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

MAHOOR, MOHAMMAD HOSSEIN (Ph.D., Electrical and Computer Engineering)
A MULTI-MODAL APPROACH FOR FACE MODELING AND RECOGNITION (December 2007)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Dr. Mohamed Abdel-Mottaleb.
No. of pages in text 33.

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 when 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 performance 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 Ph.D study.

More than everyone, I indebted to my mother for her enthusiastic encouragement, prayers and unlimited support during all stages of my life. Also, I must express my appreciation to my beloved wife, Eshrat, my brothers and 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 and Ms. Rosamund Coutts 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. Omaima Nomir, Dr. Charay Leurdwick, Hamed Pirsiavash, 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 thesis.

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.

Contents

List of Figures

List of Tables

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

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.

1.1 3-D Face Matching Based on Ridge Images

Figure 1.1 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.

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

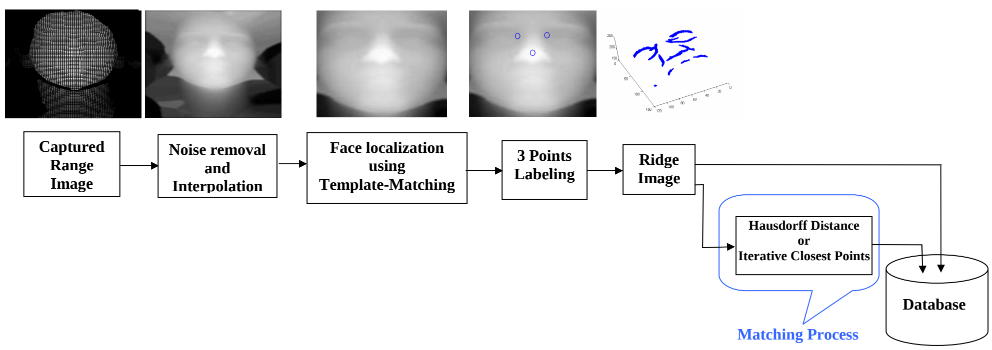

Figure 1.1: *Block diagram of our system for 3-D face recognition from range data.*

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.

1.2 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 [2], 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 called valleys when the ridges are looked from the opposite side.

There are different approaches to locate the ridges [9] 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 1.2 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.

1.2.1 Ridge Matching Using Robust Hausdorff Distance

Huttenlocher *et al.* originally proposed Hausdorff distance (HD) [6] 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 [7].

Given two sets of points $\mathcal{A} = \{a_1, a_2, ..., a_{N_\mathcal{A}}\}$ and $\mathcal{B} = \{b_1, b_2, ..., b_{N_\mathcal{B}}\}$ 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:

$$H(\mathcal{A}, \mathcal{B}) = max(h_K(\mathcal{A}, \mathcal{B}), h_K(\mathcal{B}, \mathcal{A})) \tag{1.1}$$

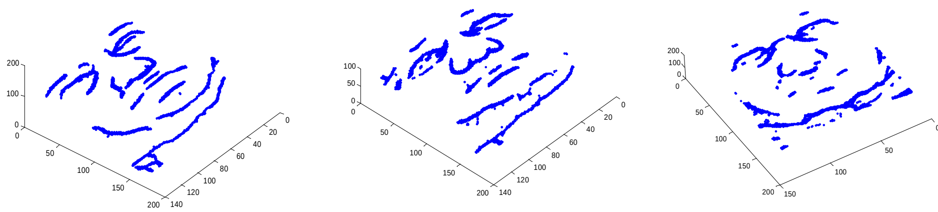

Figure 1.2: *Sample of ridge image extracted for different subjects.*

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:

$$h_K(\mathcal{A},\mathcal{B}) = K^{th}_{a\epsilon\mathcal{A}} d_{\mathcal{B}}(a), \quad h_K(\mathcal{B},\mathcal{A}) = K^{th}_{b\epsilon\mathcal{B}} d_{\mathcal{A}}(b) \tag{1.2}$$

