

Copyright

by

Shalini Gupta

2008

The Dissertation Committee for Shalini Gupta
certifies that this is the approved version of the following dissertation:

**NOVEL ALGORITHMS FOR 3D HUMAN FACE
RECOGNITION**

Committee:

Alan C. Bovik, Supervisor

Mia K. Markey, Supervisor

Ross Baldick

Joydeep Ghosh

Donald S. Fussell

Kenneth R. Castleman

**NOVEL ALGORITHMS FOR 3D HUMAN FACE
RECOGNITION**

by

Shalini Gupta, B.E.; M.S.E.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2008

Dedicated to my parents and to Paul.

Acknowledgments

Firstly, I humbly thank the higher power that *always* guided me. I thank my parents for their unshakable belief in my abilities. I am grateful to them for all their sacrifices, for giving me the best education possible, and for the values that they taught me. Despite the thousands of miles between us, I would never have accomplished this milestone without their love, care and support.

I am grateful to my dissertation advisors Dr. Alan C. Bovik and Dr. Mia K. Markey for their technical guidance, and financial and moral support. I feel very fortunate to have had such great advisors and will always hold in my heart the deepest regard for both of them. Mia, both as a female mentor and an advisor has unarguably been a great positive influence in my engineering career. Her professionalism, organizational skills, diligence, and positive drive have been an inspiration. I credit her for any professional maturity that I may have acquired as a graduate student. I am also very grateful to Dr. Bovik for his technical guidance and prompt support on all administrative matters. His astute yet encouraging and compassionate comments have been invaluable for my research.

I thank all members of my dissertation committee Dr. Baldick, Dr. Ghosh, Dr. Fussell and Dr. Castleman for their technical support. I especially acknowledge Dr. Castleman and former Advanced Digital Imaging Research, LLC, a Division of Iris International, Inc., for providing 3D facial data and financial support for this project. This work was supported in part by the Advanced Technology Program

of the National Institute of Standards and Technology. A word of acknowledgment is also in order for Dr. Aggarwal of the ECE department and Dr. Hamrick of the Mathematics department, with whom I shared numerous insightful technical discussions.

I am grateful to Paul for standing by me through all the trials and tribulations of graduate student life, and for telling me to ‘hang on’ whenever I had the urge to abandon my dissertation. I am also grateful to him for helping me with the mundane chores of my daily life and for providing to me a carefree and supportive environment outside of work. I acknowledge my sister Madhuri, who in her own unique way supported my academic endeavors. I am grateful to my friends Bhavya, Simran, Aradhana, Mike and Ravi, and former room mates Urmila and Jyotirmoy, all of whom wished me well and brought moments of cheer to my life.

I will cherish for life the bonds that I have formed with my present and former colleagues at LIVE and BMIL. I feel fortunate to have had kind and knowledgeable senior graduate mentors like Mehul, Jin and Umesh. I will also never forget the wonderful moments of laughter that I shared with Wendy, which brightened up my days and nights at BMIL. I also acknowledge the support of all system administrators at BMIL who kept the lab machines slick and running. Lastly, I wish to thank numerous people at the BME department at UT Austin for their smiles and kind words, and especially Jack Hart and Dr. Zaman for their help and encouragement.

SHALINI GUPTA

The University of Texas at Austin

August 2008

NOVEL ALGORITHMS FOR 3D HUMAN FACE RECOGNITION

Publication No. _____

Shalini Gupta, Ph.D.

The University of Texas at Austin, 2008

Supervisor: Alan C. Bovik

Co-supervisor: Mia K. Markey

Automated human face recognition is a computer vision problem of considerable practical significance. Existing two dimensional (2D) face recognition techniques perform poorly for faces with uncontrolled poses, lighting and facial expressions. Face recognition technology based on three dimensional (3D) facial models is now emerging. Geometric facial models can be easily corrected for pose variations. They are illumination invariant, and provide structural information about the facial surface. Algorithms for 3D face recognition exist, however the area is far from being a matured technology. In this dissertation we address a number of open questions in the area of 3D human face recognition. Firstly, we make available to qualified

researchers in the field, at no cost, a large Texas 3D Face Recognition Database, which was acquired as a part of this research work. This database contains 1149 2D and 3D images of 118 subjects. We also provide 25 manually located facial fiducial points on each face in this database. Our next contribution is the development of a completely automatic novel 3D face recognition algorithm, which employs discriminatory anthropometric distances between carefully selected local facial features. This algorithm neither uses general purpose pattern recognition approaches, nor does it directly extend 2D face recognition techniques to the 3D domain. Instead, it is based on an understanding of the structurally diverse characteristics of human faces, which we isolate from the scientific discipline of facial anthropometry. We demonstrate the effectiveness and superior performance of the proposed algorithm, relative to existing benchmark 3D face recognition algorithms. A related contribution is the development of highly accurate and reliable 2D+3D algorithms for automatically detecting 10 anthropometric facial fiducial points. While developing these algorithms, we identify unique structural/textural properties associated with the facial fiducial points. Furthermore, unlike previous algorithms for detecting facial fiducial points, we systematically evaluate our algorithms against manually located facial fiducial points on a large database of images. Our third contribution is the development of an effective algorithm for computing the structural dissimilarity of 3D facial surfaces, which uses a recently developed image similarity index called the complex-wavelet structural similarity index. This algorithm is unique in that unlike existing approaches, it does not require that the facial surfaces be finely registered before they are compared. Furthermore, it is nearly an order of magnitude more accurate than existing facial surface matching based approaches. Finally, we propose a simple method to combine the two new 3D face recognition algorithms that we developed, resulting in a 3D face recognition algorithm that is competitive with the existing state-of-the-art algorithms.

Contents

Acknowledgments	v
Abstract	vii
List of Tables	xii
List of Figures	xv
Chapter 1 Introduction	1
Chapter 2 Background	6
2.1 3D Facial Models	6
2.1.1 Acquisition	6
2.1.2 Representation	7
2.2 Face Recognition Tasks	8
2.2.1 Verification	8
2.2.2 Identification	10
2.3 Existing 3D Face Recognition Algorithms	11
2.3.1 Appearance Based Techniques	12
2.3.2 Facial Surface Matching	14

2.3.3	Local Facial Features	16
2.3.4	Ensemble Approaches	17
2.3.5	Expression Invariant Approaches	18
2.3.6	FRCG 2005 Algorithms	19
Chapter 3 The Texas 3D Face Recognition Database		21
3.1	Acquisition and Normalization	23
3.2	Preprocessing	25
3.3	Manual Fiducial Point Detection	26
3.4	Data Partitions	27
3.5	Availability	29
Chapter 4 3D Anthropometric Face Recognition		33
4.1	3D Anthropometric Face Recognition (AnthroFace)	35
4.1.1	Anthropometric Cranio-Facial Proportions	35
4.1.2	Manual Detection of Anthropometric Fiducial Points	37
4.1.3	Classifier	38
4.1.4	Effect of Choice of Facial Fiducial Points	39
4.1.5	Reduction of Anthropometric Facial Fiducial Points	40
4.1.6	Automatic Detection of Anthropometric Fiducial Points	41
4.2	Benchmark Algorithms	52
4.3	Performance Evaluation	53
4.3.1	Data	53
4.3.2	Automatic Fiducial Point Detection	54
4.3.3	3D Face Recognition	55
4.4	Results and Discussion	55

4.4.1	Manual 3D AnthroFace	55
4.4.2	Automatic 3D AnthroFace	61
Chapter 5 3D Face Recognition by Assessing Structural Similarity		67
5.1	Existing Approaches	68
5.1.1	Iterative Closest Point Algorithm	68
5.1.2	Existing Distance Metrics	70
5.2	3D Structural Similarity Face Recognition (SSIMFace)	74
5.2.1	Complex-wavelet Structural Similarity Index	74
5.2.2	3D Facial Representations	79
5.2.3	Recognition Algorithm	84
5.3	Benchmark Algorithms	86
5.4	Performance Evaluation	88
5.4.1	Data	88
5.4.2	3D Face Recognition	89
5.5	Results and Discussion	89
5.5.1	Discriminatory Facial Regions	89
5.5.2	3D SSIMFace	91
5.5.3	Before and After ICP	92
5.5.4	Analysis of Geometric Transformations	95
Chapter 6 Combined Algorithm		99
Chapter 7 Conclusion		104
Bibliography		107
Vita		130

List of Tables

2.1	A summary of the five 3D face recognition algorithms that were evaluated at the Face Recognition Grand Challenge 2005 on the FRGC v0.1 database.	19
3.1	A summary of the data partitions employed for developing 3D face recognition algorithms.	28
4.1	The 23 most variable anthropometric facial proportions for adult humans along with their standard deviation values [1]. The corresponding fiducial points are presented in Fig. 4.1(b). N denotes nasal proportions, O denotes orbital proportions, L denotes proportions related to the the mouth region, and F denotes facial proportions. . .	36
4.2	A summary of the data partitions employed for developing 3D face recognition algorithms.	54
4.3	The observed (a) EER, (b) AUC, and (c) rank 1 RR values and their 95% confidence intervals for the eigensurfaces, fishersurfaces, ICP, and the 3D AnthroFace algorithms based on 25 manually located anthropometric points and 25 arbitrary facial points.	56

4.4	The observed (a) EER, (b) AUC, and (c) rank 1 RR values and their 95% confidence intervals for the 3D AnthroFace algorithms based on 25 manually located Fiducial points, 10 manually located fiducial points, and 10 automatically located fiducial points.	61
4.5	The standard deviations of the positional errors of the 10 automatically located anthropometric facial fiducial points, in pixel units. The distance between two adjacent pixels was 0.32 mm.	62
5.1	A summary of the data partitions employed for training and testing 3D face recognition algorithms.	88
5.2	The observed EER, AUC, and rank 1 RR values and their 95% confidence intervals for the eyes, nose and mouth regions of the facial range images, wherein the facial sub-regions were compared using the CW-SSIM index.	90
5.3	The observed (a) EER, (b) AUC, and (c) rank 1 RR values and their 95% confidence intervals for the proposed 3D SSIMFace algorithm, wherein coarsely aligned 3D faces were compared using the CW-SSIM index, and for the benchmark eigensurfaces, fishersurfaces, and ICP algorithms.	92
5.4	The observed EER, AUC, and rank 1 RR values and their 95% confidence intervals for comparing facial surface representations with the CW-SSIM, MSE , MSE_{CP} , and pH indices, before and after 3D faces were finely registered using the ICP algorithm. The performance statistics are for all faces (neutral and expressive).	94

5.5	The maximum distortions for each of the similarity indices/3D facial representations, that resulted in scores between a 3D face and its transformed version that were less than the operating point of the corresponding indices' 3D face recognition system.	97
6.1	The observed (a) EER, (b) AUC, and (c) rank 1 RR values and their 95% confidence intervals for the 3D AnthroFace, 3D SSIMFace and a combination of both these algorithms.	100
6.2	A summary of the five 3D face recognition algorithms that were evaluated at the Face Recognition Grand Challenge 2005 on the FRGC v0.1 database, and of our final proposed 3D face recognition algorithm, which was a combination of the 3D AnthroFace and 3D SSIM-Face algorithms.	102

List of Figures

2.1	A (a) 3D face point cloud, (b) 3D triangular mesh representation, and (c) face range image/depth map.	8
2.2	Schematic diagrams for automatic face (a) verification, and (b) identification systems.	9
2.3	A typical (a) Receiver Operating Characteristic curve, and (b) a Cumulative Match Characteristic curve.	10
3.1	Color images of faces in the Texas 3D Face Recognition Database in (a) neutral, and (b) expressive modes.	22
3.2	Raw (a) color, and (b) range images of the Texas 3D Face Recognition Database.	23
3.3	Impuse noise on the surface of 3D models.	24
3.4	Preprocessed (a) color, and (b) range images of the Texas 3D Face Recognition Database.	25
3.5	Twenty-five anthropometric fiducial points (a) on a color image, and (b) on a range image.	26
3.6	The graphical user interface that was developed for manually locating anthropometric facial fiducial points.	27

3.7	Examples of (a) color images, and (b) range images from the FRGC 2005 data set.	31
4.1	The 25 facial fiducial points associated with highly variable anthropometric facial proportions on a (a) color image, and (b) on a range image, and (c) a set of 25 arbitrary facial points.	35
4.2	The reduced set of 10 anthropometric facial fiducial points for the AnthroFace algorithm on a (a) color image, and (b) range image. . .	40
4.3	An example of a 3D face in a canonical pose that was used to automatically detect facial fiducial points.	42
4.4	The 3D template face and its manually located nose tip, which was used to automatically locate the nose tips of all 3D faces.	44
4.5	The magnitudes of the Gaussian curvatures of the (a) elliptic, (b) convex elliptic regions of a facial range image, and (c) of its central 301×301 pixel region. In each image, the Gaussian curvature has been plotted as a 3D surface with the facial texture warped onto it.	45
4.6	(a) The edges detected with the LOG edge detector in sub-regions of the facial range images centered about the detected nose tip, and (b) critical points along the nose boundary and the tip of the nose. . . .	46
4.7	(a) The magnitude of the Gaussian curvature of the concave elliptic regions of a facial range image, (b) that of a smaller region defined to search for the right eye's inner corner, and (c) that of a smaller region defined to search for the left eye's inner corner. In each image the surface curvature is plotted as a 3D surface, with the facial texture warped onto it.	48

4.8	(a) The the mean curvature (H) of the mouth region of an example face plotted as a 3D surface, with the facial texture warped onto it. Notable are the distinct peaks at the corners of the mouth. (b) Regions below the nose with elliptic Gaussian curvature (non-black regions) of an example face. Notable are the regions of the upper and the lower lip.	50
4.9	The sub-regions of the facial range images that were employed for the eigensurfaces and fishersurfaces 3D face recognition algorithms. . . .	51
4.10	(a) The semi-log CMC curves, and (b) the ROC curves for the performance of the benchmark 3D face recognition algorithms, the AnthroFace algorithm that employed manually located 25 points, and the 3D AnthroFace algorithm that employed 10 automatically located points.	57
4.11	(a) The semi-log CMC curves, and (b) ROC curves for the performance of the 3D AnthroFace algorithm based on 25 manually located facial fiducial points and the algorithm based on 25 arbitrary facial points.	58
4.12	The 20 most discriminatory facial (a) Euclidean, and (b) geodesic distance features. The geodesic distances are symbolically depicted by straight lines. In reality, they are along the surface of the face. . .	59
4.13	(a) The semi-log CMC curves, and (b) ROC curves for the performance of 3D AnthroFace algorithm that employed only Euclidean distances and the 3D AnthroFace algorithm that employed both the Euclidean and the geodesic distances. These curves were for expressive faces only.	60

4.14	The 2D histograms for the errors for the two stages of the detection algorithms for (a) the tip of the nose, (b) the inner corner of the left eye (2 left images) and inner corner of the right eye (2 right images), (c) the outer corner of the left eye (2 left images) and outer corner of the right eye (2 right images), and (d) the left (2 left images) and right (2 right images) corners of the mouth.	66
5.1	An original 3D facial point cloud A and its rotated version B (a) before, and (b) after fine alignment using the ICP algorithm.	68
5.2	Examples of facial range images in the Texas 3D Face Recognition Database.	79
5.3	Examples of the eyes, nose and mouth regions, respectively (from top to bottom) of two different faces.	80
5.4	The 10 automatically detected anthropometric facial fiducial points.	81
5.5	The sub-sections of the facial range images, which contained only the nose and the eyes regions, and which were employed for the 3D SSIMFace algorithm.	82
5.6	The facial edge images that were employed for the proposed 3D SSIM-Face algorithm.	83
5.7	From left to right are shown the local (a) range, and (b) edge images of size 257×257 pixels centered about the anthropometric fiducial points <i>en-en</i> , <i>prn</i> , <i>al-al</i> , respectively.	84
5.8	The sub-regions of the facial range images that were employed for the eigensurfaces and fishersurfaces 3D face recognition algorithms.	87

5.9	(a) The semi-log CMC curves, and (b) the ROC curves for the eyes, nose and mouth regions of the facial range images, wherein the facial sub-regions were compared using the CW-SSIM index.	90
5.10	(a) The semi-log CMC curves, and (b) the ROC curves for the proposed 3D SSIMFace algorithm, wherein coarsely aligned 3D faces were compared using the CW-SSIM index, and for the benchmark eigensurfaces, fishersurfaces, and ICP algorithms.	93
5.11	The right handed 3D co-ordinate system.	95
5.12	The (a) range images, and (b) edge images (from left to right) of an original 3D face, and its transformed version for -15° rotation about the z axis, -15° rotation about the y axis, -15° rotation about the x axis, scaling of 0.75 times its original magnification, and translation of $r = 34$ mm and $\theta = \pi/4$ in the (xy) or $(r\theta)$ plane.	96
6.1	(a) The semi-log CMC curves, and (b) the ROC curves for the 3D AnthroFace, 3D SSIMFace algorithms, and a combination of both these algorithms.	101

Chapter 1

Introduction

Automated human face recognition is a non-trivial computer vision problem of considerable practical significance. It has numerous applications including automated secured access to ATM machines and buildings, automatic surveillance, forensic analysis, fast retrieval of records from databases in police departments, automatic identification of patients in hospitals, checking for fraud or identity theft, and human-computer interaction. Currently, popular techniques for automated person identification include the use of personal identification numbers, access codes or cards, bar codes, and radio frequency identification tags. All of these are susceptible to loss or theft. Personal identification numbers and access codes/cards also require substantial involvement of the human subject. Hence, they are of limited utility for identifying very young children, or seriously ill or deceased persons.

A *biometric* is defined as an anatomical or physiological measurement, or a behavioral characteristic believed to be unique to each individual. For nearly a hundred years, manual approaches based on biometrics have been employed to identify humans [2]. With the availability of superior computing power, automatic

biometric techniques for human identification have emerged. These include techniques based on the recognition of the face, the iris and retina of the eye, fingerprints, palmprints, gait, voice, and handwriting. Of these, techniques based on iris and fingerprint recognition are reported to be highly accurate [3], but they require substantial cooperation from the human subjects. They are difficult to deploy in real-time screening and surveillance applications, where minimal interaction with humans is desirable, or where the system is to be operated covertly. Automated face recognition systems, on the other hand, require less co-operation from human subjects, relative to iris and fingerprint recognition systems. They are amenable to surveillance applications, and can be developed using relatively low cost components. Hence, considerable attention has been directed, over the past several decades, towards developing reliable automatic face recognition systems.

Besides the need for automation, interest in computer algorithms for face recognition is inspired by the need to develop objective measures of facial similarity [4]. Such measures are required, for example, to construct effective lineups, of individuals similar in appearance to that of an offender, in police departments [5]. Humans alone cannot be relied upon to perform such tasks accurately, since some deficiencies in the face recognition abilities of humans are known to exist. For example, a study of DNA exonerations reported that 84% of wrongful convictions were due in part to false recognition by eyewitnesses or victims [6]. It is known that the face recognition abilities of humans are influenced by cross-racial effects and other socio-economic biases [7].

Considerable research attention has been directed, over the past few decades, towards developing reliable automatic face recognition systems that use two dimensional (2D) facial images. Commercial systems are also now available for 2D face

recognition [8]. Two dimensional face recognition systems are easy and economical to construct, but are inadequate for robust face recognition. A Face Recognition Vendors Test was conducted in the year 2002 (FRVT 2002) to assess the performance of fully automatic 2D face recognition algorithms at the time [9]. It was found that the performance of the three best algorithms dropped nearly in half for facial images with uncontrolled illumination or poses. At the same event, researchers identified that using three dimensional (3D) facial morphable models [10] to synthetically generate frontal 2D facial images could be a promising solution to the facial pose problem.

Three dimensional face recognition technology is now emerging, in part, due to the availability of improved 3D imaging devices and processing algorithms. For such techniques, 3D images of the facial surface are acquired using 3D acquisition devices. Three dimensional facial images have some advantages over 2D facial images. Their pose can be easily corrected by rigid rotations in 3D space. The shape of a 3D facial surface depends on its underlying anatomical structure. Hence, images acquired using 3D laser range finders are invariant to illumination conditions during image acquisition. Moreover, when 3D facial models are acquired along with 2D facial images, information about the direction of facial surface normals obtained from the 3D facial images can be used to correct the illumination of the 2D facial images [11]. Three dimensional facial images also provide structural information about the face (*e.g.*, surface curvature and geodesic distances), which cannot be obtained from a single 2D image.

The existing 3D face recognition algorithms can be broadly classified into three groups [12]. First, there are algorithms that are based on the *appearance* of facial range images (hereafter ‘appearance based techniques’). Then, there are

those that rigidly align and compare 3D facial surfaces (hereafter ‘surface matching based’). Finally, there are those that employ structural properties of local facial features (hereafter ‘local feature based’). The existing appearance based techniques for 3D face recognition are straightforward extensions of successful 2D face recognition techniques, *e.g.*, ‘eigenfaces’ and ‘fisherfaces’ applied to facial range images. While these have been successful to a degree at the task of 3D face recognition, intuitively understanding the discriminatory facial structural information that they encode remains an open problem [13]. Current techniques that rigidly align facial surfaces and compare them are generally computationally expensive. Despite their success with 2D face recognition [9], 3D techniques based on local facial features are poorly developed. This is due to a lack of understanding of the discriminatory structural characteristics of facial surfaces, and because of poorly developed algorithms for reliably detecting the 3D local facial features.

In this dissertation we address a number of open questions in the area of 3D human face recognition. Firstly, we make available to qualified researchers in the field, at no cost, a large Texas 3D Face Recognition Database, which was acquired as a part of this research work. This database contains 1149 2D and 3D images of 118 subjects. We also provide 25 manually located facial fiducial points on each face in this database.

Our next contribution is the development of a completely automatic novel 3D face recognition algorithm, which employs discriminatory anthropometric distances between carefully selected local facial features. This algorithm neither uses general purpose pattern recognition approaches, nor does it directly extend 2D face recognition techniques to the 3D domain. Instead, it is based on an understanding of the structurally diverse characteristics of human faces, which we isolate from the

scientific discipline of facial anthropometry [1, 14, 15]. We demonstrate the effectiveness and superior performance of the proposed algorithm, relative to existing benchmark 3D face recognition algorithms. A related contribution is the development of highly accurate and reliable 2D+3D algorithms for automatically detecting 10 anthropometric facial fiducial points. While developing these algorithms, we identify unique structural/textural properties associated with the facial fiducial points. Furthermore, unlike previous algorithms for detecting facial fiducial points, we systematically evaluate our algorithms against manually located facial fiducial points on a large database of images.

Our third contribution is the development of an effective algorithm for computing the structural dissimilarity between 3D facial surfaces, which uses a recently developed image similarity index called the complex-wavelet structural similarity index [16, 17]. This algorithm is unique in that unlike existing approaches, it does not require that the facial surfaces be finely registered before they are compared. Furthermore, it is nearly an order of magnitude more accurate than existing facial surface matching based approaches.

In Chapter 2 of this dissertation, we present some background on the problem of 3D face recognition and review existing techniques. In Chapter 3, we describe in detail the Texas 3D Face Recognition Database. In Chapters 4 and 5, we describe the two novel algorithms for 3D face recognition that we developed. In Chapter 6 we compare these two new algorithms and explain how they can be combined into a single algorithm in a simple way, thus producing even better performance. We conclude this dissertation in Chapter 7 with ideas for extending the work presented in this dissertation.

Chapter 2

Background

2.1 3D Facial Models

2.1.1 Acquisition

Three dimensional facial models can be acquired using either active or passive techniques [18]. Among the active techniques, laser range finders are widely used [19, 20, 21, 22, 23, 24, 25, 26, 13, 27, 28]. Laser Range finders project light from a laser source onto a scene and record its reflection to determine the depth at each location. They produce dense and accurate 3D models that are invariant to ambient illumination conditions during image acquisition. However, they take longer to acquire 3D images than the passive techniques, during which human subjects must remain perfectly still [29]. Hence, they are unsuitable for high throughput screening applications. Another concern is the intrusive nature of laser light on the human eye.

The passive techniques employed to acquire 3D faces are stereo imaging [30, 31, 32, 33, 34, 35, 36, 33], and approaches based on structured light [37, 38, 39, 40, 41,

11, 42, 43, 44]. In stereo imaging systems, multiple cameras simultaneously image a face from different view points. The facial depth information is resolved using camera calibration parameters and the disparity between the different facial images. For the structured light approach, a standard light pattern (*e.g.*, a light stripe pattern) is projected onto a scene and its deformation is employed to determine the depth at each location. Unlike active techniques, both 2D and 3D images can be acquired simultaneously with passive techniques. Passive techniques are also faster, safer, and cheaper than laser range finders, but they typically contain more missing data.

Some researchers have raised concerns about the illumination invariant nature of 3D facial models acquired using passive techniques [45]. They argue that since intensity images are employed to construct 3D models in passive imaging techniques, variations in illumination can alter the shape of the constructed 3D model. However, a number of studies have dispelled these concerns and have shown that the performance of 3D face recognition algorithms that employ passively acquired 3D models is unaffected by variations in illumination conditions during image acquisition [46, 43, 11].

2.1.2 Representation

Three dimensional point clouds, triangulated surface meshes, and range images are employed to represent 3D facial surfaces. The point cloud representation contains the (x, y, z) coordinates of a set of random points on the facial surface (*e.g.*, Fig. 2.1(a)). These points can be connected to their nearest neighbors via straight lines resulting in a triangulated mesh representation (*e.g.*, Fig. 2.1(b)). The 3D points in these two representations are usually unstructured, and hence they require relatively

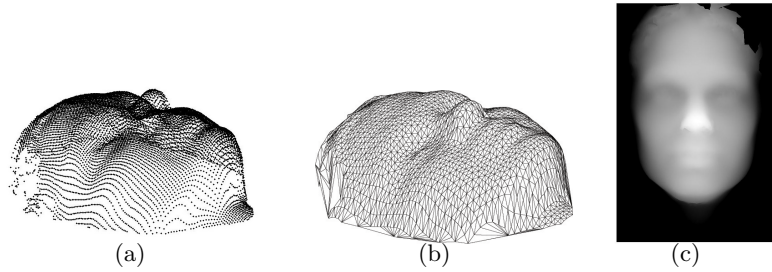


Figure 2.1: A (a) 3D face point cloud, (b) 3D triangular mesh representation, and (c) face range image/depth map.

involved processing algorithms.

