point.) Intuitively, ridges are the points that form the drainage patterns and are

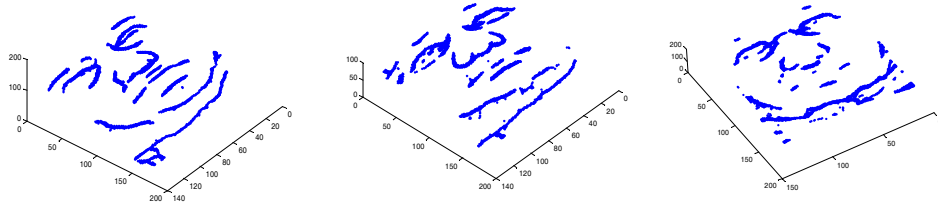


Figure 4.2: Sample of ridge image extracted for different subjects.

called valleys when the ridges are looked from the opposite side. There are different approaches to locate the ridges [103] of a 3-D surface. One of the main approaches applies thresholding the k_{max} values which is used in this work. We threshold the k_{max} values to find these points. The suitable threshold is obtained based on a small training set that is different from the images in the gallery. The suitable threshold is selected such that the highest recognition rate for that small training set is achieved. Then, in our experiments the suitable threshold (a fixed value) is used for creating the ridge images for all the facial images in the databases under evaluation.

Figure 4.2 shows few examples of the ridge images obtained by thresholding the k_{max} values. These are 3-D binary images that show the locations of the ridge lines on the surface of the face. The lines on the boundary of the face are filtered out and are not considered as feature points for recognition. To filter out the points on the boundary of the face, we ignore the points on the boundary of the matched template within a margin. In other words, after localizing the face by template matching, the points that are within a certain distance (for example 15 pixels) from the boundary of the matched face template are excluded from the process of ridge creation.

4.2.1 Ridge Matching Using Robust Hausdorff Distance

Huttenlocher *et al.* originally proposed Hausdorff distance (HD) [70] as a measure for object matching in computer vision. Unlike other shape matching methods, HD can be calculated without knowing the exact correspondences of the points in different sets. Modifications to the Hausdorff distance raise its capability to handle not only noisy points, but also missing data from occlusion and outliers [71].

Given two sets of points $\mathcal{A} = \{a_1, a_2, \dots, a_{N_{\mathcal{A}}}\}$ and $\mathcal{B} = \{b_1, b_2, \dots, b_{N_{\mathcal{B}}}\}$ of size $N_{\mathcal{A}}$ and $N_{\mathcal{B}}$, respectively, the partial Hausdorff distance between the two sets of points, \mathcal{A} and \mathcal{B} , is defined as:

$$H(\mathcal{A}, \mathcal{B}) = \max(h_K(\mathcal{A}, \mathcal{B}), h_K(\mathcal{B}, \mathcal{A})) \quad (4.1)$$

where $h_K(\mathcal{A}, \mathcal{B})$ and $h_K(\mathcal{B}, \mathcal{A})$ represent the directed distance between the two sets \mathcal{A} and \mathcal{B} . The directed distances of the partial HD are defined as:

$$h_K(\mathcal{A}, \mathcal{B}) = K_{a \in \mathcal{A}}^{th} d_{\mathcal{B}}(a), \quad h_K(\mathcal{B}, \mathcal{A}) = K_{b \in \mathcal{B}}^{th} d_{\mathcal{A}}(b) \quad (4.2)$$

where $d_{\mathcal{B}}(a)$ represents the minimum distance (e.g. Euclidean distance) value at point a to the point set \mathcal{B} , $d_{\mathcal{A}}(b)$ represents the minimum distance (e.g. Euclidean distance) value at point b to the point set \mathcal{A} , $K_{a \in \mathcal{A}}^{th}$ denotes the K^{th} ranked value of $d_{\mathcal{B}}(a)$, and $K_{b \in \mathcal{B}}^{th}$ denotes the K^{th} ranked value of $d_{\mathcal{A}}(b)$.

After Huttenlocher *et al.*'s original work, researchers have proposed many different definitions and methods to realize directed HD. Dubbisson and Jain revised

the original HD and investigated the performance of 24 different Hausdorff distance measures based on their behavior in the presence of noise [48]. They proposed the modified Hausdorff distance MHD. Sim *et al.* applied the robust statistic techniques of regression analysis to the computation of the HD measures for object matching, resulting in two robust HD measures: M-HD based on M-estimation and least trimmed square-HD (LTS-HD) based on LTS [152]. Based on the experimental matching performance of these different HD measures, robust LTS-HD based on the least trimmed square (LTS) measure [152] is adopted in our work. In the proposed LTS-HD [152], the directed distance $h_{LTS}(\mathcal{A}, \mathcal{B})$ is defined by a linear combination of order statistics:

$$h_{LTS} = \frac{1}{H} \sum_{i=1}^H d_{\mathcal{B}}(a)_{(i)} \quad (4.3)$$

where H denotes $h \times N_{\mathcal{A}}$ ($0 \leq h \leq 1$) as in the partial *HD* case, and $d_{\mathcal{B}}(x)_{(i)}$ represents the i^{th} distance value in the sorted sequence $d_{\mathcal{B}}(x)_{(1)} \leq d_{\mathcal{B}}(x)_{(2)} \leq \dots \leq d_{\mathcal{B}}(x)_{(N_{\mathcal{A}})}$. The measure $h_{LTS}(\mathcal{A}, \mathcal{B})$ is calculated by eliminating the large distance values and only keeping the h fraction of the minimum distances. In our experiments, the value of h that resulted in the best recognition rate was 0.8.

In our case, the calculation of LTS-HD is between the two point sets of two 3-D binary images, one is the ridge image of the test face image and the other is the ridge image of a gallery face image. The process of finding the best pose between a probe ridge image and a gallery ridge image can be formulated as follows:

$$\arg \min_{\alpha, \beta, \gamma, t_x, t_y, t_z, s} h_{LTS}(Tr(\mathcal{A}), \mathcal{B}) \quad (4.4)$$

- (1) Set $\hat{h}_{LTS} := +\infty$, and $t := 0$
- (2) Initially align the 3-D ridge image of the test image P , i.e. translate, rotate and scale, to the gallery image P' , by using the three labeled feature points and the auxiliary point. This similarity transformation is calculated by procrustes analysis [68].
- (3) Set $Success := 0$
- (4) Place the aligned probe ridge image, $T(P)$, over the gallery ridge image. For all the points in the aligned probe image, find the distance to the closest point in the gallery image, P' , using:

$$D_{P'}(x) = \min_{y \in P'} \|x - y\| \quad (4.5)$$

where the $\|\cdot\|$ denotes the $L2$ norm.

- (5) Sort the minimum calculated distances and then calculate the robust Hausdorff distance, h_{LTS} , using Eq. 4.3.

- (6) If ($h_{LTS} < \hat{h}_{LTS}$), set the following items:

$$\hat{h}_{LTS} := h_{LTS}$$

$$t := t + 1$$

$$Success := 1$$

- (7) Change the parameters i.e., translation, rotation, and scale, of the similarity transformation based on the optimization technique. For example, direct search in the simplex method.

- (8) If $Success = 1$ AND ($t < Max_Iterations$) goto 3.

- (9) Return h_{LTS} .

Table 4.1: Iterative algorithm to find the optimum pose in Hausdorff distance matching.

where $Tr = \begin{bmatrix} sR & T \\ 0^t & 1 \end{bmatrix}$ is a 3-D similarity transformation, s is a scale factor,

$T = [t_x \ t_y \ t_z]'$ is the 3-D translation, and R is a 3-D rotation matrix with α, β , and γ

as roll, pith, and yaw rotation angels.

The process of finding the optimum pose between a probe ridge image and a gallery ridge image is achieved by an iterative approach as shown in table 4.1. We used the Matlab optimization toolbox, i.e., *fminsearch* Matlab function, to solve this problem. The *fminsearch* uses the simplex search method of [89]. This is a direct

search method that does not use numerical or analytic gradients. This procedure is repeated to find the matching distance between a probe image and all the images in the gallery. The gallery face image that results in the minimum matching distance, is considered the best match.

4.2.2 Ridge Matching Using Iterative Closest Points

The ICP algorithm is widely used for geometric alignment of 3-D models when an initial estimate of the relative pose is known. Many variants of ICP have been proposed, where the differences are in the phases of selecting, matching the feature points, and/or the minimization strategy. In this work, we use a fast ICP variant [143]. Instead of using random sampling of the feature points as in [143], we use all the feature points in the 3-D ridge image in the matching process. Although *random sampling of the points* speeds up the matching process, it has an adverse effect on the accuracy of the final results. The details of the ICP algorithm is presented in Appendix B. For the initial alignment of a probe 3-D binary ridge image to the 3-D ridge images in gallery, the three labeled feature points, i.e., the two inner corners of the eyes and the tip of the nose and an auxiliary point are utilized. Procrustes analysis [68] is then used to estimate the parameters of the similarity transformation (scale, rotation, and translation.) After the initial alignment, we use the aforementioned ICP algorithm to finely align a 3-D ridge probe image with a given 3-D ridge gallery image and compute the Mean Square Error (MSE) between the points. The smaller the MSE the closer the probe image to the gallery image.

4.2.3 Computational Complexity of Ridge Matching

Compared to other 3-D matching approaches for face recognition such as [108, 145, 37], i.e., the second category in our classification, our approach is faster and requires less computations. This reduction in computations is due to the fact that we only deal with the ridge lines around the important facial regions on the face, i.e., the eyes, the nose, and the mouth and ignore the surface patches on the face during the matching process. In other words, instead of matching the entire surface of two faces (a probe image and a gallery image), we only match the ridge lines on the face that are detected based on the principal curvature. In this work, the number of the points in the ridge images that represent the lines around the facial regions are $14\% \pm 2\%$ of the total number of points that cover the face. The computational complexity for both the Hausdorff distance ICP in finding the closest point is $O(PQ)$ with Euclidean distance calculations as the elementary operations, where P and Q are the number of the points in the probe and the gallery, respectively. Also, by employing the K-D tree for searching the closest points, the computational complexity is reduced to $O(Q \log(P))$. By using the ridge lines, only a fraction of the points on the surface of the face are used for face recognition. If we assume that the ridge points are only a fraction, i.e., $\rho < 1$, of the entire points on face surface (e.g., $14\% \pm 2\%$ in our work), then the computational complexity reduces to $\rho^2 O(PQ)$ for the regular scheme and to $\rho O(Q \log(\rho P))$ for the accelerated scheme.

4.3 Experiments and Results

We use the Gavab database [115] and the FRGC V2.0 [133] 3-D face databases for our experiments. In the following subsections we review these two databases and present our experiments and results.

4.3.1 Experiments on Gavab Database

As described in Section 3.4, the Gavab database contains range images of 61 subjects. In our experiments, we used the two neutral frontal images (the 1st and the 2nd captures), the two neutral looking up and down images (the 5th and the 6th captures), the frontal images with smile expression (the 7th capture), the frontal images with laughing expression (the 8th capture), and a frontal image with random gesture (the 9th capture). The images in the 2nd capture are used as gallery images and the images in the 1st, 3rd, 4th, and 7 – 9th captures are used as the probe images for recognition.

Figure 4.3 shows three samples of the original images in the Gavab database, results of noise removal and interpolation, face localization, feature points labeling, and ridge images.

For recognition, we compared between the robust Hausdorff distance and the ICP techniques. Table 4.2 presents the results of the experiments. For neutral frontal images, only four subjects out of 61 subjects were not identified by both algorithms (93.5% rank-one identification rate). In another experiment, we projected the frontal ridge images to 2-D (ignoring the 3rd dimension) and the recognition process was

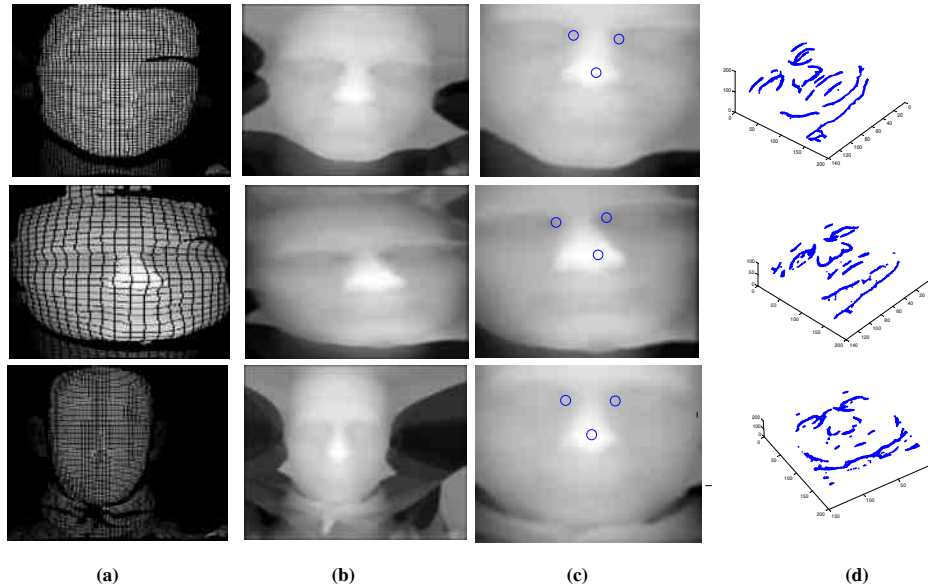


Figure 4.3: Samples of range images in the gallery and the results of preprocessing (a) Original range images, (b) Noise removal and interpolation, (c) Face localization and three feature points labeling, (d) Ridge images

tested. By ignoring the 3^{rd} dimension, we obtained rank-one identification rate of 82.0% and 86.9% using robust Hausdorff distance and ICP, respectively. This result supports the opinion that 3-D data has more potential for face recognition than 2-D data.

For faces with expressions, we considered only the upper part of the face, i.e., the 3-D ridge lines around the eyes and the nose, for recognition and excluded the lower part of the face, i.e., the mouth, which is affected by the expression. We achieved a recognition rate of 83.6% using the ICP technique and 82.0% using the robust Hausdorff distance for the smiling expression.

Furthermore, we evaluated the performance of our approach for recognition of facial images with pose (looking up/down) based on both the ICP and the robust Hausdorff distance techniques. The recognition rates for the facial images with the

Facial Expression	1 st Rank Recognition (%)	
	Robust H.D.	ICP
Neutral (3-D)	93.5	95.0
Smiling (3-D)	82.0	83.6
Laughing (3-D)	73.8	68.9
Random Gesture (3-D)	63.4	63.4
Looking Up (3-D)	75.4	88.6
Looking Down (3-D)	70.5	85.3

Table 4.2: Results of first-ranked recognition rate on the Gavab face database using range data.

looking up (down) pose are 88.6% (85.3%) and 75.4% (70.5%) using the ICP technique and the robust Hausdorff distance, respectively. Figure 4.4 and 4.5 show the performance of the system in term of Cumulative Match Characteristic (CMC) curve and Receiver Operating Characteristic (ROC), respectively, for the neutral versus neutral frontal facial images using the ICP and the robust Hausdorff distance matching techniques. Our experiments show that the performance of the ICP is better than the robust Hausdorff distance for matching (except for the laughing expression.)

We compared our algorithm with three different approaches for 3-D face recognition that were presented by Moreno *et al.* in [118, 117, 116] based on the Gavab dataset. In [118], they segmented the range images into isolated subregions using the mean and the Gaussian curvatures. Then, they extracted 86 descriptors such as the areas, the distances, the angles, and the average curvatures of the subregions. They selected 35 best features and utilized them for face recognition based on the minimum Euclidean distance classifier. They achieved a first ranked recognition rate of 78.0% for neutral frontal images and 62% for images with smile expression (only 60 subjects out of 61 from the database were utilized). In [117], they selected a set of

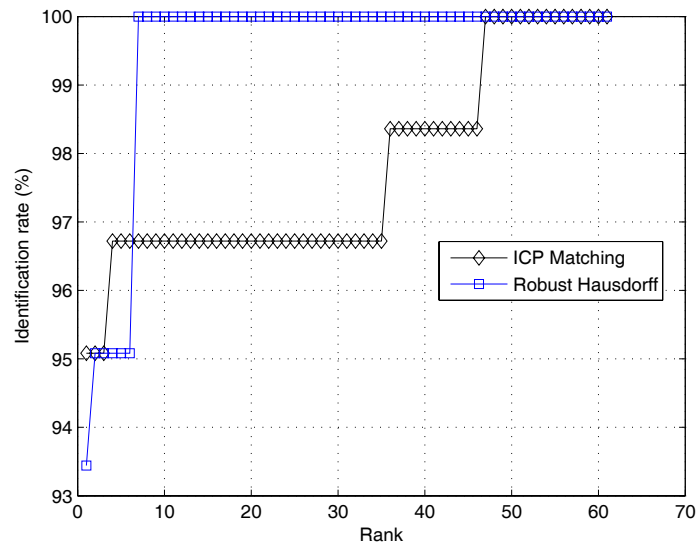


Figure 4.4: CMC curves for frontal neutral images in Gavab database based on ridge images matched using the ICP and Hausdorff distance techniques.

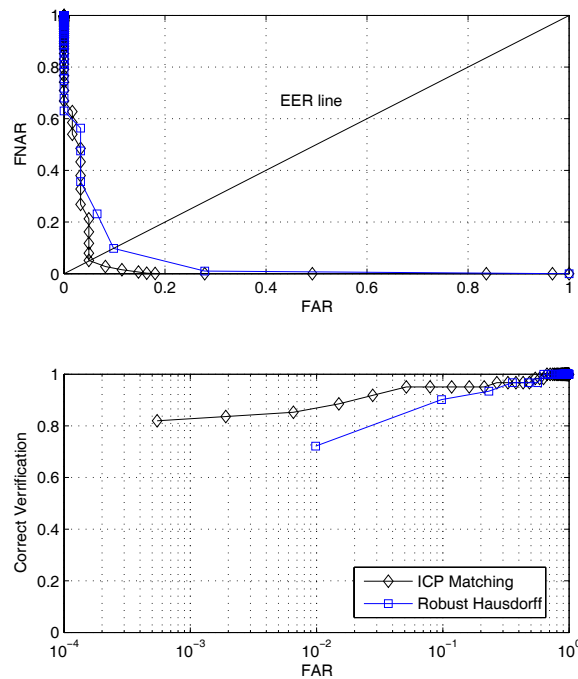


Figure 4.5: ROC curves for frontal neutral images in Gavab database based on ridge images matched using the ICP and Hausdorff distance techniques.