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 $\mathcal{A}$, $K^{th}_{a\epsilon\mathcal{A}}$ denotes the K^{th} ranked value of $d_{\mathcal{B}}(a)$, and $K^{th}_{b\epsilon\mathcal{B}}$ 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 Hausdorrf distance measures based on their behavior in the presence of noise [4]. They proposed the modified Haussdorff 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 [18]. Based on the experimental matching performance of these different HD measures, robust LTS-HD based on the least trimmed square (LTS) measure [18] is adopted in our work. In the proposed LTS-HD [18], the directed distance $h_{LTS}(\mathcal{A},\mathcal{B})$ is defined by a linear combination of order statistics:

$$h_{LTS} = \frac{1}{H} \sum_{i=1}^{H} d_{\mathcal{B}}(a)_{(i)} \tag{1.3}$$

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 $d_{\mathcal{B}}(x)_{(1)} \leq d_{\mathcal{B}}(x)_{(2)} \leq \cdots \leq d_{\mathcal{B}}(x)_{(N_{\mathcal{A}})}$. 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:

$$\arg\min_{\alpha,\beta,\gamma,t_x,t_y,t_z,s} h_{LTS}(Tr(\mathcal{A}), \mathcal{B}) \tag{1.4}$$

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 1.1.

We used the Matlab optimization toolbox, i.e., *fminsearch* Matlab function, to solve this problem. The *fminsearch* uses the simplex search method of [8]. 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

(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 [5].
(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:

$$D_{P'}(x) = min_{y \in P'} ||x - y|| \tag{1.5}$$

where the $||.||$ denotes the $L2$ norm.
(5) Sort the minimum calculated distances and then calculate the robust Hausdorff distance, h_{LTS}, using Eq. 1.3.
(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 1.1: *Iterative algorithm to find the optimum pose in Hausdorff distance matching.*

matching distance, is considered the best match.

1.2.2 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 [16]. Instead of using random sampling of the feature points as in [16], 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 a 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 [5] 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.

1.2.3 Computational Complexity of Ridge Matching

Compared to other 3-D matching approaches for face recognition such as [10, 17, 3], 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 robust Hausdorff distance and the ICP 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. By employing the K-D tree for searching the closest points, the computational complexity is reduced to $O(log(P))$. So, by using the ridge lines, only 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., 0.14 ± 0.02 in our work, then the computational complexity reduces to $\rho^2 O(PQ)$ for the regular scheme and to $O(log(P) - log(1/\rho))$ for the accelerated scheme. This means that by using the ridge points, the computational complexity of matching is reduced two orders of magnitude.

1.3 Experiments and Results

We use the Gavab database [11] and the FRGC V2.0 [15] 3-D face databases for our experiments. In the following subsections we review these two databases and present our experiments and results.

1.3.1 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 1^{st} and the 2^{nd} captures), the two neutral looking up and down images (the 5^{th} and the 6^{th} captures), the frontal images with smile expression (the 7^{th} capture), the frontal images with laughing expression (the 8^{th} capture), and a frontal image with random gesture (the 9^{th} capture). The images in the 2^{nd} capture are used as gallery images and the images in the $1^{st}, 3^{rd}$, 4^{th}, and $7-9^{th}$ captures are used as the probe images for recognition.

Figure 1.3 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 1.2 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 3^{rd} dimension) and the recognition process was tested. By ignoring the 3^{rd} dimension, we obtained rank-one identification rate of