A range image, also referred to as a 2.5D surface or depth map, consists of (x, y) points on a regular rectangular grid. Each (x, y) point is associated with a z value of the point on the 3D surface closest to the acquisition device (*e.g.*, Fig. 2.1(c)). Range images can be captured directly using laser range finders or can be generated from meshes or 3D point clouds by projecting them orthographically. As points in a range image are placed along a regular rectangular grid, they can be processed via relatively straightforward image processing algorithms. Range images can also be easily transformed into 3D point clouds or surface meshes.

2.2 Face Recognition Tasks

The two main tasks performed by an automatic human face recognition system are verification/authentication and identification [47].

2.2.1 Verification

Verification/authentication is a one-to-one matching task, wherein a person claims to be a specific entity known to the system (Fig. 2.2(a)). The database of people

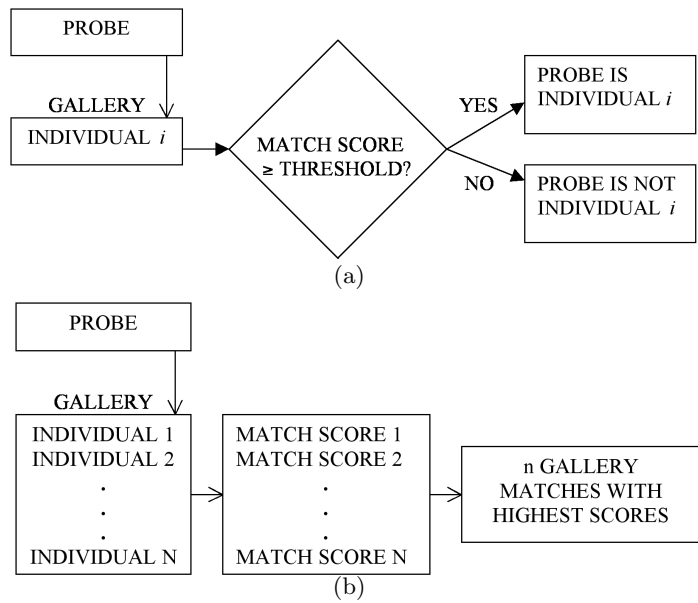


Figure 2.2: Schematic diagrams for automatic face (a) verification, and (b) identification systems.

known to the system is referred to as the ‘gallery’. The individual whose identity is verified/authenticated by the system is referred to as a ‘probe’. A facial representation of the probe is captured in real-time and compared against the gallery representation of the claimed entity. If the similarity score between the two is greater than a predefined threshold, the individual is verified as the claimed entity; otherwise, he or she is rejected as an imposter. An example of a verification scenario is a system for automated secured access to a building.

The performance of a verification system is evaluated in terms of a Receiver Operating Characteristic (ROC) curve [48]. An ROC curve is a plot of the false acceptance rate (FAR) versus the false rejection rate (FRR) (Fig. 2.3(a)). FAR is defined as the proportion of comparisons between two different individuals that are falsely accepted by the system [49]. FRR is the proportion of comparisons

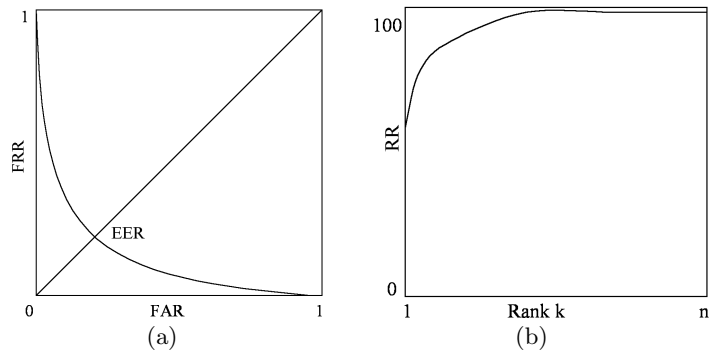


Figure 2.3: A typical (a) Receiver Operating Characteristic curve, and (b) a Cumulative Match Characteristic curve.

between two different images of the same individual that are falsely rejected by the system [49]. Both FAR and FRR vary as the system's decision threshold is varied [50]. A single performance metric typically reported for verification systems is the equal error rate (EER), where $FAR = FRR$ (Fig. 2.3(a)). For an ideal system $EER = 0\%$. The area under the ROC curve (AUC) is also sometimes reported as a measure of performance. It ranges from 0 for an ideal system to 0.5 for a system with chance performance. It should be noted that while the EER quantifies the verification performance of a 3D face recognition system at only *one* operating point, the AUC is an overall measure of the system's verification performance across the whole gamut of operating points. Parametric [49] or non-parametric methods [51], which are based on data sampling techniques [52], are employed to obtain statistical confidence intervals for these performance statistics.

2.2.2 Identification

Identification is a one-to-many matching task wherein an unknown individual's identity is established by comparing his/her probe face against a gallery of faces of known

individuals (Fig. 2.2(b)). The closest matches in the gallery are found and ranked in descending order of their similarity scores [9]. The probe is assigned the identity of its closest matched face in the gallery.

The performance of an identification system is evaluated in terms of a Cumulative Match Characteristic (CMC) curve [47, 53]. The CMC curve is a plot of the recognition rate (RR) versus the top ranked k database matches considered (Fig. 2.3(b)). The rank k RR value is the ratio of the number of probes, for which the correct gallery match is present among the top k matches, to the total number of probes that query the system. This formulation assumes a ‘closed universe’ model wherein all individuals that query the system are assumed to be present in the gallery. If the closed universe assumption is false, *i.e.*, some probes are not present in the gallery, the maximum RR achieved by the system is less than 100%. Non-parametric methods based on sampling techniques are generally employed to obtain statistical error bounds on the recognition rates [54, 55].

2.3 Existing 3D Face Recognition Algorithms

The existing 3D face recognition algorithms can be broadly classified into three groups [12]. First, there are algorithms that are based on the *appearance* of facial range images (‘appearance based techniques’). Then, there are those that rigidly align and compare 3D facial surfaces (‘surface matching based’). Finally, there are those that employ structural properties of local facial features (‘local feature based’). Before reviewing these existing techniques, we would like to point out that it is difficult to directly compare the performance of the existing 3D face recognition algorithms for several reasons. First, they have been evaluated on different private data sets not accessible to other researchers. A few 3D face recognition algorithms

have been evaluated on a common publicly available database as a part of a technology evaluation initiative called the Face Recognition Grand Challenge (FRGC) 2005 [13]. These algorithms can be regarded as the state-of-the-art in 3D face recognition. Second, the existing 3D face recognition algorithms vary considerably with regards to their reported performance statistics and experimental design protocols. Third, most studies do not report the statistical error bounds for the observed results.

2.3.1 Appearance Based Techniques

Three dimensional appearance based techniques are straightforward extensions of successful 2D appearance based techniques [56, 57, 58, 59, 60, 61, 62] to facial range images. The preprocessing and normalization steps they require are segmentation of faces from 3D images; removal of surface noise and holes; normalization of faces to a canonical frontal pose; and the generation of range images in that pose (*e.g.*, Fig. 2.1(c)).

Statistical subspace projection methods including principal component analysis (PCA) or ‘eigensurfaces’ [37, 30, 21, 63, 64, 65, 66, 19, 67, 68], independent component analysis (ICA) [69], and linear discriminant analysis (LDA) or ‘fishersurfaces’ [31, 32, 38] are the prominent 3D appearance based techniques. The underlying philosophy of subspace projection techniques is to regard a facial range image as an instance in N dimensional feature space, where N is the number of pixels in the image. All human faces are modeled to lie on a linear subspace of this feature space [70]. A statistical learning technique is employed to learn the linear subspace from an ensemble of facial images. All facial images are projected onto the learned subspace and are compared by means of a suitable distance metric in that subspace.

The various statistical learning techniques differ in the objective criteria that they optimize. For PCA, a set of orthogonal directions that maximize the variance of facial range images are obtained by eigen decomposition of their scatter matrix [70]. PCA minimizes the mean squared error between an original image and the corresponding image reconstructed from the PCA eigen directions. LDA projects data onto novel directions so as to maximize the ratio of the between class scatter and the within class scatter [71]. ICA considers not only the linear relationships between pixels in a facial image, but also higher order relationships. It projects data linearly onto a set of new basis vectors that are as statistically independent as possible [72].

Of all techniques for 3D face recognition, PCA has been explored most extensively, partly due to the success of the ‘eigenfaces’ technique for 2D face recognition [56, 57]. For the FRGC 2005, the performance of 3D PCA was regarded as the baseline [73]. PCA is reported to perform well with small 3D facial databases of less than 100 subjects, but poorly for larger data sets. Analogous to 2D face recognition, 3D techniques based on ICA [69], and LDA [31, 32, 41], have been reported to perform better than 3D PCA. For subspace projection algorithms in general, employing gradient images derived from facial range images is reported to produce better results than employing the range images directly.

Besides the subspace projection techniques, other appearance based 2D face recognition techniques based on Hidden Markov Models (HMM) [61], and embedded HMM [62] have also been extended to facial range images [37, 74, 11, 75, 76, 43]. These techniques exploit the fact that facial features naturally occur in a fixed order from top to bottom and from left to right, irrespective of changes in illumination, pose, and facial expression. For these techniques, different facial components are

modeled as states in statistical Markov models that are learned from an ensemble of facial range images.

Appearance based techniques for 3D face recognition are generally regarded as holistic techniques, in that they employ information from large facial image regions. Their performance is generally greatly affected by the presence of outliers, cluttered backgrounds, occlusions, noise, and variations in facial expression and pose. It is hard to interpret the discriminatory facial structural information that the subspace basis vectors encode [13]. Furthermore, in order to reliably learn the facial subspaces, these techniques require training images of many subjects under diverse imaging conditions.

2.3.2 Facial Surface Matching

The surface of the human face is an example of a *free form* 3D object that cannot be recognized as either planar or naturally quadric [77]. A class of 3D face recognition algorithms exists, wherein the shapes of the free form facial surfaces are compared by means of a suitable distance metric. The general philosophy of such approaches is to compute the structural dissimilarity between two facial surfaces after rigidly registering them either coarsely, or finely. For coarse alignment the gross pose and position of each facial surface in 3D space is computed and it is transformed to a canonical position [20, 33, 22, 78, 23]. For fine alignment, the Iterative Closest Point (ICP) [79] algorithm, wherein one 3D model is rotated and translated iteratively in space until its distance from the other model converges to a minimum, has been explored most extensively [34, 80, 81, 82, 83, 27, 84, 85, 86, 24, 35, 87, 88, 29].

Metrics including the mean squared error (MSE), point-to-closest-point mean squared error (MSE_{CP}) [79], point-to-closest-surface mean squared error [89],

the Hausdorff distance, and the partial Hausdorff distance (pH) [90], have been employed to compute the structural dissimilarity between pairs of facial surfaces. The pH metric rejects a fraction of the largest distances between points on the two surfaces and hence is robust to outliers caused by holes, noise, and occlusions, but its performance depends on the fraction of points rejected. For ICP, the point-to-closest-surface mean squared error metric is less susceptible to local minima problems than MSE_{CP} [89].

Within this class of 3D face recognition algorithms, the algorithms based on the ICP procedure have been the most successful. They are reported to be robust to variable facial poses [24], and illumination conditions during 3D image acquisition [46]. They are also reported to perform better than 3D PCA [81, 82]. However, they are computationally expensive. For example, it is reported that comparison of a pair of 3D facial surfaces required 20 seconds on an average Pentium IV 2.8 Ghz CPU, for 3D facial models containing approximately 18,000 effective points [24]. The iterative ICP procedure to align each pair of facial surfaces, is also not guaranteed to converge to a global minimum. The existing facial surface matching techniques that do not finely align facial surfaces, however, are not adequate for robust face recognition. Their accuracy depends on that of the MSE , MSE_{CP} , or pH metrics, which in turn are very sensitive to alignment errors between surfaces. Facial surface matching based algorithms are also affected by the presence of variable facial expressions, which are non-rigid deformations of the facial surface [86, 24, 35, 29]. For expression invariant facial surface matching, researchers have investigated using only those parts of the face, which deform less than others (*e.g.*, the nose region [87, 88]). However, this is reported to lower the overall recognition performance.

Some attempts have also been made to match facial surfaces using the statis-

tics of the orientation of surface normals as measured by extended Gaussian images [91, 92, 93], and phase Fourier transforms [94, 95]. Although such representations of surfaces are scale and rotation invariant [95], they have not been explored extensively for 3D face recognition.

2.3.3 Local Facial Features

Local 3D geometric characteristics including facial profile curves, and properties of facial landmarks and their relationships to each other have been employed previously for 3D face recognition. Facial profile curves can be easily obtained by intersecting 3D models by appropriately located 2D planes. For 3D face recognition, various profile curves have been rigidly aligned and compared [26, 96, 39, 97, 36, 98, 83, 42, 41]. Similar to the rigid surface matching algorithms, the performance of rigid profile matching algorithms is also adversely affected by the presence of variable facial expressions [36]. In numerous facial profile matching studies, the central vertical facial profile has been reported to be highly discriminatory for identifying individuals [26, 96, 39, 36]. Hence, techniques have also been investigated to automatically locate this natural axis of bilateral symmetry of the human face. Cartoux *et al.* automatically detected it by iteratively minimizing the difference between the principal surface curvatures on either sides of the face [98]. Others have located it by iteratively aligning a facial surface to its mirror image [83, 42, 82, 36].

Local geometric characteristics of facial sub-regions including their positional coordinates, surface areas and curvatures, and 3D Euclidean distances, ratios of distances, and angles between them have been employed previously for 3D face recognition [99, 100, 25, 101]. The shapes of facial landmarks have been quantified by Gaussian and mean curvature values [99, 25], Gaussian-Hermite moments [44],

‘point signatures’ [102, 103], and by 2D and 3D Gabor filter coefficients [104, 100]. A successful 2D face recognition technique called local feature analysis (LFA) [59], has also been applied to 3D facial images [38]. LFA exploits the inherent correlations and redundancies between neighboring pixels of an image to statistically derive topographical kernels that capture variations in sub-regions of the facial images [59].

Automatic 3D facial landmark detection algorithms are not very well developed. A few attempts have been made to automatically locate facial landmarks on 3D models using their characteristic surface [87, 105, 106, 20, 99, 100, 24, 102] or profile [101, 36] curvature properties, or by aligning 3D models rigidly [44], or non-rigidly [81] to generic facial templates with known landmarks. The tip of the nose has been detected as the most prominent point for 3D facial models in canonical frontal poses [26, 22]. However, this heuristic fails for faces in arbitrary poses. None of these studies, however, have reported the accuracy of their feature localization techniques against any form of ‘ground truth’ data.

A number of techniques based on local facial features have been reported to perform better than 3D PCA [102, 107, 108, 81]. Some have also been reported to perform better than profile matching techniques [44, 81]. They have been reported to be less affected by global changes in the appearance of facial range images including variable facial expressions, poses, and the presence of noise and occlusions than the holistic techniques [102, 103]. They have also helped to identify the local discriminatory structural characteristics of 3D facial surfaces. Nevertheless, 3D face recognition techniques based on local facial features have been explored much less than the holistic techniques.

2.3.4 Ensemble Approaches

Despite the existence of rich literature on ensemble approaches for combining classifiers [109], only a few have been applied to 3D face recognition. These include a combination of scores of 3D LDA, an algorithm based on surface normals, and a profile matching algorithm using a non-linear rank sum method [41]; a hierarchical combination of an ICP based algorithms and 3D LDA [24]; a combination of features from the whole face, and from the mouth, nose and orbital regions [44]; and a combination of the output of a facial surface matching algorithm and of a central profile matching algorithm using the ‘maximum’ rule [83, 42]. All these ensemble approaches have been reported to perform better than their individual constituent algorithms.

2.3.5 Expression Invariant Approaches

Achieving invariance to facial expression variations is an open problem in 3D face recognition. While the pose of 3D facial models can be easily corrected, changes in facial expression, which are non-rigid transformations of the facial surface, are not as trivial to rectify. Numerous studies have demonstrated that the performance of existing 3D face recognition techniques is reduced for facial models with arbitrary (other than neutral) expressions [24, 29, 35, 107, 108, 36, 104].

In order to solve this problem, researchers have proposed different ‘expression invariant representations’ of the facial surface. Bronstein *et al.* [110, 40] assumed different facial expressions to be isometric deformations of the facial surface. They isometrically embedded facial surfaces into expression invariant canonical forms using multidimensional scaling. However, their assumption of isometric deformations does not hold true for faces with open mouths. Wang and Chua [104] employed

Author	Algorithm	Data		Performance
		Gallery	Probe	
Chang [65]	PCA	200	676	$RR = 94.5\%$
Pan [112]	Isometric flattening + PCA	–	–	$EER = 2.83\%$
Russ[84]	ICP, $MSE_{CP} + pH$	198	745	$FRR = 6.5\%$ $FAR = 0.1\%$ $RR = 98.5\%$
Koudelka [88]	ICP, pH	198	198	$RR = 94\%$
Kakadiaris [107]	AFM	152	608	$FRR \sim 3.1\%$ $FAR = 0.1\%$ $RR = 99.3\%$

Table 2.1: A summary of the five 3D face recognition algorithms that were evaluated at the Face Recognition Grand Challenge 2005 on the FRGC v0.1 database.

2D/3D Gabor coefficients calculated from range images at facial fiducial points to generate expression invariant representations of 3D human faces. Lu and Jain [111] proposed a technique, wherein 3D models were first matched rigidly using ICP, and then non-rigidly deformed using thin-plate spline deformation to generate a displacement vector image. The displacement vector images were then classified into intra-person and inter-person deformations. Both these latter techniques met with limited success. Chang *et al.* [87] investigated matching the ‘rigid’ nose and the nose bridge regions using ICP. This, however, tended to lower the overall recognition performance.

2.3.6 FRGC 2005 Algorithms

Lastly, in Table 6.2 we summarize the five 3D face recognition algorithms that were evaluated on the publicly available FRGC v0.1 database during the FRGC 2005 evaluation [13]. These algorithms represent the current state-of-the-art in 3D face recognition. The FRGC v0.1 database contained 943 images of 198 subjects, which

were acquired using the Minolta Vivid 900/910 series (Konika Minolta Holdings, Inc., Tokyo, Japan) 3D laser scanning system. At the FRGC 2005, researchers were free to define their own training, and test data partitions. Hence, as is evident from the third and fourth columns of Table 6.2, each of these algorithms was evaluated on different data partitions of the FRGC v0.1 database.

Among these algorithms, two [65, 112] were appearance based algorithms that employed 3D PCA. Another two [84, 88] were surface matching based algorithms that used the ICP procedure to align 3D faces. The last was an ‘Annotated Face Model’ (AFM) algorithm [107], wherein multiscale representations of local facial regions were employed. Among these five existing algorithms, the AFM algorithm performed the best with rank 1 $RR = 99.3\%$, and $FRR \sim 3.1\%$ at $FAR = 0.1\%$.

To summarize, in this chapter we presented some background on 3D face recognition and reviewed existing algorithms. It is evident that there is considerable room for improvement within various domains of 3D face recognition. These open problems provide rich avenues for further research and investigation, and motivate the work for this doctoral dissertation.

Chapter 3

The Texas 3D Face Recognition Database

To promote serious research in 3D face recognition and related scientific disciplines, we are pleased to make available to qualified researchers in the field, at no cost the Texas 3D Face Recognition Database. This large database of 2D and 3D facial models was acquired at the company Advanced Digital Imaging Research (ADIR), LLC (Friendswood, TX), formerly a subsidiary of Iris International, Inc. (Chatsworth, CA), with assistance from research students and faculty from the Laboratory for Image and Video Engineering (LIVE) at the University of Texas at Austin. The project was sponsored by the Advanced Technology Program of the National Institute of Standards and Technology. The database of images and information about the subjects' gender, ethnicity, facial expression, *etc.*, will be hosted and managed on LIVE's web-servers, in an easy to browse, query and downloadable format. All requests for access to the database will also be received and processed by LIVE. This database will be a valuable resource to the 3D face recognition research community.

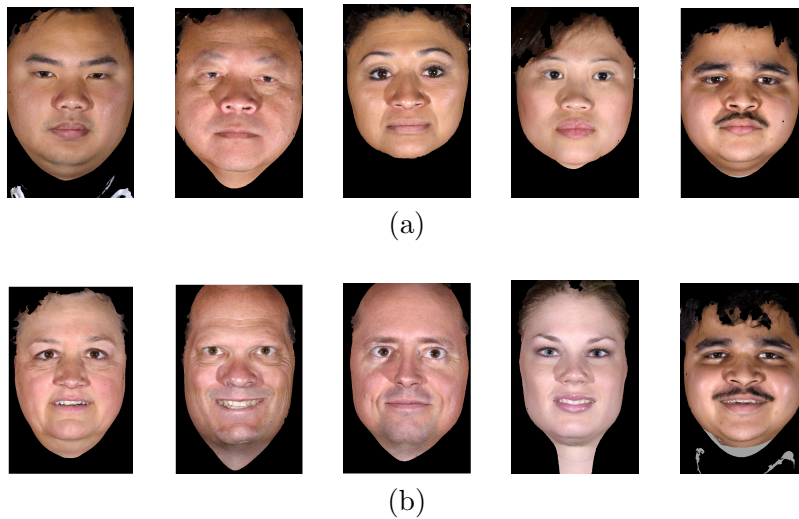


Figure 3.1: Color images of faces in the Texas 3D Face Recognition Database in (a) neutral, and (b) expressive modes.

It will be the largest publicly available database of 3D facial images that have been acquired using a stereo imaging system.

The database contains 1149 3D models of 118 adult human subjects. The number of images for each subject varies from 1 per subject to 89 per subject. The subjects' ages range from $\sim 22 - 75$ years. The database includes images of both males and females from the major ethnic groups of Caucasians, Africans, Asians, East Indians and Hispanics. The faces are in neutral and expressive modes (Fig. 3.1). The facial expressions present are smiling or talking faces with open/closed mouths and/or closed eyes. The neutral faces are emotionless. All subjects were requested to remove hats and eye-glasses prior to image acquisition.

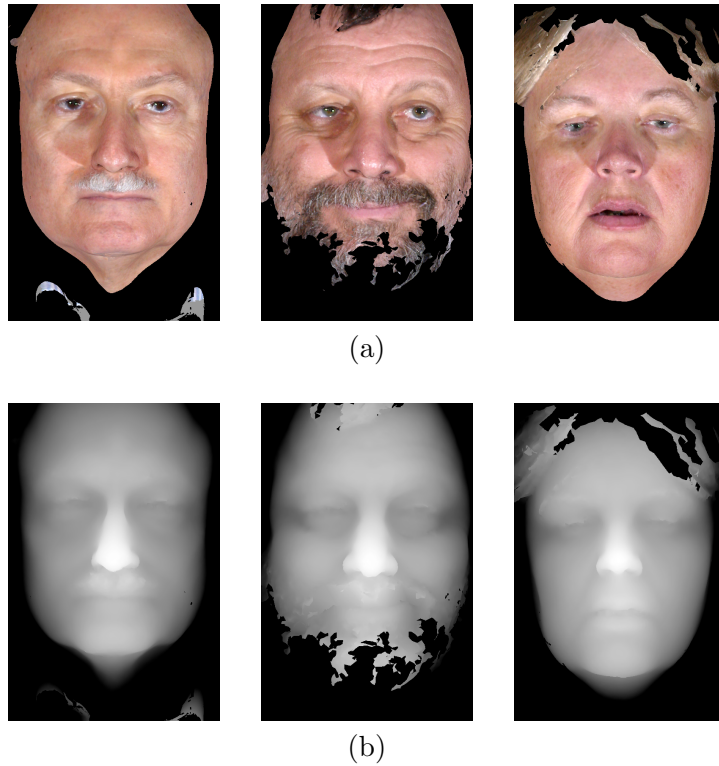


Figure 3.2: Raw (a) color, and (b) range images of the Texas 3D Face Recognition Database.

3.1 Acquisition and Normalization

The 3D models in the Texas 3D Face Recognition Database were acquired using an MU-2 stereo imaging system [113, 114] manufactured by 3Q Technologies Ltd. (Atlanta, GA). All subjects were requested to stand at a known distance from the camera system. At the beginning of each acquisition session, and at regular intervals during the session (a session is defined as a set of images acquired one a particular day) the stereo system was calibrated against a target image containing a known pattern of dots on a white background [115]. This ensured that each 3D facial model had the same dimensions as the actual real-world dimensions of the face. The stereo

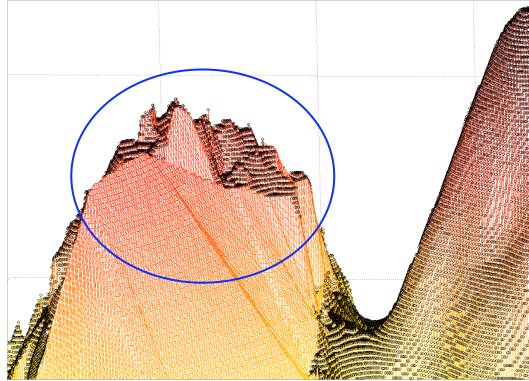


Figure 3.3: Input noise on the surface of 3D models.

system acquired both the shape and the color images of the face simultaneously. Hence, all pairs of range and color images for a particular acquisition in the Texas 3D Face Recognition Database are perfectly aligned.

The acquired 3D models were successfully transformed to a frontal orientation with the forehead tilted back by 10° to the vertical axis. This was achieved by iteratively aligning facial models in arbitrary poses to a template face in a canonical frontal pose using the ICP algorithm [79]. Tilting the forehead of the 3D face back by 10° to the vertical axis ensured that each (x, y) location was associated with a unique z value, and hence the facial surface could be represented as $z = f(x, y)$.