30 features out the 86 features and obtained recognition rates of 82.0% and 90.16%, when the images are frontal views with neutral expression using Principal Component Analysis (PCA) and Support Vector Machines (SVM), respectively. The recognition rates decreased to 76.2% and 77.9%, using PCA and SVM matching schemes, respectively, when using probe images with expressions and slight face rotation. In [116], the authors represented the face using 3-D voxels. Experiments were performed on both images with neutral expression and images with either pose variations or facial expressions. The best recognition rate that they achieved was 90.16% for the images with neutral expression and 77.9% for the images with pose and facial expressions. In addition Ansari [11] applied a 3-D mesh modeling technique to represent the data in Gavab database and used Euclidean distance and voting for comparing 3-D meshes. As a result he achieved 90.16% Table 4.3 summaries their results as well as ours. As the results show, our method based on ridge images and the ICP technique for matching has a better recognition performance for images with neutral expression, with expressions, and with pose variations.

4.3.2 Experiments on FRGC V2.0 Face Database

The FRGC V2.0 database [133] consists of 50,000 recordings divided into training and validation partitions. The training partition is designed for training algorithms and the validation partition is for assessing the performance of a system in a laboratory setting. FRGC V2.0 consists of six experiments, where the third experiment measures the performance of 3-D face recognition. In the third experiment, the gallery

Approach	Neutral-Neutral%	Neutral-Pose/Expr. %	Technique
Moreno <i>et al.</i> [118]	78.0	62.0	35 features (e.g., areas of subregions)
Moreno <i>et al.</i> [117]	82.0 90.16	76.2 77.9	30 features and PCA 30 features and SVM
Moreno <i>et al.</i> [116]	90.16	77.9	3-D voxels
Ansari [11]	90.16	-	3-D Mesh
Our results	93.5 95.0	75.4/82.0 88.6/83.6	Robust Hausdorff distance Iterative Closest Points

Table 4.3: Comparing the results by Moreno *et al.* [118, 117, 116] and Ansari [11] on Gavab database and our work.

and probe data sets consist of both range and texture images for each subject. The 3-D images were acquired by a Minolta Vivid 900/910 series sensor. There are 4007 pairs of images (range and texture) for 466 subjects in the validation set. The set contains images from 1 to 22 sessions per subject, including images with neutral expression and images with other expressions. 370 subjects have at least two images with neutral expression and 432 subjects have at least one neutral image. Figure 4.6 shows few samples of range images along with the range data for different subjects in the FRGC V2.0 database.

We investigated the performance of our algorithm on the neutral 3-D face images of the FRGC V2.0 database. First, we compared the performance of the robust Hausdorff distance and the ICP technique for matching the ridge images. There are 370 subjects that have at least two neutral images captured in different sessions. For some of the subjects, there are more than two captured neutral images with a time laps of one week between them. We chose the two farthest captured images for each

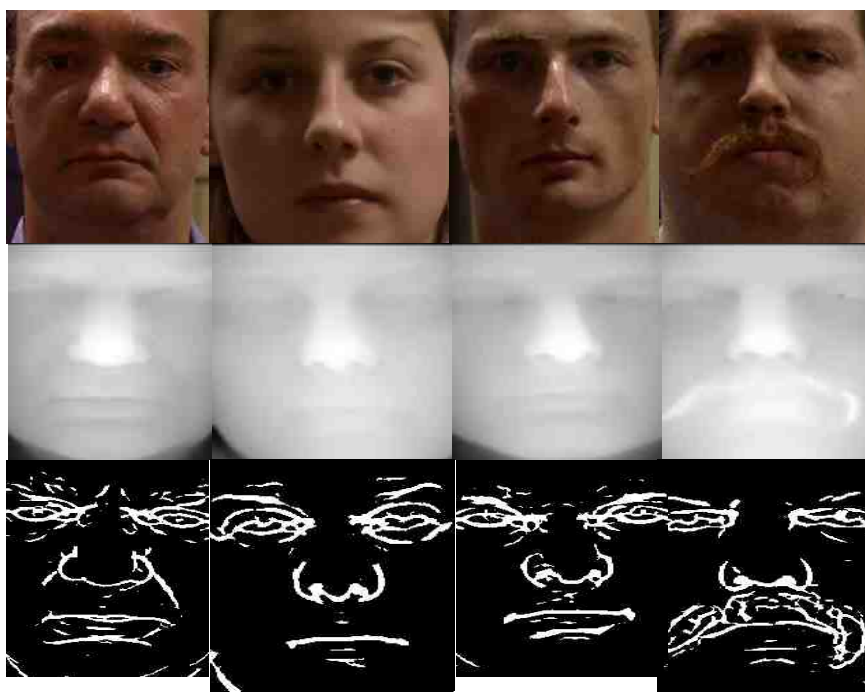


Figure 4.6: Samples of facial images in the FRGC V2.0 database (texture, range, and extracted ridge images.)

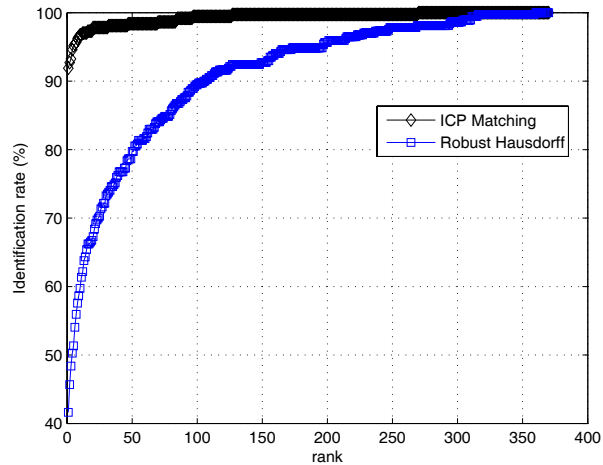


Figure 4.7: CMC curve for 370 subjects in the FRGC V2.0 database based on ridge images matched with the ICP and Hausdorff techniques.

subject and considered the oldest one as the gallery and the most recent captured as the probe. The result of rank-one identification using the robust Hausdorff distance on this selected dataset is 41.62% while the result of the ICP technique for matching is 91.8%. This means that the ICP based matching approach not only gives the best performance, but also it is robust with the increase in the size of the database. To remind the readers, for a small size database such as Gavab, the performance of the Hausdorff matching and the ICP matching were comparable (ICP was slightly better.) This conclusion made by Yan and Bowyer in [169], where they compared ICP and Hausdorff for ear surface matching: The ICP outperforms the Hausdorff distance for shape matching.

Database	Ridge Points	Random Points	Complete Surface
Gavab (61 subjects)	95.0	67.2	95.0
FRGC V2.0 (370 subjects)	91.8	10.0	93.7

Table 4.4: Comparison between the ridge points, random points selection, and entire surface based on the ICP matching technique; Results are in term of rank-one identification rate (%).

4.3.3 Recognition Using the Ridge Points, Random set of Points, and Entire Face Surface

In order to justify that the ridge points are the important points on the face for 3-D face matching, we experimented with 3-D face recognition using a random selection of points from the entire surface of the face and used the ICP method for matching. Here the x and y coordinates are randomly selected from a uniform distribution and the z value of each point is taken from the existing depth values of the probe and the gallery image. The number of the random points is equal to the average number of the ridge points in the ridge image. Then, these random points are used for matching based on the ICP technique. In addition we have used all the points on the face surface for matching based on the ICP technique. The results of these experiments are summarized in table 4.4 for the data GavabDB and FRGC V2.0.

The result of our experiment shows that the ridge points are robust for 3-D face recognition versus random point selection for matching. Specially, when the size of the database is large enough (370 subjects in FRGC V2.0), the performance of matching based on random points selection is very low (10% for the FRGC V2.0 dataset.) Compared to the matching scheme using all the surface points, as long as the size of the database is small, the ridge points and the entire points on the surface

of the face have the same performance. With the increase in the size of the database, the recognition rate does not drop significantly (91.8% compared to 93.7% on the FRGC2.0 dataset.) In conclusion, there is a tradeoff between the performance and computational complexity in shape matching, our experiments show that ridge points are very promising in surface matching. More precisely, the use of ridge points result in negligible performance deterioration while reducing the computational complexity of matching two orders of magnitude.

4.3.4 More Results on FRGC V2.0

In another experiment, we evaluated the capability of the ridge images for face verification on FRGC V2.0 face database. Since ICP has a better performance than robust HD technique for matching, only the ICP technique is used in the rest of our experiments. Figure 4.8 shows the result of the verification experiment for the neutral facial images (total of 2365 facial images for 432 individual subjects). The results are presented using an ROC curve. As the ROC curve shows the performance of 3-D face recognition based on ridge images and the ICP technique for matching is 88.5% verification at 0.1% False Acceptance Rate (FAR). Table 4.5 breaks down the results of the 3-D for verification at 0.1% FAR in terms of three different ROCs. In ROC I all the data are within the semesters (Fall 2003 and Spring 2004), in ROC II the data are within the year, and in ROC III the images are between the semesters (see Figure 4.9). This means that the experiment that produced ROC III is the toughest experiment in term of time laps between the images. The table represents the results

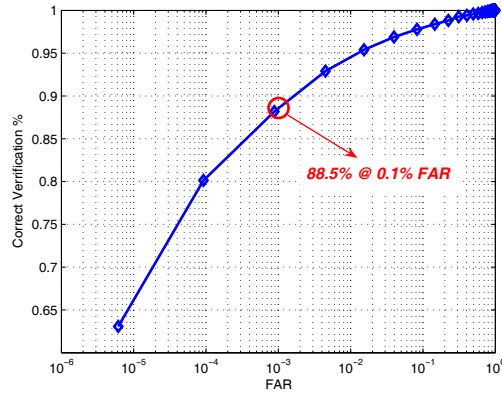


Figure 4.8: ROC curve for the neutral faces versus neutral in FRGC2.0 database (total of 2365 facial images for 432 individual subjects); ICP technique is used for matching the ridge images.

Neutral	3-D (%)	FRGC Baseline (%)
ROC I	90.69	90.00
ROC II	88.5	86.01
ROC III	85.75	81.58

Table 4.5: Verification rates (%) at 0.1% FAR for the ROC I, II, and III of the neutral v.s. neutral images.

of verification for neutral v.s. neutral images in FRGC V2.0 dataset (2365 images of 432 subjects). The last column of the table shows the results of the FRGC baseline for the three ROCs. The baseline algorithm for the 3-D scans in FRGC consists of applying PCA on the shape and texture channels separately and then fusing the results. Compared to the FRGC baseline, our approach has a better performance. Furthermore, comparison between our results in the three ROCs, shows that the performance of our system does not drop significantly under the effect of aging and time laps between the capturing sessions. This validates our claim that the ridge lines have great potential for face recognition under the effect of aging.

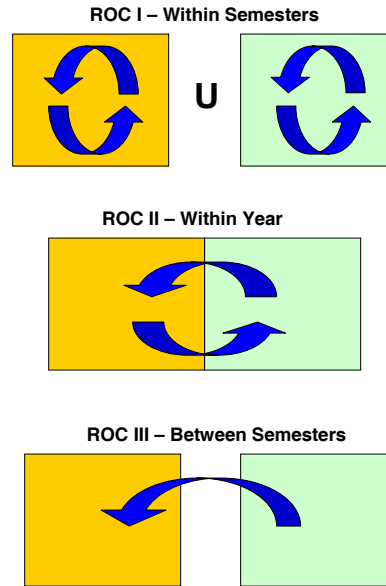


Figure 4.9: ROC I: the data are within the semesters. ROC II: the data are within the year. ROC III: the images are between the semesters.

4.4 Summary

In this chapter, we have presented an approach for 3-D face recognition from frontal range data based on the ridge lines on the face surface. We have used the principal curvature, k_{max} , to represent the face image as a 3-D binary image called ridge image. The ridge image shows the locations of the ridge points around the important facial regions on the face, i.e., the eyes, the nose, and the mouth. We have utilized the robust Hausdorff distance and the Iterative Closest Points (ICP) for matching the ridge image of a given probe image to the ridge images of the facial images in the gallery. To test the performance of our approach for 3-D face recognition, we have performed experiments on GavabDB face database (a small size database) and Face Recognition Grand Challenge V2.0 (a large size database). The results of the experiments have shown that the ridge lines have great capability for 3-D

face recognition. In addition, we have found that as long as the size of the database is small, the performance of the ICP based matching and the robust Hausdorff matching are comparable. But, when the size of the database increases, ICP based matching outperforms the robust Hausdorff matching technique.

Chapter 5

Multi-modal Face Recognition

In biometrics, the classical definition of “multi-modal” refers to the use of more than one modality or multiple sensors in order to increase the accuracy and robustness of the biometric recognition system. In this chapter, our goal is to develop a multi-modal algorithm for face recognition. The two modalities are the 3-D shape and the 2-D texture of the face.

When the 2-D and 3-D face data are registered and thus correspondences can be established between the points in 2-D and 3-D data, we develop a multi-modal scheme for face recognition based on Attributed Relational Graphs (ARG). In this technique, the shape and texture are integrated in one graph model. When the 2-D and 3-D face data are not registered, e.g., due to time lapse between the data acquisition processes such as in case of FRGC data, we model the 2-D and 3-D face data separately; ridge images for shape and ARG model for texture. We then fuse the match results at the score level. In other words, the process of face recognition in each modality is accomplished separately.

The ARG is a geometric graph with nodes and edges, where the nodes represent the facial landmarks on the face and the edges connect the nodes based on the Delau-

may triangulation. At each node of the graph, a set of attributes are extracted from both the shape and the texture based upon log-Gabor filters. Besides, we define a set of mutual relations between the sides of the triangles in the graph. The results of face recognition using the 2-D attributes, 3-D attributes, and the mutual relations are fused together in the score level using various techniques such as the weighted sum rule and the Dempster-Shafer (DS) theory of evidence.

5.1 A Graph Approach For 2-D and 3-D Face Recognition

Figure 5.1 shows the general block diagram of our system for 2-D and 3-D face recognition based on the ARG model. We assume that the shape and texture images for each individual subject in the database are registered. This means that if we extract the facial landmarks from 2-D images, they match also to the facial landmarks on the 3-D data. The ARG model represents both the 2-D and the 3-D information within a geometric graph model.

Based on the availability of data from each modality, i.e., 2-D and 3-D, the recognition based on only one modality or fusing data from the two modalities can be handled easily. More importantly, the proposed 3-D ARG model provide a unified framework to approach various face recognition such as 2-D versus 3-D, 3-D versus 3-D, and multi-modal. Furthermore, this multi-modal approach has the potential of minimizing the problems of pose variations. Specifically, the pose variation problems can be handled effectively by 3-D information.

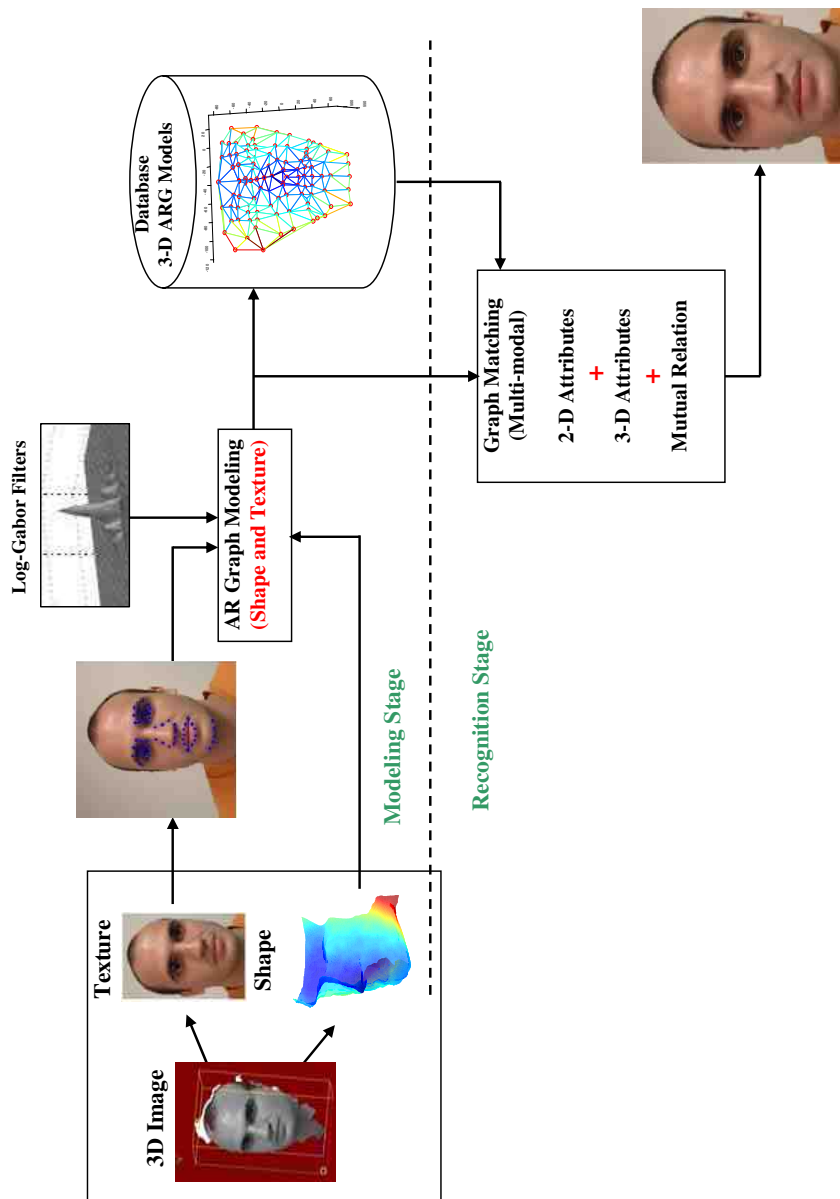


Figure 5.1: General block diagram of our system for face modeling and recognition based on Attributed Relational Graph.

5.2 Graph Modeling

Elastic Bunch Graph Matching (EBGM) has shown to be successful in representing complex objects [163, 29]. Wiskott *et al.* [163] modeled 2-D facial images, with pose variations and intensity changes, for recognition using the EBGM technique. They represented a facial image by a labeled graph called bunch graph; edges are labeled with distance information and nodes are labeled with Gabor filter responses, called jets. In addition, bunch graphs are treated as combinatorial entities in which, for each fiducial point, a set of jets from different sample faces are combined, thus creating a highly adaptable model. This model is matched to new facial images to reliably find the fiducial points in the image.

There are pros and cons with the EBGM technique. This graph representation has shown to be successful in face recognition with good performance rate. But, it is computationally expensive in extracting the facial features of a new given face image. The edge representation of the graph in EBGM (i.e., which is simply the distance between the nodes) is considered only in face detection process. Furthermore, the constructed graph is ad-hoc and requires manual intervention to establish correspondences between the nodes.