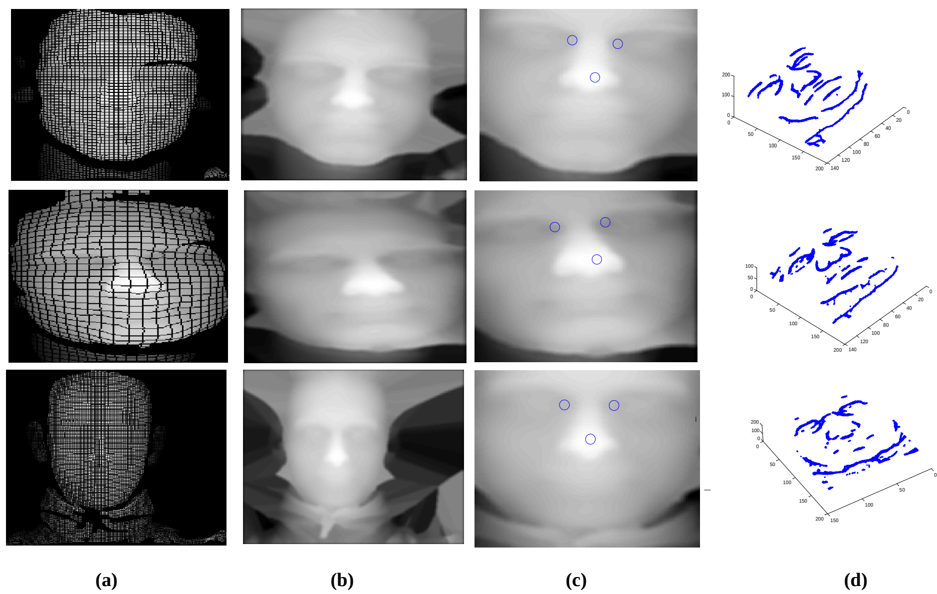

Figure 1.3: *Samples of range images in the gallery and the results of preprocessing (a) Original range images, (b) Noise removal and interpolation, (c) Face localization and three feature points labeling, (d) Ridge images*

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

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 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 1.4 and 1.5 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

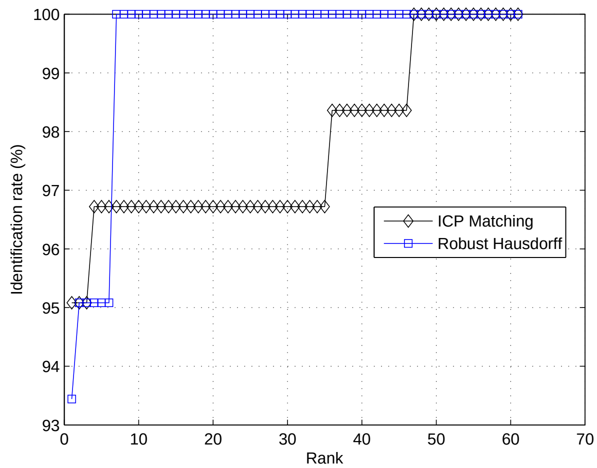

Figure 1.4: *CMC curves for frontal neutral images in Gavab database based on ridge images matched using the ICP and Hausdorff distance techniques.*

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 [14, 13, 12] based on the Gavab dataset. In [14], 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 [13], they selected a set of 30 features out the 86 features and obtained recognition rates of 82.0% and 90.16%, when the images

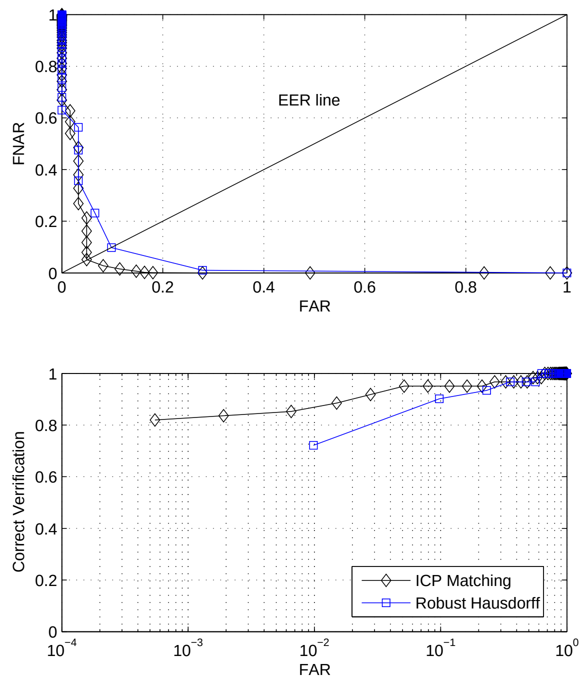

Figure 1.5: *ROC curves for frontal neutral images in Gavab database based on ridge images matched using the ICP and Hausdorff distance techniques.*

Approach	**Neutral-Neutral%**	**Neutral-Pose/Expr. %**	**Technique**
Moreno *et al.* [14]	78.0	62.0	35 features(e.g., areas of subregions)
Moreno *et al.* [13]			
	82.0	76.2	30 features and PCA
	90.16	77.9	30 features and SVM
Moreno *et al.* [12]	90.16	77.9	3-D volxels
Ansari [1]	90.16	-	3-D Mesh
Our results	93.5	75.4/82.0	Robust Hausdorff distance
	95.0	88.6/83.6	Iterative Closest Points

Table 1.3: *Comparing the results by Moreno et al. [14, 13, 12] and Ansari [1] on Gavab database and our work.*

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 [12], 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 [1] 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 1.3 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.