The final representation of each face in the database is a pair of range and color images in the canonical frontal pose that are perfectly aligned to each other (*e.g.*, Fig. 3.2). The range images were constructed by orthographically projecting the pose normalized 3D models onto a regularly spaced rectangular grid. The corresponding color images were constructed by obtaining the color information at each point in the range image. The tip of the nose of each model is located at the center of the image. The range images are of size 751×501 pixels with a resolution

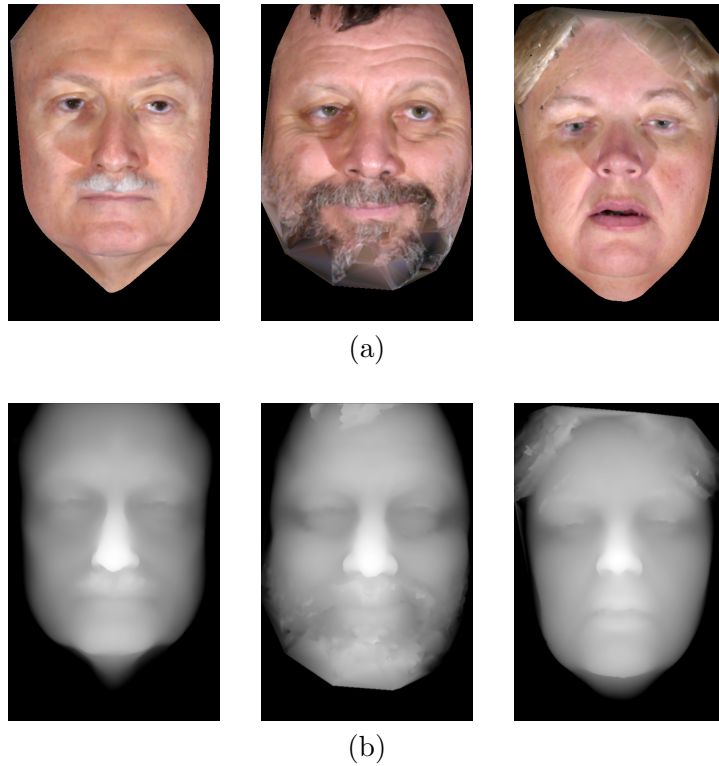


Figure 3.4: Preprocessed (a) color, and (b) range images of the Texas 3D Face Recognition Database.

of 0.32 mm along the x , y , and z dimensions. Each z value is represented in an 8 bit format with the highest value of 255 assigned to the tip of the nose and a value of 0 assigned to the background. The color images are similarly of size $751 \times 501 \times 3$ pixels represented in an uncompressed 8 bit RGB format.

3.2 Preprocessing

We further preprocessed the raw range and color images to convert them into a form useful for face recognition. We removed small extraneous regions that were not attached to the face region, *e.g.*, shirt collars in the leftmost image in Fig.

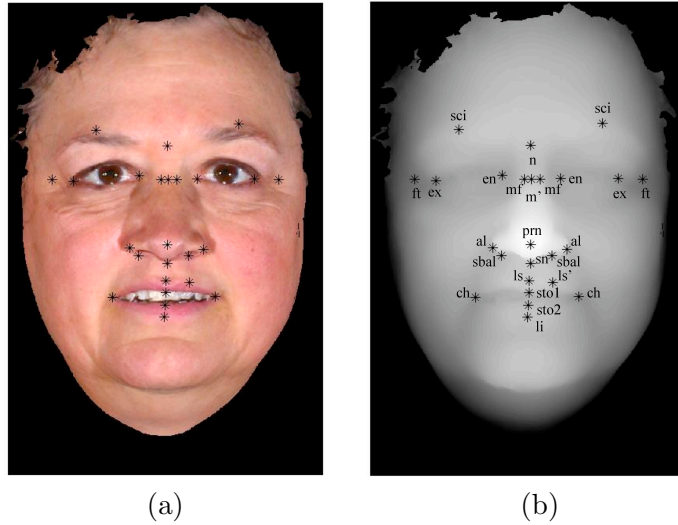


Figure 3.5: Twenty-five anthropometric fiducial points (a) on a color image, and (b) on a range image.

3.2 (a). For this, we detected the face region as the largest connected region of non-zero z values in range images and retained it. All other regions were removed. We eliminated small amounts of impulse noise present in the range images (*e.g.*, Fig. 3.3) by median filtering them with a square window of size 3×3 pixels. We interpolated the range images using bi-cubic interpolation to remove large holes and finally smoothed them by applying a Gaussian window with $\sigma = 1$ pixel. These steps were also applied to each of the R, G and B channels of the color images. The preprocessed versions of the raw images shown in Fig. 3.2 are presented in Fig. 3.4.

3.3 Manual Fiducial Point Detection

We annotated all 1149 images in the Texas 3D Face Recognition Database with the positions of 25 anthropometric facial fiducial points (Fig. 3.5). These fiducial points are associated with facial anthropometric proportions [1] that are reported

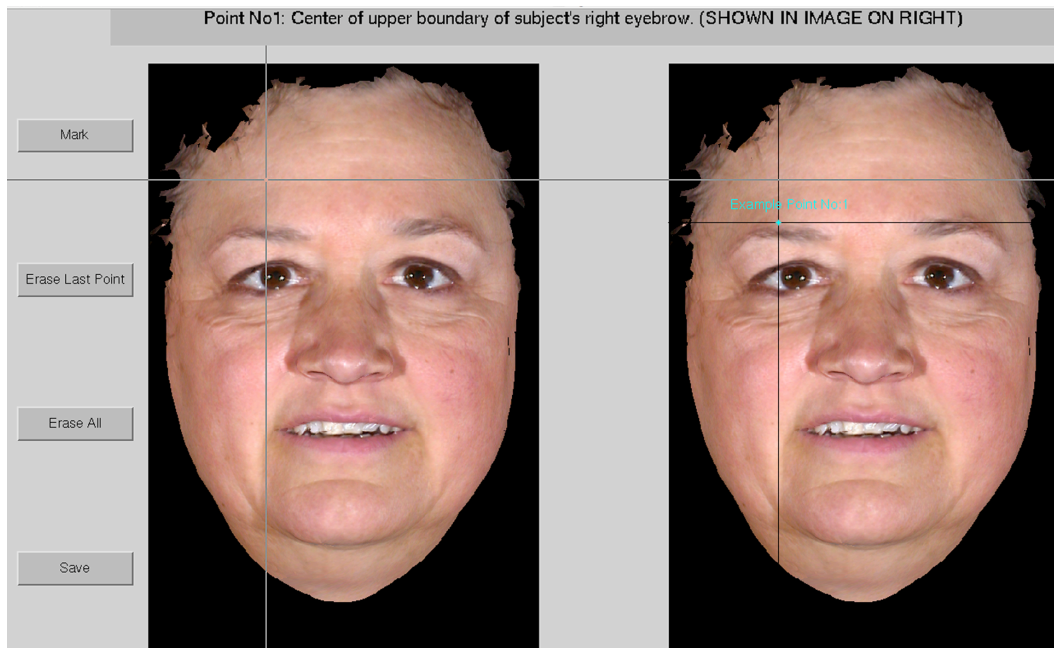


Figure 3.6: The graphical user interface that was developed for manually locating anthropometric facial fiducial points.

to be variable for human populations [116, 117]. We located the fiducial points manually on the color images by clicking at appropriate locations with a mouse and a computer based graphical user interface (Fig. 3.6). The locations of the facial fiducial points on the facial range and color images are the same as the two types of images are perfectly aligned.

3.4 Data Partitions

With the intended goal of developing 3D face recognition algorithms, we partitioned this database into a training data set and a test data set (Table 3.1). The training data set contained a set of randomly selected 360 images of 12 subjects (30 images per subject) in neutral or expressive modes. For all the 3D face recognition

Partition	No. of Subjects	No. of Images		
		Neutral	Expressive	Total
Training	12	228	132	360
Test	Gallery	105	0	105
	Probes	95	480	663
Remaining	13	0	21	21

Table 3.1: A summary of the data partitions employed for developing 3D face recognition algorithms.

algorithms that we developed, steps such as automatic facial fiducial point detection, classifier feature selection and classifier optimization were performed using the training data set only. The trained classifier was evaluated on the independent test data set, which did not overlap with the training data set.

The test data set included 768 images of 105 subjects. This test set was further partitioned into a gallery set and a probe set. Consistent with the evaluation protocol of the FRVT 2002 [118] and the FRGC 2005 [13], the gallery set contained one range image each of 105 subjects with a neutral facial expression. The probe set contained another set of 663 images of 95 of the gallery subjects with a neutral or an arbitrary facial expression. In the probe set, the number of images of each subject varied from 1 to 55. In accordance with the widely accepted ‘closed universe’ model for the evaluation of face recognition algorithms [118], every subject in the probe data set was represented in the gallery data set. After partitioning the entire database of 1149 images into the training and test data sets, 21 images of 13 subjects remained. All of these were faces with an arbitrary facial expression. In the remainder of this document we refer to this set of images as the ‘Remaining’ set.

3.5 Availability

In order to facilitate further research in the area of 3D face recognition, and other related scientific disciplines, *e.g.*, facial animation and graphics, and facial surgical planning, we have made the Texas 3D Face Recognition Database available to qualified researchers in the field at no cost. The database of images is accompanied with information about the gender, ethnicity, facial expression, and the locations of the 25 manually detected anthropometric facial fiducial points. Information about the specific data partitions that we employed for developing 3D face recognition algorithms has also been made available. This will enable researchers to directly compare the results of future 3D face recognition algorithms against ours. The database of images along with its ancillary information is hosted and managed on the web server maintained by the Laboratory for Image and Video Engineering at the University of Texas at Austin. The database is available in an easy to browse, search and downloadable format. All requests for access to the database are also received and processed by LIVE.

For developing 3D face recognition algorithms, the Texas 3D Face Recognition Database will complement the publicly available and widely used FRGC 2005 database [13]. The Texas 3D Face Recognition Database differs from the FRGC 2005 database in a number of respects. First, it is the largest (in terms of the number of images and subjects) publicly available 3D facial database that has been acquired using a stereo imaging system. Images in the Texas 3D Face Recognition Database have been acquired at a high resolution (0.32 mm) along the x , y and the z dimensions. In comparison, images in the FRGC 2005 database have been acquired using a 3D laser scanning system Minolta Vivid 900/910 series (Konika Minolta Holdings, Inc., Tokyo, Japan), at a lower average resolution of 0.98 mm along the x

ad y dimensions and 0.5 mm along the z dimension [65]. Hence, the Texas 3D Face Recognition Database is a natural choice for researchers interested in experimenting with high resolution 3D facial images acquired using a stereo imaging system.

Second, all images in the Texas 3D Face Recognition Database have been acquired at the same scale and the dimensions of the 3D faces correspond to their actual real-world dimensions. The scale of faces in the FRGC database varies considerably (*e.g.*, Fig. 3.7). Third, as a consequence of using a stereo imaging system, which acquires the color and shape information of a scene simultaneously, the pairs of color and range images in the Texas 3D Face Recognition Database are perfectly aligned. In contrast, the color and range images in the FRGC 2005 database were acquired a few seconds apart, and hence are not perfectly aligned [13]. For the same reason, certain pairs of range and color images in the FRGC 2005 database have inconsistent facial expressions [29]. As the acquisition time for the Minolta 3D scanner is greater than 100 ms, certain 3D meshes in the FRGC 2005 database are also reported to be distorted due to the subjects' motion during image acquisition [29].

The FRGC 2005 database is four times larger than the Texas 3D Face Recognition Database in terms of the number of subjects and images, and contains greater diversity of facial expressions. However, images in the FRGC 2005 database require considerable amounts of non-trivial initial preprocessing, including hair, clothing and background elimination, and facial pose and scale normalization (Fig. 3.7). These steps are not required for the Texas 3D Face Recognition Database and hence, it provides a good alternative for researchers focussed specifically on developing and evaluating novel algorithms for *3D face recognition*, without regard to the initial preprocessing of 3D images. Furthermore, the Texas 2D Face Recognition

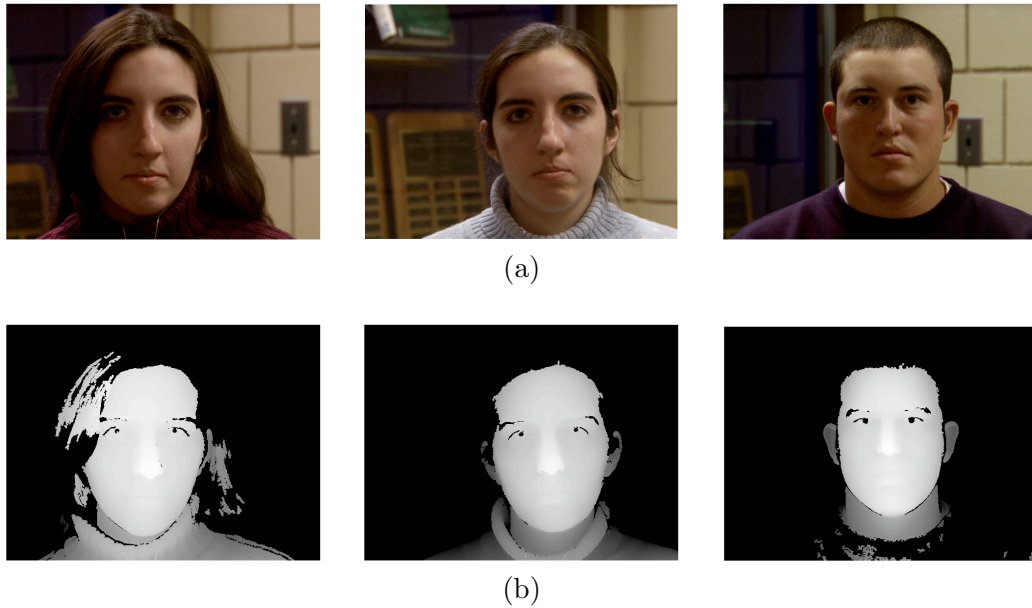


Figure 3.7: Examples of (a) color images, and (b) range images from the FRGC 2005 data set.

Database contains adequate variability to model a controlled real world operating environment of co-operative users. The Texas 3D Face Recognition Database also has the advantage of allowing researchers to evaluate face recognition algorithms in a timely manner using modest computational resources. By comparison, evaluating algorithms on the FRGC 2005 database may require the use of super computing facilities.

The Texas 3D Face Recognition Database is also unique in that it contains the positions of 25 manually annotated anthropometric facial fiducial points for every image (Fig. 3.5). Currently, the locations of such a large number of manually located facial fiducial points are not available for any 2D or 3D public database of comparable size. Hence, none of the currently reported studies of 3D or 2D+3D facial fiducial point detection report detection errors relative to any form of ‘ground

truth data' [87, 105, 106, 20, 99, 100, 24, 102, 100, 33, 63, 103, 104, 119]. The availability of this information as a part of the Texas 3D Face Recognition Database will thus be a very valuable resource for researchers working in the area of 2D and 3D facial feature detection, face recognition, and facial processing and anthropometry.

Chapter 4

3D Anthropometric Face Recognition

Face recognition algorithms based on local facial features can be successful. At the FRVT 2002 [9], two of the top three 2D face recognition algorithms, namely local feature analysis [59], and elastic bunch graph matching (EBGM) [120] employed local facial features. In EBGM, a face is represented as an ‘elastic bunch graph’, comprised of Gabor wavelet coefficients computed at facial fiducial points, and 2D facial Euclidean distances between them. Hüsken *et al.* developed a successful 2D+3D face recognition technique called ‘hierarchical graph matching’, which combined scores of 2D EBGM and 3D EBGM (EBGM applied to facial range images) [100]. Their technique was one of the top performers at the FRGC 2005 among the algorithms that were evaluated on the FRGC v0.2 database.

Despite this, techniques for 3D face recognition based on local facial features have been investigated much less than holistic face recognition techniques for two reasons. First and foremost, there has not been any attempt to systemati-

cally identify the local discriminatory structural characteristics of the human face for automatic face recognition purposes. Among the few reported 3D face recognition techniques that employ local facial features, the choice of facial landmarks has either been *ad hoc* [99, 25, 101, 103, 81], or has been a straightforward extension of local 2D techniques to range images [100]. Hüsken *et al.* observed that 2D EBGM outperformed 3D EBGM, and acknowledged that merely extending local 2D techniques to range images may not be the best way to discover discriminatory structural characteristics of the human face.

Second, local 3D face recognition techniques require robust and accurate automatic detection of facial fiducial points. Techniques for automatic detection of facial fiducial points using 3D images remain poorly developed. This is further complicated by the fact that the ‘ground truth’ locations of the facial fiducial points are not available for any of the publicly available 3D face databases. It is no surprise that all of the current studies of 3D or 2D+3D facial fiducial point detection [87, 105, 106, 20, 99, 100, 24, 102, 100, 33, 63, 103, 104, 119] report results of *visual inspection* only and do not report errors statistics for comparison against any form of ‘ground truth’ data.

We address these two fundamental open problems that may significantly impact the design of effective 3D face recognition algorithms. We obtain information about the structural diversity of human faces from the related scientific discipline of facial anthropometry [1, 121], and identify discriminatory structural characteristics of the face. We develop algorithms to automatically and accurately detect facial fiducial points associated with these characteristics. Lastly, we develop a novel completely automatic 3D Anthropometric Face Recognition (3D AnthroFace) algorithm that employs these local discriminatory facial characteristics. On a database of 1149

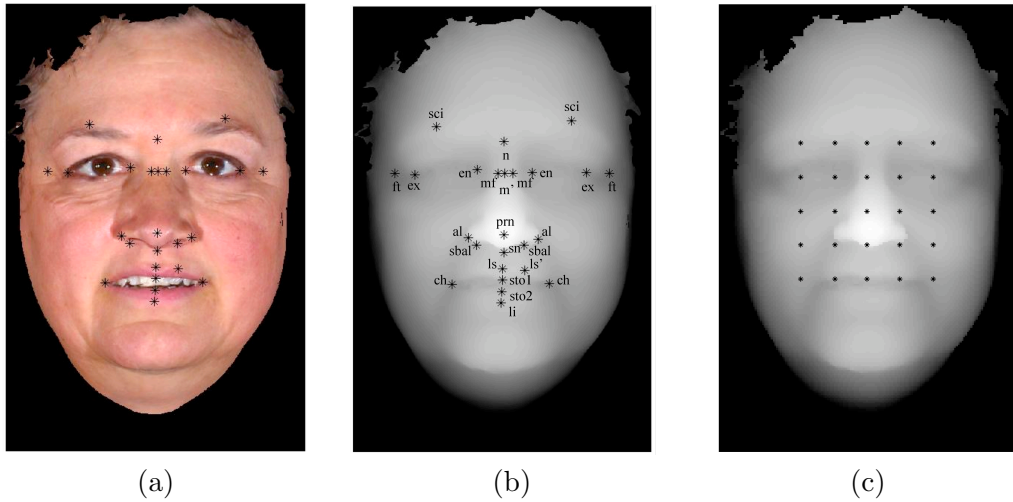


Figure 4.1: The 25 facial fiducial points associated with highly variable anthropometric facial proportions on a (a) color image, and (b) on a range image, and (c) a set of 25 arbitrary facial points.

3D faces, we demonstrate the potential and significantly superior performance of our proposed AnthroFace algorithm relative to three of the existing state-of-the-art 3D face recognition algorithms [116, 117].

4.1 3D Anthropometric Face Recognition (AnthroFace)

4.1.1 Anthropometric Cranio-Facial Proportions

Anthropometric cranio-facial proportions [1] are ratios of pairs of straight-line and/or along-the-surface distances between specific cranial and facial fiducial points (*e.g.*, Fig. 4.1(a) and Fig. 4.1(b)). For example, the most commonly used nasal index $N1$ is the ratio of the horizontal nose width to the vertical nose height ($N1 = (al - al)/(n - sn)$ from Fig. 4.1(b)). The scientific discipline of cranio-facial anthropometry has existed for nearly three centuries. Over the years numerous an-

S. No	Anthropometric Proportion	σ
1.	$O3 = (ex - en, l)/(en - en)$	7.75
2.	$O10 = (en - en)/(al - al)$	8.29
3.	$O12 = (en - en)(ch - ch)$	6.02
4.	$F32 = (n - sto1)/(ex - ex)$	5.30
5.	$N1 = (al - al)/(n - sn)$	5.81
6.	$N2 = (mf - mf)/(al - al)$	7.08
7.	$N4 = (sbal - sn, l + r)/(al - al)$	8.80
8.	$N6 = (ex - m'_{sag}, l)/(mf - mf)$	14.6
9.	$N7 = (sn - prn)/(al - al)$	6.28
10.	$N8 = (sn - prn)/(sbal - sn, l + r)$	12.8
11.	$N15 = (en - m'_{sag}, l)/(sn - prn)$	11.2
12.	$N16 = (en - m'_{sag}, l)/(en - m, l)$	7.26
13.	$N30 = (mf - mf)/(en - en)$	6.06
14.	$N31 = (ex - m'_{sag}, l)/(en - en)$	7.01
15.	$N32 = (al - al)/ch - ch$	5.04
16.	$N33 = (sn - prn)/(sn - sto1)$	13.8
17.	$L1 = (sn - sto1)/(ch - ch)$	5.40
18.	$L4 = (sn - ls)/(sbal - ls', l)$	10.2
19.	$L5 = (sn - ls)/(sn - sto1)$	5.97
20.	$L6 = (ls - sto1)/(sn - sto1)$	7.10
21.	$L7 = (ls - sto1)/(sn - ls)$	13.3
22.	$L9 = (ls - sto1)/(sto2 - li)$	16.9
23.	$L14 = (sn - sto1)/(n - sn)$	5.10

Table 4.1: The 23 most variable anthropometric facial proportions for adult humans along with their standard deviation values [1]. The corresponding fiducial points are presented in Fig. 4.1(b). N denotes nasal proportions, O denotes orbital proportions, L denotes proportions related to the the mouth region, and F denotes facial proportions.

thropometric facial proportions have been proposed and researchers have collected, recorded and analyzed their values on various human populations.

Cranio-facial proportions are widely employed in art and sculpture as neo-classical canons to aid in the creation of well-proportioned faces; in anthropology for analyzing prehistoric human remains [122]; for quantifying facial attractiveness

[15]; for analyzing facial disproportionality in anomalies or after facial injury as an aid to planning facial cosmetic and reconstructive surgery [1, 123]; and recently for creating parametric models of human faces in computer graphics [124]. As far back as 1939, Hrdlička [125] emphasized the importance of anthropometric facial proportions for comparing groups of people or populations. However, they have not been employed previously to aid in the design 3D face recognition algorithms.

Farkas and Munro consolidated a list of 155 cranio-facial anthropometric proportions that are used for planning facial reparative and cosmetic surgery [1]. By means of physical measurements, they also computed the mean (μ) and standard deviation (σ) values of these proportions for a population of 2564 healthy adult human subjects belonging to diverse ethnic, gender, and age groups [1, 14]. From among these 155 proportions, we isolated 70 anthropometric proportions associated with the facial region that can be computed automatically from the frontal 3D facial models normally acquired by 3D imaging devices. We identified a third (23) of these 70 facial proportions with the highest standard deviation values (Table 4.1) as being representative of discriminatory facial structural characteristics. It is reasonable to hypothesize that characteristics that display wide variation between individuals are likely to be most useful for distinguishing them. Associated with these 23 most variable anthropometric proportions are 25 anthropometric face fiducial points (Fig. 4.1(a) and Fig. 4.1(b)). This information about the structural diversity of human faces forms the basis of our proposed 3D Anthropometric Face Recognition algorithm [116, 117].

4.1.2 Manual Detection of Anthropometric Fiducial Points

We manually located the 25 anthropometric facial fiducial points associated with the diverse anthropometric measurements on all 1149 color images (Fig. 4.1(a)) in the Texas 3D Face Recognition Database (Chapter 3) that we employed. Since the images in this database were acquired using a stereo imaging system, the range and the color images for a particular acquisition were perfectly aligned. Hence, the locations of the fiducial points for the range images (Fig. 4.1(b)) were the same as their locations for the color images (Fig. 4.1(a)).

It is reasonable to assume that the human observer is *perfect* at detecting anthropometric facial fiducial points, since they are defined simply by convention of human observation. Hence, manual identification of the fiducial points served two purposes. First, it provided the ‘ground truth’ data for assessing the performance of the algorithms that we developed for automatically detecting the facial fiducial points. Second, in a loose sense it helped to establish an upper bound on the expected performance of 3D face recognition algorithms for perfectly detected anthropometric facial fiducial points. At first we developed all 3D face recognition algorithms using manually located facial fiducial points. We reasoned that only upon establishing the potential of our proposed technique for manually detected points, would it be worthwhile to investigate approaches to automatically locate the facial fiducial points.

4.1.3 Classifier

We employed 300 3D Euclidean distances, and 300 geodesic distances between all pairs of the 25 anthropometric face fiducial points (Fig. 4.1(b)) as features for our proposed 3D AnthroFace algorithm [117]. We computed geodesic distances along

the facial surface using Dijkstra’s shortest path algorithm [126, 127]. While 3D Euclidean distances between facial fiducial points have been employed previously as features for 3D face recognition [99, 100, 25, 101], using anthropometric geodesic distances is novel to this work [117]. Studies have shown that geodesic distances are better at representing ‘free-form’ 3D objects than 3D Euclidean distances [128]. Furthermore, a recent study suggested that changes in facial expressions may be modeled as isometric deformations of the facial surface [40]. When a surface is deformed isometrically, intrinsic properties of the surface, including Gaussian curvature and geodesic distances are preserved [129]. Hence, algorithms based on geodesic distances are likely to be robust to changes in facial expressions.

We identified the 106 and 117 most discriminatory Euclidean and geodesic distances, respectively, using the stepwise linear discriminant analysis [130] procedure (‘stepdisc’, SAS Institute Inc., NC, USA). We pooled these 106 Euclidean and 117 geodesic anthropometric distances together, and using a second stage stepwise linear discriminant analysis procedure, we identified the final combined set of 123 most discriminatory distance features. We then trained a Fisher’s linear discriminant analysis classifier [70] that linearly projected these 123 anthropometric distance features onto 11 dimensions (11D). The classifier was trained using a training data set and all images in an independent test data set were projected onto the learned LDA directions. The final metric for comparing a pair of 3D faces in the 11D LDA sub-space was the L_2 norm.