In mathematics, a geometric graph is one where the vertices or edges are associated with geometric objects or configurations [17, 136]. A geometric graph, where the vertices are embedded as points in the Euclidean plane and the edges are embedded as non-crossing line segments, is called a planar straight line graph. A triangulation is a planar straight line graph to which no more edges can be added. A special case

is the Delaunay triangulation, a graph defined from a set of points, P , in the plane by defining triangles such that no point in P is inside the circumscribed circle of any triangle. The Delaunay triangulation has some useful properties:

- It maximizes the minimum angle of all the angles of the triangles in the triangulation.
- The triangulation process is invariant under similarity transformation.
- Local perturbing in the node positions does not affect the entire configuration of the graph (This property is useful when representing non-rigid objects like the face by a graph model.)

In particular, the first property guarantees that the triangles are fat triangles with maximum area. The second property guarantees pose variation would not change the constructed graph model. Compared to the EBGM, the ARG model has some advantages. First, the edges of the graph are defined based on the Delaunay triangulation. Second, a set of features are extracted and represented using the edge information. Third, unlike the EBGM, we use an enhanced version of Active Shape Model technique to find the facial landmarks in the face, which are used as the nodes of the graph. In the following sections, we present in detail our algorithm for face recognition based on the ARG model.

5.2.1 Building the Attributed Graph

The ARG plays an important role in characterizing the content of an object [129]. It consists of a set of nodes, edges, and mutual relations between them. Let us denote

the ARG by $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$, where $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ is the set of nodes of the graph and $\mathcal{E} = \{e_1, e_2, \dots, e_N\}$ is the set of edges. In this dissertation, nodes of the graph represent the extracted facial features on the face, and \mathcal{R} is a set of mutual relations between the three edges of each triangle in the Delaunay triangulation. Mathematically, we write $\mathcal{R} = \{r_{ijk} | e_i, e_j, e_k \in \mathcal{D}_t\}$, where \mathcal{D}_t is a set of triangles in Delaunay triangulation. Recall that a Delaunay triangulation, $D_t(p)$, for a set P of points is one where no point in P is inside the circumcircle of any triangle in $D_t(P)$.

Geometric graphs that are created by Delaunay triangulation are invariant to scale, in-plane rotation, and translation. In other words, if an object contains a set of vertices and a set of edges are defined between the vertices (i.e., edges of the graph) using Delaunay triangulation, then under the similarity transformation the set of edges that compose the graph will be the same.

Unlike the work presented by Park *et al.* [129], where the nodes of the ARG do not have labels and they rely on stochastic analysis to find the feature correspondences, the nodes of the ARGs have direct one-to-one correspondences in our study.

Shape and Texture Attributes

Gabor filters represent a popular choice for obtaining localized frequency information. Gabor filters have similar shapes as the receptive fields of simple cells in the visual cortex of vertebrate animals [137, 77, 46] and can be statistically derived from the images of natural scenes, at least qualitatively [123] [20].

Gabor filters have two main limitations. They are not optimal if one is seeking broad spectral information with maximal spatial localization and the maximum band-

width is limited to approximately one octave and Gabor filters. An alternative to the Gabor function is the Log-Gabor function proposed by Field [52]. Log-Gabor filters can be constructed with arbitrary bandwidth and the bandwidth can be optimized to produce a filter with minimal spatial extent. It is worth noting that a Log-Gabor filter with a three-octave bandwidth has the same spatial width as a one-octave Gabor filter, demonstrating the ability of the filter to capture broad spectral information with a compact spatial filter. In order to cover the frequency spectrum effectively, a range of both scales and orientations must be considered. The overall aim is to provide an even coverage of the frequency components of interest while maintaining a minimum overlap between filters so as to achieve a measure of independence between the extracted coefficients. The Filter is defined by

$$G(w) = e^{-\frac{(\log(w/w_0))^2}{2*(\log(\sigma/w_0))^2}} \quad (5.1)$$

where w_0 is the filter's center frequency and σ is a scaling factor of the bandwidth. To obtain constant shape ratio filters, the term σ/w_0 must be held constant for varying w_0 . For example choosing $\sigma/w_0 = .74$ will result in a filter bandwidth of approximately one octave, .55 corresponds to two octaves, and .41 will produce three octaves. Figure 5.2 shows a set of 1-D log-Gabor filters seen in logarithmic and linear frequency scales.

There are two important characteristics to note. First, log-Gabor functions, by definition, always have no DC component, and secondly, the transfer function of the log-Gabor function has an extended tail at the high frequency end. Field's studies of

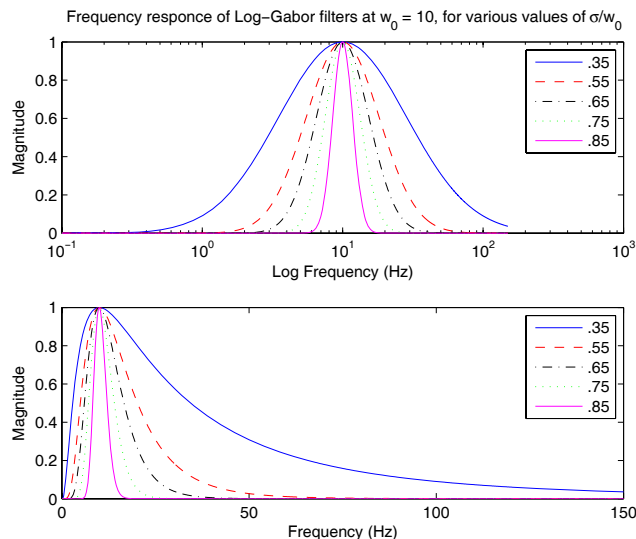


Figure 5.2: A set of 1-D log-Gabor filters seen in logarithmic and linear frequency scales.

the statistics of natural images indicate that natural images have amplitude spectra that fall off at approximately $1/w$. To encode images having such spectral characteristics one should use filters with similar spectra. Field suggests that log-Gabor functions, having extended tails, should be able to encode natural images more efficiently than, say, ordinary Gabor functions, which would over-represent the low frequency components and under-represent the high frequency components in any encoding. Another point in support of the log-Gabor function is that it is consistent with the measurements on mammalian visual systems where the cell responses are symmetric on the log frequency scale.

Unfortunately, due to the singularity of the log function at the origin, one cannot construct an analytic expression for the shape of the log-Gabor function in the spatial domain. Instead, one designs the filters in the frequency domain and then perform a numerical inverse Fourier Transform. Their general appearance is similar to the

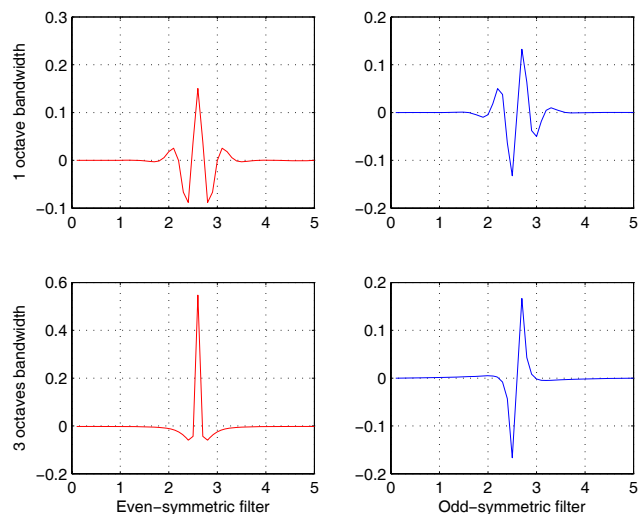


Figure 5.3: Two pairs of log-Gabor wavelets, odd-symmetric and even-symmetric (all tuned to the same frequency, $w_0 = 1$, but having bandwidths of 1 and 3 octaves respectively.)

Gabor functions though their shape becomes much ‘sharper’ as the bandwidth is increased. The shapes of log-Gabor and Gabor functions are almost identical for bandwidths less than one octave. Figure 5.3 shows two log-Gabor wavelets of one and two octaves bandwidths all tuned to the same center frequency.

The 2-D log-Gabor filters are constructed in terms of two components, the radial and angular components control the frequency band and orientation that the filter responds to, respectively. The two components are multiplied together to construct the overall filter. The radial component is defined in Equation 5.1. The angular component controlling the orientation selectivity of the filter is simply a Gaussian with respect to the polar angle around the center frequency. Figure 5.4 shows a sample angular component and Figure 5.5 shows an example of the even-symmetric component and odd-symmetric component of 2-D log-Gabor filters in spatial domain.

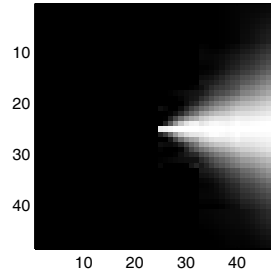


Figure 5.4: Angular component at angle = 0 and sigma = 1.5.

In our experiments, we use log-Gabor filters with 12 orientations and four scales. We set $\sigma/w_0 = 0.41$, which results in filters with three octaves bandwidth.

Mutual Relation

For the relation between the nodes, we define a set of mutual relations between every three nodes that are described by the Delaunay triangulation. We define the mutual relations based on the angle between the sides of the triangle and the ratio of the side lengths. These relations are invariant to translations, rotations, and scale changes [101], and have been shown to be powerful for face recognition by Park *et al.* [129]. Specifically, referring to Figure 5.6, the mutual relations used in this work are defined to be:

$$r_{ijk}(1) = \alpha_1$$

$$r_{ijk}(2) = \alpha_2$$

$$r_{ijk}(3) = \alpha_3$$

$$r_{ijk}(4) = (L_1 + L_2)/(L_1 + L_2 + L_3) \tag{5.2}$$

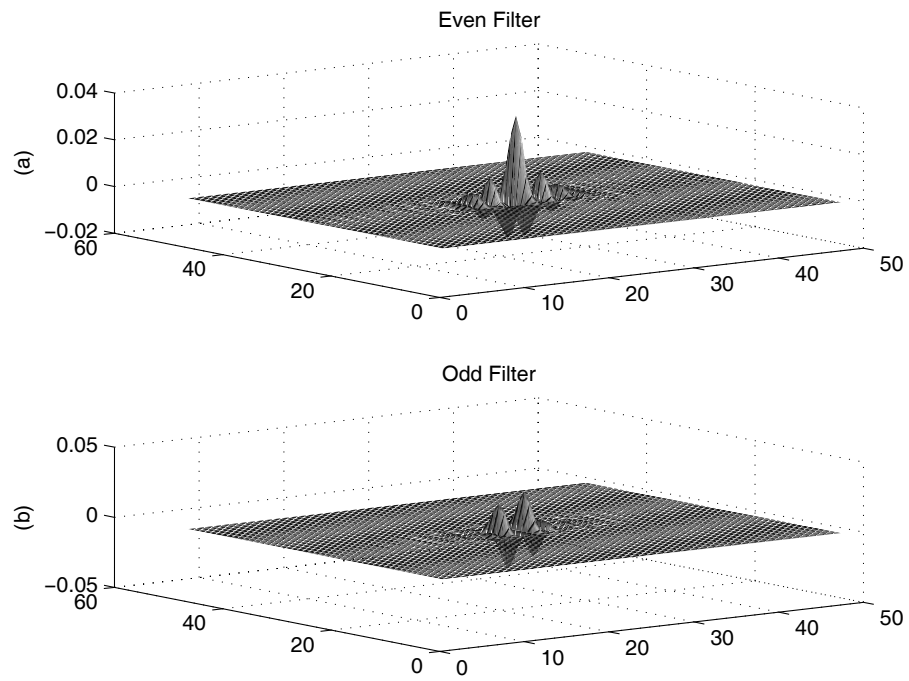


Figure 5.5: Even-symmetric and Odd-Symmetric 2-D log-Gabor filters at wavelength = 1 with 3 octaves bandwidth.

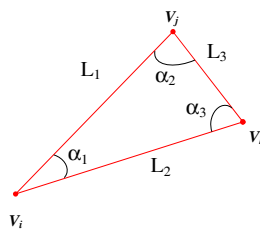


Figure 5.6: The mutual relation between the nodes.

Since the order of the nodes is important in calculating the mutual relations, a predefined and fixed order of the nodes is used during the computation of the mutual relations between the nodes.

5.2.2 Feature Selection

The number of the features (attributes) play an important role in the performance of the graph representation for face modeling and recognition. In this work, we initially extract 57 landmark points, (x, y) values using the improved ASM technique presented in chapter three. We then use a standard template comprising 109 vertices to include more landmark points at certain positions on the face, such as the cheek and the points on the ridge of the nose. Extracting these points using the ASM technique is difficult because of lack of texture in these regions. Since the graph is a 3-D graph, the z values of the locations of the nodes are picked up from the 3-D shape data. Figure 5.7 shows a sample face in the gallery along with the points for building the ARG model. We designed an experiment based on the Sequential Floating Forward Selection (SFFS) [139] that searches through the log-Gabor attributes, 109 vertices \times 4 scales \times 12 orientations = 5232 attributes for each modality, to identifying a set

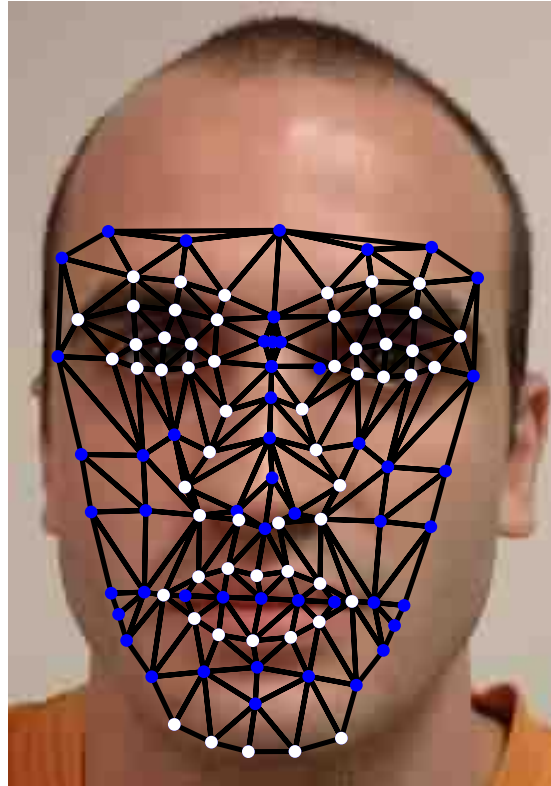
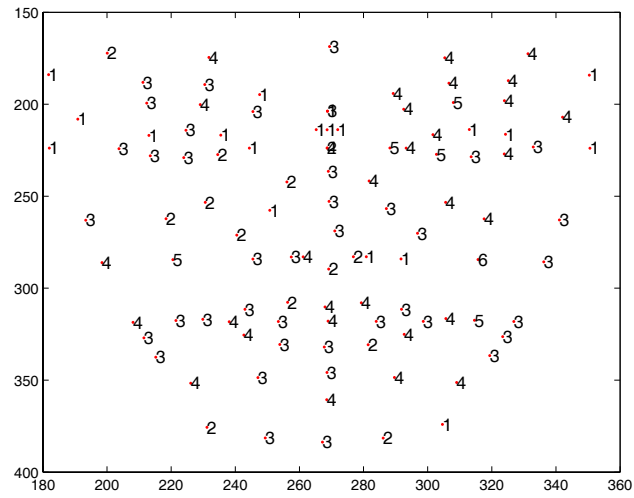


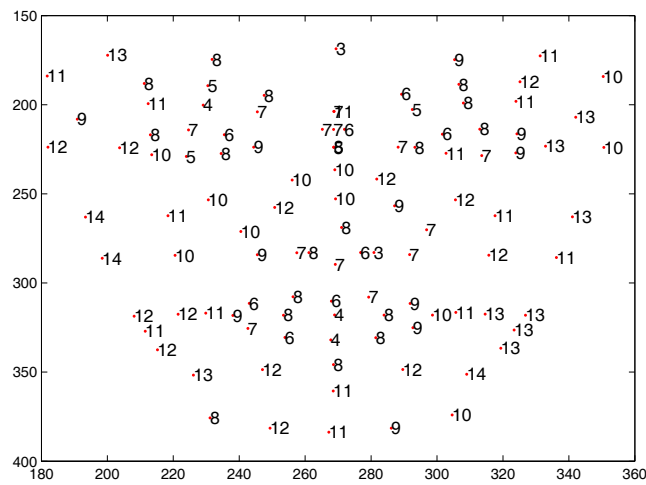
Figure 5.7: Extracted landmark points. 57 landmark points (white color) are extracted by the ASM method and the 52 landmark points (blue color) are added by aligning a standard template to the facial points.

of features which results in the highest performance of the system. The details of the SFFS algorithm are given in Appendix C.

Figure 5.9 shows the results of the feature selection based on SFFS algorithm. Based on this analysis, 316 and 999 Gabor attributes are selected for 2-D and 3-D attributes, respectively. In addition, every node of the graph has at least one feature which is selected for face recognition. Figure 5.8 shows how many 2-D and 3-D attributes are selected from each node of the graph.



(a)



(b)

Figure 5.8: Maps that show number of the selected (a) 2-D and (a) 3-D attributes at each node of the ARG model.

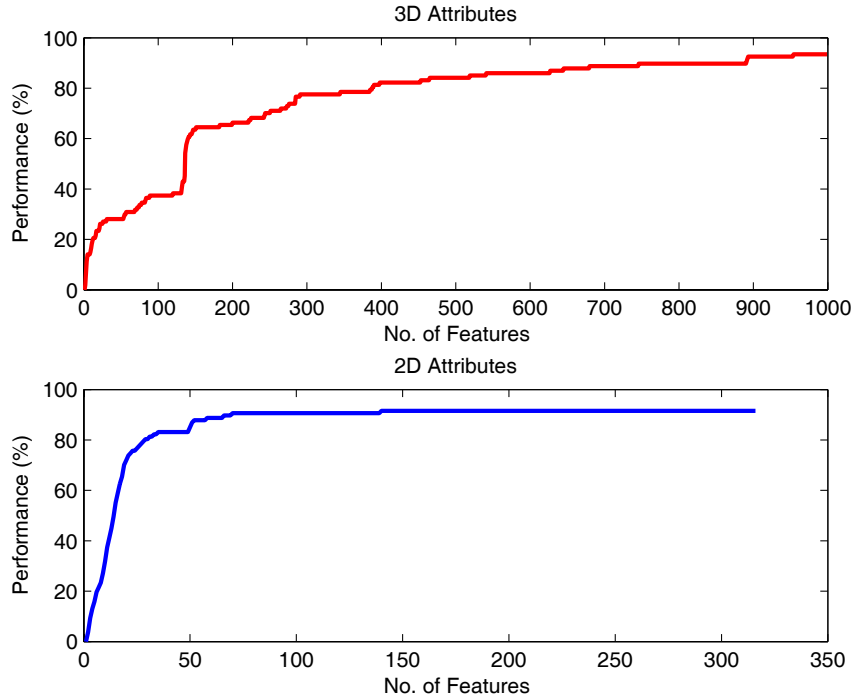


Figure 5.9: Results of feature selection for 3-D and 2-D attributes in the ARG model.