1.3.2 Experiments on FRGC V2.0 Face Database

The FRGC V2.0 database [15] 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 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 1.6 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 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

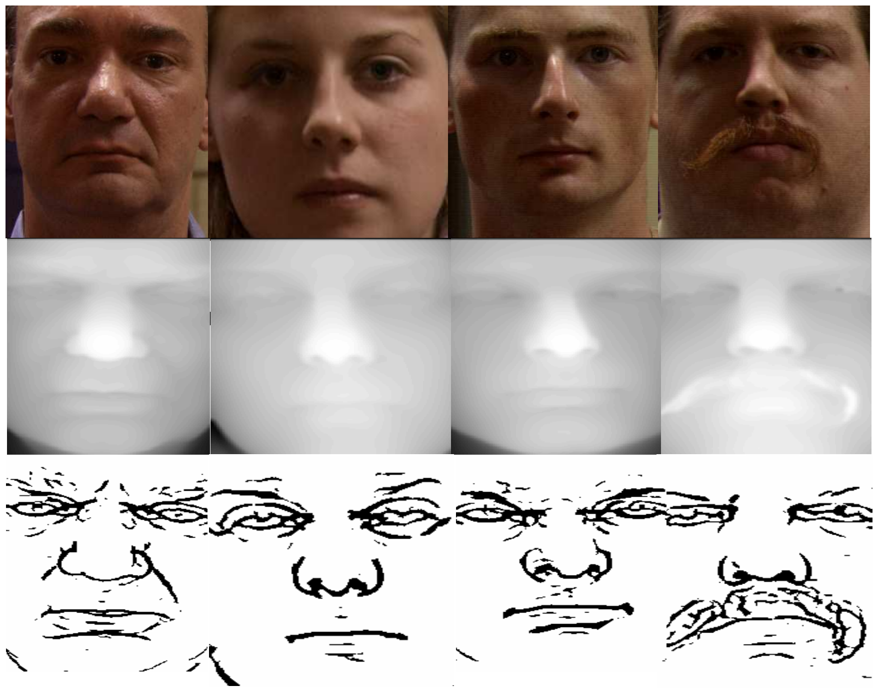

Figure 1.6: *Samples of facial images in the FRGC V2.0 database (texture, range, and extracted ridge images.)*

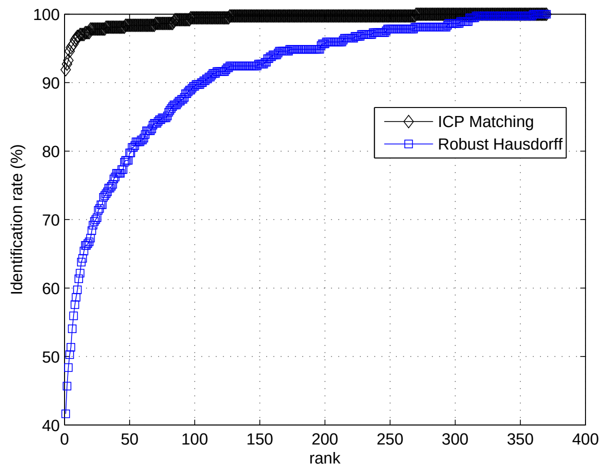

Figure 1.7: *CMC curve for 370 subjects in the FRGC V2.0 database based on ridge images matched with the ICP and Hausdorff techniques.*

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 [19], where they compared ICP and Hausdorff for ear surface matching: The ICP outperforms the Hausdorff distance for shape matching.

1.3.3 Comparison Between 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 random selection of

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 1.4: *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 (%).*

the points from the entire surface of the face and used the ICP method for matching. A set of random points was selected from the range images such that the x and y coordinates are randomly selected using a random generator with uniform distribution and the z value of each point comes 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 used all the points on the face surface for matching based on the ICP technique. The results of these experiments are summarized in table 1.4 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 and our experiments show that ridge points are very promising in surface matching. In other words, ridge points have good performance rate while reducing the computational complexity of matching two orders of magnitude.

1.3.4 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 1.8 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 1.5 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 1.9). 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 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

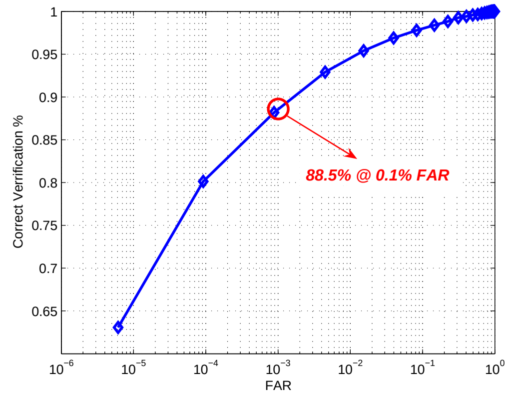

Figure 1.8: *ROC curve for the neutral faces versus neutral in FRGC2.0 database (total of 2365 facial images for 432 individual subjects); ICP technique is used for matching the ridge images.*

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

Table 1.5: *Verification rates (%) at 0.1% FAR for the ROC I, II, and III of the neutral v.s. neutral images.*

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.