4.1.4 Effect of Choice of Facial Fiducial Points

We also investigated the effect of the choice of face fiducial points on the performance of the proposed 3D AnthroFace algorithm [116]. We repeated the steps of the 3D

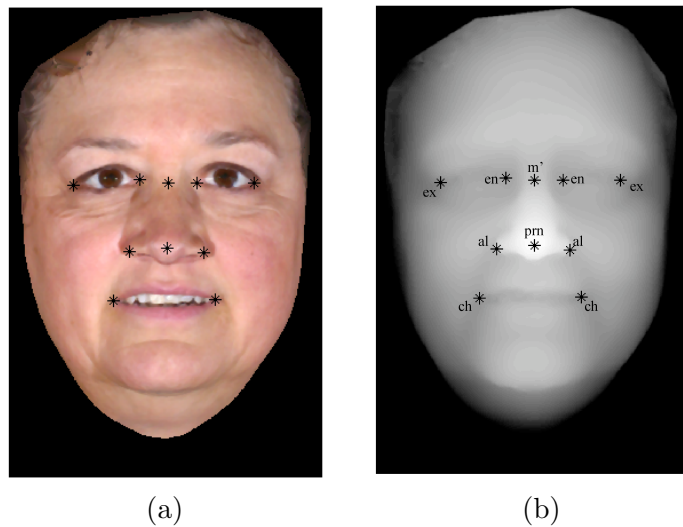


Figure 4.2: The reduced set of 10 anthropometric facial fiducial points for the AnthroFace algorithm on a (a) color image, and (b) range image.

AnthroFace algorithm with Euclidean and geodesic distances between 25 arbitrary facial points instead of the 25 anthropometric fiducial points. These points were located in the form of a 5×5 rectangular grid positioned over the primary facial features of each face (Fig. 4.1(c)). We chose these arbitrary facial points as they measure distances between the significant facial landmarks, including the eyes, nose and the mouth regions, without requiring localization of specific fiducial points. A similar set of facial points was also employed in a previous 3D face recognition algorithm for aligning 3D facial surfaces using the Iterative Closest Point (ICP) algorithm [24].

4.1.5 Reduction of Anthropometric Facial Fiducial Points

Next, we determined if a sub-set of the 25 manually located anthropometric facial fiducial points could be employed for the 3D AnthroFace algorithm, without a sig-

nificant loss in its performance. This was an important step towards completely automating the proposed 3D AnthroFace algorithm. Clearly, the task of automatically detecting all of the 25 anthropometric fiducial points with a high accuracy is non-trivial and may even be redundant. To isolate this sub-set of points, we first removed individual points (*e.g.* *prn* from Fig. 4.1(b)), or pairs of symmetric points (*e.g.* *al-al* from Fig. 4.1(b)) from the overall 3D AnthroFace algorithm and re-evaluated its performance. Interestingly, the removal of none of the individual points/pairs resulted in a statistically significant loss in the performance of the 3D AnthroFace algorithm. This indicated that some points in the set of 25 anthropometric points were clearly redundant. We then proceeded to remove larger groups of fiducial points associated with the orbital, nasal, and mouth regions and re-evaluated the overall performance of the 3D AnthroFace algorithm. Finally, we isolated a sub-set of 10 anthropometric facial fiducial points (Fig. 4.2) that resulted in statistically equivalent recognition performance to that of the algorithm that employed 25 fiducial points. Hence, our final proposed 3D AnthroFace algorithm employed only these 10 anthropometric facial fiducial points (Fig. 4.2), instead of the 25 points that we initially proposed.

4.1.6 Automatic Detection of Anthropometric Fiducial Points

To completely automate our proposed 3D AnthroFace algorithm, we developed algorithms to automatically detect the 10 anthropometric facial fiducial points (Fig. 4.2) that we isolated in the previous analysis. We automatically detected 3 of these points (*prn*, and *al-al* in Fig. 4.2) using only the 3D shape information of the face, while to locate the remaining seven points we employed both the 2D and the 3D information of the face. Furthermore, to locate all the points except for the tip of

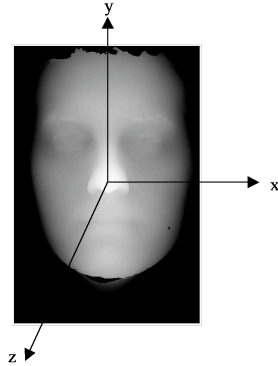


Figure 4.3: An example of a 3D face in a canonical pose that was used to automatically detect facial fiducial points.

the nose (prn), we assumed that each face was in a frontal upright position, with its natural axes of bilateral symmetry roughly along the vertical dimension (*e.g.*, Fig. 4.3).

The overall algorithm for detecting the 10 anthropometric facial fiducial points proceeded in a sequence of cascaded steps. Each stage in the sequence utilized fiducial point locations found in the previous stages to assist in locating the current fiducial points. The logical sequence used was to begin with the most reliable and easy to detect feature, proceeding to the features that were less reliable and harder to detect. However, as it turns out, all facial fiducial points were quite reliable.

The sequence of stages for automatically detecting the fiducial points started with the detection of the tip of the nose (prn). The location of this point was then employed to detect the nose width points $al-al$. These three points were then employed to detect the inner corners of the eyes (points $en-en$) and the center of the nose root (point m'). The points prn and $al-al$ were also employed to detect the corners of the mouth (points $ch-ch$). Lastly, the locations of the points $en-en$ were used to locate the outer corners of the eyes (points $ex-ex$).

While developing algorithms for detecting these fiducial points, an underlying concept guided us. By their definition of being *fiducial points* or *landmarks*, they had unique structural and/or textural properties that differentiated them from their surrounding regions. When humans search for these facial fiducial points, either by visual or tactile inspection, it is these characteristics that inherently guide their search. We reasoned that the key to accurately locating these anthropometric fiducial points was to isolate their unique structural and/or textural characteristics and to search for them in an appropriately constrained region of the face. In the following sections, we describe in detail the steps that we employed to automatically detect each of the 10 anthropometric fiducial points.

4.1.6.1 Nose Tip (*prn*)

To locate the tip of the nose (point *prn* in Fig. 4.2) of each face in our database, we employed a 3D template face with a manually located nose tip (Fig. 4.4). The template face was one of the faces in the training data partition of our database. It was a relatively noise free and symmetric face with no facial or surrounding hair, and a neutral facial expression. We finely registered the entire surface of every 3D face in our database to the surface of the template face using the ICP algorithm [79]. After aligning the face to the template, we found the point on its surface that was closest to the tip of the nose of the template face. This was our initial estimate (ICP estimate) of the tip of the nose.

Although the ICP estimate was not very accurate (the standard deviation of errors from the manually located nose tips was $\sigma_x = 6.271$ and $\sigma_y = 9.415$ pixels), the ICP procedure served two purposes. First, it helped to transform 3D models in arbitrary poses to a frontal upright canonical pose, which was required for detecting

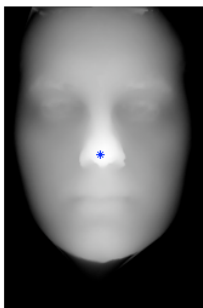


Figure 4.4: The 3D template face and its manually located nose tip, which was used to automatically locate the nose tips of all 3D faces.

all the other points. Second, since the ICP estimate was in the central region of all faces, we limited our search for the tip of the nose in the next stage to a window of 301×301 pixels about the ICP estimate.

For facial range images of the form $(x, y, z(x, y))$, the Gaussian surface curvature (K), the mean surface curvature (H), and two principal curvatures (κ_1, κ_2) can be computed from their first and second partial derivatives as [129]

$$K = \frac{z_{xx}z_{yy} - z_{xy}^2}{(1 + z_x^2 + z_y^2)^2}, \quad (4.1)$$

$$H = \frac{z_{xx}(1 + z_y^2) + z_{yy}(1 + z_x^2) - 2z_xz_yz_{xy}}{(1 + z_x^2 + z_y^2)^{3/2}}, \quad (4.2)$$

$$\kappa_1, \kappa_2 = H \pm \sqrt{H^2 - K}, \quad (4.3)$$

where z_x and z_y are the first partial derivatives of $z(x, y)$ w.r.t x and y , respectively, and z_{xx} , z_{yy} and z_{xy} are the second first partial derivatives of $z(x, y)$ w.r.t x and y . Furthermore, the Gaussian curvature $K = \kappa_1\kappa_2$, and the mean curvature $H = (\kappa_1 + \kappa_2)/2$. We computed these partial derivatives, and the Gaussian and mean curvature values for the facial range images, using a method developed by Besl [131].

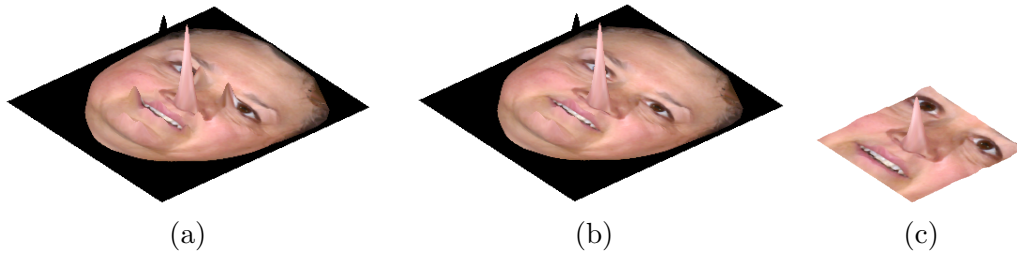


Figure 4.5: The magnitudes of the Gaussian curvatures of the (a) elliptic, (b) convex elliptic regions of a facial range image, and (c) of its central 301×301 pixel region. In each image, the Gaussian curvature has been plotted as a 3D surface with the facial texture warped onto it.

The signs of the Gaussian and the mean curvatures define differently shaped regions of a surface [132]. The regions with $K > 0$ are ‘elliptic’, those with $K < 0$ are ‘hyperbolic’, and those with $K = 0$ are either ‘planar’ or are ‘cylindrical’. For the right-handed 3D co-ordinate system defined in Fig. 4.3, regions of the surface with $H > 0$ are ‘concave’, while those with $H < 0$ are ‘convex’.

Researchers in the past have noted that the sub-parts of the human face have distinct surface curvature properties [99, 25]. We observed that in fact, of all the regions on the facial surface, the region surrounding the tip of the nose has the highest elliptic Gaussian curvature (Fig. 4.5(a)), and more specifically the highest convex elliptic Gaussian curvature (Fig. 4.5(b)). Hence, we employed a very simple procedure, which reliably and accurately detected the tip of the nose for *all* 3D faces in our database. Within the central region of each face of size 301×301 pixel (Fig. 4.5(c)), which we had isolated previously using the ICP procedure, we searched for the point with the maximum elliptic Gaussian curvature. This was the location of the final automatically located tip of the nose. For apparent reasons (Fig. 4.5(b)), we coined the term ‘The Pinocchio Feature’ for this reliable facial fiducial point.

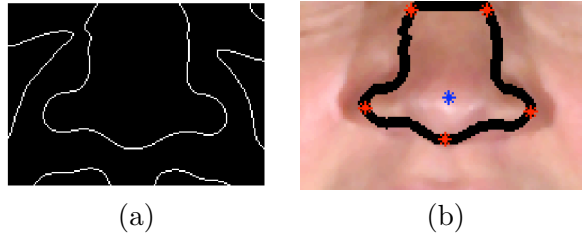


Figure 4.6: (a) The edges detected with the LOG edge detector in sub-regions of the facial range images centered about the detected nose tip, and (b) critical points along the nose boundary and the tip of the nose.

4.1.6.2 Nose Width Points (*al-al*)

To detect the anthropometric facial fiducial points (*al - al* in Fig. 4.2(b)), which define the anthropometric measurement of ‘nose width’, we restricted our search to sub-regions of the range images of size 42×50 mm centered about the automatically detected nose tip. We determined the size of this search region, using information about statistics of the height and width of noses of adult human males [121]. On an average, human males are reported to have wider ($\mu = 35$ mm, and $\sigma = 2.5$ mm) and taller ($\mu = 53$ mm with $\sigma = 3.4$ mm) noses than females. Hence, to account for variations in the human population, we fixed the width of the search region for points *al-al* about the tip of the nose at the $\mu + 6\sigma$ value for the width of noses of human males. Similarly, we fixed the height of the search region at the $0.6(\mu + 6\sigma)$ value for the height of noses of human males.

We detected edges of the range image sub-regions using a Laplacian of Gaussian edge detector [133], with $\sigma = 7$ pixels. Since the human nose is a distinct protrusion in the facial surface, we observed that the left and right boundaries of the nose were always clearly delineated in the edge maps of all faces (Fig. 4.6(a)). From this edge map, we further isolated the left and right boundaries of the nose

by traversing outwards horizontally in both directions from the tip of the nose, and by retaining the first curves that were encountered.

We then detected all the ‘critical’ points (points of high curvature) with negative curvature values [134], which were present along the nose boundary contours. We traversed all the nose boundary curves in the clockwise direction. For *all* faces in our database, the points *al - al* were among these critical points (shown in Fig. 4.6(b)). From among the critical points, we isolated the points *al - al*, by searching for the leftmost and rightmost critical points, that were closest to the tip of the nose along the vertical direction.

4.1.6.3 Inner Eye Corners (*en-en*) and Center of Nose Root (*m'*)

To automatically locate the inner corners of the eyes (points *en-en*) we observed that for *all* faces, these points were located in regions of the face that were distinctly concave elliptic (Fig. 4.7(a)). We located the peaks of Gaussian curvature of these two regions, as the initial estimates (curvature estimates) of the locations of the inner corners of the eyes. In order to define the search regions for these peaks, we employed the locations of automatically detected points *prn*, *al-al*, the location of the highest point of each 3D model (*v*), and knowledge of the established horizontal and vertical proportions of a normal adult human face [1].

For an average adult, the vertical distance between the inner corners of the eyes and the tip of the nose is ~ 0.3803 times the vertical distance between the top of the head and the tip of the nose [1]. To account for variations in human populations, we fixed the upper limits of our search regions to $(prn_y + 0.3803 \times 1.5 \times |prn_y - v_y|)$, and their lower limits to $(prn_y + 0.3803 \times 0.33 \times |prn_y - v_y|)$, where prn_y is the vertical co-ordinate of the tip of the nose and v_y is the vertical co-ordinate of the

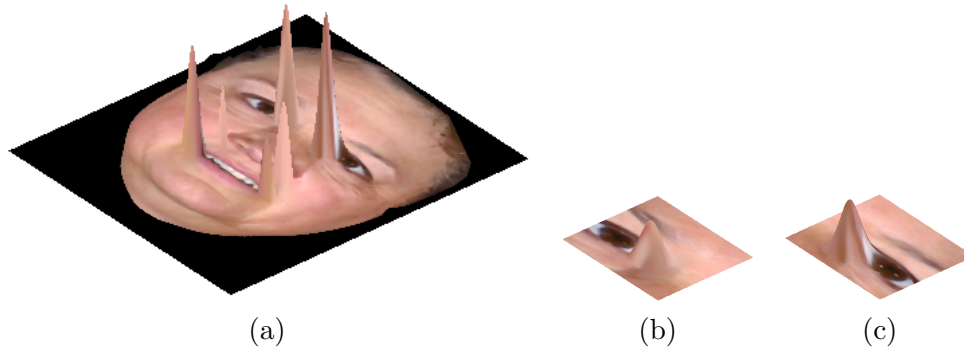


Figure 4.7: (a) The magnitude of the Gaussian curvature of the concave elliptic regions of a facial range image, (b) that of a smaller region defined to search for the right eye’s inner corner, and (c) that of a smaller region defined to search for the left eye’s inner corner. In each image the surface curvature is plotted as a 3D surface, with the facial texture warped onto it.

highest point of the 3D model.

For the horizontal limits of the two search regions, we employed the fact that for an average face, the ratio of the horizontal distance between the inner corners of the eyes to nose width (the distance between points $al-al$) is unity [1]. We searched between prn_x , the horizontal co-ordinate of the the tip of the nose and $(al_{x,left} + 0.5 \times |al_{x,left} - al_{x,right}|)$ for the curvature estimate of the inner corner of the subject’s left eye. Similarly, for the curvature estimate of the inner corner of subject’s right eye we searched between the horizontal limits $(al_{x,right} - 0.5 \times |al_{x,left} - al_{x,right}|)$ and prn_x . These search regions for the two inner eye corners are shown in Fig. 4.7(b) and Fig. 4.7(c), respectively. Within these two regions, we determined the locations of the points with the highest Gaussian curvatures as the curvature estimates for the inner corners of the eyes.

We obtained the final positions of points $en-en$ by further searching in a region of size 63×63 pixels about the curvature estimates, using a recently developed 2D+3D EBGM algorithm [135]. Briefly, in the 2D+3D EBGM algorithm the fiducial

points of interest are located manually on a set of 2D and 3D example images. Forty 2D and 3D Gabor coefficients (at 5 scales and 8 orientations) are computed for the manual fiducial points of the example images. The Gabor filters correspond to a carefully designed version [136, 137] that are widely used in the area of 2D face detection and recognition [120, 103, 104, 119]. We also carefully selected 68 images of 12 subjects from the training partition of our database, and 21 images of 13 subjects from the Remaining partition (Table 4.2), as example images for the 2D+3D EBGM algorithm. We included faces with various expressions, *e.g.*, open/closed eyes, and neutral/smiling with open/closed mouths in this set of example images.

In the 2D+3D EBGM algorithm, for a face with unknown fiducial points, the same 2D and 3D Gabor coefficients are computed for every point within a search window. The point, within this search window, that has Gabor coefficients most similar to the Gabor coefficients of any example image, is regarded as the final location of the fiducial point. We could have applied the 2D+3D EBGM algorithm directly to the search windows (Fig. 4.7(b) and Fig. 4.7(c)) that we employed to find the initial curvature estimates of the points *en-en*. However, this produced a significant number of false positives at the locations of the inner corners of the eyebrows, which have textural characteristics similar to the points *en-en*.

The center of the root of the nose (point m' in Fig. 4.2), was located at the algebraic mean of the positions of the automatically detected inner corners of the two eyes.

4.1.6.4 Outer Eye Corners (*ex-ex*)

To automatically locate the outer corners of the eyes (points *ex-ex* in Fig. 4.2), we employed the positions of the automatically detected inner corners of the eyes

(*en-en*). For an average human adult, the distance between the inner and the outer corner of an eye is approximately equal to the distance between the inner corners of the two eyes [1]. With this anthropometric information, we computed the initial estimate for the position of the outer corner of the subject’s left eye as $(en_{x,left} + |en_{x,left} - en_{x,right}|, (en_{y,left} + en_{y,right})/2)$, and that of the outer corner of the right eye as $(en_{x,right} - |en_{x,left} - en_{x,right}|, (en_{y,left} + en_{y,right})/2)$. We then used the 2D EBGM algorithm to search within a rectangular window of size 63×105 pixels about these initial estimates for the final positions of the points *ex-ex*. The set of example images for this 2D EBGM algorithm were the same as those for detecting the point *en-en*. Note, that since the outer corners of the eyes did not have distinct surface curvature characteristics, we used 2D EBGM instead of 2D+3D EBGM.

4.1.6.5 Mouth Corners (*ch-ch*)

We examined the curvature of facial surfaces regions located below the nose, and observed that for *all* faces, the outer corners of the mouth were distinct concavities, *i.e.*, regions of high positive mean curvature (H) (Fig. 4.8(a)). Hence, the peaks of mean curvature (H) in this region served as the initial estimates (curvature estimates) for the locations of the points *ch-ch*. To find these peaks of mean curvature, we defined appropriate search regions as follows.

We observed that for all faces (including those with beards, mustaches, and arbitrary facial expressions) regions of the upper lip and lower lip had elliptic Gaussian curvature (Fig. 4.8(b)). By detecting these upper and lower lip regions below the tip of the nose, we determined the vertical limits of the search regions for the curvature estimates of the corners of the mouth. Furthermore, we employed the locations of the automatically detected points *al-al* to horizontally constrain these

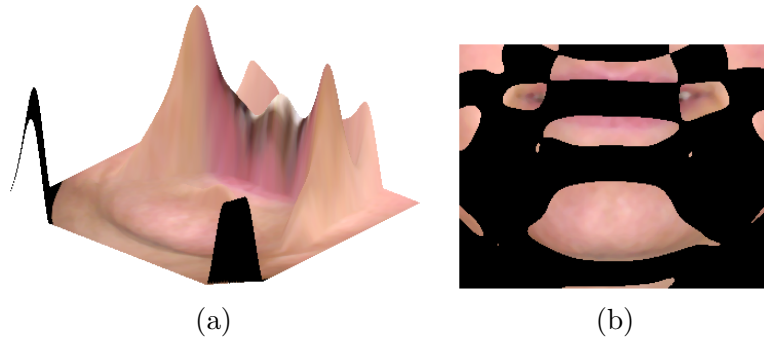


Figure 4.8: (a) The the mean curvature (H) of the mouth region of an example face plotted as a 3D surface, with the facial texture warped onto it. Notable are the distinct peaks at the corners of the mouth. (b) Regions below the nose with elliptic Gaussian curvature (non-black regions) of an example face. Notable are the regions of the upper and the lower lip.

search regions. For ch_{left} , we searched to the left of al_{left} between the horizontal positions $al_{x,left}$ and $(al_{x,left} + 0.7 \times |al_{x,left} - al_{x,right}|)$, and for ch_{right} we searched to the right of al_{right} between $al_{x,right}$ and $(al_{x,right} - 0.7 \times |al_{x,left} - al_{x,right}|)$.

Finally, we refined the positions for the corners of the mouth by searching in a window of size 33×93 pixels about the initial curvature estimates using the 2D+3D EBGGM algorithm. We employed the same set of example images that we had employed to detect the points *en-en*. This second stage, which refined the positions of the corners of the mouth, eliminated a significant number of errors that resulted in the first stage on open mouths.

4.2 Benchmark Algorithms

We compared the performance of the proposed 3D AnthroFace algorithm to that of three existing state-of-the-art automatic 3D face recognition algorithms. We implemented the eigensurfaces algorithm [19], the fishersurfaces algorithm [31, 38],



Figure 4.9: The sub-regions of the facial range images that were employed for the eigensurfaces and fishersurfaces 3D face recognition algorithms.

and a 3D face recognition algorithm based on the ICP procedure [24, 84] on our database. For the eigensurfaces and the fishersurfaces algorithms, we employed sub-sections of the facial range images between the pixels 147 and 553 along the vertical, and 38 and 478 along the horizontal (*e.g.*, Fig. 4.9). These limits corresponded to the extrema of the uppermost, bottommost, leftmost, and rightmost co-ordinates, respectively, of the 25 manually located anthropometric fiducial points across all faces in the database. In setting these limits to the extrema of all faces, we ensured that the main facial features were not excluded for any face.

Both the eigensurfaces and the fishersurfaces algorithms were trained and tested on exactly the same data sets, which were employed to train and test the AnthroFace algorithm. For the eigensurfaces algorithm, we learned 69 eigen directions that accounted for 99% of the variance of the data, and linearly projected all faces in the test data set onto these eigen directions. The final metric for the comparison of 3D facial surfaces in the eigen sub-space was the L_1 norm. For the fishersurfaces algorithm, we first reduced the dimensionality of the range images to 348 using principal component analysis (PCA) [70]. This was done to ensure that the within-class scatter matrix employed in the LDA computations was non-singular. We then learned 11 LDA directions from the 348 PCA features and projected all faces in the

test data set onto these LDA directions and compared them using the L_2 norm.

The 3D face recognition algorithm based on ICP did not require training. Hence, we implemented it only on the test data set and compared all faces in the probe set to all faces in the gallery set. We finely registered each pair of facial surfaces using the ICP algorithm. We implemented the ICP procedure on complete 3D facial surfaces, which were sub-sampled down to 0.1 times of their original magnifications. The rotation and translation matrices learned from the ICP procedure were employed to align the *original* sized 3D facial surfaces. These were finally compared by means of the partial Hausdorff distance metric [90].

4.3 Performance Evaluation

4.3.1 Data

We employed the Texas 3D Face Recognition Database (Chapter 3). This database contained 1149 2D and 3D images of 118 human subjects. We partitioned this database into a training data set and a test data set (Table 4.2). The training data set contained 360 randomly selected images of 12 subjects (30 images per subject) in neutral or expressive modes. For all the 3D face recognition algorithms that we developed, steps such as automatic facial fiducial point detection, classifier feature selection and classifier optimization were performed using the training data set only. The trained classifier was evaluated on the independent test data set, which did not overlap with the training data set.

The test data set included 768 images of 105 subjects. This test set was further partitioned into a gallery set and a probe set. Consistent with the evaluation protocol of the FRVT 2002 [118] and the FRGC 2005 [13], the gallery set contained one range image each of 105 subjects with a neutral facial expression. The probe

Partition	No. of Subjects	No. of Images		
		Neutral	Expressive	Total
Training	12	228	132	360
Test	Gallery	105	0	105
	Probes	95	480	663
Remaining	13	0	21	21

Table 4.2: A summary of the data partitions employed for developing 3D face recognition algorithms.

set contained another 663 images of 95 of the gallery subjects with a neutral or an arbitrary facial expression. In the probe set, the number of images of each subject varied from 1 to 55. In accordance with the widely accepted ‘closed universe’ model for the evaluation of face recognition algorithms [118], every subject in the probe data set was represented in the gallery data set.

After partitioning the entire database of 1149 images into the training and test data sets, 21 images of 13 subjects remained. All these were of faces with an arbitrary facial expression. We employed this Remaining set of images along with 68 images of 12 subjects from the training data set as example images in the 2D/2D+3D EBGm algorithms for detecting facial fiducial points.

4.3.2 Automatic Fiducial Point Detection

In order to evaluate the performance of the automatic facial fiducial point detection algorithms that we developed, we regarded the positions of the manually located points as the ‘ground truth’. We computed the positional errors between the automatically and manually located fiducial points along the x and y dimensions. As a comprehensive measure of the detection performance of the algorithm for each of the 10 fiducial points, we computed the standard deviations (σ_x and σ_y , respectively) of both types of errors for 1060 facial images. This set of images included all the

images in our database except for the 89 example images, which were employed in the EBGM fiducial point detection algorithms.