5.3 Recognition; Matching the ARG Models

Assume that the ARG models of two faces, \mathcal{G} and $\hat{\mathcal{G}}$, are given. The similarity between these two ARGs is defined based on the similarity between the assigned 2-D and 3-D attributes, and the similarity (dissimilarity) between the mutual relations.

The similarity between two given set of attributes (either 2-D or 3-D) for the graph \mathcal{G} and $\hat{\mathcal{G}}$ is defined as:

$$S_v(\mathcal{V}, \hat{\mathcal{V}}) = \frac{\sum_{j=1}^N a_j \hat{a}_j}{\sqrt{\sum_{j=1}^N a_j^2 \sum_{j=1}^N \hat{a}_j^2}} \quad (5.3)$$

where a_j is the magnitude of the set of complex coefficients of the log-Gabor attributes, obtained at the j^{th} node of the graph.

Euclidean distance is used to measure the dissimilarity $D_{\mathcal{R}}(\cdot)$ between the two sets of the mutual relation vectors \mathcal{R} and $\hat{\mathcal{R}}$. Then, the dissimilarity is converted to similarity by subtracting it from one:

$$S_{\mathcal{R}} = 1 - D_{\mathcal{R}} \quad (5.4)$$

5.3.1 Pose Normalization

Prior to building the attributed relational graph, we normalize the orientation of each deformed 3-D model in the database to be the same as a reference 3-D face model. Figure 5.10 shows our reference 3-D model with zero degree in pitch, roll, and yaw angles. We use the extracted facial landmarks by the ASM technique to estimate the rigid facial pose. Any change in the pitch, yaw, and roll angles between the captured 3-D model and the reference model can be accurately computed and used to bring the deformed model's pose to that of the reference model. Scale normalization in 3-D is related to the distance of the object from the camera. The farther the object is, the smaller it is in the captured image. We want to bring all the models to comparable distance from the capturing camera, hence maintaining comparable scales in the images. Therefore, we translate the pose normalized 3-D models in depth, such that their depth centers would coincide with the mean depth (obtained from all the deformed 3-D models in the database.) This insures that all 2-D facial images are at comparable scales when projecting from the 3-D model. Figure 5.11 shows the textured 3-D model of a subject in the UM database before and after normalization.

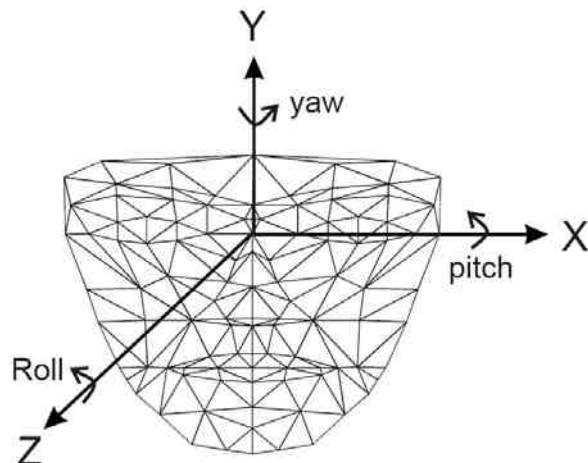


Figure 5.10: Neutral 3-D model used to estimate and normalize the pose of a deformed 3-D model. The reference model is at angles (pitch, yaw, roll) = (0,0,0).

5.4 Fusing the Information: the 2-D and 3-D Attributes and the Mutual Relation

In a multi-modal biometric system, fusion is accomplished by utilizing the information available from different modalities. Figure 5.12 shows various types of fusion that can be used in the context of a biometric system. They can be categorized as 1) fusion prior to matching and 2) fusion after matching [140]. The fusion prior to matching integrates information from multiple sources and can take place either at the sensor level or at the feature level. In the sensor level fusion, the raw data from the sensors are combined while in the feature level fusion, different extracted sets of features from multiple sources are combined.

Schemes for integrating information after the classification/matching stage can be divided into four categories: dynamic classifier selection, fusion at the decision level,

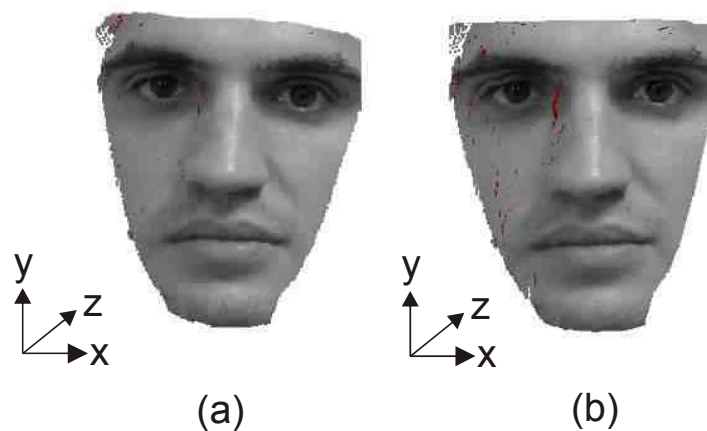


Figure 5.11: 3-D textured model (a) before normalization and (b) after normalization.

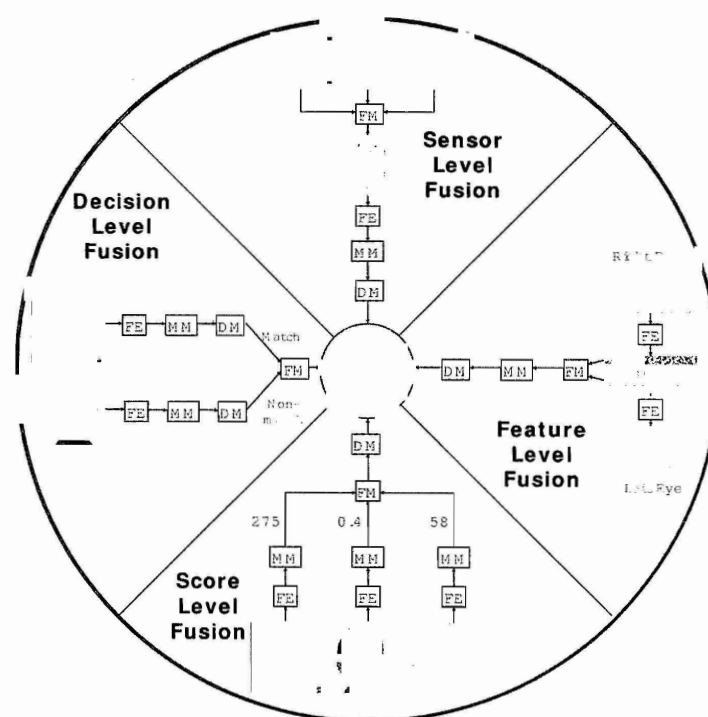


Figure 5.12: Fusion can be accomplished at various levels in a biometric system [140].

fusion at the rank level and fusion at the match score level. A dynamic classifier selection scheme chooses the classifier which is most likely to give the correct decision for the specific input pattern. Integration of information at the abstract or decision level can take place when each biometric system independently makes a decision about the identity of the user (in an identification system) or determines if the claimed identity is true or not (in a verification system). When each biometric system or modality generates a match score indicating the proximity of the input data to a template, integration can be done at the match score level. This is also known as fusion at the measurement level or confidence level. Next to the feature vectors, the matching scores produced by biometric matchers contain the richest information about the input pattern. Also, it is relatively easy to access and combine the scores generated by the different matchers. Consequently, integration of information at the matching score level is the most common approach in multi-biometric systems. When the output of each biometric system is a subset of possible matches (i.e., identities) sorted in decreasing order of confidence, the fusion can be done at the rank level.

5.4.1 Score Level Fusion

In this dissertation, our aim is to combine the two modalities, 2-D and 3-D at the score level. At the score level, we have the flexibility to fuse the match scores from various modalities upon their availability. We use two different approaches for score fusion. The first technique is based on the Dempster-Shafer theory of evidence and the second approach is the weighted sum. Both approaches are in the category

of transform-based techniques (i.e., based on the classification presented in [140].) In practical multi-biometric systems, a common fusion method is to directly combine the match scores provided by different matchers without converting them into posteriori probabilities. However, the combination of the match scores is meaningful only when the scores of the individual matchers are comparable. Hence, a score normalization is applied to transform the match scores obtained from the different matchers into a common domain. This refers to changing the locations and scale parameters of the match score distributions at the outputs of the individual matchers. Afterwards, the normalized scores of different matchers can be fused using different rules such as *Dempster-Shafer theory of evidence*, *weighted sum of scores*, *maximum scores*, and *minimum scores*.

Fusion Techniques

One of the well known fusion techniques used in biometrics is the weighted sum technique:

$$S_f = \sum_{j=1}^R w_j * s_j^n \quad (5.5)$$

where s_j and w_j are the score and weight of the j^{th} modality, respectively, with the condition $\sum_{j=1}^R w_j = 1$. In our case, the weights w_i , $i = 1, 2, 3$ are for the 3-D and 2-D, and mutual relations, respectively.

The weights can be assigned to each matcher by exhaustive search or based on their individual performance [140]. The exhaustive search has shown to be not robust as the second approach where the weights are derived based on individual performance

of each matcher. In this scenario, we assign for each matcher a weight, which is a function of the matcher's performance derived using a set of training data. The weights are calculated as follows:

$$w_i = \frac{1 - (FAR_i + FRR_i)}{3 - \sum_{j=1}^3 (FAR_j + FRR_j)} \quad (5.6)$$

where w_i is the weight for matcher i , FAR_i and FRR_i are the false acceptance and the false rejection rates for matcher i , respectively, and $i = 1, 2, 3$. The values for FAR_i , and FRR_i are threshold dependent. Thus, when the threshold is changed, the weights assigned to the individual matcher will be suitably modified.

Another fusion algorithm that is applied to combine the results of the 2-D and 3-D face recognition is based on Dempster-Shafer theory [150]. A brief overview of DS theory is given in Appendix E.

In a verification problem (in contrast to identification), we have two possible focal elements: 'accept' and 'reject'. Therefore, the match score calculated by the face recognition system in the verification mode after normalization is considered as the mass assignment in DS theory. Let s_1 and s_2 be the mass assignment computed from two face recognition algorithms. Then the final result based on the Dempster rule of combination is obtained as:

$$S_f = \frac{s_1 s_2}{1 + 2 s_1 s_2 - s_1 - s_2} \quad (5.7)$$

The above combination rule can be generalized by iteration to fuse more than two match scores (see Appendix D.)

Although the above technique works for data fusion, it is more realistic to add uncertainty while making the decision. To use a biometric system in the verification mode, usually we have a certain threshold (i.e., the operating point) and if the match score is above that threshold, we accept the proposition otherwise we reject the proposition. However, when the match scores are close to the operating point threshold (e.g., within $\pm 5\%$ from the threshold) the decision might not be accurate. Therefore, for match scores close to the operating point of the system, we impose uncertainty to our decision (see Figure 5.13). In this case, we define two thresholds, T_{low} and T_{high} around the operating point. If the match score is below the T_{low} , then we reject the proposition that the two given faces belong to the same identity. If the match score is above the T_{high} , then when we accept the proposition that the two given faces belong to the same identity. Otherwise, the match score is in the uncertainty margin and we assign a value to the mass $m('accept', 'reject')$ based on the non-linear functions in Figure 5.13. The amount of the uncertainty and hence the new mass assignments in the BoE are calculated as follows.

$$m('accept', 'reject') = a \cdot |s - T_{opt}| + b \quad (5.8)$$

where a and b are the slope and y-intercept of the uncertainty function, respectively in Figure 5.13. After assigning this uncertainty, we normalize the mass assignment in the BoEs to ensure that the condition $\sum_{A \in 2^\Theta} m(A) = 1$ is achieved.

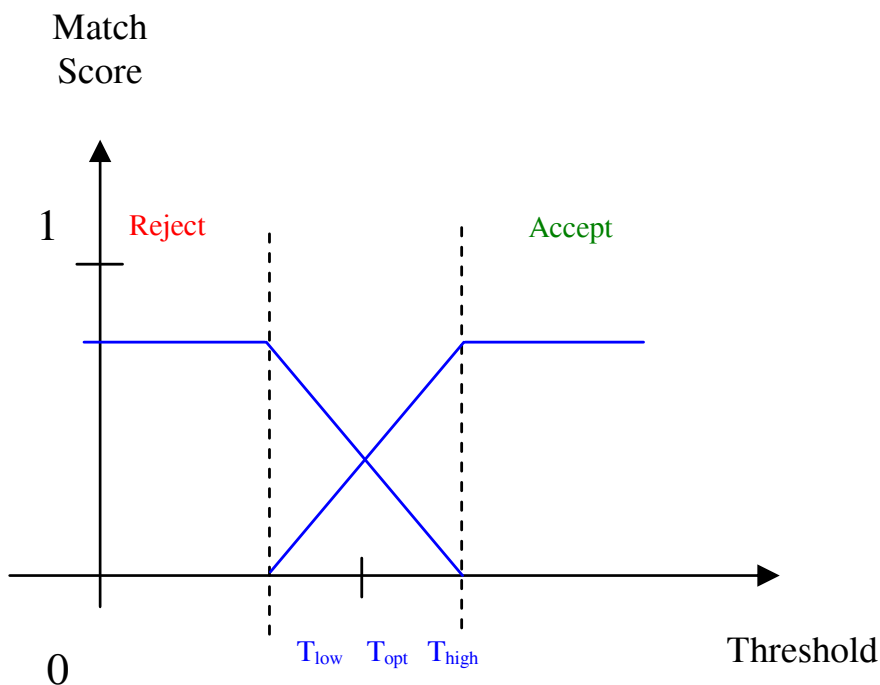


Figure 5.13: Dual threshold scenarios and uncertainty in a verification mode.

5.5 Experiments and Results

In this Section, we present the experimental results of our methods using the University of Miami and the FRGC2.0 face databases.

5.5.1 Experiments on the University of Miami Face Database

The University of Miami (UM) face database consisting of 2-D and dense 3-D facial images of 107 subjects (with neutral expression) were captured during academic years 2005-2007. A stereo-based system was designed and developed for 3-D face modeling (3-D dense reconstruction). At least two sets of images with the 3-D models exist in the UM database. The first set is used as the gallery and the second set is used as

the probe. For the details of this database, see [11]. Figure 5.14 shows few subjects in the UM database including the 2-D images and the 3-D textured models.

Results

We tested the performance of our system (ARG modeling) using the UM database. The results of our experiments are reported in the form of the Cumulative Match Characteristic (CMC) and Receiver Operating Characteristic curves (ROC) for identification and verification, respectively. Figure 5.15 shows the CMC curves for the 2-D and 3-D modalities along with the two fusion techniques based on the weighted sum and DS theory. The weighting factors in the weighted sum technique are 0.35, 0.34, and 0.31 for the 2-D, 3-D, mutual relations, respectively calculated based on Equation 5.6. The FAR and FRR in Equation 5.6 are selected at the Equal Error Rate (EER) operating point. As Figure shows, the performance of the system is boosted significantly by fusing the mutual relations with the 2-D and 3-D attributes. In addition, the rank-one identification based on the DS combination rule and the weighted sum of scores for fusion have the same performance, 99%. However, for the verification experiments, the DS rule of combination outperforms the weighted sum rule (90% verification at 0.1% FAR for the DS rule compared to 81% verification at 0.1% FAR for the weighted sum.) Figure 5.16 compares the results of rank-one identification with those in [11] which is based on 3-D mesh modeling. Our approach has 99% rank-one identification rate, compared to their best performance rate of 96.4%. In addition, we applied our technique for ridge images to the UM face database (see Chapter 4). Figure 5.16 shows the results of rank-one identification.

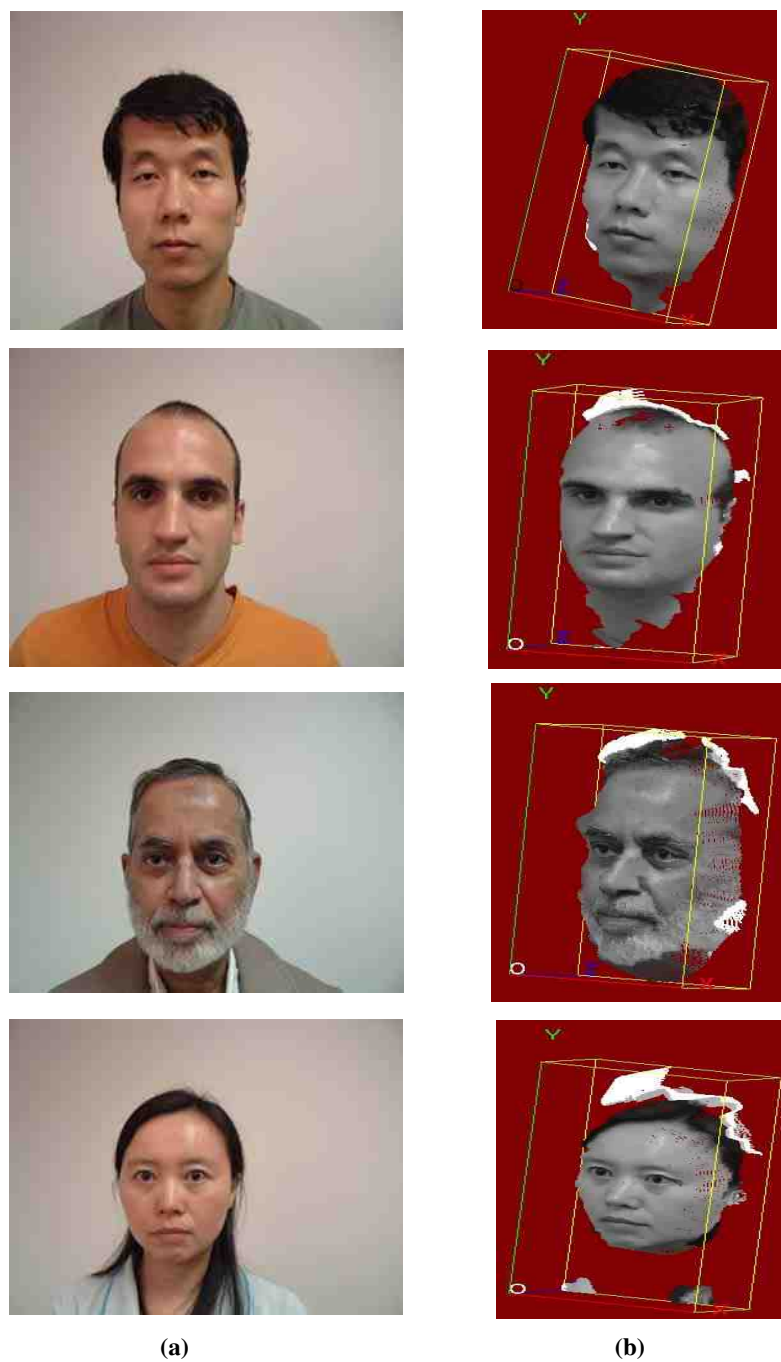


Figure 5.14: Examples of subjects in the UM database: (a) 2-D images and (b) 3-D textured models.