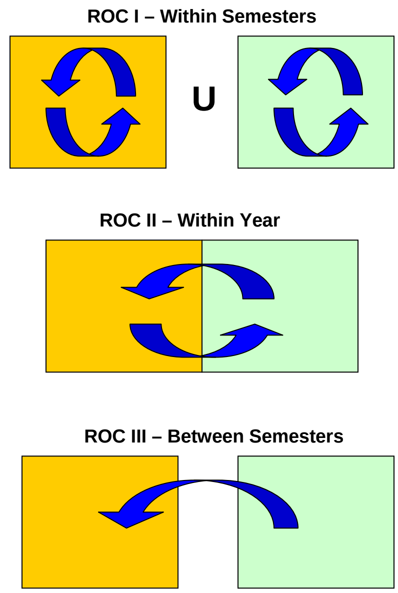

Figure 1.9: *ROC I: the data are within the semesters. ROC II: the data are within the year. ROC III: the images are between the semesters.*

1.4 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 2

Conclusion and Future Research Direction

2.1 Summary and Contributions

In this dissertation, we have introduced a multi-modal scheme, i.e., 2-D (texture) + 3-D (shape), for face modeling and recognition. Recently, with advancements in computer vision technology and capturing devices, scanning or constructing a 3-D dense model of the face is easy. 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 face recognition such as pose variations and illumination variations can be easily handled. On the other hand, 2-D face recognition algorithms have been investigated for few decades and is in a mature stage. However, the 2-D and the 3-D data have their own limitations and we believe that only a multi-modal scheme can provide a 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 points extraction, 3-D face 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 algorithms for 2-D and 3-D face recognition. For 3-D face recognition, we have developed a novel algorithm based on ridge images in 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 such that the nodes of the graph represent a set of landmark points (extracted by an improved active shape model technique developed in Chapter three and Appendix A) and the edges are defied based on 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 the 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 four) to handle the 3-D face recognition.

Then, we 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 an algorithm 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. The main difficulty with extracting facial landmarks from 3-D data is the lack of texture. We used the extracted facial features for the initial alignment of the ridge images during the matching process.

- We have developed an algorithm for 3-D face modeling and recognition based upon 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 an algorithm for multi-modal (3-D + 2-D) face recognition based on attributed relational graphs (ARG). The nodes of the graph represent the locations of the facial landmark points. 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 features that define the geometric relations between the edges of the graph are extracted and used in the representation to improve the performance rate of face recognition.

- 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 thesis, we have achieved promising results using the developed techniques.

2.2 Future Work

Face recognition is an important area of research that is continuously progressing. Possible future developments, expansions, and improvements of our presented algorithms in this thesis are as follows:

- Improving the developed method for 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 technique 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, there is another set of ravine points that can be extracted using the k_{min} principal curvature. It will be interesting to extract and use these points either alone or in conjunction with the other ridge points (extracted by thresholding the k_{max} principal curvature) for 3-D face recognition.

Bibliography

[1] 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.

[2] 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 Comtputer Software and Applications Conference*, 2004.

[3] 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.

[4] M. Dubisson and A. K. Jain. A modified hausdorff distance for object matching. In *12th Intertaional Conference Pattern Recognition (ICPR), Jerusalem, Israel*, Oct. 1994.

[5] 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.

[6] 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.

[7] D. P. Huttenlocher, R. Lilien, 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] A. Belen Moreno, Angel Sanchez, Jose Fco. Velez, and Fco. Javier Dkaz. 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.

[14] Ana Belen Moreno, Angel Sanchez, Jose Fco. Velez, and Fco. Dkaz. Face recognition using 3d surface-extracted descriptors. In *Irish Machine Vision and Image Processing Conference 2003 (IMVIP'03)*, September 2003.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] P. Yan and K.W. Bowyer. Empirical evaluation of advanced ear biometrics. pages III: 41–41, 2005.

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