4.3.3 3D Face Recognition

We evaluated the verification performance of all 3D face recognition algorithms using the Receiver Operating Characteristic (ROC) methodology [50], and observed the values of the Equal Error Rates (EER) and the Areas Under the ROC Curves (AUC). The identification performance of the algorithms was evaluated using Cumulative Match Characteristic (CMC) curves, and the rank 1 Recognition Rates (RR) were observed. Statistical 95% confidence intervals for the EER, AUC, and the rank 1 RR values were obtained empirically using bootstrap sampling. All performance statistics were observed separately for neutral faces, for expressive faces, and for both types of faces in the probe data set.

4.4 Results and Discussion

4.4.1 Manual 3D AnthroFace

The EER, AUC and rank 1 RR values for the 3D AnthroFace algorithm based on 25 manually detected anthropometric fiducial points (Fig. 4.1(b)), and for the algorithm based on 25 arbitrary facial points (Fig. 4.1(c)) are presented in Tables 4.3(a), 4.3(b) and 4.3(c), respectively. The same Tables also contain the performance statistics for the benchmark eigensurfaces, fishersurfaces, and ICP algorithms. The ROC and the CMC curves for these algorithms are presented in Fig. 4.10 and Fig. 4.11, respectively.

For all faces in the probe data set the proposed 3D AnthroFace algorithm based on 25 manually located anthropometric fiducial points performed well ($EER =$

EER [Confidence Interval] %						
Algorithms	Neutral		Expressive		All	
Eigensurfaces	24.0	[21.3 26.3]	23.6	[19.2 26.2]	24.0	[21.8 26.6]
Fishersurfaces	8.11	[6.39 10.9]	3.60	[2.00 6.43]	6.69	[5.27 8.16]
ICP	7.97	[6.85 9.95]	9.92	[6.71 14.7]	9.03	[7.67 10.2]
3D AnthroFace (25 anthro)	0.84	[0.53 1.14]	1.58	[0.64 2.67]	1.00	[0.64 1.45]
3D AnthroFace (25 arbitrary)	8.78	[6.58 10.9]	5.10	[3.37 8.61]	7.65	[6.16 10.1]

(a)

AUC [Confidence Interval] $\times 10^{-2}$						
Algorithms	Neutral		Expressive		All	
Eigensurfaces	16.7	[14.1 18.8]	14.9	[11.9 18.0]	16.3	[14.8 18.7]
Fishersurfaces	2.88	[2.12 3.71]	1.32	[0.32 2.49]	2.40	[1.79 2.96]
ICP	2.97	[2.16 4.12]	4.39	[2.18 7.80]	3.44	[2.56 4.27]
3D AnthroFace (25 anthro)	0.07	[0.03 0.12]	0.08	[0.04 0.12]	0.08	[0.04 0.12]
3D AnthroFace (25 arbitrary)	3.00	[2.23 3.95]	2.08	[0.68 4.61]	2.70	[2.03 3.65]

(b)

Rank 1 RR [Confidence Interval] %						
Algorithms	Neutral		Expressive		All	
Eigensurfaces	58.1	[54.0 62.7]	52.5	[45.4 60.1]	56.6	[52.9 60.2]
Fishersurfaces	91.7	[89.4 94.0]	95.1	[91.8 97.8]	92.6	[90.6 94.4]
ICP	88.5	[85.6 91.5]	86.3	[80.9 91.0]	87.9	[85.5 90.2]
3D AnthroFace (25 anthro)	98.8	[97.7 99.6]	95.6	[92.4 98.4]	97.9	[96.8 98.9]
3D AnthroFace (25 arbitrary)	86.0	[82.9 89.0]	91.3	[87.4 95.1]	87.5	[84.9 89.9]

(c)

Table 4.3: The observed (a) EER, (b) AUC, and (c) rank 1 RR values and their 95% confidence intervals for the eigensurfaces, fishersurfaces, ICP, and the 3D AnthroFace algorithms based on 25 manually located anthropometric points and 25 arbitrary facial points.

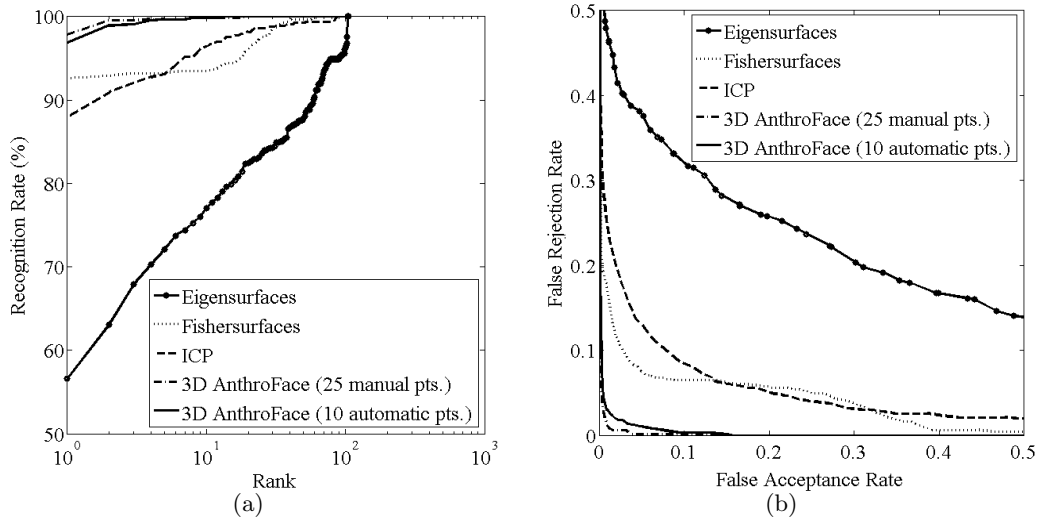


Figure 4.10: (a) The semi-log CMC curves, and (b) the ROC curves for the performance of the benchmark 3D face recognition algorithms, the AnthroFace algorithm that employed manually located 25 points, and the 3D AnthroFace algorithm that employed 10 automatically located points.

1% and rank 1 $RR = 97.9\%$). It also performed significantly better than the three holistic benchmark eigensurfaces, fishersurfaces, and ICP algorithms (Table 4.3, Fig. 4.10). These results clearly establish the potential of our proposed 3D AnthroFace algorithm. The performance of the manual 3D AnthroFace algorithm can be also regarded as an upper bound on the expected performance of the automated 3D AnthroFace algorithm for perfectly located anthropometric facial fiducial points. The results also point towards the superiority of local feature based 3D face recognition algorithms over holistic techniques. Interestingly, some studies in the cognitive sciences suggest that similar to the 3D AnthroFace algorithm, humans also process facial information using relational information between parts of the face [138]. They acquire this information by means of sequential eye movements between the different facial features [139].

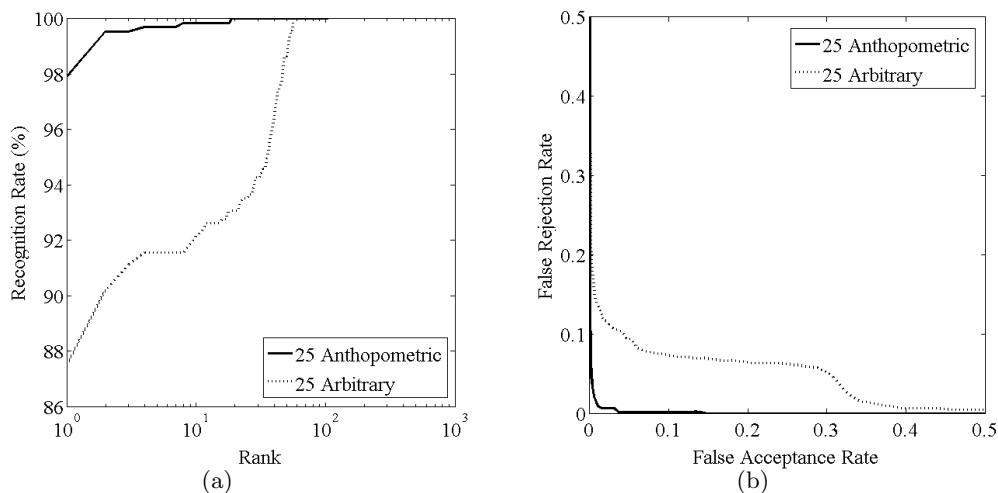


Figure 4.11: (a) The semi-log CMC curves, and (b) ROC curves for the performance of the 3D AnthroFace algorithm based on 25 manually located facial fiducial points and the algorithm based on 25 arbitrary facial points.

It may be non-trivial to automatically and accurately locate all the 25 anthropometric facial fiducial employed in the manual 3D AnthroFace algorithm. Nonetheless, its high performance is an indication of the potential of 3D face recognition algorithms that are carefully designed to incorporate knowledge about the structural diversity and statistical distribution of anthropometric measurements of human faces. This is further supported by the fact that the 3D AnthroFace algorithm based on 25 anthropometric fiducial points performed significantly better than the algorithm based on 25 arbitrary facial points (Table 4.3 and Fig. 4.11).

We attempted to gain further insight into the discriminatory structural information contained in the variable anthropometric facial proportions that we initially selected (Table 4.1). We separately ranked the anthropometric Euclidean and geodesic distance between the 25 manually located facial fiducial points in descending order of their individual Fisher’s ratio [70] values. The 20 most discriminatory

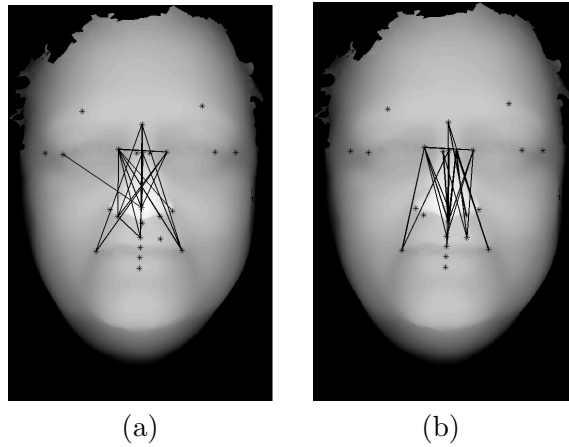


Figure 4.12: The 20 most discriminatory facial (a) Euclidean, and (b) geodesic distance features. The geodesic distances are symbolically depicted by straight lines. In reality, they are along the surface of the face.

facial Euclidean and geodesic distances are presented in Fig. 4.12. Interestingly, these distances were predominantly associated with the nasal region of the face, as were a majority (12 out of 23) of the variable anthropometric facial proportions (Table 4.1). Furthermore, 17 (O_{10} , O_{12} , N_1 , N_6 , N_7 , N_8 , N_{15} , N_{16} , N_{30} , N_{31} , N_{33} , L_1 , L_4 , L_5 , L_6 , L_7 , and L_{14}) of the 23 facial proportions that we selected have also been reported to be significantly different between the two sexes by Farkas [14], and one (N_7) has been reported to be significantly different for various ethnic groups [15]. Clearly, all these factors contributed to the success of the 3D AnthroFace algorithm.

We compared the performance of the proposed 3D AnthroFace algorithm, which employed Euclidean and geodesic distances between the 25 anthropometric facial fiducial points to a similar algorithm that employed *only* Euclidean distances between the same fiducial points. We observed that for expressive faces, the verification performance of the algorithm that employed both Euclidean and geodesic

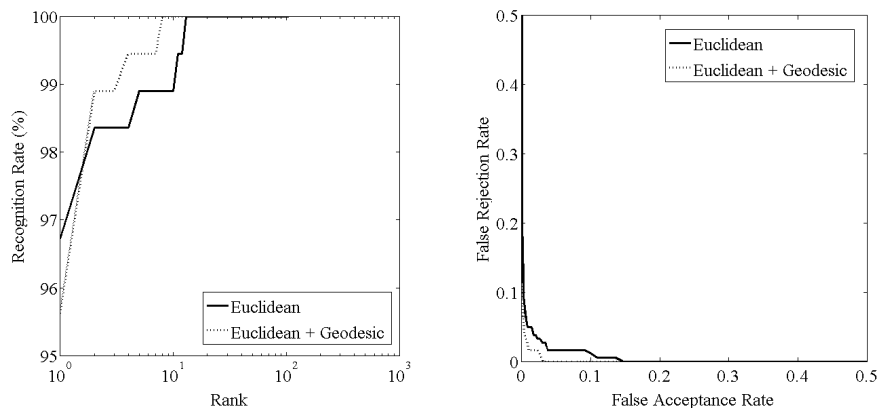


Figure 4.13: (a) The semi-log CMC curves, and (b) ROC curves for the performance of 3D AnthroFace algorithm that employed only Euclidean distances and the 3D AnthroFace algorithm that employed both the Euclidean and the geodesic distances. These curves were for expressive faces only.

distances ($AUC = 0.0008$, with a confidence interval of $[0.0004 \ 0.0012]$), was significantly better than the performance of the algorithm that employed only anthropometric Euclidean distances ($AUC = 0.0015$, with a confidence interval of $[0.0009 \ 0.0023]$) (Fig. 4.13(a)). This suggests that facial geodesic distances may be useful for expression invariant 3D face recognition. It further supports Bronstein *et al.*'s [40] proposition that different facial expressions may be isometric deformations of the facial surface.

The performance of the 3D AnthroFace algorithm based on the reduced set of 10 manually located anthropometric fiducial points (Fig. 4.2) is presented in Table 4.4. For all faces in the probe set, this algorithm resulted in $EER = 1.68\%$ and $AUC = 0.0014$. By comparison, the 3D AnthroFace algorithm based on 25 manually located fiducial points performed slightly better with $EER = 1.00\%$ and $AUC = 0.0008$ (Table 4.4(b)). The recognition performance of the two algorithms was comparable.

EER [Confidence Interval] %			
Algorithms	Neutral	Expressive	All
3D AnthroFace (25 manual)	0.84 [0.53 1.14]	1.58 [0.64 2.67]	1.00 [0.64 1.45]
3D AnthroFace (10 manual)	1.10 [0.65 1.96]	2.34 [1.01 3.10]	1.68 [1.10 2.24]
3D AnthroFace (10 automatic)	1.65 [1.11 2.28]	2.81 [1.27 4.30]	1.98 [1.37 2.88]

(a)

AUC [Confidence Interval] $\times 10^{-2}$			
Algorithms	Neutral	Expressive	All
3D AnthroFace (25 manual)	0.07 [0.03 0.12]	0.08 [0.04 0.12]	0.08 [0.04 0.12]
3D AnthroFace (10 manual)	0.12 [0.04 0.30]	0.18 [0.05 0.30]	0.14 [0.07 0.25]
3D AnthroFace (10 automatic)	0.14 [0.08 0.23]	0.25 [0.11 0.42]	0.18 [0.11 0.28]

(b)

Rank 1 RR [Confidence Interval] %			
Algorithms	Neutral	Expressive	All
3D AnthroFace (25 manual)	98.8 [97.8 99.6]	95.6 [92.4 98.4]	97.9 [96.8 98.9]
3D AnthroFace (10 manual)	98.8 [97.7 99.6]	96.2 [93.4 98.9]	98.0 [97.0 98.9]
3D AnthroFace (10 automatic)	97.3 [95.8 98.5]	95.6 [92.4 98.4]	96.8 [95.3 98.0]

(c)

Table 4.4: The observed (a) EER, (b) AUC, and (c) rank 1 RR values and their 95% confidence intervals for the 3D AnthroFace algorithms based on 25 manually located Fiducial points, 10 manually located fiducial points, and 10 automatically located fiducial points.

4.4.2 Automatic 3D AnthroFace

The standard deviations of the positional errors of the 10 automatically located anthropometric facial fiducial points (Fig. 4.2) from their respective manual ‘ground

Fiducial Point	Error σ_x	Error σ_y	Radial Error σ
<i>prn</i>	3.265	5.250	6.183
<i>al_{left}</i>	2.254	5.172	5.642
<i>al_{right}</i>	2.495	5.144	5.718
<i>en_{left}</i>	4.650	3.892	6.064
<i>en_{right}</i>	4.232	4.199	5.962
<i>m'</i>	4.234	5.658	7.067
<i>ex_{left}</i>	5.609	4.018	6.900
<i>ex_{right}</i>	6.643	4.326	7.930
<i>ch_{left}</i>	6.088	2.917	6.751
<i>ch_{right}</i>	6.174	3.267	6.985

Table 4.5: The standard deviations of the positional errors of the 10 automatically located anthropometric facial fiducial points, in pixel units. The distance between two adjacent pixels was 0.32 mm.

truth' locations are presented in Table 4.5. All 10 anthropometric facial fiducial points were detected very accurately. The radial standard deviation of error for each of the 10 fiducial points was less than 8 pixels or 2.65 mm. The average radial standard deviation across all the 10 fiducial points was 6.52 pixels or 2.09 mm.

The tip of the nose was detected reliably and accurately for *all* faces in our database as the point in the central region of the face with the highest convex elliptic Gaussian curvature. This characteristic property also corresponds well with the intuitive definition of the *tip* of the nose. The outer corners of the eyes (*ex-ex*) were detected least accurately, followed by the corners of the mouth (*ch-ch*). A large majority of the false positives for the corners of the eyes were located at the edges of the irises. This was not surprising as the edges of the irises can have textural properties similar to the corners of the eyes. We also observed that many of the false positive detections for the mouth corners occurred on smiling faces.

To automatically detect each of the fiducial points *prn*, *en-en*, *ex-ex* and *ch-ch* we employed two stages. For each point, we obtained an initial estimate of

its location using the first stage, and then searched within a window surrounding this estimate for the final location of the point. The 2D histograms of the errors for the two stages (ICP based and curvature based, respectively) of the algorithm to detect the tip of the nose are presented in Fig. 4.14(a). The histograms for the two stages (surface curvature based and 2D+3D EBGGM based) of the inner eye corner (*en-en*) detection algorithm, and those of the mouth corner (*ch-ch*) detection algorithm are shown in Fig. 4.14(b) and Fig. 4.14(d), respectively. The histograms for the errors of the two stages of the algorithm for detection of the outer corners of the eyes (*ex-ex*) are shown in Fig. 4.14(c). It is clear that for each of these fiducial points, the second detection stage helped to considerably reduce the errors of the first stage. It is also instructive to note that for the points *en-en*, *ex-ex* and *ch-ch* a combination of information from the 2D and 3D images resulted in the best overall detection performance.

In the future it may be interesting to study the sensitivity of our proposed fiducial point detection algorithms to small in and out of plane rotations of the 3D face. In our analysis, we computed the curvature of the frontal upright facial surfaces from range images. Hence, it may also be useful to employ techniques to compute surface curvature directly from arbitrarily oriented 3D point clouds or 3D meshes, which might eliminate the need to operate on frontal upright faces.

Lastly, the performance of the 3D AnthroFace algorithm based on these 10 automatically detected fiducial points is presented in Table 4.4 and Fig. 4.10. Overall, the verification performance ($EER = 1.98\%$ and $AUC = 0.0018$) of this algorithm was statistically the same as that of the 3D AnthroFace algorithm based on 10 manually detected fiducial points ($EER = 1.68\%$ and $AUC = 0.0014$). The recognition performance of the two algorithms was also similar (Table 4.4(c)). It

can be concluded that the 3D AnthroFace algorithm is robust to detection errors with a minimum radial $\sigma = 2.65$ mm.

The verification performance of the 3D AnthroFace algorithm based on the 10 automatically detected points was slightly lower than the verification performance of the 3D AnthroFace algorithm based on 25 manually detected fiducial points (Table 4.4(a) and (b), and Fig. 4.10(b)). On closely examining Table 4.4 it is evident that this decrease in performance was due to the reduction of anthropometric fiducial points from 25 to 10, rather than a result of errors in detection of the fiducial points. The recognition performance of the two 3D AnthroFace algorithms based on 25 manually detection fiducial points, and 10 automatically detected fiducial, respectively, was comparable (Table 4.4(c) and Fig. 4.10(a)). The 3D AnthroFace algorithm based on 10 automatically detected facial fiducial points also performed significantly better than the benchmark eigensurfaces, fishersurfaces, and ICP algorithms (Fig. 4.10).

In conclusion, our proposed 3D AnthroFace algorithm presents a number of novel contributions to the field of 3D face recognition. In the larger context, we presented a novel way of thinking about the problem. Rather than employing general purpose pattern recognition algorithms for the task, we introduced the concept of employing domain specific knowledge about the structural diversity of faces to design effective 3D face recognition algorithms. We presented a practical method for isolating this knowledge from the scientific discipline of facial anthropometry and developed successful, fully automatic algorithms based on this knowledge. We obtained an upper bound for the expected performance of the proposed algorithm for manually detected facial fiducial points and also demonstrated its performance for automatically detected points.

In the process, we also made significant strides towards solving the largely unexplored and challenging problem of reliable and accurate 2D+3D facial fiducial point detection. We envision the detection algorithms to be useful not only for 3D face recognition, but also for other disciplines that require facial anthropometric measurements including facial surgical planning and computer graphics. Unlike many of the previously reported studies of automatic facial fiducial point detection, we systematically assessed the performance of our algorithms against manually detected points and verified their reliability and accuracy.

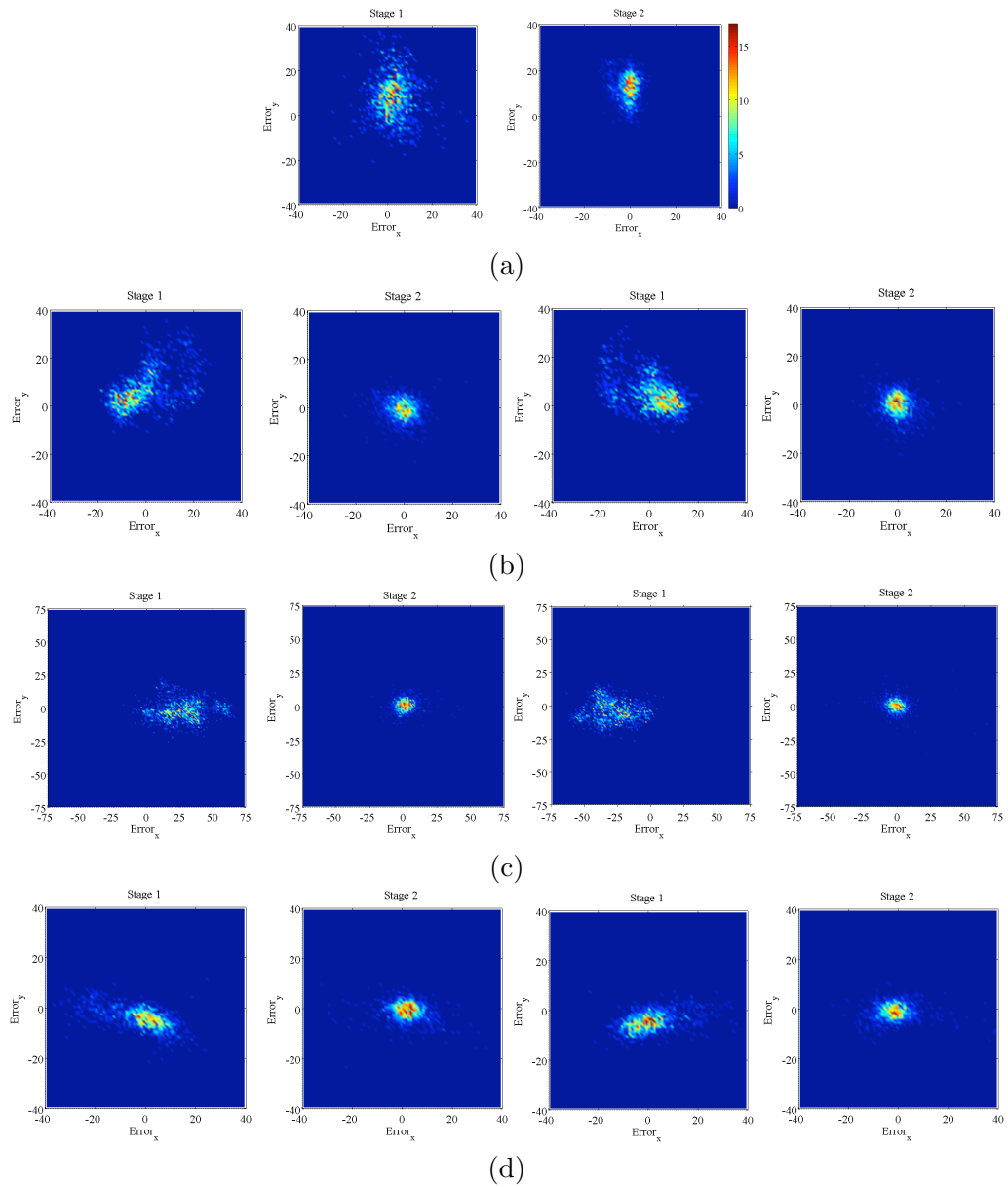


Figure 4.14: The 2D histograms for the errors for the two stages of the detection algorithms for (a) the tip of the nose, (b) the inner corner of the left eye (2 left images) and inner corner of the right eye (2 right images), (c) the outer corner of the left eye (2 left images) and outer corner of the right eye (2 right images), and (d) the left (2 left images) and right (2 right images) corners of the mouth.

Chapter 5

3D Face Recognition by Assessing Structural Similarity

The surface of the human face can be regarded as a *free form* 3D object that is neither planar nor naturally quadric [77]. A class of 3D face recognition algorithms exists, wherein two free form 3D facial surfaces are aligned and their shapes are compared by means of a suitable distance metric. This paradigm can be regarded as a template matching procedure, wherein the templates are the 3D facial surfaces.