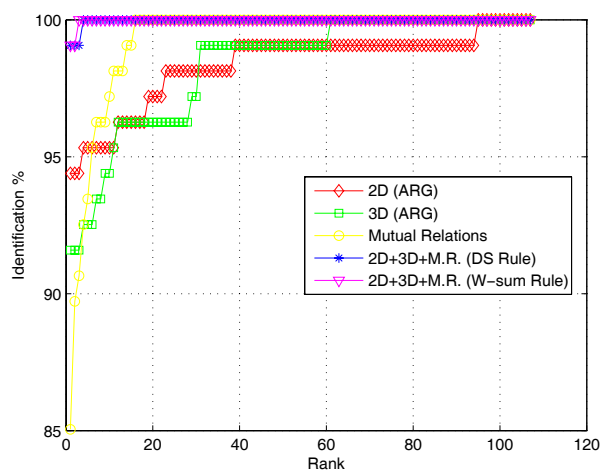


Figure 5.15: Cumulative Match Characteristic curves using the ARG modeling and two different fusion techniques on the UM database.

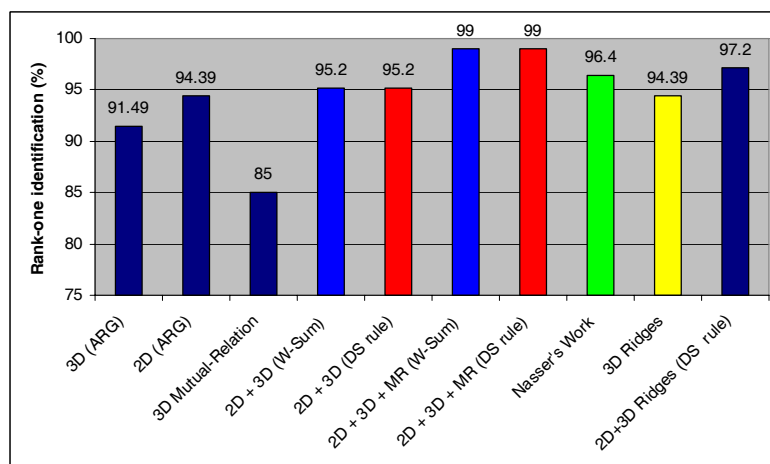


Figure 5.16: Results of rank-one identification using the ARG modeling and two different fusion techniques on the UM database and the results of other matching techniques applied to the UM database.

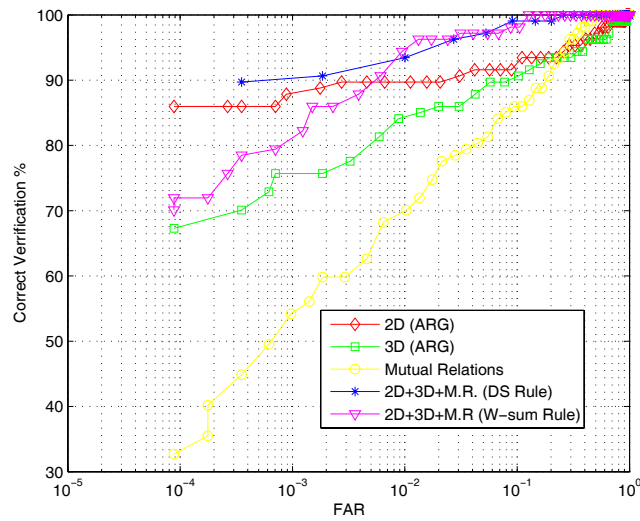


Figure 5.17: Receiver Operating Characteristic curves (ROC) using the ARG modeling and two different fusion techniques on the UM database.

Figure 5.17 shows the results of verification experiments on the UM database. The best performance of our system, 91% at 0.1% False Acceptance Rate (FAR), is achieved by the fusion of different modalities.

In another experiment, we test the improvement in the performance of our system by adding uncertainty as described by Equation 5.8. At the operating point of our system (e.g., the Equal Error Rate point) we define the uncertainty region as 5% above and below this threshold. As a result of adding this uncertainty and applying the Dempster's combination rule to fuse the three match scores obtained from matching the 2-D and 3-D attributes and the mutual relations, the Equal Error rate was improved by 1.5%. More specifically as shown in Figure 5.18, the Equal Error rate was 3.5% without uncertainty and decreased to 2.0% by imposing the uncertainty.

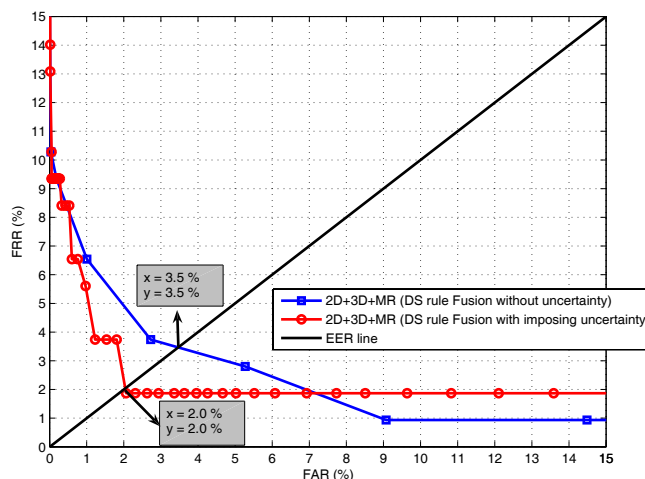


Figure 5.18: Imposing uncertainty and then applying DS rule of combination to fuse the three match scores resulted from ARG modeling (the UM face database), the Equal Error Rate was improved by 1.5%.

5.5.2 Accuracy Analysis of the ARG Model For Face Recognition

We recall that the nodes of the constructed graph in our system rely on the locations of the feature points and their correspondences. Thus any error in the extraction of landmark points, either by the ASM technique, pose estimation or alignment deteriorate the accuracy of the node location/representation, and consequently the system performance. We have designed an experiment to analyze the effect of errors in the location of the nodes on the system performance. We add Gaussian noise, $N(0, \sigma^2)$, to the x and y coordinates of each node of the graph, which are extracted by the improved ASM technique. We then measure the performance of the system for face recognition based on the constructed ARG models from noisy data. The results of rank-one identification reported for various values of σ are in Figure 5.19. As this Figure shows, the 2-D attributes are more robust than the 3-D attributes and the

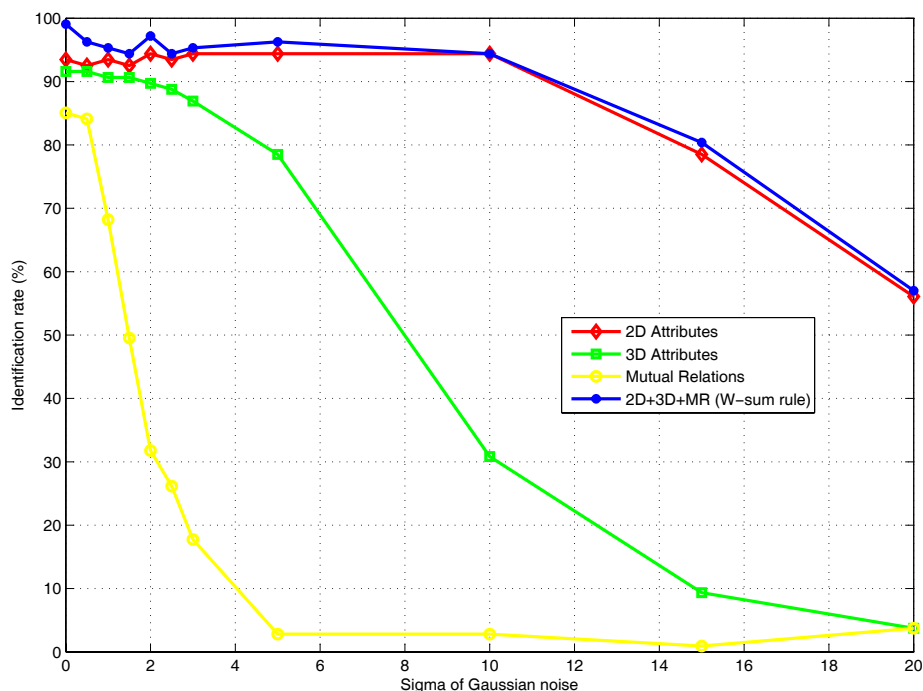


Figure 5.19: Analyzing the robustness of the ARG model in face recognition with respect to added noise in location of the nodes. Results of rank-one identification are represented for 2-D attributes, 3-D attributes, mutual relations, and multi-modal fusion with respect to various values of σ of the Gaussian noise.

mutual relations under the effect of noise and the performance of the face recognition does not drop with even $\sigma = 10$ pixels. For the 3-D features, the performance drops where added noise has a σ greater than three pixels. We also learn from this analysis that the mutual relation features are very sensitive to the noise and the recognition rate drops rapidly with the increase in σ .

5.5.3 Experiments on the FRGC V2.0 Face Database

We next analyze the results for the FRGC V2.0 face database (For a description of the FRGC V2.0 face database, see Section 4.3.2). Because of the time lapse between

acquisitions of the 3-D and 2-D data for each subjects, the 2-D and 3-D images in this database are poorly registered. It is expected that, the extracted locations of the landmark points in the 2-D images do not correspond to the locations of the landmark points in 3-D range images. Therefore, instead of modeling the 3-D (shape) data based on the ARG technique, we use the technique in chapter four for 3-D face matching based on the ridge images. For the 2-D (texture) data, we model the face based on 2-D ARG model (utilizing 2-D attributes and mutual relations.) Then, the results are fused at the score level to obtain a multi-modal system.

As stated, the matching scores for the 3-D ridge images are not in the same domain as the other matching scores, and thus need to be normalized. After score normalization, the matching scores are fused based on the DS combination rule and the weighted sum rule (see Section 5.4.1).

Score Normalization

Two normalization techniques of the matching score, known to have high efficiency and robustness, are used in this work: the *Z-Score* and the *Min-Max* normalization functions [140]. The *z-score* normalization is most commonly used, and is defined as

$$s_j^n = \frac{s_j - \mu_j}{\sigma_j} \quad (5.9)$$

where s_j and s_j^n are the scores before and after normalization, respectively. The arithmetic mean μ_j and the standard deviation σ_j for the j^{th} matcher are estimated using a training set from the UM database.

The *Min-Max* normalization is defined as

$$s_j^n = \frac{s_j - \min(s_j)}{\max(s_j) - \min(s_j)} \quad (5.10)$$

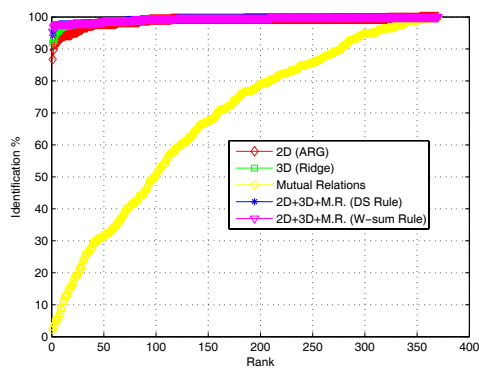
For our system, the best performance is achieved by using the *Min – Max* normalization technique.

Results

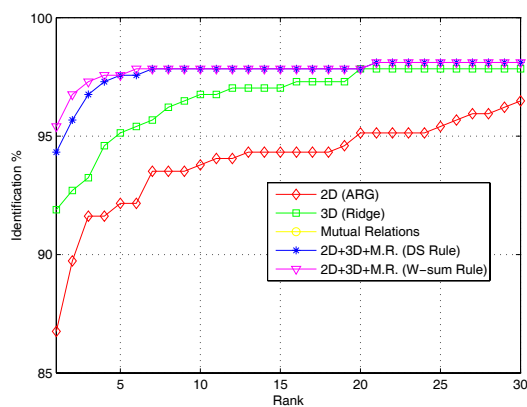
Figure 5.20.a shows the identification performance in term of CMC curves on the FRGC V2.0 database (the details of the CMC curves are given in 5.20.b.) The results are for the 3-D face recognition based on ridges images, 2-D ARG modeling, and fusion using the DS combination rule and weighted sum rule. The results for Rank-one identification are illustrated in Figure 5.21.

Figure 5.22 shows the results of the verification experiment for the neutral facial images versus the neutral facial images. These are presented as Receiver Operating Curves (ROC) for the two modalities along with the fusion of the 2-D and 3-D modalities by the weighted sum rule and the DS combination rule. As the ROC curve shows, the 3-D modality has better performance than the 2-D modality (86.5% v.s. 74.50% verification at 0.1% FAR). In terms of fusion (3-D, 2-D, and Mutual relations), the weighted sum rule has a comparable performance with the DS combination rule (verification rates of 92.5% and 92.05% at 0.1% FAR, respectively.)

The result of face recognition using the mutual relations for the FRGC V2.0 is not impressive because the 2-D (texture) images in the FRGC V2.0 (the dataset in



(a)



(b)

Figure 5.20: CMC curves. (a)for 2-D, 3-D, Mutual relations features, and fusion using DS rule and weighted sum rule for the neutral images versus neutral in FRGC V2.0 data-set; (b) the details in (a).

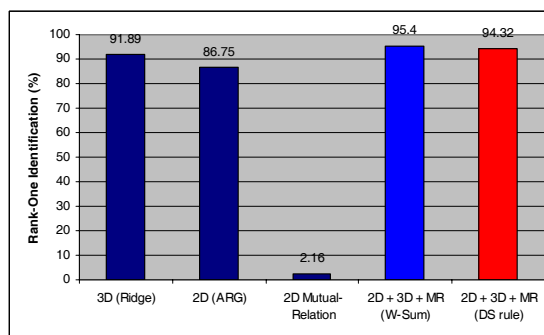


Figure 5.21: Results of rank-one identification using the 2-D ARG modeling, 3-D ridge images, and two different fusion techniques on the FRGC database.

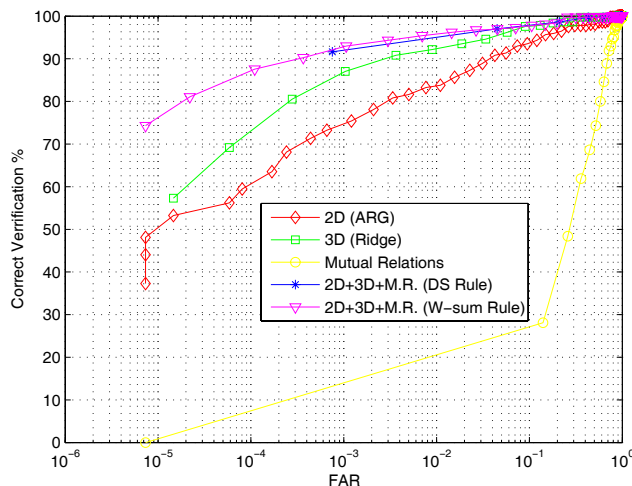


Figure 5.22: ROC curves for 2-D, 3-D, Mutual relations features, and fusion using DS rule and weighted sum rule for the neutral images versus neutral in FRGC V2.0 data-set.

experiment three of the FRGC contest contains both 2-D and 3-D data) have poor quality due to bad lighting condition, facial hair, closed eyes, and the occlusion of the facial regions in these images (Figure 5.23 shows few samples of the 2-D images with poor quality in the FRGC V2.0 database.) Therefore, locations of the extracted facial features from these poor quality images are not accurate. We discussed in Section 5.5.2 that the ARG model and in particular the mutual relations are sensitive to the errors in the facial features locations and thus the performance of face recognition degrades significantly.