The existing approaches for 3D face recognition based on this philosophy [20, 33, 22, 78, 23, 34, 80, 81, 82, 83, 27, 84, 85, 86, 24, 35, 87, 88, 29] essentially employ two steps. The first of these is to align the two facial surfaces. This normally proceeds in two stages, namely *coarse* alignment and *fine* alignment. For coarse alignment, each 3D model is individually transformed to a canonical (frontal upright in most cases) pose. As a consequence, the two models to be compared become coarsely aligned by virtue of being in the same canonical pose. Fine alignment involves further alignment of the two coarsely aligned 3D facial surfaces. The

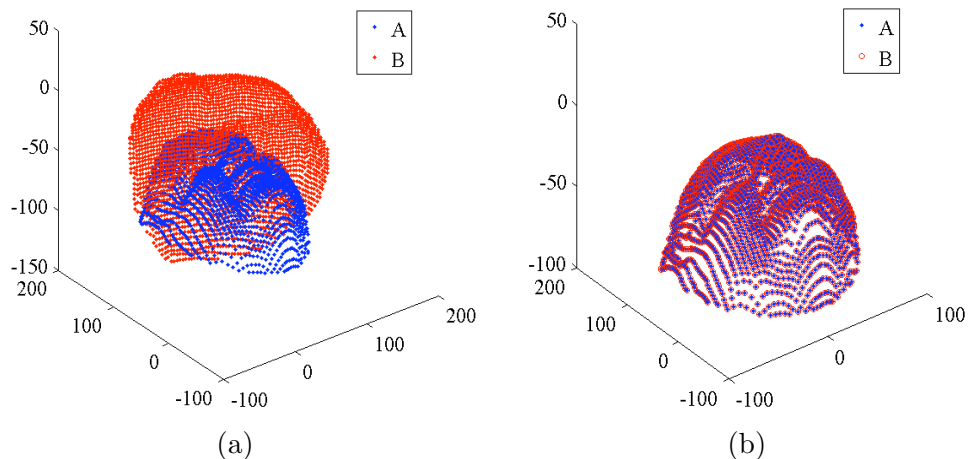


Figure 5.1: An original 3D facial point cloud A and its rotated version B (a) before, and (b) after fine alignment using the ICP algorithm.

primary algorithm employed to finely align 3D facial models is the iterative closest point (ICP) algorithm [79], wherein one 3D facial model is rotated and translated in space, in an iterative manner, until its distance from the other model converges to a minimum. Since the ICP algorithm is currently the *de facto* standard for all 3D face recognition algorithms that belong to this class of algorithms, we describe its mathematical details in the following section.

5.1 Existing Approaches

5.1.1 Iterative Closest Point Algorithm

Given two sets of 3D points $A = \{a_1, a_2, a_3 \dots a_M\}$ and $B = \{b_1, b_2, b_3 \dots b_N\}$, containing M and N points, respectively, we wish to rigidly align the point set B to the point set A (*e.g.*, Fig. 5.1). In other words, we wish to find the orthogonal rotation matrix R and the translation matrix T which minimize the distance between

the two point sets. Let $A' = \{a'_1, a'_2, a'_3 \dots a'_N\}$ be the set of points in A that are closest in space (in terms of the L_2 norm) to the points in set B . The optimization problem for ICP is formulated as a least squares problem of the form

$$\min_{R,T} f(R, T) = \min_{R,T} \sum_{i=1}^N \|a'_i - (Rb_i + T)\|^2 \quad (5.1)$$

$$R^*, T^* = \arg \min_{R,T} f(R, T) \quad (5.2)$$

subject to the constraints $RR^T = I$ and $\det(R) = 1$.

On substituting a'_i and b_i with $\alpha_i = a'_i - \bar{a}'$ and $\beta_i = b_i - \bar{b}$, respectively, where $\bar{a}' = \frac{1}{N} \sum_{i=1}^N a'_i$ and $\bar{b} = \frac{1}{N} \sum_{i=1}^N b_i$, equations (5.1) and (5.2) reduce to

$$\min_R f(R) = \min_R \sum_{i=1}^N \|\alpha_i - R\beta_i\|^2, \quad (5.3)$$

$$R^* = \arg \min_R f(R), \quad (5.4)$$

respectively, and the translation matrix is computed as

$$T^* = \bar{a} - R^*\bar{b}. \quad (5.5)$$

Suppose that C and D are matrices of size $3 \times N$, with 3D points α_i and β_i , respectively, along their columns. The solution to equation (5.3) [140, 141], can be obtained using the singular value decomposition of the matrix

$$H = CD^T = U\Lambda V^T, \quad (5.6)$$

where U and V are orthogonal matrices of size 3×3 and Λ is a diagonal matrix of

size 3×3 , as

$$R^* = UPV^T, \quad (5.7)$$

where P is a diagonal matrix with values $\{1, 1, \det(UV^T)\}$ along the principal diagonal.

The iterative optimization procedure of the ICP algorithm begins by assigning the initial values $R_0 = I$ and $T_0 = [0 \ 0 \ 0]^T$ to the overall rotation and translation matrices, respectively. At each k^{th} iteration, the set of points A'_k that are closest to the points in set $R_k B + T_k$ are found. Equations (5.7) and (5.5) are employed to compute the new values for R^* and T^* , and the overall rotation and translation matrices are updated as

$$R_{k+1} = R_k R^* \quad (5.8)$$

$$T_{k+1} = R^* T_k + T^*, \quad (5.9)$$

until the mean squared error between the two 3D models converges to a minimum or some other termination criteria are satisfied. The ICP procedure is normally applied to coarsely aligned 3D models. This makes the optimization procedure less susceptible to being trapped in a local minimum.

5.1.2 Existing Distance Metrics

After the two 3D facial models are aligned, the dissimilarity between their shapes is computed by means of a suitable distance metric. In the existing literature on 3D face recognition, three main distance metrics have been employed for this task. The simplest of these is the mean squared error (*MSE*) metric applied to the z values of two aligned facial range images [20, 33, 22, 78, 23]. Suppose Z_A and Z_B are two facial range images, both of the same size $R \times S$ pixels. The *MSE* between them is

defined as

$$MSE = \frac{1}{RS} \sum_{i=1}^R \sum_{j=1}^S [z_A(i, j) - z_B(i, j)]^2 \quad (5.10)$$

where $z_A(i, j)$ and $z_B(i, j)$ are the z values of the two images at the same image location indexed by (i, j) .

Two other metrics have been commonly employed for comparing the shapes of aligned facial surfaces. They are the point-to-closest-point mean squared error (MSE_{CP}) [34, 46, 24, 35], and the partial Hausdorff distance (pH) [82, 84]. Given two sets of 3D points $A = \{a_1, a_2, a_3 \dots a_M\}$ and $B = \{b_1, b_2, b_3 \dots b_N\}$, containing M and N points, respectively, the directed $MSE_{CP}(A \rightarrow B)$ from set A to set B is defined as

$$MSE_{CP}(A \rightarrow B) = \frac{1}{M} \sum_{i=1}^M \|a_i - b'_i\|^2, \quad (5.11)$$

where $b'_i \in B$ is the closest point to $a_i \in A$. The directed $MSE_{CP}(B \rightarrow A)$ is similarly computed. This distance metric is the same as the distance that is computed between a pair of 3D models during each iteration of the ICP algorithm. The undirected MSE_{CP} between the two surfaces is defined as

$$MSE_{CP} = \max\{MSE_{CP}(A \rightarrow B), MSE_{CP}(B \rightarrow A)\}. \quad (5.12)$$

The partial Hausdorff distance metric is computed in a manner similar to the MSE_{CP} metric [90]. Given two sets of 3D points $A = \{a_1, a_2, a_3 \dots a_M\}$ and $B = \{b_1, b_2, b_3 \dots b_N\}$, containing M and N points, respectively, the partial directed Hausdorff distance ($h_P(A \rightarrow B)$) is computed by sorting in ascending order the distances of all points $a_i \in A$ to their closest points $b'_i \in B$ and by finding the P^{th}

largest distance in the ordered set

$$h_P(A \rightarrow B) = P^{th} \max_{a \in A} \|a_i - b'_i\|. \quad (5.13)$$

The directed distance $h_Q(B \rightarrow A)$ is similarly computed by considering the Q^{th} largest distance from point set B to point set A, and the undirected partial Hausdorff distance (pH) between the two point sets is defined as

$$pH = \max\{h_P(A \rightarrow B), h_Q(B \rightarrow A)\}. \quad (5.14)$$

The pH distance metric is more robust to outliers produced by holes, noise and occlusions, than are MSE and MSE_{CP} , as it rejects a fraction of the largest distances between the pair of 3D models. The selection of the parameters P and Q that govern the fraction of the distances that are rejected between the 3D models, is often determined using prior knowledge about the expected extent of overlap between them.

The MSE metric is computationally and analytically simple, but is very sensitive to rotational and translational errors between facial range images. Both the MSE_{CP} and the pH distance metrics differ from MSE , in that they do not assume perfect correspondence between the two 3D models. For MSE_{CP} and pH , correspondence information between the two models is determined by searching for pairs of closest points on the two models, using a computationally expensive procedure ($O(MN)$ for two models with M and N points, respectively).

Overall, these metrics have yielded poor results for comparing the shapes of 3D faces that are coarsely aligned [20, 33, 22]. This is due to their high sensitivity to registration errors between the models. Currently, the successful 3D face recognition

algorithms that employ one or the other of these metrics to compare the shapes of 3D faces, first *finely* align the faces using the ICP algorithm [34, 46, 24, 35, 82, 84]. The performance of these ICP based algorithms is reported to decrease significantly for faces with arbitrary (other than neutral) facial expressions [86, 24, 35, 29]. Some researchers have suggested employing sub-parts of the face that deform less than the others during changes in facial expression (*e.g.*, the nasal region), as the solution to the problem of facial expressions [87, 88]. However, this has been reported to lower the overall recognition performance of ICP based 3D face recognition algorithms, relative to when they employ the entire facial surface.

Within this class of 3D face recognition algorithms that directly compare the shapes of facial surfaces, we propose a novel algorithm called 3D Structural Similarity Face Recognition (3D SSIMFace) [142]. The proposed 3D SSIMFace algorithm computes the structural similarity between *coarsely* aligned 3D facial surfaces using a recently developed image similarity index called the complex-wavelet structural similarity (CW-SSIM) index [17]. The proposed algorithm is significantly superior to existing ICP based algorithms that employ the $MSE/MSE_{CP}/pH$ metrics in a number of respects. First, since the CW-SSIM index is robust to small translations and rotations between images, the 3D SSIMFace algorithm completely eliminates the need to *finely* align facial models using the ICP algorithm, before they are compared. Second, the CW-SSIM metric is inspired by the pattern recognition capabilities of humans [17, 136, 137]. Furthermore, it computes the *structural* similarity between images, by comparing their relative phases and is significantly more accurate than the general purpose $MSE/MSE_{CP}/pH$ metrics at discriminating between the 3D facial surfaces of different human subjects.

We also propose a number of simple representations of the 3D facial surfaces,

which when employed in conjunction with the CW-SSIM index are highly effective at the task of 3D face recognition. These representations include edge maps derived from the zero-crossings of Laplacian of Gaussian (LOG) filtered facial range images, sub-sections of the facial range images that contain the eyes and the nose regions only, and local regions of the facial range and edges images centered about five anthropometric facial fiducial points. We empirically demonstrate the higher accuracy of our proposed 3D SSIMFace algorithm relative to algorithms based on ICP and $MSE/MSE_{CP}/pH$, and other benchmark algorithms including eigensurfaces [19] and fishersurfaces [31, 38]. Furthermore, we quantify the extent to which each of the CW-SSIM, MSE , the MSE_{CP} and pH indices is robust to geometric transformations of 3D facial models, for the task of 3D face recognition.

5.2 3D Structural Similarity Face Recognition (SSIM-Face)

5.2.1 Complex-wavelet Structural Similarity Index

The CW-SSIM index [17, 143] is an extension of the highly successful structural similarity index (SSIM) [16, 144] to the complex-wavelet domain. The SSIM index was developed for predicting human preferences in evaluating image quality. However, the utility of the CW-SSIM and SSIM indices is not limited to perceptual image quality assessment, and they can be employed for pattern recognition tasks as well. SSIM compares local structure and variation about a pixel, independent of the mean intensity and contrast of images. It operates in the spatial domain and has been shown to greatly surpass MSE both at predicting perceptual image quality and at pattern recognition tasks [144, 17, 145]. The SSIM index, however, is not

very robust to small registration errors and scale differences between images. The CW-SSIM index was designed to improve these properties of the SSIM index [17]. Previous successful applications of the CW-SSIM index include accurate recognition of handwritten digits [17], and the assessment of observer variability in human segmentations of medical images [146].

To compute the CW-SSIM index between a pair of images, the images are first decomposed using a family of symmetric complex wavelets that are dilated/contracted and translated versions of a ‘mother wavelet’ [17, 143]. In one dimension (1D), the mother wavelet, which is the modulated version of a slowly varying and symmetric low-pass filter $g(t)$, and can be written as

$$w(t) = g(t)e^{j\omega_c t}, \quad (5.15)$$

where ω_c is the center frequency of the modulating band-pass filter. The family of wavelets can be derived from the mother wavelet as

$$w_{s,p}(t) = \frac{1}{\sqrt{s}} w\left(\frac{t-p}{s}\right) = \frac{1}{\sqrt{s}} g\left(\frac{t-p}{s}\right) e^{j\omega_c(t-p)/s}, \quad (5.16)$$

where $s \in R^+$ is the scale factor, and $p \in R$ is the translation factor. The continuous wavelet transform of a given real 1D signal $x(t)$ can be written as [147]

$$X(s,p) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) \sqrt{s} G(s\omega - \omega_c) e^{j\omega p} d\omega, \quad (5.17)$$

where $X(\omega)$ and $G(\omega)$ are the Fourier transforms of $x(t)$ and $g(t)$, respectively. The wavelet transform of a real 2D signal can be analogously computed. The discrete wavelet coefficients are sampled versions of the continuous wavelet transform.

Let $\mathbf{c}_A = \{c_{A,i} | i = 1, \dots, N\}$ and $\mathbf{c}_B = \{c_{B,i} | i = 1, \dots, N\}$, respectively, be the

two sets of discrete wavelet coefficients extracted at the same spatial location in the same wavelet sub-bands of the two images being compared. The CW-SSIM index between the images is defined as

$$\tilde{S}(\mathbf{c}_A, \mathbf{c}_B) = \frac{2 \left| \sum_{i=1}^N c_{A,i} c_{B,i}^* \right| + K}{\sum_{i=1}^N |c_{A,i}|^2 + \sum_{i=1}^N |c_{B,i}|^2 + K}, \quad (5.18)$$

where c^* denotes the complex conjugate of c , and K is a small positive constant. The CW-SSIM index is computed locally using a 7×7 sliding window that moves across each of the wavelet subbands. Finally, the resulting CW-SSIM index map for a particular subband is combined into a scalar similarity measure using a weighted summation. The weighting function is obtained using a Gaussian profile located at the center of the image with a standard deviation equaling a quarter of the image size. The CW-SSIM scores from the various subbands are averaged together to obtain a single CW-SSIM similarity value for a pair of images. The value of the CW-SSIM index ranges from 0 to 1, where 1 denotes perfect similarity between the two images and 0 denoted complete dissimilarity.

In order to understand the CW-SSIM index better, equation (5.18) can be decomposed into the product of two components as

$$\tilde{S}(\mathbf{c}_A, \mathbf{c}_B) = \frac{2 \sum_{i=1}^N |c_{A,i}| |c_{B,i}| + K}{\sum_{i=1}^N |c_{A,i}|^2 + \sum_{i=1}^N |c_{B,i}|^2 + K} \cdot \frac{2 \left| \sum_{i=1}^N c_{A,i} c_{B,i}^* \right| + K}{2 \sum_{i=1}^N |c_{A,i} c_{B,i}^*| + K}. \quad (5.19)$$

The first component in equation (5.19) is completely determined by the magnitudes of the wavelet coefficients. It achieves a value of 1 if $|c_{A,i}| = |c_{B,i}|$ for all i . This term is equivalent to the SSIM index applied to the magnitudes of the complex wavelet coefficients and is related to CW-SSIM's insensitivity to luminance and contrast variations between images [17, 143].

The second component in equation (5.19) is completely determined by the relative phases of the complex wavelet coefficients of the two images. It achieves a value of 1 if for all i , the phases of $c_{A,i}$ and $c_{B,i}$ differ by a constant value. This phase component is related to CW-SSIM’s insensitivity to translation, rotation and scale variations between images. To understand this, consider a signal shifted in the time/space domain. The magnitudes of the Fourier transform coefficients of the signal are unchanged as a result of the shift. However, the phases of its Fourier transform coefficients undergo a linear shift. Similarly, for the case of the complex wavelet transform, small consistent translations of an image (compared to the size of the wavelet envelope), produce an approximately constant shift in the phase of all coefficients of a particular wavelet subband [17, 143]. Hence, the magnitude component of the CW-SSIM index computed for an image and its slightly translated version is 1, and the phase component is ~ 1 , making the CW-SSIM index insensitive to small translational errors. Furthermore, the net effect of small scale changes and rotations can be approximated as small translations [17, 143]. Hence, the CW-SSIM index is simultaneously insensitive to small translations, rotations and scales differences between images.

A number of characteristics of the CW-SSIM index make it well suited for computing the structural similarity between 3D facial surfaces. First, as mentioned above, CW-SSIM is insensitive to small translations, rotations and scale differences between images [17]. Hence, when applying the CW-SSIM index to coarsely aligned 3D facial surfaces, which have small registration errors between them, it is unnecessary to further finely align them using the computationally expensive ICP procedure. We will also demonstrated this empirically. For the same reason, CW-SSIM is also attractive for comparing pairs of 3D facial surfaces that have been finely aligned, as

one can always expect small misalignment errors to exist even after the ICP procedure. Such errors are more pronounced when the ICP algorithm fails to converge to the global minimum and does not correctly register the facial surfaces. Similarly, the CW-SSIM index is also a good choice for comparing 3D faces acquired at slightly different scales, *e.g.*, when the 3D acquisition devices are not precisely calibrated.

Second, the CW-SSIM index explicitly compares the phases of facial range images, which is important for assessing the structural similarity between them. It has been shown that much of the structural information of local image features of typical natural images (such as those of human faces) exists in the relative phase patterns of their wavelet coefficients [148]. Similar ideas of correlation between wavelet phases of images have been successfully employed in a number of applications of machine vision and image processing [149, 150], including image alignment [151], feature localization [152, 153], texture description [154], and blur detection [147]. The idea of comparing the relative phases of facial images has also been successfully employed previously for 2D and 3D facial feature detection [120, 100, 119, 104, 103, 135].

Third, the CW-SSIM index employs a wavelet decomposition. This is inspired by the remarkable pattern recognition and hence face recognition capabilities of the human visual system. Similar to the CW-SSIM index, neurons in the primary visual cortex have been successfully modeled as multi-scale band-pass oriented filters, that decompose an image into multiple channels [136, 137]. It has also been suggested that these multiple channels representations may be useful for the numerous pattern recognition tasks (including that of facial recognition) that humans perform successfully [155].

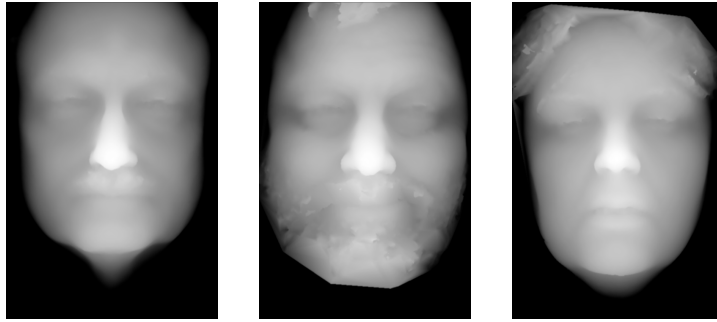


Figure 5.2: Examples of facial range images in the Texas 3D Face Recognition Database.

5.2.2 3D Facial Representations

For the proposed 3D SSIMFace algorithm, we compared the shapes of coarsely aligned 3D faces. Each face was in a canonical frontal upright pose, with its nose tip located at the origin of the coordinate system (*e.g.*, Fig. 5.2). We employed a number of representations of the 3D face. Two of these representations were holistic, in that they included large regions of the facial surface. These included the ‘holistic range’ and ‘holistic edge’ images. The other representations were smaller ‘local facial range and edge images’, centered about five anthropometric facial fiducial points. These representations are described in detail in the following sections.

5.2.2.1 Holistic Range Images

We first evaluated how well each of the eyes, nose and the mouth regions (*e.g.* Fig. 5.3) of the facial range images, individually discriminated between humans, when they were compared using the CW-SSIM index. As described in Section 4.1.6, we automatically located 10 anthropometric facial fiducial points (Fig. 5.4) on all faces in the Texas 3D Face Recognition Database. We employed the locations of these fiducial points to segment the eyes, nose and the mouth regions of each face. We

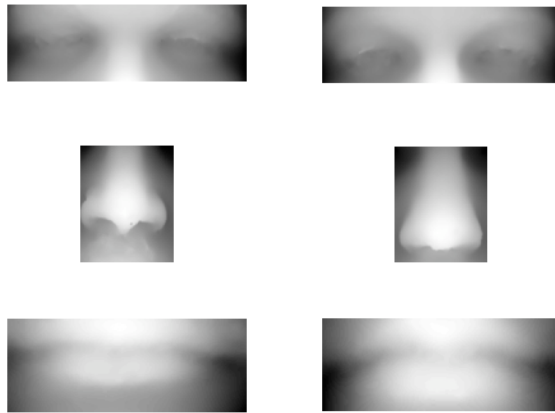


Figure 5.3: Examples of the eyes, nose and mouth regions, respectively (from top to bottom) of two different faces.

computed the means and the standard deviations of the boundaries of each of these regions across all images in the database. The final values of these boundaries were fixed at their respective $\mu + 3\sigma$ or $\mu - 3\sigma$ values for the entire database of images. For all faces, this created sub-regions that were of the same size, which is required for computing the CW-SSIM index.

We evaluated the face recognition performance of each of the individual eyes, nose and mouth sub-regions of the facial range images, that of the eyes and the nose regions together, and that of the complete facial range images on a smaller training data set. This data set did not overlap with the test data set, which we eventually employed to evaluate all the other 3D face recognition algorithms. This data set contained 360 facial range images of 12 subjects (30 images per subject). It was randomly partitioned into a gallery set, which contained one image each of the 12 subjects with a neutral facial expression, and a probe set. The probe set contained the remaining 348 images of the 12 subjects with a neutral or an arbitrary facial expression. We compared all faces in the probe data set to all faces in the gallery

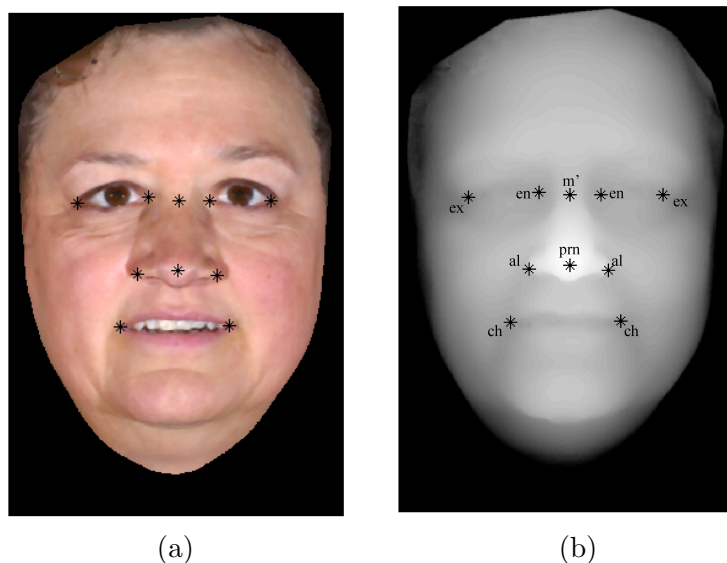


Figure 5.4: The 10 automatically detected anthropometric facial fiducial points.

data set and computed the Equal Error Rates (EER), Areas Under the Receiver Operating Characteristic (ROC) curves (AUC), and the rank 1 Recognition Rates (RR) for the face recognition performance of each of the eyes, nose and mouth facial sub-regions.

We found that when the eyes and the nose regions of the 3D facial range images were compared using the CW-SSIM index, they were good at discriminating between human faces. However, the mouth region was extremely poor at the task, with a rank 1 recognition rate of nearly half of that for the eyes and nose regions. We also observed that the facial sub-regions that contained both the eyes and the nose of 3D faces, were also more discriminatory than the complete facial range images (*e.g.*, Fig. 5.2).

Hence, as the first holistic representation of the 3D facial surfaces (hereafter ‘holistic range images’) for the proposed 3D SSIMFace algorithm, we employed sub-

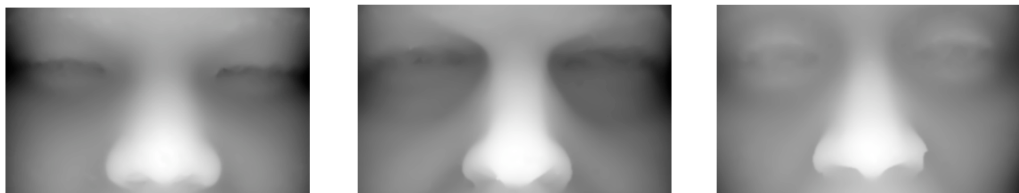


Figure 5.5: The sub-sections of the facial range images, which contained only the nose and the eyes regions, and which were employed for the 3D SSIMFace algorithm.

sections of facial range images, which only contained the eyes and nose regions of the 3D faces (*e.g.*, Fig. 5.5). For each range image this sub-section was located between pixels 200 and 425 along the vertical, and 75 and 425 along the horizontal. We employed the locations of the 10 automatically detected anthropometric facial fiducial points (Fig. 5.4) on all faces in our database to decide these limits. They corresponded to the $\mu + 3\sigma$ or $\mu - 3\sigma$ values of the boundaries of these region for all faces in the database.

5.2.2.2 Holistic Edge Images

Edges of facial range images are physically meaningful in that they are the locations of the ridges and valleys of the facial surface. Furthermore, they are compact representations of facial surfaces, which are rich in information about the ‘detailed’ components (regions of high curvature) of the surface [156]. In the past, similar edge images have been successfully employed to match 3D aerial terrain images [134]. For the eigensurfaces and fishersurfaces [38, 31, 30, 32] algorithms too, gradient images derived from facial range images have been reported to perform better than range images.

Hence, the second representation of 3D faces (hereafter ‘holistic edge images’) that we employed for the 3D SSIMFace algorithm, were binary edge maps (*e.g.* Fig.



Figure 5.6: The facial edge images that were employed for the proposed 3D SSIM-Face algorithm.