5.6 Summary

We have developed a technique called Attributed Relational Graph (ARG) for the representation of the 2-D (texture) and 3-D (shape) modalities of the face. By using

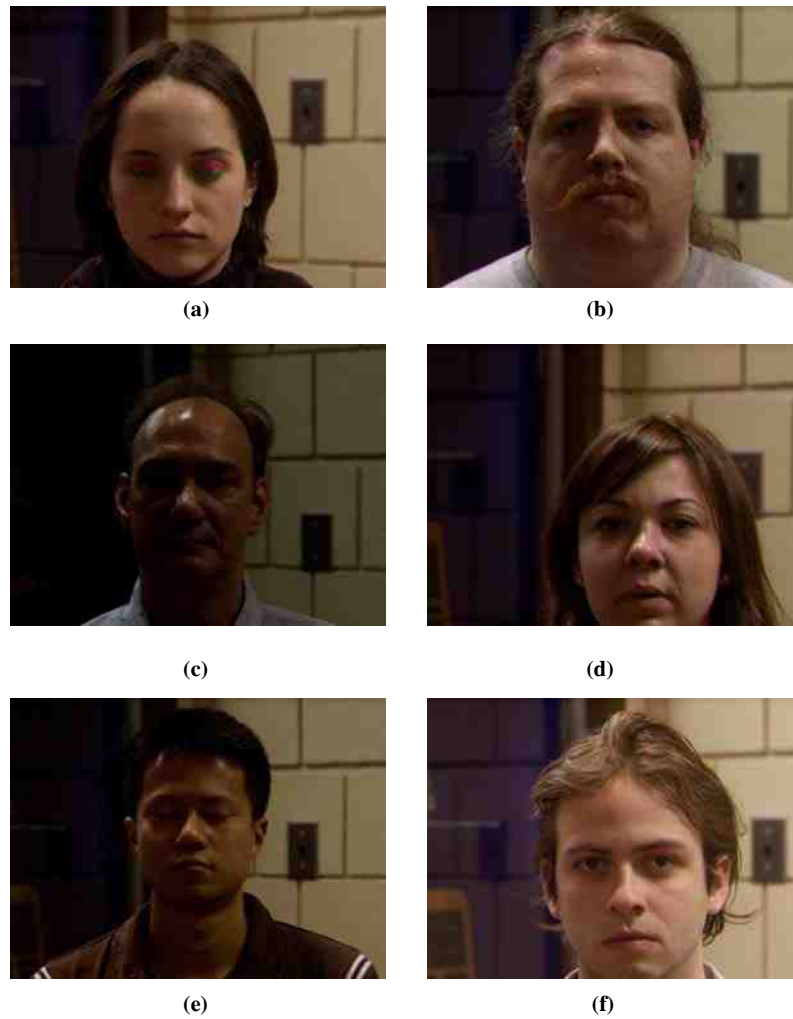


Figure 5.23: Samples of facial images in FRGC V2.0 database show various factors for poor quality of the dataset: (a) closed eyes, (b) facial hair, (c) poor illumination, (d) occluded facial regions (e) closed eyes and poor illumination and occlusion of facial regions, (f) occluded facial regions.

this technique, the shape and texture are integrated in one graph model. Here, a face is modeled by a 3-D geometric graph with nodes and edges, where the nodes of the graph are the locations of the facial landmarks on the face. A set of attributes are extracted at the location of each facial landmark of the face image from both 2-D (texture) and 3-D (shape) images and assigned to the nodes of the graph. These attributes are extracted using a set of log-Gabor filters. In addition to these attributes, a set of geometric features, i.e., mutual relations are extracted based upon the relations between the three sides of the triangles constructed by the Delaunay triangulation. These mutual relations are invariant under similarity transformation. Therefore, each 3-D ARG model of a face contains three sets of features: 2-D and 3-D attributes at each node and mutual relations. In order to have a multi-modal system, we have fused the matching results of the attributes and mutual relations at the score level. We have utilized the DS theory of evidence and weight sum rule for fusion. We have tested the performance of the technique using the University of Miami face database. Our experiments show excellent results for face recognition based on the ARG model. For the case where the 2-D (texture) and 3-D (shape) facial images are not registered, we have modeled the 3-D face data using the ridge images (described in Chapter 4). We have fused the results of the 2-D face recognition based on ARG models and 3-D face recognition based on the ridge images. Our experiments show that the two fusion techniques (weighted sum rule and DS combination rule) utilized in this study have almost the same performance.

Chapter 6

Conclusion and Future Research Direction

6.1 Summary and Contributions

In this dissertation, we have introduced a novel multi-modal scheme for face modeling and recognition from 2-D texture and 3-D shape images. Recently, the scanning or construction of a 3-D dense model of the face has become easy with the advancements in computer vision and image acquisition technologies. One of the main approaches to improve the face recognition technology is to use 3-D data. By using the 3-D data, some of the known problems that affect the performance of 2-d face recognition such as pose variations and illumination variations can be readily handled. On the other hand, 2-D face recognition algorithms have been investigated for few decades and is now in a mature stage. However, the 2-D and the 3-D data have their own limitations and we believe that a multi-modal scheme can provide a very robust solution for the problem of face recognition. In this regard, we have developed a multi-modal system and addressed various issues in building such a system. In particular, we have studied and provided solutions for the following problems: 2-D and 3-D

facial landmark extraction, 3-D facial shape matching, 2-D and 3-D face modeling and recognition, fusion at the score level and experiments on public databases (FRGC V2.0, 3-D Gavab, and UM face databases.)

We have developed novel techniques for 2-D and 3-D face recognition. For 3-D face recognition, we have developed a novel method based on ridge images extracted from the facial range images. For multi-modal face recognition, we have introduced a technique based on Attributed Relation Graph (ARG) that represent both the 3-D (shape) and 2-D (texture) data in a single model. The model is a geometric graph model with nodes and vertices, where the nodes of the graph represent a set of landmark points (extracted by an improved active shape model technique developed in Chapter 3 and Appendix A) and the edges are defined by Delaunay triangulation. A set of attributes are extracted from the shape and texture data using the log-Gabor filters and assigned to each node of the graph. In addition, a set of geometric features are extracted from the edges of the graph (mutual relations). The graph models (i.e., the ARGs) are matched by calculating the similarity between the 2-D and 3-D attributes as well as the mutual relations. The matching scores are fused using two different techniques including the Dempster-Shafer theory of evidence and a weighted sum rule. The ARG model has the capability of integrating both the 2-D and 3-D data in a single model. However, in case where the 2-D and 3-D are not registered, we cannot integrate the 3-D information in the same graph model. Therefore, we have used the algorithm based on ridge images (Chapter 4) to address the 3-D face recognition. We then fuse the results of 2-D face recognition based on ARGs and the results of 3-D face recognition based on ridge images. We summarize the major

contributions of this research as follows:

- We have improved the Active Shape Model approach for 2-D facial features extraction from color images. We have presented solutions to solve some of the limitations of the Active Shape Model approach.
- We have developed a method for 3-D facial features extraction from range data (the inner corners of the eyes and tip of the nose.) Compared to 2-D facial feature extraction, extracting facial features from 3-D range images is more difficult due to the lack of texture. We have used the extracted facial features for the initial alignment of the ridge images during the matching process.
- We have developed a technique for 3-D face modeling and recognition by utilizing the ridge images. The ridge lines on the range image data carry the most important distinguishing information of the 3-D face and have high potential for face recognition. For matching the ridge images of two faces (probe and gallery), the Hausdorff and the Iterative Closest Points methods have been utilized. By using the ridge images for shape matching, the computational complexity of 3-D face matching have been reduced by two orders of magnitude.
- We have developed a technique for multi-modal (3-D and 2-D) face recognition based on the attributed relational graphs (ARG). The nodes of the graph represent the locations of the facial landmark points. The edges of the graph are defined based on Delaunay triangulations. A set of attributes are extracted from the shape and the texture of the face using log-Gabor filters and are assigned to each node of the graph. Also, a set of mutual relations that define

the geometric relations between the edges of the graph are extracted and used in the representation to improve the performance rate of the face recognition. The attributes and the mutual relations are fused at the score level.

- We have developed a fusion technique based on the Dempster-Shafer theory of evidence for fusion at the score level.
- We have evaluated the performance of our developed algorithms and techniques for multi-modal face recognition using various databases such the FRGC v2.0 face database, the 3-D Gavab face database, and the University of Miami (UM) face database. As we have shown through the dissertation, we have achieved promising results using the developed techniques.

6.2 Future Work

Face recognition is an important area of research that is continuously progressing. Possible future developments, expansions, and improvements to our contributions are as follows:

- Improving the extracting the 3-D facial features (the two inner corners of the eyes and the tip of the nose.) Statistical modeling approaches have shown to be successful in extracting facial features from 2-D textured images. Applying a statistical approach such as the Active Shape Model to select the best facial features from candidate points extracted by thresholding the Gaussian curvature is a promising approach.

- Expanding the current technique for 3-D facial landmark extraction or developing a new technique to extract more than three facial landmarks from the 3-D data (either range images or stereo-based reconstructed data.) Extracting 3-D facial landmarks is one of the most challenging problems that requires more attention and research.
- Improving the Active Shape Model technique for facial landmark extraction by utilizing nonlinear models such as Kernel PCA or manifold learning. This makes the process of landmark points extraction under pose variations in 2-D facial images more robust.
- Extending the ARG approach to recognize faces with expressions. In order to handle facial expressions, a good approach can be sub-graph matching. In fact, instead of using the whole graph for matching, the ARG graph can be matched partially (e.g., the upper part of the face including the eyes, the eyebrows and the nose).
- Improving the technique for extracting ridge images for recognition. For example, by using preprocessing techniques, ridge images can be refined and as a result a set of smooth ridges could be extracted and used for 3-D face matching.
- Besides the extracted ridge points in this dissertation, a set of ravine points can be extracted using the k_{min} principal curvature. It will be interesting to use these points either alone or with the ridge points (extracted by thresholding the k_{max} principal curvature) for 3-D face recognition.

Appendix A

Enhancement of Active Shape Model Using Color Information

The ASM approach represents a target structure by a parameterized statistical model. By choosing the model parameters, different variations of a target shape can be obtained. In this subsection, first we review the theoretical background of the ASM and subsequently we will present our method for improving the technique.

In the ASM technique, the location of n landmark points (e.g. facial features in our study), are annotated on a set of training images by a human expert. This set of points is represented by a vector $X = (x_1, y_1, \dots, x_n, y_n)^T$, where x_i and y_i are the coordinates of the i^{th} landmark. Then, a model that incorporates the variations in shape over the training set is represented as follows:

$$X \approx \bar{X} + P b \tag{A.1}$$

The vector \bar{X} contains the mean values of the coordinates of the annotated data, P is a matrix of the first t eigenvectors of the covariance matrix of the annotated data, and b is a vector that defines the model parameters. The variance of the i^{th} parameter,

P_i , across the training set is given by the corresponding eigenvalue λ_i . By limiting the parameter b_i in the range of $\pm 3\sqrt{\lambda_i}$, we ensure that the generated shape is similar to those in the original training set. To apply the created shape model to a given target shape, we need to find a transformation to move from the model coordinate system to the image coordinate system. Typically, this is achieved by a similarity transformation defining the translation (X_t, Y_t) , rotation θ , and scale s . Therefore, the position of the model points, X , in the image are given by:

$$X = T_{X_t, Y_t, \theta, s}(\bar{X} + P b) \quad (\text{A.2})$$

For a given new image, the ASM is performed to find where the target object lies on the image. Therefore, we need to find the optimum parameters of the ASM that best fit of the model to the target structure. Generally, this optimization problem is solved iteratively [45]. At the first step, the model is initialized by the mean shape. Afterwards, a region of the image around each feature point is examined to find the best nearby match (e.g., searching along the profile line for the edge locations). In the next step, the parameters X_t, Y_t, s, θ , and b are updated to best fit the new found points. Then, the constraint $|b_i| < 3\sqrt{\lambda_i}$ is applied to the parameters b_i . These steps are repeated until there is no significant change in the shape parameters.

A.1 Shape Model Initialization and Face Alignment

As we mentioned in subsection 3.2.1, the initialization of the ASM is very important. With poor initialization, the search process may either fail or become slow. Therefore,

a good initialization would help in finding the optimum solution in less iterations. We use our algorithm in [63] to find the centers of the mouth and the two eyes. Figure A.1 shows the extracted locations of the eyes and the mouth in a given color face image using this method. We use these three points to obtain the affine parameters in Equation A.3 to initialize the shape model for the image.

For a 3-D object like the face, since it may have some pose variations in the captured images, a similarity transformation with 4 degrees of freedom is not effective especially when there are large pose variations. When the head is rotated to the left or right, this problem is more severe. To solve this problem, we use a 2-D affine transformation with 6 degrees of freedom, given by Equation A.3, to align the extracted feature points in the image coordinate system with the points represented by the model coordinate system.

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (\text{A.3})$$

To find the parameters of the 2-D affine transformation, we need at least three corresponding points, which are not collinear.

A.2 Improvement of The Local Structure Model

In the original ASM, the local structure of a feature point is modeled by assuming that the normalized first derivative of the pixel intensity values along a profile line satisfies a multivariate Gaussian distribution. This gives a statistical model for the

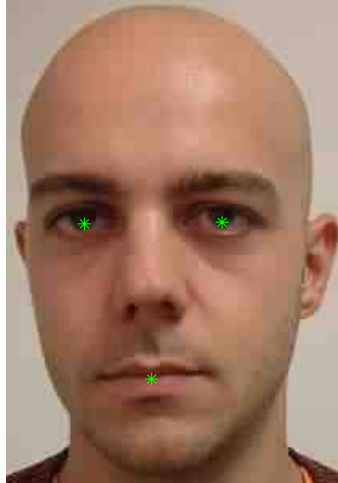


Figure A.1: Detecting the center of the mouth and the eyes as the salient features for initialization.

profile around the point. As shown in Figure A.2, a sampled profile, g_s , is matched to a reference model by searching along the profile line that goes through the point and finding the best fit. This is achieved by calculating the Mahalanobis distance:

$$f(g_s) = (g_s - \bar{g})^T \Sigma^{-1} (g_s - \bar{g}) \quad (\text{A.4})$$

where \bar{g} is the mean value and Σ is the covariance matrix.

In this dissertation, we assumed that the normalized first derivative of three channel values (i.e., Hue, Saturation, and Value) along a profile line for each individual channel satisfies a multivariate Gaussian distribution. Then, we use a weighted sum of Mahalanobis distances for the three color channels to find the best match for the feature points. Similar to Equation A.4, the best matching of a probe sample, g_{hsv} ,

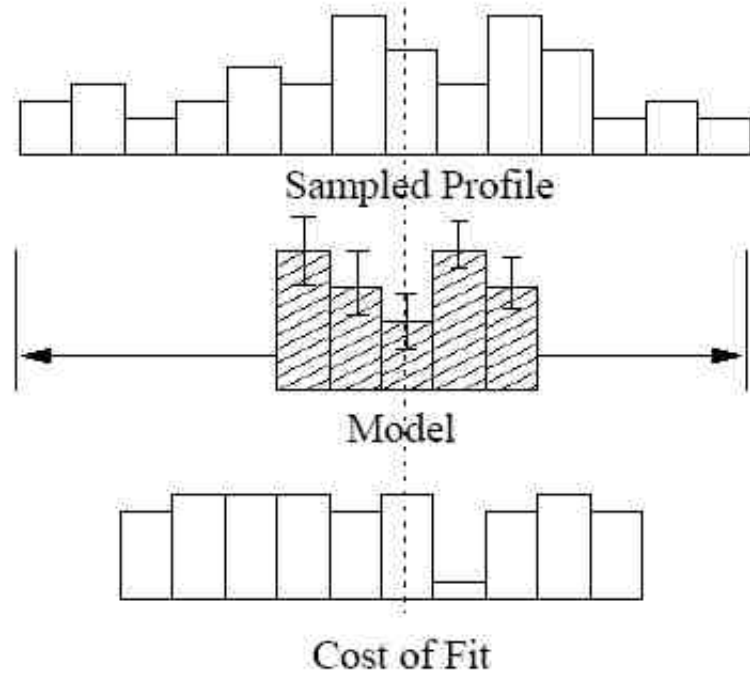


Figure A.2: Searching along a sampled profile to find the best fit.

in HSV color space to a reference model is carried out by:

$$f(g_{hsv}) = \sum_{i \in \{h,s,v\}} w_i \cdot (g_i - \bar{g}_i)^T \Sigma_i^{-1} (g_i - \bar{g}_i) \quad (\text{A.5})$$

where w_i is the weighting factor for the i^{th} component of the Gaussian model with unit sum.

A.3 Enhancement of the Location of Features Around the Lips

To accurately localize the feature points around the lips, we segment the lips region from the facial skin region. This is achieved by classifying each pixel, $I(h,s,v)$, in HSV color space, either as facial skin or lips. This classification starts by applying

the Fisher Discriminant Analysis (FDA) technique. FDA is a data analysis technique that provides more class separability than Principal Component Analysis (PCA), which provides better data representation. By applying FDA to each pixel of the color image, $I(h,s,v)$, we obtain a scalar function that can be used to discriminate between the two classes: facial skin and lips. This function is calculated by using the within-class scatter matrix and is defined as:

$$Fisher(I) = W.I^T \quad (A.6)$$

where I is a given pixel value. The projection vector W , is calculated by:

$$W = S_W^{-1}(m_1 - m_2) \quad (A.7)$$

The within class scatter matrix S_W , is:

$$S_W = S_1 + S_2 \quad (A.8)$$

and

$$S_i = \sum (I - m_i)(I - m_i)^T \quad (A.9)$$

The sample mean vector of each class, m_i , is defined as:

$$m_i = \frac{1}{n_i} \sum_{I \in D_i} I \quad (A.10)$$

Where D_i is the set of the pixels in i^{th} class and n_i is the number of the pixels in the class.

To learn the projection matrix W , we obtain a color database of facial images and manually extract patches of lips and facial skin regions. Then, the matrices S_i and m_i are calculated for the two classes of facial skin and lips. For a given test image, we apply the FDA function (Eq. A.6) and apply threshold to the result to segment the lips from the facial skin. Figure A.3 shows the results of the different steps for lips detection applied to a sample image in our database using FDA. Figure A.3(a) is the original image, Figure A.3(b) shows the result of applying FDA, Figure A.3(c) shows the result of thresholding. The value of the threshold is obtained by a small training set different from the images used in the experiment Section. The result of applying morphological operators to remove noise and fill holes is shown in A.3(d). The following summaries our iterative approach for facial features extraction using the enhanced active shape model:

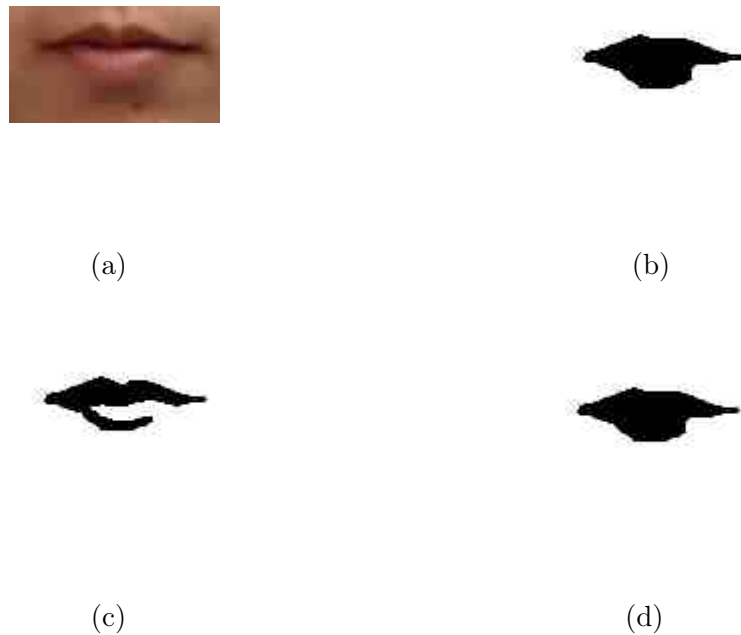


Figure A.3: Lips detection using FDA. (a) Original image.(b) Result of FDA classification. (c) Result of thresholding. (d) Result of applying morphological operators.

1. Extract the centers of the eyes and the mouth [63].
2. Initialize the shape model based on the extracted three points from step 1.
3. Calculate the shape parameters, b .
4. Examine a region of the image around each point, (x_i, y_i) , to find the best nearby match for that point using the color model.
5. Use lips detection to tune the feature points around the mouth.
6. Update the parameters of the affine transformation $(a_{11}, a_{12}, a_{21}, a_{22}, t_x, t_y)$ to best fit the new found locations of the instance, X , of the target model.
7. Apply the constraints to the parameters, b , to ensure reasonable shapes (i.e. $|b_i| \leq 3\sqrt{\lambda_i}$).
8. Go to step 3 and repeat until convergence.

Appendix B

The Iterative Closest Point Algorithm

The Iterative Closest Point (ICP) algorithm has become the dominant method for aligning three-dimensional models based purely on the geometry, and sometimes color, of the meshes. The algorithm is widely used for registering the outputs of 3D scanners, which typically only scan an object from one direction at a time. ICP starts with two meshes and an initial guess for their relative rigid-body transform, and iteratively refines the transform by repeatedly generating pairs of corresponding points on the meshes and minimizing an error metric. The general framework of the algorithm is represented in Table B.1. Generating the initial alignment may be done by a variety of methods, such as tracking scanner position, identification and indexing of surface features [51, 154], “spin-image” surface signatures [75, 67], computing principal axes of scans [47], exhaustive search for corresponding points [39], or user input.

Since the introduction of ICP by Chen and Medioni [40] and Besl and McKay [22], many variants have been introduced on the basic ICP concept. According to [143], these variants affect one of six stages of the algorithm:

```

begin  $E' \leftarrow +\infty$ ;
(Rot,Trans)  $\leftarrow$  In Initialize-Alignment(Scene,Model);
  repeat
     $E \leftarrow E'$ ;
    Aligned-Scene  $\leftarrow$  Apply-Alignment(Scene,Rot,Trans);
    Pairs  $\leftarrow$  Return-Closest-Pairs(Aligned-Scene,Model);
    (Rot,Trans, $E'$ )  $\leftarrow$  Update-Alignment(Scene,Model,Pairs,Rot,Trans);
  Until  $|E' - E| < Threshold$ 
  return (Rot,Trans);
end

```

Table B.1: The iterative closest point algorithm.

1. Selection of some set of points in one or both meshes.
2. Matching these points to samples in the other mesh.
3. Weighting the corresponding pairs appropriately.
4. Rejecting certain pairs based on looking at each pair individually or considering the entire set of pairs.
5. Assigning an error metric based on the point pairs.
6. Minimizing the error metric.

Rusinkiewicz [144] has constructed a high-speed ICP algorithm with consideration of the accuracy of the final answer by combining some of the variants of the algorithm. The implementation (C++ code) of this algorithm is available on the web (<http://www.cs.princeton.edu/gfx/proj/trimesh2/>). We wrote a wrapper to use this implementation of ICP in Matlab environment for 3D facial image registration.

Appendix C

Sequential Floating Forward Selection Procedure

Feature selection (FS) constitutes an important component in building a pattern recognition system. Basically, it is the process of choosing the input to the pattern recognition system. Accordingly, the main goal of FS is to select a subset of d features from the given set of D measurements, $d < D$, without significantly degrading (or possibly even improving due to the “peaking phenomenon” [49]) the performance of the recognition system. Assuming that a suitable criterion function has been chosen to evaluate the effectiveness of feature subsets, FS is reduced to a search problem that selects an optimal feature subset based on the criterion function.

The well-known Sequential Forward Selection (SFS) and Sequential Backward Selection (SBS) are step-optimal only since the best (the worst) feature is always added (discarded) in SFS and SBS, respectively. This results in nested feature subsets without any chance to correct the decision in later steps., causing the performance to be often far from optimal.

A definitive improvement can be obtained by combining SFS and SBS to avoid the nesting effect. One of the efficient algorithms that combines the SFS and SBS is

SFFS Algorithm
<p><i>Input:</i> $Y = \{y_j j = 1, \dots, D\}$ //available measurements//</p> <p><i>Output:</i> $X_k = \{x_j j = 1 \dots k; x_j \in Y\}; k = 0, 1, \dots, D$</p> <p><i>Initialization:</i> $X_0 := ; k := 0$ (in practice one can begin with $k = 2$ by applying SFFS twice)</p> <p><i>Termination:</i> Stop when k equals the number of features required</p> <p>Step 1 (Inclusion) $x^+ := \arg \max_{x \in Y - X_k} J(X_k + x)$ //the most significant feature with respect to X_k $X_{k+1} := X_k + x^+; k := k + 1$</p> <p>Step 2 (Conditional Exclusion) $x^- := \arg \max_{x \in X_k} J(X_k - x)$ // the least significant feature in X_k if $J(X_k - \{x^-\}) > J(X_{k-1})$ then $X_{k-1} := X_k - x^-; k := k - 1$ go to Step 2 else go to Step 1</p>

Table C.1: Sequential Floating Forward Selection algorithm for feature selection [139].

Sequential Floating Forward Selection (SFFS.) The SFFS consists of applying after each forward step of adding a feature, a number of backward steps (removing features) as long as the resulting subsets are better than the previously evaluated step [139]. Consequently, there are no backward steps at all if the performance cannot be improved. The SFFS method is described algorithmically as follows. The SFFS algorithm allows a “self-controlled backtracking” so it can eventually find good solutions by adjusting the trade-off between forward and backward steps dynamically.

Appendix D

Dempster-Shafer Theory of Evidence

We briefly review the Dempster-Shafer theory of evidence [78]. Let $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ be a set of mutually exclusive and exhaustive propositions that a classifier may discern. This is referred to as frame of discernment (FoD); it signifies the “scope” of expertise. A proposition θ_i , referred to as a singleton, represents the lowest level of discernible information. We use $|\Theta|$ and 2^Θ to denote the cardinality and the power set of Θ , respectively. Elements in 2^Θ form all the propositions of interest in DS theory. We use \bar{A} to denote all singletons in Θ that are not included in A . The “support” for proposition A is provided via a basic belief assignment (BBA) or mass assignment.

The mapping $m : 2^\Theta \mapsto [0, 1]$ is a BBA for the FoD Θ such that: 1) $m(\emptyset) = 0$ and 2) $\sum_{A \in 2^\Theta} m(A) = 1$.

The mass assigned to a proposition is free to move into the individual singletons that constitute the composite proposition thus generating the notion of ignorance. The set of propositions \mathcal{F} that possesses nonzero BBAs forms the focal elements of Θ ; the triple $\{\Theta, \mathcal{F}, m\}$ is the corresponding body of evidence (BoE). The quantity $m(A)$ measures the support assigned to proposition A only. The belief assigned to A on the other hand takes into account the supports for all proper subsets of A as well.

Given a BoE, $\{\Theta, \mathcal{F}, m\}$, and $A \subseteq \Theta$, belief is defined as

$$Bl : 2^\Theta \mapsto [0, 1] \quad \text{where } Bl(A) = \sum_{B \subseteq A} m(B) \quad (\text{D.1})$$

and plausibility is defined as

$$Pl : 2^\Theta \mapsto [0, 1] \quad \text{where } Pl(A) = 1 - Bl(\bar{A}). \quad (\text{D.2})$$

In other words, in DS theory, $Bl(A)$ represents the total support that can move into A without any ambiguity; and $Pl(A)$ accounts for all the observations that do not rule out the proposition A . When each focal set contains only one element, i.e., $m(A) = 0, \forall |A| \neq 1$, belief functions become probability functions. Then, the BBA, belief and plausibility all reduce to probability [138].

In DS theory, the Dempster's evidence combination function allows one to combine the evidence generated by several BoEs spanning the same FoD. Consider BoE1, $\{\Theta_1, \mathcal{F}_1, m_1\}$, and BoE2, $\{\Theta_2, \mathcal{F}_2, m_2\}$, where $\Theta_1 = \Theta_2 = \Theta$. Then, the Dempster's combination rule generates the BBA $m(\cdot) : 2^\Theta \mapsto [0, 1]$ for a new BoE as:

$$m(A) = m(B) \oplus m(C) = \frac{\sum_{B \cap C = A} m(B)_{\Theta_1} m(C)_{\Theta_2}}{1 - \sum_{B \cap C = \emptyset} m(B)_{\Theta_1} m(C)_{\Theta_2}}, \quad \forall A, B, C \subseteq \Theta. \quad (\text{D.3})$$

This combining rule can be generalized by iteration to fuse more than two source

of information. For example, fusion of three evidences is defined as:

$$m_{final} = m(B) \oplus m(C) \oplus m(D) \quad (\text{D.4})$$

where \oplus is the DS combination rule in Eq. D.3.

Bibliography

- [1] <http://cvc.yale.edu>.
- [2] <http://www.facekey.com>.
- [3] <http://www.identix.com>.
- [4] <http://www.imagistechnologies.com/>.
- [5] <http://www.maximus.com/corporate/pages>.
- [6] <http://www.premierelect.co.uk>.
- [7] <http://www.viisage.com>.
- [8] International biometric group: <http://www.identix.com>.
- [9] Neven vision, inc. nevenvision machine vision technology, <http://www.nevenvision.com/>.
- [10] Face recognition vendor test 2006. Available online at <http://www.frvt.org>, 2006.
- [11] Nasser A-Ansari. *3D Face Mesh Modeling With Applications to 3D and 2D face Recognition*. Ph.d. dissertation, University of Miami, Coral Gables, FL, 2007.
- [12] M. Abdel-Mottaleb. Image retrieval based on edge representation. In *The 7th International Conference on Image Processing*, September 2000.
- [13] M. Abdel-Mottaleb, N. Dimitrova, R. Desai, and J. Martino. Conivas: Content-based image and video access system. In *ACM Multimedia96, Boston*, 1996.
- [14] E.V. Anoshkina, A.G. Belyaev, R. Huang, T.L. Kunii, and O.G. Okuncv. S. Takahashi. Hierarchic shape description via singularity and multiscaling. In *Proceedings of IEEE Computer Software and Applications Conference*, 2004.
- [15] A. N. Ansari and M. Abdel-Mottaleb. Automatic facial feature extraction and 3d face modeling using two orthogonal views with application to 3d face recognition. *Pattern Recognition*, 38(12):2549–2563, December 2005.
- [16] B. Moghaddam B. and A. Pentland. Probabilistic visual learning for object representation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 19(7):696–710, July 1997.
- [17] Hans-Jürgen Bandelt and Victor Chepoi. Metric graph theory and geometry: a survey. In *Contemporary Mathematics, Special issue of Proceedings of the Joint Summer Research Conference on Discrete and Computational Geometry*, 2006.

- [18] M. S. Bartlett, H. M. Lades, and T. Sejnowski. Independent component representation for face recognition. In *In Proceedings, SPIE Symposium on Electronic Imaging: Science and Technology*, pages 528–539, 1998.
- [19] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Patt. Anal. Mach. Intell.*, 19:711–720, 1997.
- [20] A.J. Bell and T. J. Sejnowski. The independent components of natural scenes are edge filters. *Vis. Res.*, 37:3327–3338, 1995.
- [21] J. BenjArie and D. Nandy. A volumetric/iconic frequency domain representation for objects with application for pose invariant face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(5):449–457, May 1998.
- [22] P. besl and N. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.
- [23] C. Beumier and M. Acheroy. Automatic 3d face authentication. *Image Vision Computing*, 18(4):315–321, 2000.
- [24] C. Beumier and M. Acheroy. Face verification from 3d and grey level cues. *Pattern Recognition*, 22:1321–1329, 2001.
- [25] I. Biederman and P. Kalocsai. *Neural and psychophysical analysis of object and face recognition*, volume Springer-Verlag, Berlin, Germany. In *Face Recognition: From Theory to Applications*, H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie, and T. S. Huang, 1998.
- [26] D. Blackburn, M. Bone, and P. J. Phillips. Face recognition vendor test 2000. technical report, 2001.
- [27] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *In Proc. of ACM SIGGRAPH 99*, pages 187–194, 1999.
- [28] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, (25):1063–1074, Sep. 2003.
- [29] David S. Bolme. Elastic bunch graph matching. Msc thesis, Colorado State University, Summer 2003.
- [30] K. Bowyer, K. Chang, and P. Flynn. A survey of approaches to three-dimensional face recognition. In *Proc. of International Conference on Pattern Recognition, Cambridge, England*, pages 358–361, August 2004.

- [31] Kevin W. Bowyer, K. Chang, and Patrick Flynn. A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition. *Computer Vision and Image Understanding*, 101:1–15, 2006.
- [32] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. Three-dimensional face recognition. *International Journal of Computer Vision*, pages 5–30, 2005.
- [33] V. Bruce. *Recognizing faces*. Lawrence Erlbaum Associates, London, U.K., 1988.
- [34] J. Buhmann, M. Lades, and C. V. D. Malsburg. Size and distortion invariant object recognition by hierarchical graph matching. In *In Proceedings of International Joint Conference on Neural Networks*, pages 411–416, 1990.
- [35] H. Chan and W. Bledos. A man-machine facial recognition system: some preliminary results. Technical report, Technical report, Panaromic Research Inc, California, 1965.
- [36] K. I. Chang, K. Bowyer, and P. Flynn. Face recognition using 2d and 3d facial data. In *Multimodal User Authentication Workshop*, pages 25–32, December 2003.
- [37] K. I. Chang, K. W. Bowyer, and P. J. Flynn. Adaptive rigid multi-region selection for handling expression variation in 3d face recognition. In *IEEE Workshop on Face Recognition Grand Challenge Experiments*, June 2005.
- [38] R. Chellappa, C. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5):705–740, 1995.
- [39] C. Chen, Y. Hung, and J. Cheng. Ransac-based darces: A new approach to fast automatic registration of partially overlapping range images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 21(11), 1999.
- [40] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. In *Proc. IEEE Conf. on Robotics and Automation*, 1991.
- [41] A. Chowdhury and R. Chellappa. Face reconstruction from monocular video using uncertainty analysis and a generic model. *Computer Vision and Image Understanding*, 91:188–213, August 2003.
- [42] C. Chua, F. Han, and Y. Ho. 3d human face recognition using point signature. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 233–238, 2000.
- [43] I. Cohen, N. Sebe, F. Cozman, M. Cirelo, and T. Huang. Learning bayesian network classifiers for facial expression recognition using both labeled and unlabeled data. In *Proceedings of IEEE International conference on Computer Vision and Pattern Recognition*, 2003.

- [44] J. A. Cook, V. Chandran, and C. Fookes. 3d face recognition using log-gabor templates. In *The 17th British Machine Vision Conference (BMVC)*, September 2006.
- [45] T.F. Cootes, D. Cooper, C.J. Taylor, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, Jan. 1995.
- [46] R. L. DeValois and K. K. DeValois. *Spatial Vision*. Oxford Press, 1988.
- [47] C. Dorai, J. Weng, and A. Jain. Optimal registration of object views using range data. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(10), 1997.
- [48] M. Dubisson and A. K. Jain. A modified hausdorff distance for object matching. In *12th International Conference Pattern Recognition (ICPR)*, Jerusalem, Israel, Oct. 1994.
- [49] R. O. Duda, P.E. Hart, and D. G. Stork. *Pattern Classification*. New York: Wiley, 2nd edition, 2000.
- [50] K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images. *J. Opt. Soc. Am.*, 14:1724–1733, 1997.
- [51] O. Faugeras and M. Hebert. The representation, recognition, and locating of 3-d objects. *Int. J. Robotic Res.*, 5(3), 1986.
- [52] D. Fields. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of Optical Society of America*, 4(12):2379–2394, 1987.
- [53] Frost and Sullivan. World face recognition biometrics market, 2006.
- [54] K. Fukunaga. *Statistical Pattern Recognition*. Academic Press, New York, NY., 1989.
- [55] Simson Garfinkel. *Database Nation: The Death of Privacy in the 21st Century*. O'Reilly and Associates, 2000.
- [56] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone model for face recognition under variable lighting and pose. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23(6), 2001.
- [57] B. Van Ginneken, A.F. Frangi, J.J. Staal, B.M. Ter Haar Romeny, and M.A. Viergever. A non-linear gray-level appearance model improves active shape model segmentation. In *IEEE workshop on Mathematical Models in Biomedical Image Analysis*, pages 205–212, 2001.

- [58] A. Godil, S. Ressler, and P. Grother. Face recognition using 3d facial shape and color map information comparison and combination. *Biometric Technology for Human Identification, SPIE*, 5404:351–361, April 2005.
- [59] G. Gordon. Face recognition from depth maps and surface curvature. In *in Proc. of SPIE, Geometric Methods in Computer Vision, San Diego*, volume 1570, July 1991.
- [60] A. Gunduz and H. Krim. Facial feature extraction using topological methods. In *International Conference on Image Processing(ICIP)*, pages I: 673–676, 2003.
- [61] P. W. Hallianan. A low-dimensional representation of human faces for arbitrary lighting conditions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 995–999, 1994.
- [62] R. Houghton. Named faces: putting names to faces. *IEEE Intelligent Systems Magazine*, 14(5):45–50, 1999.
- [63] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain. Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(5):696–706, May 2002.
- [64] R.L. Hsu and A.K. Jain. Face modeling for recognition. In *Proc. Int'l Conf. Image Processing (ICIP)*, October 2001.
- [65] C. Hu, R.S. Feris, and M. Turk. Active wavelet networks for face alignment. In *British Machine Vision Conference*, 2003.
- [66] Yongli Hu, Baocai Yin, and Dehui Kong. A new facial feature extraction method based on linear combination model. In *International Conference on Web Intelligence*, 2003.
- [67] D. Huber. Automatic 3d modeling using range images obtained from unknown viewpoints. In *Proc. 3DIM*, 2001.
- [68] J. R. Hurley and R. B. Cattell. The procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behavioral Science*, pages 258–262, 1962.
- [69] Michael Husken, Michael Brauckmann, Stefan Gehlen, and Christoph Von der Malsburg. Strategies and benefits of fusion of 2d and 3d face recognition. In *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, page 174, 2005.
- [70] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 15(9):850–863, 1993.