5.6) derived from the original facial range images (*e.g.* Fig. 5.2). We obtained these holistic edge images by convolving the facial range images with a Laplacian of Gaussian (LOG) [133] kernel with $\sigma = 7$ pixels, and by detecting all zero-crossings of the filtered images. For detecting edges, we employed the zero crossings of the LOG filtered range images, because several theoretical studies have supported the uniqueness of zero-crossings as a representation of certain band-limited two-dimensional signals [157, 158].

5.2.2.3 Local Range and Edge Images

The third representation of 3D facial surfaces (hereafter ‘local range and edge images’) that we employed for the 3D SSIMFace algorithm were local patches of the facial range and edge images (*e.g.*, Fig. 5.7). The local images were of size 257×257 pixels and were centered about 5 automatically detected anthropometric facial fiducial points. The 5 fiducial points were the two inner corners of the eyes, the tip of the nose, and the two points that define the ‘nose width’, *i.e.*, points *en-en*, *prn*,

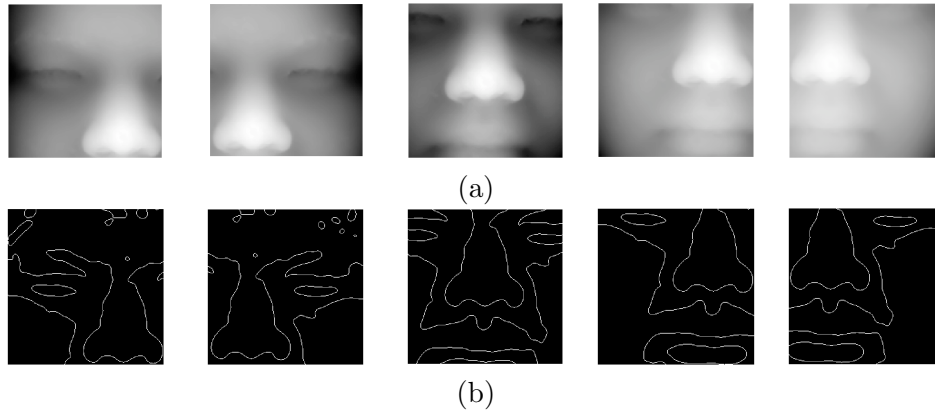


Figure 5.7: From left to right are shown the local (a) range, and (b) edge images of size 257×257 pixels centered about the anthropometric fiducial points $en-en$, prn , $al-al$, respectively.

and $al-al$, respectively in Fig. 5.4(b). We selected these 5 points, because of all the 10 anthropometric facial fiducial points, the regions about these 5 points were individually the most discriminating.

The automatic fiducial point detection algorithms that we employed had a maximum standard deviation of error from manually located fiducially points of less than 8 pixels or 2.65 mm. These fiducial point detection errors translate to small translational errors between the local range and edge images that we employed for the 3D SSIMFace algorithm. Hence, the CW-SSIM index, which is insensitive to small translations between images, is an ideal choice for comparing the local range and edge images.

5.2.3 Recognition Algorithm

For the final 3D SSIMFace algorithm, we computed the structural similarity between the holistic range images ($s_{R,holistic}$), holistic edge images ($s_{E,holistic}$), the five local range images ($s_{R,local}^i$, where $i = \{1, 2, 3, 4, 5\}$), and the five local edge images

($s_{E,local}^i$, where $i = \{1, 2, 3, 4, 5\}$) for pairs of coarsely aligned 3D faces using the CW-SSIM index. All images were decomposed into their wavelet coefficients using a complex version [159] of a 6-scale, 16-orientation steerable pyramid wavelet transform [160]. This is a type of redundant wavelet transform that avoids aliasing in the sub-bands and hence is a good choice for analyzing signals.

In all we obtained 12 (2 holistic and 10 local) similarity scores for each pair of 3D faces. We converted the similarity scores into dissimilarity values using $d = \sqrt{2(1 - s)}$ [161]. We computed an overall local dissimilarity score (d_{local}) as the arithmetic average of 10 local dissimilarity scores $d_{R,local}^i$ and $d_{E,local}^i$. Finally, we combined the scores of the two holistic representations $d_{R,holistic}$ and $d_{E,holistic}$, and the local dissimilarity score d_{local} to obtain the final dissimilarity score (d) between a pair of 3D facial surfaces as

$$d = \omega_1 d_{R,holistic} + \omega_2 d_{E,holistic} + \omega_3 d_{local}. \quad (5.20)$$

The weights (ω_i) in equation (5.20) were computed using the areas under the ROC curves of each of the individual representations as

$$\omega_i = \frac{\max[AUC_i]}{AUC_i}, \quad (5.21)$$

where i corresponds to $\{R, holistic\}$, $\{E, holistic\}$ and $\{local\}$. These AUC values were obtained using a smaller data set of 360 images of 12 subjects, which was independent of the data set on which the algorithms were finally evaluated. By computing the weights in this manner, we ensured that the scores of the individual representations that performed better than the others, *i.e.*, produced lower AUC values, were weighted more heavily.

5.3 Benchmark Algorithms

We compared the performance of our proposed 3D SSIMFace algorithm to three existing state-of-the-art 3D face recognition algorithms. The first of these was an ICP based algorithm [24, 84], wherein we first finely registered each pair of coarsely aligned 3D facial models in our database. We implemented the ICP procedure on complete 3D facial surfaces (*e.g.* Fig. 5.1), which were reduced to 0.1 times of their original magnifications. The rotation and translation matrices learned from the ICP procedure on the sub-sampled 3D models were employed to align the *original* 3D facial models. We compared the finely aligned 3D facial models with each of the MSE , MSE_{CP} and the pH distance metrics. For computing the pH distance metric we rejected 10% of the largest distances between the two 3D models as outliers to account of variations in the sizes of 3D faces.

For the ICP based algorithms we compared the shapes of entire 3D facial surfaces. It should be noted that this is different from the SSIMFace algorithm, wherein we compared parts of the facial surfaces that contained the eyes and nose only. For the ICP based approaches we found that excluding the mouth region lowered their overall performance, relative to when the entire facial surface was employed. Similar results have been reported by other researchers in the past [87, 88]. Hence, we implemented the version of the ICP algorithm that resulted in the best overall performance, and which employed the entire 3D facial surface.

We also implemented two other benchmark 3D face recognition algorithms, namely eigensurfaces [19], and fishersurfaces [31, 38]. For these algorithms, we employed sub-sections of the facial range images between the pixels 147 and 553 along the vertical, and 38 and 478 along the horizontal (*e.g.*, Fig. 5.8). These limits corresponded to the extrema of the uppermost, bottommost, leftmost, and



Figure 5.8: The sub-regions of the facial range images that were employed for the eigensurfaces and fishersurfaces 3D face recognition algorithms.

rightmost co-ordinates, respectively, of 25 manually located anthropometric fiducial points (Section 3.3) across all faces in the Texas 3D Face Recognition Database. In setting these limits to the extrema of all faces, we ensured that the main facial features were not excluded for any face.

For the eigensurfaces algorithm, we learned 69 eigen directions that accounted for 99% of the variance of the data, and linearly projected all faces in the test data set onto these eigen directions. The final metric for the comparison of 3D facial surfaces in the eigen sub-space was the L_1 norm. For the fishersurfaces algorithm, we first reduced the dimensionality of the range images to 348 using principal component analysis (PCA) [70]. This was done to ensure that the within-class scatter matrix employed in the LDA computations was non-singular. We then learned 11 LDA directions from the 348 PCA features and projected all faces in the test data set onto these LDA directions before comparing them by means of the L_2 norm. The eigensurfaces and the fishersurfaces algorithms were trained on a training data set and were evaluated on a separate test data set. The proposed 3D SSIMFace and the ICP based algorithms, which did not require training were also evaluated on the test data that was employed to evaluate the eigensurfaces and the fishersurfaces algorithms.

Partition	No. of Subjects	No. of Images		
		Neutral	Expressive	Total
Training	12	228	132	360
Test	Gallery	105	0	105
	Probes	95	480	663

Table 5.1: A summary of the data partitions employed for training and testing 3D face recognition algorithms.

5.4 Performance Evaluation

5.4.1 Data

We employed the Texas 3D Face Recognition Database (Chapter 3). This database contained 1149 2D and 3D images of 118 human subjects. We partitioned the database into a training data set and a test data set (Table 5.1). The training data set contained 360 randomly selected images of 12 subjects (30 images per subject) in neutral or expressive modes. For all the 3D face recognition algorithms that we developed, steps such as automatic facial fiducial point detection, classifier feature selection and classifier optimization were performed using the training data set only. The trained classifier was evaluated on the independent test data set, which did not overlap with the training data set.

The test data set included 768 images of 105 subjects. This test set was further partitioned into a gallery set and a probe set. For all 3D face recognition algorithms, faces in the probe set were compared to faces in the gallery set, which resulted in a matrix of 105×663 comparison scores. Consistent with the evaluation protocol of the FRVT 2002 [118] and the FRGC 2005 [13], the gallery set contained one range image each of 105 subjects with a neutral facial expression. The probe set contained another 663 images of 95 of the gallery subjects with a neutral or an

arbitrary facial expression. In the probe set, the number of images of each subject varied from 1 to 55. In accordance with the widely accepted ‘closed universe’ model for the evaluation of face recognition algorithms [118], every subject in the probe data set was represented in the gallery data set.

5.4.2 3D Face Recognition

We evaluated the verification performance of all 3D face recognition algorithms using the Receiver Operating Characteristic methodology [50], and observed the values of the Equal Error Rates and the Areas Under the ROC Curves. The identification performance of the algorithms was evaluated using Cumulative Match Characteristic (CMC) curves, and the rank 1 Recognition Rates were observed. Statistical 95% confidence intervals for the EER, AUC, and the rank 1 RR values were obtained empirically using bootstrap sampling. All performance statistics were observed separately for neutral faces, for expressive faces, and for both types of faces.

5.5 Results and Discussion

5.5.1 Discriminatory Facial Regions

The recognition performance of the individual eyes, nose, and mouth regions of the facial range images when they were compared using the CW-SSIM index are presented in Table 5.2. The performance statistics reported are for both neutral and expressive faces. The ROC and CMC curves for the verification and identification performance of these regions are shown in Fig. 5.9.

We found that for comparisons using the CW-SSIM index, the orbital region of the 3D face was the most discriminatory, closely followed by the nasal region (Table 5.2, and Fig. 5.9). The mouth region, however, was very poorly discriminating

Facial Region	EER (%)	AUC $\times 10^{-2}$	Rank 1 RR (%)
Eyes	5.09 [3.74 6.04]	0.87 [0.55 1.05]	95.7 [93.4 97.7]
Nose	8.58 [6.91 10.1]	2.56 [1.62 3.23]	92.8 [89.9 95.4]
Mouth	29.0 [25.2 32.8]	22.7 [19.1 27.0]	52.0 [47.1 57.6]

Table 5.2: The observed EER, AUC, and rank 1 RR values and their 95% confidence intervals for the eyes, nose and mouth regions of the facial range images, wherein the facial sub-regions were compared using the CW-SSIM index.

($EER = 29\%$ and rank 1 $RR = 52\%$). The poor performance of the mouth region could be explained by greater variation in its shape for different facial expressions. In conclusion, the mouth region of the facial range images was clearly not very accurate for matching 3D faces with the CW-SSIM index. Hence, we were justified in excluding this region from the final holistic range image representation (Section 5.2.2.1) that we employed for the 3D SSIMFace algorithm.

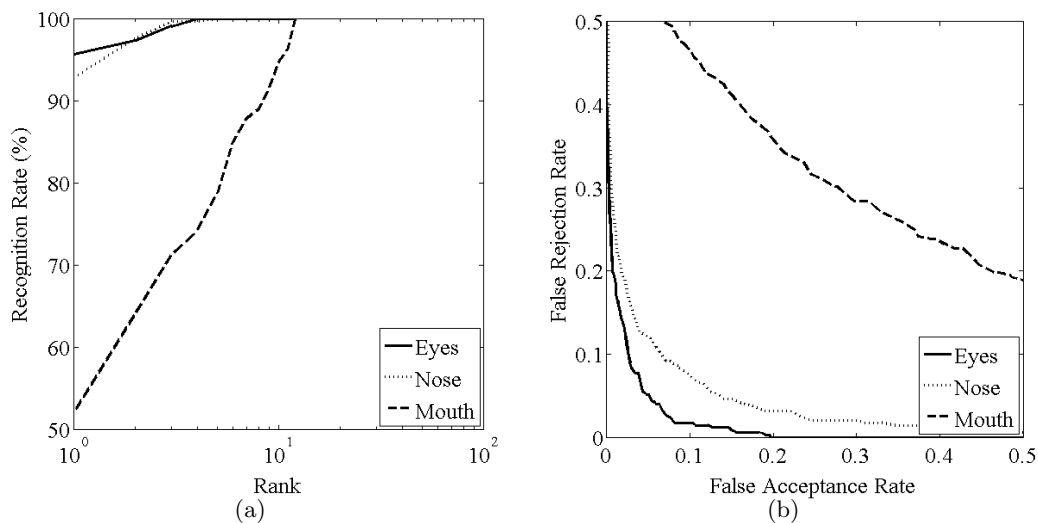


Figure 5.9: (a) The semi-log CMC curves, and (b) the ROC curves for the eyes, nose and mouth regions of the facial range images, wherein the facial sub-regions were compared using the CW-SSIM index.

5.5.2 3D SSIMFace

The performance of the proposed 3D SSIMFace algorithm, and the benchmark eigensurfaces, fishersurfaces, and ICP algorithms is presented in Table 5.3. The ROC and CMC curves for these algorithms are shown in Fig. 5.10. For all faces (neutral and expressive), the 3D SSIMFace algorithm performed well both at the tasks of facial verification ($EER = 1.36\%$ and $AUC = 0.0019$), and facial identification (rank 1 $RR = 99.7\%$). It also performed significantly better than all the three benchmark eigensurfaces and fishersurfaces, and ICP algorithms.

Note, that the performance statistics reported in Table 5.3 for the 3D SSIMFace algorithm are for the case when coarsely aligned 3D facial surfaces were compared. By comparison, the performance statistics reported for the ICP based algorithm are for the case when the same coarsely aligned facial surfaces were further finely aligned using the ICP algorithm and were then compared using the pH distance metric. In Table 5.3, we have reported the results of the ICP algorithm with the pH metric only, because of the three MSE , MSE_{CP} and pH metrics, comparison of the finely registered facial surfaces using the pH metric resulted in the best performance. It is evident that even after finely registering the 3D faces, the best ICP based approach was nearly an order of magnitude less accurate at discriminating between them than the proposed 3D SSIMFace algorithm was at discriminating between their coarsely aligned versions. Hence, the 3D SSIMFace algorithm, which is based on the CW-SSIM index clearly obviates the need to finely register 3D faces that are already in a canonical frontal upright pose.

EER [Confidence Interval] %						
Algorithms	Neutral		Expressive		All	
Eigensurfaces	24.0	[21.3 26.3]	23.6	[19.2 26.2]	24.0	[21.8 26.6]
Fishersurfaces	8.11	[6.39 10.9]	3.60	[2.00 6.43]	6.69	[5.27 8.16]
ICP + pH	7.97	[6.85 9.95]	9.92	[6.71 14.7]	9.03	[7.67 10.2]
3D SSIMFace	0.84	[0.27 1.21]	1.68	[0.49 3.81]	1.36	[0.94 1.79]

(a)

AUC [Confidence Interval] $\times 10^{-2}$						
Algorithms	Neutral		Expressive		All	
Eigensurfaces	16.7	[14.1 18.8]	14.9	[11.9 18.0]	16.3	[14.8 18.7]
Fishersurfaces	2.88	[2.12 3.71]	1.32	[0.32 2.49]	2.40	[1.79 2.96]
ICP + pH	2.97	[2.16 4.12]	4.39	[2.18 7.80]	3.44	[2.56 4.27]
3D SSIMFace	0.17	[0.02 0.49]	0.18	[0.04 0.37]	0.19	[0.06 0.46]

(b)

Rank 1 RR [Confidence Interval] %						
Algorithms	Neutral		Expressive		All	
Eigensurfaces	58.1	[54.0 62.7]	52.5	[45.4 60.1]	56.6	[52.9 60.2]
Fishersurfaces	91.7	[89.4 94.0]	95.1	[91.8 97.8]	92.6	[90.6 94.4]
ICP + pH	88.5	[85.6 91.5]	86.3	[80.9 91.0]	87.9	[85.5 90.2]
3D SSIMFace	99.8	[99.4 100]	99.5	[99.4 100]	99.7	[99.2 100]

(c)

Table 5.3: The observed (a) EER, (b) AUC, and (c) rank 1 RR values and their 95% confidence intervals for the proposed 3D SSIMFace algorithm, wherein coarsely aligned 3D faces were compared using the CW-SSIM index, and for the benchmark eigensurfaces, fishersurfaces, and ICP algorithms.

5.5.3 Before and After ICP

We further verified whether the CW-SSIM index was indeed robust to small rotations and translations that were present between the pose normalized 3D models in our database. We compared pairs of holistic range images, holistic edge images, and local range and edge images of 3D faces, using the CW-SSIM index, before and after they were finely aligned using the ICP procedure. Similarly, we compared

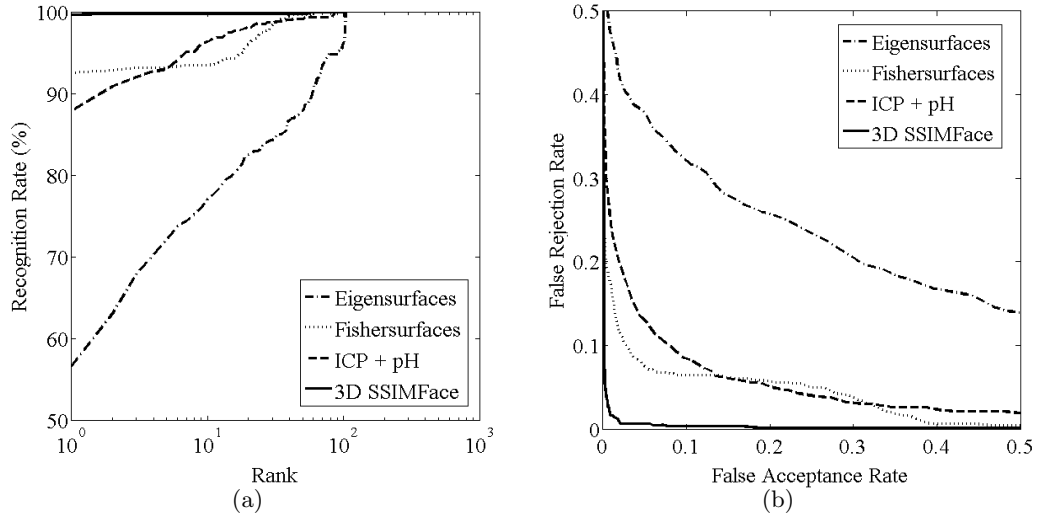


Figure 5.10: (a) The semi-log CMC curves, and (b) the ROC curves for the proposed 3D SSIMFace algorithm, wherein coarsely aligned 3D faces were compared using the CW-SSIM index, and for the benchmark eigensurfaces, fishersurfaces, and ICP algorithms.

pairs of facial surfaces, using the MSE , MSE_{CP} and pH metrics, before and after they had been finely aligned. The face recognition performance of these comparisons are presented in Table 5.4.

On comparing the first three columns of Table 5.4(a) to the corresponding columns of Table 5.4(b), it can be observed that for the CW-SSIM index, the performance was largely statistically equivalent before and after the 3D faces were finely registered. This was true for all the three representations of 3D faces, namely holistic range images, holistic edge images, and local range and edge images. By comparison, the recognition performance of the MSE metric increased significantly from a rank 1 $RR = 48.7\%$ before fine alignment of the 3D faces using ICP, to a rank 1 $RR = 60.9\%$ after fine alignment. Similarly, the verification and recognition performance of the MSE_{CP} and pH metrics also significantly improved after the

Before ICP							
Index	EER (%)			AUC $\times 10^{-2}$		Rank 1 RR (%)	
$d_{R,holistic}$	6.69	[4.93	7.88]	1.44	[1.05	1.97]	91.4 [89.3 93.4]
$d_{E,holistic}$	3.58	[2.90	4.60]	0.53	[0.33	0.97]	96.8 [95.3 98.0]
d_{local}	2.11	[1.44	2.89]	0.40	[0.11	0.68]	97.6 [96.4 98.6]
MSE	21.4	[19.2	23.4]	13.0	[11.9	14.5]	48.7 [44.9 52.6]
MSE_{CP}	15.6	[13.6	17.1]	7.32	[6.39	8.56]	58.2 [54.6 61.9]
pH	14.4	[12.3	16.0]	6.15	[5.32	7.43]	65.6 [61.8 69.3]

(a)

After ICP							
Index	EER (%)			AUC $\times 10^{-2}$		Rank 1 RR (%)	
$d_{R,holistic}$	5.73	[4.36	6.84]	1.17	[0.72	1.54]	93.7 [91.7 95.5]
$d_{E,holistic}$	3.96	[3.15	4.78]	0.57	[0.36	0.95]	96.2 [94.7 97.6]
d_{local}	2.12	[1.46	2.74]	0.34	[0.17	0.57]	98.5 [97.6 99.4]
MSE	21.6	[19.0	23.2]	12.4	[11.3	14.1]	60.9 [56.9 64.7]
MSE_{CP}	12.6	[11.0	14.1]	5.57	[4.69	6.63]	77.1 [73.8 80.1]
pH	9.03	[7.67	10.2]	3.44	[2.56	4.27]	87.9 [85.5 90.2]

(b)

Table 5.4: The observed EER, AUC, and rank 1 RR values and their 95% confidence intervals for comparing facial surface representations with the CW-SSIM, MSE , MSE_{CP} , and pH indices, before and after 3D faces were finely registered using the ICP algorithm. The performance statistics are for all faces (neutral and expressive).

coarsely aligned faces were finely registered using ICP. This analysis confirmed that, unlike the MSE , MSE_{CP} and pH metrics, the CW-SSIM index was indeed robust to the small registration errors that were present between the pairs of frontal upright 3D faces in our database.

As an aside, from Table 5.4 it is interesting to note that out of the three representations that we employed for the 3D SSIMFace algorithm, the combination of local range and edge images resulted in the best performance, followed closely by the holistic edge images. In a sense the edge images are also local representations of the facial surfaces in that they only contain information about specific *local* regions

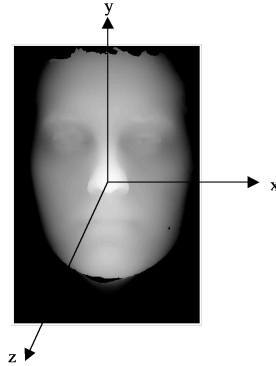


Figure 5.11: The right handed 3D co-ordinate system.

of high surface curvature of the face. Hence, these results are interesting in that they point towards the superior performance of local feature based approaches for 3D face recognition relative to holistic approaches. Furthermore, on comparing Table 5.3 and Table 5.4, it is evident that for the 3D SSIMFace algorithm a combination of the various holistic and local representations of 3D faces resulted in significantly superior performance relative to either of them individually. This corroborates with similar findings in previous 3D face recognition studies [44, 83, 42].

5.5.4 Analysis of Geometric Transformations

Lastly, to gain a better understanding of the workings of the CW-SSIM, MSE , MSE_{CP} , and pH indices with different facial representations, we evaluated the extent to which each of them was robust to geometric transformations (rotations, translations and scaling) of 3D facial surfaces. In the context of 3D face recognition, the presence of geometric distortions between pairs of facial surfaces can result in two types of errors, namely false positives and false negatives. We analyzed the robustness of each index with regard to false negative errors. A similar procedure

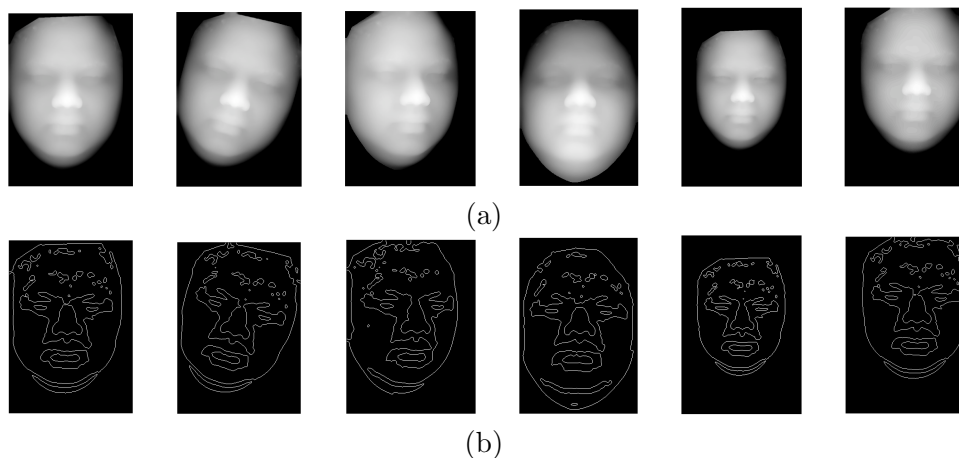


Figure 5.12: The (a) range images, and (b) edge images (from left to right) of an original 3D face, and its transformed version for -15° rotation about the z axis, -15° rotation about the y axis, -15° rotation about the x axis, scaling of 0.75 times its original magnification, and translation of $r = 34$ mm and $\theta = \pi/4$ in the (xy) or $(r\theta)$ plane.

was adopted in a previous study on 3D face recognition [162]. We determined the extent to which a true positive 3D face could be geometrically transformed without it being misclassified as a false negative. The robustness of an index defined in this manner depends on the operating point of its corresponding 3D face recognition system. For all indices, we fixed this operating point at the false acceptance rate of 1%, which is a typical operating point for face recognition systems [13].