- [71] D. P. Huttenlocher, R. Liliën, and C. Olson. View-based recognition using an eigenface approximation to the hausdorff measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 21(9):951–955, 1993.
- [72] Paul A. Griffin J. Atick and A. Norman Redlich. Statistical approach to shape from shading: Reconstruction of 3d face surfaces from single 2d images. *Neural Computation*, 1997.
- [73] A. K. Jain, Y. Zhang, and M.P. Dubuisson-Jolly. Deformable template models: A review. *Journal of Signal Processing*, 71(2):109–129, 1998.
- [74] F. Jiao, S. Li, H.Y. Shum, and D. Schuurmans. Face alignment using statistical models and wavelet features. In *Computer Vision and Pattern Recognition*, pages I: 321–327, 2003.
- [75] A. Johnson and M. Hebert. Surface registration by matching oriented points. In *Proc. 3DIM*, 1997.
- [76] A. Johnson and M. Hebert. Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 21(5):433–449, 1999.
- [77] J. P. Jones and L. A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *J. of Neurophysiology*, 58:1233–1258, 1987.
- [78] K. Premaratne K., D.A. Dewasurendra, and P.H. Bauer. Evidence combination in an environment with heterogeneous sources. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 37(3):298–309, May 2007.
- [79] I.A. Kakadiaris, G. Passalis, G. Toderici, N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis. 3d face recognition in the presence of facial expressions: An annotated deformable model approach. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 29(4):640–649, 2007.
- [80] T. Kanade. *Computer Recognition of Human Faces*. Birkhauser, 1977.
- [81] M. D. Kelly. Tech. rep. ai-130, visual identification of people by computer. Technical report, Stanford AI Project, Stanford, CA, 1970.
- [82] Kyung-A Kim, Se-Young Oh, and Hyun-Chul Choi. Facial feature extraction using pca and wavelet multi-resolution. In *Face and Gesture Recognition*, pages 439–444, 2004.
- [83] M. Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Trans. on Patt. Anal. Mach. Intell.*, 12, 1990.

- [84] Seiji Kobayashi and Shuji Hashimoto. Automated feature extraction of face image and its applications. In *International Workshop on Robot and Human Communication*, pages 164–169, 1995.
- [85] N. Kruger, M. Potzsch, and C.V.D. Malsburg. Determination of face position and pose with a learned representation based on labelled graphs. *Image Vision Computing*, 15:665–673, 1997.
- [86] A. Kuchinsky, C. Pering, M. Creech, D. Freeze, B. Serra, and J. Gvvezdka. Fotofile: A consumer multimedia organization and retrieval system. In *the CHI'99 Conference on Human Factors in Computing Systems*, Pennsylvania, U.S.A, 1999.
- [87] S. Y. Kung and J. S. Taur. Decision-based neural networks with signal/image classification applications. *IEEE Trans. on Neural Network*, 6:170–181, 1995.
- [88] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. V. D. Malsburg, R. Wurtz, and W. Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.*, 42:300–311, 1993.
- [89] J.C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright. Convergence properties of the nelder-mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9(1):112–147, 1998.
- [90] Kin-Man Lam and Yan-Lin Li. An efficient approach for facial feature detection. In *4th International Conference on Image Processing*, volume 2, 1998.
- [91] A. Lanitis, C. J. Taylor, and T. F. Cootes. An automatic face identification system using flexible appearance models. *Image Vis. Comput.*, 13:393–401, 1995.
- [92] A. Lanitis, C.J. Taylor, and T.F. Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 19(7):743–756, July 1997.
- [93] S. Lao, Y. Sumi, M. Kawade, and F. Tomita. 3d template matching for pose invariant face recognition using 3d facial model built with iso-luminance line based stereo vision. *International Conference on Pattern Recognition (ICPR 2000)*, pages 911–916, 2000.
- [94] C.M. Lau, W.K. Cham, H.T. Tsui, and K.N. Ngan. An energy function for facial feature extraction. In *International Symposium on Intelligent Multimedia, Video and Speech Processing*, volume 1, pages 348–351, 2001.
- [95] J. C. Lee and E. Milios. Matching range images of human faces. In *Proc. IEEE Soc. 3rd Int. Conf. on Computer Vision (ICCV)*, 1990.

- [96] M. Lee and S. Ranganath. Pose invariant face recognition using 3d deformable model. *Pattern Recognition*, 36(8), 2003.
- [97] Y. Lee, H. Song, H. Shin U. Yang, and K. Sohn. Local feature based 3d face recognition. In *in: International Conference on Audio- and Videobased Biometric Person Authentication (AVBPA 2005*, volume 3546, pages 909–918, July 2005.
- [98] Y. Lee, D. Terzopolous, and K. Waters. Realistic modeling for facial animation. In *Proc. SIGGRAPH, Los Angeles*, pages 55–61, August 1995.
- [99] R. Lengagne, J. P. Tarel, and O. Monga. From 2d images to 3d face geometry. In *Proc. IEEE Int'l Conf. Automatic Face and Gesture Recognition, AFGR96*, Oct. 1996.
- [100] Charay Lerdsudwichai, Mohamed Abdel-Mottaleb, and A-Nasser Ansari. Tracking multiple people with recovery from partial and total occlusion. *Pattern Recognition*, 38(7):1059–1070, July 2005.
- [101] S. Z. Li. Matching: Invariant to translations, rotations and scale changes. *Pattern Recognition*, 25(6):583–594, 1992.
- [102] S. H. Lin, S. Y. Kung, and L. J. Lin. Face recognition/detection by probabilistic decisionbased neural network. *IEEE Trans. on Neural Network*, 8:114–132, 1997.
- [103] A. M. Lopez, F. Lumbreras, J. Serrat, and J. J. Villanueva. Evaluation of methods for ridge and valley detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 21(4):327–335, APRIL 1999.
- [104] X. Lu. *3D Face Recognition across Pose and Expression*. Ph.d. dissertation, Michigan State University, 2006.
- [105] X. Lu and A. K. Jain. Deformation analysis for 3d face matching. In *7th IEEE Workshop on Applications of Computer Vision*, 2005.
- [106] X. Lu and A. K. Jain. Integrating range and texture information for 3d face recognition. In *7th IEEE Workshop on Applications of Computer Vision (WACV '05)*, pages 155–163, 2005.
- [107] X. Lu and A. K. Jain. Multimodal facial feature extraction for automatic 3d face recognition. In *Technical Report MSU-CSE-05-22, Department of Computer Science, Michigan State University, East Lansing, Michigan*, August 2005.
- [108] Xiaoguang Lu, Anil K. Jain, and Dirk Colbry. Matching 2.5d face scans to 3d models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):31–43, 2006.

- [109] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni. Performance of geometrix activeid 3d face recognition engine on the frgc data. In *IEEE Workshop on Face Recognition Grand Challenge Experiments*, June 2005.
- [110] T. Maurer and C. V. D. Malsburg. Single view based recognition of faces rotated in depth. In *In Proceedings, International Workshop on Automatic Face and Gesture Recognition*, pages 176–181, 1996.
- [111] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 2(1):91–108, 1996.
- [112] G. Medioni and R. Waupotitsch. Face recognition and modeling in 3d. In *In Proc. Workshop on Analysis and Modeling of Faces and Gestures*, pages 232–233, October 2003.
- [113] K. Messer, J. Matas, J. Kittler, J. Luetttin, and G. Maitre. Xm2vtsd: The extended m2vts database. In *In Proceedings, International Conference on Audio- and Video-Based Person Authentication*, pages 72–77, 1999.
- [114] H. Moon and P. J. Phillips. Computational and performance aspects of pca-based face recognition algorithms. *Perception*, 30:301–321, 2001.
- [115] A. B. Moreno and A. Sanchez. Gavab: A 3d face database. In *2nd COST Workshop on Biometrics on the Internet: Fundamentals, Advances and Applications*, 2004.
- [116] A. B. Moreno, Angel Sanchez, and Jose Fco. Velez. Voxel-based 3d face representations for recognition. In *12th International Workshop on Systems, Signals and Image Processing (IWSSIP'05)*, September 2005.
- [117] A. Belen Moreno, Angel Sanchez, Jose Fco. Velez, and Fco. Javier Dkz. Face recognition using 3d local geometrical features: Pca vs. svm. In *Fourth International Symposium on Image and Signal Processing and Analysis (ISPA 2005)*, September 2005.
- [118] Ana Belen Moreno, Angel Sanchez, Jose Fco. Velez, and Fco. Dkz. Face recognition using 3d surface-extracted descriptors. In *Irish Machine Vision and Image Processing Conference 2003 (IMVIP'03)*, September 2003.
- [119] Kenji Nagao. Bayesian approach with nonlinear kernels to feature extraction. In *17th International Conference on Pattern Recognition (ICPR04)*, 2004.
- [120] Francesco G.B. De Natale, Daniele D. Giusto, and Fabrizio Maccioni. A symmetry-based approach to facial features extraction. In *13th International Conference on DSP*, volume 2, pages 521–525, July 1997.

- [121] A. Nikolaidis, C. Kotropoulos, and I. Pitas. Facial feature extraction using adaptive hough transform, template matching and active contour models. In *13th International Conference on DSP*, volume 2, pages 865–868, July 1997.
- [122] K. Okada, J. Steffans, T. Maurer, H. Hong, E. Elagin, H. Neven, and C. V. D. Malsburg. *The Bochum/USC Face Recognition System and how it fared in the FERET Phase III Test*, volume Springer-Verlag, Berlin, Germany. In *Face Recognition: From Theory to Applications*, H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie, and T. S. Huang, 1998.
- [123] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.
- [124] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. *ACM Trans. on Graphics*, 21(4):807–832, 2002.
- [125] Z. Wen P. Hong and S. Huang. Real-time speech-driven expressive synthetic talking faces using neural networks. *IEEE Transaction on Neural Networks*, 13(4):916–927, 2002.
- [126] G. Pan, S. Han, Z. Wu, and Y. Wang. 3d face recognition using mapped depth images. In *IEEE Workshop on Face Recognition Grand Challenge Experiments*, June 2005.
- [127] G. Pan, Y. Wu, Z. Wu, and W. Liu. 3d face recognition by profile and surface matching. In *Proc. IEEE International Joint Conference on Neural Networks*, pages 2168–2174, 2003.
- [128] T. Papatheodorou and D. Reuckert. Evaluation of automatic 4d face recognition using surface and texture registration. In *Sixth International Conference on Automated Face and Gesture Recognition, AFGR04*, pages 321–326, May 2004.
- [129] Bo-Gun Park, Kyoung-Mu Lee, and Sang-Uk Lee. Face recognition using face-arg matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1982–1988, 2005.
- [130] G. Passalis, I. Kakadiaris, T. Theoharis, G. Toderici, and N. Murtuza. Evaluation of 3d face recognition in the presence of facial expressions: an annotated deformable model approach. In *IEEE Workshop on Face Recognition Grand Challenge Experiments*, June 2005.
- [131] P. Penev and J. Attick. Local feature analysis: A general statistical theory for object representation. *Network : Computation in Neural Systems*, 5:477–500, 1996.

- [132] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *In Proceedings, IEEE Conference on Computer Vision and Pattern Recognition*, 1994.
- [133] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [134] P. J. Phillips, P. J. Grother, R. J. Michaels, D. M. Blackburn, E. Tabassi, and J. M. Bone. Face recognition vendor test 2002: Evaluation report, 2003.
- [135] P. J. Phillips, H. Moon S. Rizvi, and P. Pauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22, 2000.
- [136] Tomaz Pisanski and Milan Randic. Bridges between geometry and graph theory. *Geometry at Work: Papers in Applied Geometry*, (4):174–194, 2000.
- [137] D. A. Pollen and S. F. Ronner. Phase relationship between adjacent simple cells in the visual cortex. *Science*, 212:1409–1411, 1981.
- [138] K. Premaratne, J. Zhang, and K. K. R. G. K. Hewawasam. Decision making in distributed sensor networks: A belief-theoretic bayes-like theorem. In *IEEE Int. MWSCAS*, volume 2, pages 497–500, July 2004.
- [139] P. Pudil, J. Novovicova, and J. Kittler. Floating search methods in feature-selection. *Pattern Recognition Letters*, 15(11):1119–1125, November 1994.
- [140] Arun A. Ross, Karthik Nandakumar, and Anil K. Jain. *Handbook of Multibiometrics (International Series on Biometrics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [141] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 20(1):23–28, Jan. 1998.
- [142] D. Ruderman. The statistics of natural images. *Network : Computation in Neural Systems*, 5:598–605, 1994.
- [143] S. Rusinkiewicz, , and M. Levoy. Efficient variants of the icp algorithm. In *Proc. 3rd Int. Conf. on 3-D Digital Imaging and Modeling (3DIM)*, pages 145–152, 2001.
- [144] Szymon Marek Rusinkiewicz. *Real-Time Acquisition and Rendering of Large 3D Models*. Ph.d. dissertation, University of Stanford, CA, August 2001.

- [145] T. D. Russ, K. W. Koch, and C.Q. Little. A 2d range hausdorff approach for 3d face recognition. In *IEEE Workshop on Face Recognition Grand Challenge Experiments*, June 2005.
- [146] W. Gao S. Shan, Y. Chang and B. Gao. Curse of mis-alignment in face recognition: Problem and a novel mis-alignmnet learning solution. In *In Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 314–320, 2004.
- [147] F. Samaria. *Face recognition using hidden markov models*. PhD thesis, University of Cambridge, Cambridge, U.K, 1994.
- [148] F. Samaria and S. Yount. Hmm based architecture for face identification. *Image Vision Computing*, 12:537–583, 1994.
- [149] Kap-Ho Seo, Won Kim, Changmok Oh, and Ju-Jang Lee. Face detection and facial feature extraction using color snake. In *IEEE International Symposium on Industrial Electronics*, volume 2, pages 457– 462, 2002.
- [150] G. Shafer. *A Mathematical Theory of Evidence*. NJ: Princeton Univ. Press, 1976.
- [151] A. Shashua. *Geometry and photometry in 3D visual recognition*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1994.
- [152] Dong-Gyu Sim, Oh-Kyu Kwon, and Rae-Hong Park. Object matching algorithms using robust hausdorff distance measures. *IEEE Transactions on Image Processing*, 8(3):425–429, March 1999.
- [153] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human face. *Journal of Opt. Soc.*, 4:519–524, 1987.
- [154] F. Stein and G. Medioni. Structural indexing: Efficient 3-d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 14(2), 1992.
- [155] D. L. Swetes and J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Trans. Patt. Anal. Mach. Intell.*, 18:831–836, 1996.
- [156] H. T. Tanaka, M. Ikeda, and H. Chiaki. Curvature-based face surface recognition using spherical correlation principal directions for curved object recognition. In *Third International Conference on Automated Face and Gesture Recognition*, 1998.
- [157] F. Tsalakanidou, S. Malassiotis, and M. Strintzis. Integration of 2d and 3d images for enhanced face authentication. In *Sixth International Conference on Automated Face and Gesture Recognition AFGR04*, pages 266 – 271, May 2004.

- [158] F. Tsalakanidou, D. Tzocaras, and M. Strintzis. Use of depth and colour eigenfaces for face recognition. *Pattern Recognition Letters*, 24:1427 – 1435, 2003.
- [159] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 1991.
- [160] Paul Viola and Michael J. Jones. Rapid object detection using a boosted cascade of simple features. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2001.
- [161] W. Wang, S. Shan, W. Gao, B. Cao, and B. Yin. An improved active shape model face alignment. In *International Conference on Multimedia*, pages 523–528, 2002.
- [162] Y. Wang, C. Chua, and Y. Ho. Facial feature detection and face recognition from 2d and 3d images. *Pattern Recognition Letters*, 23:1191–1202, 2002.
- [163] L. Wiskott, J. M. Fellous, N. Kruger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.
- [164] Z. Wu, Y. Wang, and G. Pan. 3d face recognition using local shape map. In *Proc. IEEE International Conference on Image Processing*, volume 3, pages 2003–2006, October 2004.
- [165] D. Colbry X. Lu and A. K. Jain. Matching 2.5d scans for face recognition. In *International Conference on Pattern Recognition (ICPR 2004)*, pages 362–366, 2004.
- [166] D. Xi and S.W. Lee. Face detection and facial component extraction by wavelet decomposition and support vector machines. In *4th International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, pages 199–207, 2003.
- [167] Z. Xue, Stan Z. Li, Dinggang Shen, and Eam Khwang Tcoh. A novel bayesian shape model for facial feature extraction. In *Seventh International Caderenee on Control, Automation, Robotics And Vision (ICMCV02)*, 2002.
- [168] Yasushi Yagi. Facial feature extraction’from frontal face image. In *International Confernce on Signal Processing*, 2000.
- [169] P. Yan and K.W. Bowyer. Empirical evaluation of advanced ear biometrics. pages III: 41–41, 2005.
- [170] Y. Yan and J. Zhang. Rotationinvariant 3d recognition for face recognition. In *Proc. IEEE Int’l Conf. Image Processing*, volume I, pages 156–160, 1998.

- [171] Ming-Hsuan Yang, David Kriegman, and Narendra Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.
- [172] G. Yen and N. Nithianandan. Facial feature extraction using genetic algorithm. In *Congress on Evolutionary Computation*, 2002.
- [173] B. Zhang, W. Gao, S. Shan, and W. Wang. Constraint shape model using edge constraint and gabor wavelet based search. In *4th International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA)*, pages 52–61, 2003.
- [174] C. Zhang and F. S. Cohen. 3-d face structure extraction and recognition from images using 3-d morphing and distance mapping. *IEEE Transactions on Image Processing*, 11(11):1249–1259, November 2002.
- [175] C. Zhang and F.S. Cohen. Face shape extraction and recognition using 3d morphing and distance mapping. In *Proc. IEEE Int’l Conf. Automatic Face and Gesture Recognition, AFGR00*, pages 28–33, 2000.
- [176] W. ZHAO. *Robust Image Based 3D Face Recognition*. Ph.d. dissertation, University of Maryland, College Park, MD, 1999.
- [177] W. Zhao, R. Chellappa, and A. Krishnaswamy. Discriminant analysis of principal components for face recognition. In *In Proceedings of International Workshop on Automatic Face and Gesture Recognition*, pages 336 – 341, 1998.
- [178] W. Zhao, R. Chellappa, and P. J. Phillips. Subspace linear discriminant analysis for face recognition. tech. rep. car-tr-914. Technical report, Center for Automation Research, University of Maryland, College Park, MD, 1999.
- [179] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, December 2003.
- [180] W.Y. Zhao and R. Chellappa. 3d model enhanced face recognition. In *Proc. IEEE Int’l Conf. Image Processing*, 2000.

VITA

Mohammad Hossein Mahoor was born in Estahban, Fars Province, Iran, on January 27, 1975. He received his elementary education at Shahid Faghihi Elementary School, his secondary education at Dr. Shariati Middle School, and his high school education at Shahid Beheshti High School. In September 1992, he was admitted to the University of Petroleum Industry (Former Abadan Institute of Technology, A.I.T.) in Ahwaz, Iran, from which he was graduated with the B.S. degree in Electrical Engineering with first-class honor in September 1995. He continued his graduate studies in Sharif University of Technology and was awarded M.S. degree in Biomedical Engineering in October 1998.

In August 2003, he was admitted to the Graduate School of the University of Miami, where he was granted the degree of Doctor of Philosophy in Electrical and Computer Engineering in December 2007.

Permanent Address: 1251 Memorial Dr. #EB406, Coral Gables, Florida 33124