We compared 3D facial range and edge images (*e.g.*, Fig. 5.12) using the CW-SSIM index, and entire 3D facial surfaces (*e.g.*, Fig. 5.1) using the MSE , MSE_{CP} and pH metrics. We randomly selected one (neutral or expressive) face of each of 105 subjects from the test data set of the Texas 3D Face Recognition Database. We rotated each 3D face from -30° to 30° in steps of 3° individually about the x , y and z axes (defined in Fig. 5.11). We also scaled each face from 0.75 to 1.25 of its original magnification in steps of 0.025. In the real world, translation of the

Distortion	CW-SSIM (range)	CW-SSIM (edge)	MSE	MSE_{CP}	pH
R_z (degrees)	-9, 9	-15, 15	-9, 9	-9, 9	-9, 9
R_y (degrees)	-9, 9	-21, 21	-3, 3	-3, 3	-6, 6
R_x (degrees)	-12, 12	-24, 27	-3, 3	-3, 6	-3, 3
Scale	0.90, 1.10	0.83, 1.20	0.95, 1.0	0.93, 1.07	0.93, 1.07
r (mm)	13.6	20.4	6.79	6.79	6.79

Table 5.5: The maximum distortions for each of the similarity indices/3D facial representations, that resulted in scores between a 3D face and its transformed version that were less than the operating point of the corresponding indices’ 3D face recognition system.

subject/camera system along the z (optical) axis changes the scale of the acquired facial images. Lastly, we translated the 3D faces in the (xy) or its corresponding $(r\theta)$ plane. For this, we fixed $\theta = \pi/4$ and varied r from 0 mm to 34 mm in steps of 2.26 mm. For all transformations, we assumed that the origin of the co-ordinate system was located at the tip of the nose of the original 3D face. Using each of the CW-SSIM, MSE , MSE_{CP} , and pH indices, we compared each transformed face to its original version. We generated range and edge images of the 3D faces in the transformed poses/scales (*e.g.*, Fig. 5.12) for computing the CW-SSIM scores. For every transformation, we obtained the average and the 95% confidence interval of the scores of a particular index across all 105 faces.

The greatest geometric transformations that each index could accommodate before the score between a face and its transformed version became statistically greater than the operating threshold of the corresponding index’s 3D face recognition system, are listed in Table 5.5. It can be observed that the CW-SSIM index with facial range images was more robust to rotations about the y and x axes, and to translations along the r axis, than the MSE , MSE_{CP} , and pH metrics. Moreover, it was only slightly more robust to scale changes than the MSE , MSE_{CP} ,

and pH metrics. The CW-SSIM index with facial edge images, on the other hand was considerably more robust to all geometric transformations, than all the other indices/3D facial representations. It is also interesting to note that the CW-SSIM index, with both range and edge images, was more or equally robust to out-of-plane rotations (about the x and y axes) as it was to in-the-plane rotations (about the z axis). Clearly, these factors contribute to the overall success of the proposed 3D SSIMFace algorithm.

In conclusion, within the class of 3D face recognition algorithms that directly compare facial surfaces, we proposed a novel 3D SSIMFace algorithm which employs the CW-SSIM index. We demonstrated the superior performance of the 3D SSIMFace algorithm, both in terms of recognition accuracy and robustness to registration errors, relative to existing benchmark ICP based approaches. For comparing facial surfaces, we introduced three novel ideas. The first was to use an index which is robust to small registration errors, and which obviates the need to finely register 3D faces before they are compared. This represents a major paradigm shift from the current *de facto* practice of employing the ICP algorithm before comparing 3D faces. The second was to use a structural similarity index, which compares the phases of facial range images, and is significantly more accurate at matching 3D faces than the general purpose MSE , MSE_{CP} , pH metrics. The third was to employ information-rich edge images derived from facial range images with the CW-SSIM index. Edge images are simple to compute and yet surprisingly effective at distinguishing between 3D faces, as we demonstrated. In a more general context, the methods that we developed for comparing 3D images using the CW-SSIM index are not limited to facial range images. They could potentially be applied to other 3D pattern recognition tasks as well.

Chapter 6

Combined Algorithm

As our final proposed 3D face recognition algorithm, we combined the two 3D AnthroFace (Chapter 4) and 3D SSIMFace (Chapter 5) algorithms that we developed. The dissimilarity scores between the pairs of 3D faces for this combined algorithm were the sum of the scores of the individual 3D AnthroFace and 3D SSIMFace algorithms. We evaluated the performance of the combined algorithm on the test data partition of the Texas 3D Face Recognition Database (Chapter 3). Its performance statistics along with those of the 3D AnthroFace and 3D SSIMFace algorithms are listed in Table 6.1. The ROC and CMC curves for the 3D AnthroFace, 3D SSIMFace and the combined algorithm are presented in Fig. 6.1.

It is interesting to observe from Table 6.1 that the verification performance of the 3D AnthroFace and the 3D SSIMFace algorithms was statistically equivalent. The facial identification performance of the 3D SSIMFace algorithm (rank 1 $RR = 99.7\%$ for all faces), however, was significantly better than that of the 3D AnthroFace algorithm (rank 1 $RR = 96.8\%$ for all faces). Moreover, a simple approach of summing the scores of the two algorithms resulted in a very accurate algorithm

EER [Confidence Interval] %						
Algorithms	Neutral		Expressive		All	
3D AnthroFace	1.65	[1.11 2.28]	2.81	[1.27 4.30]	1.98	[1.37 2.88]
3D SSIMFace	0.84	[0.27 1.21]	1.68	[0.49 3.81]	1.36	[0.94 1.79]
Combined	0.42	[0.14 0.64]	1.62	[0.54 3.01]	0.77	[0.47 1.10]

(a)

AUC [Confidence Interval] $\times 10^{-2}$						
Algorithms	Neutral		Expressive		All	
3D AnthroFace	0.14	[0.08 0.23]	0.25	[0.11 0.42]	0.18	[0.11 0.28]
3D SSIMFace	0.17	[0.02 0.49]	0.18	[0.04 0.37]	0.19	[0.06 0.46]
Combined	0.05	[0.01 0.11]	0.09	[0.01 0.21]	0.07	[0.02 0.16]

(b)

Rank 1 RR [Confidence Interval] %						
Algorithms	Neutral		Expressive		All	
3D AnthroFace	97.3	[95.8 98.5]	95.6	[92.4 98.4]	96.8	[95.3 98.0]
3D SSIMFace	99.8	[99.4 100]	99.5	[99.4 100]	99.7	[99.2 100]
Combined	99.8	[99.4 100]	99.5	[99.4 100]	99.7	[99.2 100]

(c)

Table 6.1: The observed (a) EER, (b) AUC, and (c) rank 1 RR values and their 95% confidence intervals for the 3D AnthroFace, 3D SSIMFace and a combination of both these algorithms.

with $EER = 0.77\%$, $AUC = 0.0007$, and rank 1 $RR = 99.7\%$. The verification performance of this combined algorithm was also significantly better than that of either of the 3D AnthroFace and 3D SSIMFace algorithms. This suggests that the proposed 3D AnthroFace and 3D SSIMFace algorithms embody uncorrelated or complementary discriminatory facial structural information. The AnthroFace algorithm employs local discriminatory anthropometric facial distances. The 3D SSIMFace algorithm is a combination of holistic and local approaches, wherein the shapes of entire facial surfaces or their sub-regions are quantified and compared.

The fact that the two algorithms are complementary is further supported

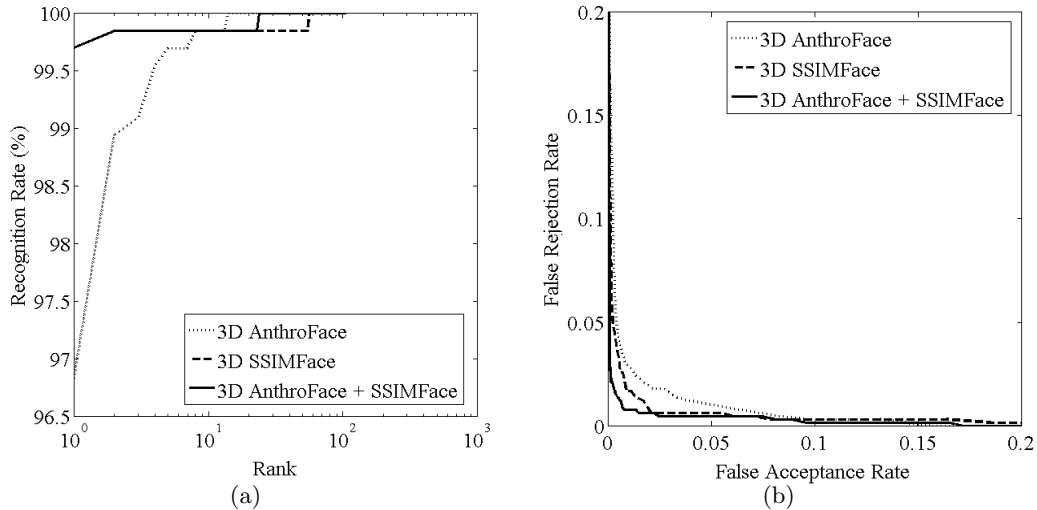


Figure 6.1: (a) The semi-log CMC curves, and (b) the ROC curves for the 3D AnthroFace, 3D SSIMFace algorithms, and a combination of both these algorithms.

by the observation that for the 3D AnthroFace algorithm, the anthropometric distances associated with the nose region of the face were the most discriminatory. By comparison, for the 3D SSIMFace algorithm the surface of the orbital region was more discriminatory than the surface of the nose region. The 3D AnthroFace and 3D SSIMFace algorithms were also based on very different design philosophies. The 3D AnthroFace algorithm was specifically designed to incorporate domain specific knowledge about the structural diversity of human faces. The holistic facial surface matching approaches that we proposed as a part of the 3D SSIMFace algorithm, on the other hand, are not limited to facial images only and can be applied to other 3D surface matching applications.

A number of results in this dissertation also point towards the potentially superior performance of the local feature based 3D face recognition algorithms relative to holistic approaches. These include the superior performance of the proposed

Author	Algorithm	Data		Performance
		Gallery	Probe	
Chang [65]	PCA	200	676	$RR = 94.5\%$
Pan [112]	Isometric flattening + PCA	–	–	$EER = 2.83\%$
Russ[84]	ICP, $MSE_{CP} + pH$	198	745	$FRR = 6.5\%$ $FAR = 0.1\%$ $RR = 98.5\%$
Koudelka [88]	ICP, pH	198	198	$RR = 94\%$
Kakadiaris [107]	AFM	152	608	$FRR \sim 3.1\%$ $FAR = 0.1\%$ $RR = 99.3\%$
Gupta <i>et al.</i>	3D AnthroFace + SSIMFace	105	663	$FRR = 2.56\%$ $FAR = 0.1\%$ $RR = 99.7\%$ $EER = 0.77\%$

Table 6.2: A summary of the five 3D face recognition algorithms that were evaluated at the Face Recognition Grand Challenge 2005 on the FRGC v0.1 database, and of our final proposed 3D face recognition algorithm, which was a combination of the 3D AnthroFace and 3D SSIMFace algorithms.

3D AnthroFace algorithm in comparison to the holistic eigensurfaces, fishersurfaces and ICP algorithms; the better performance of the local representations of the 3D SSIMFace algorithm in comparison of the holistic 3D facial representations; and the statistically equivalent verification performance of the purely local feature based 3D AnthroFace algorithm to the 3D SSIMFace algorithm, which employed both holistic and local facial representations.

Lastly, we compare the performance of the proposed combined algorithm to that of the five 3D face recognition algorithms that were evaluated on the publicly available FRGC v0.1 database as a part of the Face Recognition Grand Challenge 2005 [13] (Table 6.2). These algorithms represent the current state-of-the-art in 3D face recognition. Among these five existing algorithms, the ‘Annotated Face

Model' (AFM) algorithm proposed by Kakadiaris *et al.* [107], performed the best with rank 1 $RR = 99.3\%$, and at $FAR = 0.1\%$, $FRR = 3.1\%$. It is clear that the performance of the proposed 3D face recognition algorithm (rank 1 $RR = 99.8\%$, and at $FAR = 0.1\%$, $FRR = 2.56\%$), which combined the 3D AnthroFace and SSIMFace recognition algorithms, on the Texas 3D Face Recognition was competitive with the performance of the state-of-the-art algorithms on the FRGC v0.1 database.

Chapter 7

Conclusion

In conclusion, automatic human face recognition is a non-trivial computer vision problem of considerable practical significance. Three dimensional face recognition, which is now emerging as a significant technology, has numerous open problems that are yet to be resolved. In this dissertation, we developed novel 3D face recognition algorithms based on ideologies that are different from the existing paradigms. Firstly, we scientifically analyzed the discriminatory facial anthropometric distances and developed successful fully automatic 3D face recognition algorithms that employ these distances. We developed 2D+3D algorithms to automatically locate 10 anthropometric facial fiducial points with a high precision. In order to locate facial fiducial points, we again employed evidence about the established proportions of facial features from the literature on anthropometric facial proportions. Secondly, we developed a highly accurate algorithm for comparing the shapes of 3D facial surfaces using the CW-SSIM index, which is also significantly more robust to registration errors between 3D faces than the existing 3D face recognition algorithms.

On a large database of 3D images, we demonstrated the effectiveness and su-

perior performance of our proposed algorithms relative to the existing benchmark 3D face recognition algorithms. Furthermore, unlike many previously reported works of 3D face recognition, all algorithms in this dissertations were evaluated and compared using rigorous statistical hypothesis testing. We envision that the ideas proposed in this dissertation would not only advance the field of 3D face recognition, but would also be applicable to the related scientific disciplines of facial computer graphics, facial surgical planning, and general 3D object recognition.

In the future, a number of areas of this research work can be extended. For example, the fiducial point detection algorithms can be extended to automatically locate more than the 10 anthropometric fiducial points that we detected. It may also be interesting to study the sensitivity of the proposed fiducial point detection algorithms to small in and out of plane rotations of 3D faces. For detecting fiducial points, we computed the surface curvature of 3D faces using frontal upright facial range images. It may also be useful to investigate techniques for computing surface curvature directly from arbitrarily oriented 3D point clouds or 3D meshes. This may eliminate the need to operate on frontal upright faces. Similarly, for the 3D SSIMFace recognition algorithm, we computed the CW-SSIM index between pairs of facial range images. A 3D version of the CW-SSIM index could be developed, which could operate directly on 3D models instead of on range images. In a future study, it may also be instructive to analyze the extend to which the CW-SSIM, MSE , MSE_{CP} , and pH indices are robust to the presence of noise and holes in 3D faces.

Lastly, all the algorithms that we developed employed only 3D or structural features of facial surfaces for *recognition*. The facial intensity/color information was not employed for recognition. Numerous studies in the past have demonstrated that

a combination of the 2D and 3D imaging modalities for face recognition, results in superior performance relative to either of them individually. Hence, a natural extension of this work would be to compare the performance of our proposed 3D face recognition algorithms to state-of-the-art 2D face recognition algorithms and to investigate techniques for combining it with 2D face recognition algorithms.

Bibliography

- [1] L. Farkas, *Anthropometric Facial Proportions in Medicine*. Thomas Books, 1987.
- [2] S. L. Rogers, *The Personal Identification of Living Individuals*. Charles C Thomas, Springfield, Illinois, 1986.
- [3] N. K. Ratha, A. Senior, R. M. Boile, S. Singh, N. Murshed, and W. Kropatsch, “Automated biometrics,” in *Advances in Pattern Recognition - ICAPR 2001. Second International Conference. Proceedings (Lecture Notes in Computer Science Vol.2013)*, S. Singh, N. Murshed, and W. Kropatsch, Eds. IBM Thomas J. Watson Res. Center, Yorktown Heights, NY, USA: Springer-Verlag, 2001, pp. 445–453.
- [4] S. S. Rakover and B. Cahlon, *Face Recognition: Cognitive and Computational Processes*. John Benjamins Publishing Company, Amsterdam, 2001.
- [5] G. L. Wells and E. P. Seelau, “Eyewitness identification: Psychological research and legal policy on lineups.” *Psychology, Public Policy, and Law*, vol. 1, pp. 765–791, 1995.
- [6] B. Scheck, P. Neufeld, and J. Dwyer, *Actual Innocence*. Doubleday, New York, 2000.

- [7] R. S. Malpass and J. Kravitz, "Recognition for faces of own and other race," *Journal of Personality and Social psychology*, vol. 13, pp. 330–334, 1969.
- [8] H. Wechsler, P. J. Phillips, V. Bruce, F. F. Soulie, , and T. S. Huang, Eds., *Face Recognition: From Theory to Applications*. Springer-Verlag, Berlin, 1998.
- [9] P. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone, "Frvt 2002: Overview and summary," available at www.frvt.org, March 2003.
- [10] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 9, pp. 1063–1074, 2003.
- [11] S. Malassiotis and M. G. Strintzis, "Robust face recognition using 2d and 3d data: Pose and illumination compensation," *Pattern Recognition*, vol. 38, no. 12, pp. 2537–2548, 2005.
- [12] S. Gupta, M. K. Markey, and A. C. Bovik, *Pattern Recognition Research Horizons*. Nova Science Publishers, Inc, New York, 2007, ch. Advances and Challenges in 3D and 2D+3D Human Face Recognition, p. to appear.
- [13] 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 *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, 2005, pp. 947–954 vol. 1.
- [14] L. G. Farkas, *Anthropometry of the Head and Face in Medicine*. New York: Elsevier, 1981.

- [15] L. G. Farkas, M. I. R., and K. J. C., “An attempt to define the attractive face: an anthropometric study,” in *18th Annual meeting of the American Society for Aesthetic Plastic Surgery*, Boston, April 15 1985.
- [16] Z. Wang and A. C. Bovik, “A universal image quality index,” *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.
- [17] Z. Wang and E. P. Simoncelli, “Translation insensitive image similarity in complex wavelet domain,” in *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, vol. 2, 2005, pp. 573–576.
- [18] P. J. Besl, “Active, optical range imaging sensors,” *Machine Vision and Applications*, vol. 1, no. 2, pp. 127–152, 1988.
- [19] K. I. Chang, K. W. Bowyer, and P. J. Flynn, “An evaluation of multimodal 2d+3d face biometrics,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 4, pp. 619–624, 2005.
- [20] G. G. Gordon, “Face recognition based on depth maps and surface curvature,” in *SPIE Geometric methods in Computer Vision*, vol. 1570, 1991, pp. 234–247.
- [21] C. Heshner, A. Srivastava, and G. Erlebacher, “Principal component analysis of range images for facial recognition,” in *International Conference on Imaging Science, Systems, and Technology (CISST), in Proceedings of*, 2002.
- [22] Y. Lee and T. Yi, “3d face recognition using multiple features for local depth information,” in *Proceedings EC-VIP-MC 2003. 4th EURASIP Conference focused on Video/Image Processing and Multimedia Communications (IEEE*

- Cat. No.03EX667*), M. Grgic and S. Grgic, Eds. Zagreb: Faculty of Electrical Eng. & Comput, 2003, vol. vol.1, pp. 429–434.
- [23] Y.-H. Lee and J.-C. Shim, “Curvature based human face recognition using depth weighted hausdorff distance,” in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, vol. 3, 2004, pp. 1429–1432 Vol. 3.
- [24] X. Lu, A. K. Jain, and D. Colbry, “Matching 2.5d face scans to 3d models,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 1, pp. 31–43, 2006.
- [25] A. B. Moreno, A. Sanchez, J. Fco, V. Fco, and J. Diaz, “Face recognition using 3d surface-extracted descriptors,” in *Irish Machine Vision and Image Processing Conference (IMVIP 2003)*, September 2003.
- [26] T. Nagamine, T. Uemura, and I. Masuda, “3d facial image analysis for human identification,” in *Pattern Recognition, 1992 . Vol.1. Conference A: Computer Vision and Applications, Proceedings., 11th IAPR International Conference on*, 1992, pp. 324–327.
- [27] T. D. Russ, M. W. Koch, and C. Q. Little, “3d facial recognition: a quantitative analysis,” in *Security Technology, 2004. 38th Annual 2004 International Carnahan Conference on*, 2004, pp. 338–344.
- [28] A. Srivastava, X. Liu, and C. Heher, “Face recognition using optimal linear components of range images,” 2003, accepted for publishing in *Image and Vision Computing*.
- [29] T. Maurer, D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev, D. West, and G. Medioni, “Performance of geometrix activeidTM 3d face recognition

- engine on the frgc data,” in *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on*, vol. 3, 2005, pp. 154–154.
- [30] T. Heseltine, N. Pears, and J. Austin, “Three-dimensional face recognition: an eigensurface approach,” in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, vol. 2, 2004, pp. 1421–1424 Vol.2.
- [31] T. Heseltine, N. Pears and J. Austin, “Three-dimensional face recognition: A fishersurface approach,” in *International Conference on Image Analysis and Recognition, In Proc. of the*, ser. ICIAR 2004, A. Campilho and M. Kamel, Eds., vol. LNCS 3212. Springer-Verlag Berlin Heidelberg, 2004, pp. 684–691.
- [32] T. Heseltine, N. Pears, and J. Austin, “Three-dimensional face recognition using surface space combinations,” in *British Machine Vision Conference*, 2004.
- [33] S. Lao, Y. Sumi, M. Kawade, and F. Tomita, “3d template matching for pose invariant face recognition using 3d facial model built with isoluminance line based stereo vision,” in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 2, 2000, pp. 911–916 vol.2.
- [34] G. Medioni and R. Waupotitsch, “Face modeling and recognition in 3-d,” in *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, 2003, pp. 232–233.
- [35] T. Papatheodorou and D. Rueckert, “Evaluation of automatic 4d face recognition using surface and texture registration,” in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, 2004, pp. 321–326.
- [36] L. Zhang, A. Razdan, G. Farin, J. Femiani, M. Bae, and C. Lockwood, “3d

- face authentication and recognition based on bilateral symmetry analysis,” *Visual Computer*, vol. 22, no. 1, pp. 43–55, 2006.
- [37] B. Achermann, X. Jiang, and H. Bunke, “Face recognition using range images,” in *Proceedings. International Conference on Virtual Systems and MultiMedia*. Inst. of Comput. Sci. & Appl. Math., Bern Univ., Switzerland: Int. Soc. Virtual Syst. & MultiMedia (VSMM), 1997, pp. 129–136.
- [38] C. BenAbdelkader and P. A. Griffin, “Comparing and combining depth and texture cues for face recognition,” *Image and Vision Computing*, vol. 23, no. 3, pp. 339–352, 2005.
- [39] C. Beumier and M. Acheroy, “Automatic 3d face authentication,” *Image and Vision Computing*, vol. 18, no. 4, pp. 315–321, 2000.
- [40] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, “Three-dimensional face recognition,” *International Journal of Computer Vision*, vol. 64(1), pp. 5–30, 2005.
- [41] B. Gökberk, A. A. Salah, and L. Akarun, “Rank-based decision fusion fro 3d shape-based face recognition,” in *LNCS, International Conference on Audio- and Video-based Biometric Person Authentication*, vol. 3546, 2005, pp. 1019–1028.
- [42] G. Pan, Y. Wu, and Z. Wu, “Investigating profile extracted from range data for 3d face recognition,” in *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, vol. 2, 2003, pp. 1396–1399 vol.2.
- [43] F. Tsalakanidou, S. Malassiotis, and M. G. Strintzis, “Face localization and

- authentication using color and depth images,” *Image Processing, IEEE Transactions on*, vol. 14, no. 2, pp. 152–168, 2005.
- [44] C. Xu, Y. Wang, T. Tan, and L. Quan, “Automatic 3d face recognition combining global geometric features with local shape variation information,” in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, 2004, pp. 308–313.
- [45] K. W. Bowyer, K. Chang, and P. J. Flynn, “A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition,” *Computer Vision and Image Understanding*, vol. 101, pp. 1–15, 2006.
- [46] E. P. Kukula, S. J. Elliott, R. Waupotitsch, and B. Pesenti, “Effects of illumination changes on the performance of geometrix facevision/spl reg/ 3d frs,” in *Security Technology, 2004. 38th Annual 2004 International Carnahan Conference on*, 2004, pp. 331–337.
- [47] P. J. Phillips, H. Moon, P. Rauss, and S. A. Rizvi, “The feret evaluation methodology for face-recognition algorithms,” in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, 1997, pp. 137–143.
- [48] S. A. Rizvi, H. Moon, and P. J. Phillips, “The feret verification testing protocol for face recognition algorithms,” in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, 1998, pp. 48–53.
- [49] W. Shen, M. Surette, and R. Khanna, “Evaluation of automated biometrics-based identification and verification systems,” *Proceedings of the IEEE*, vol. 85, no. 9, pp. 1464–1478, 1997.

- [50] J. P. Egan, *Signal Detection Theory and ROC Analysis*. Academic Press, New York, 1975.
- [51] R. M. Bolle, N. K. Ratha, and S. Pankanti, “Evaluating authentication systems using bootstrap confidence intervals,” in *15th International Conference on Pattern Recognition, In Proceedings of*, 2000.
- [52] B. Efron and R. J. Tibshirani, *An Introduction to the bootstrap*. Chapman & Hall, New York, 1993.
- [53] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, “The feret evaluation methodology for face-recognition algorithms,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [54] J. R. Beveridge, K. She, B. A. Draper, and G. H. Givens, “A nonparametric statistical comparison of principal component and linear discriminant subspaces for face recognition,” in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, 2001, pp. I-535–I-542 vol.1.
- [55] R. J. Micheals and T. E. Boult, “Efficient evaluation of classification and recognition systems,” in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, 2001, pp. I-50–I-57 vol.1.
- [56] M. Kirby and L. Sirovich, “Application of the karhunen-loeve procedure for the characterization of human faces,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 1, pp. 103–108, 1990.

- [57] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, p. 7186, 1991.
- [58] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 711–720, 1997.
- [59] P. S. Penev and J. J. Atick, "Local feature analysis: a general statistical theory for object representation," *Network: Computation in Neural Systems*, vol. 7, pp. 477–500, 1996.
- [60] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *Neural Networks, IEEE Transactions on*, vol. 13, no. 6, pp. 1450–1464, 2002.
- [61] F. Samaria and S. Young, "Hmm-based architecture for face identification," *Image and Vision Computing*, vol. 12, no. 8, pp. 537–543, 1994.
- [62] A. V. Nefian and I. Hayes, M. H., "An embedded hmm-based approach for face detection and recognition," in *Acoustics, Speech, and Signal Processing, 1999. ICASSP '99. Proceedings., 1999 IEEE International Conference on*, vol. 6, 1999, pp. 3553–3556 vol.6.
- [63] S. Tsutsumi, S. Kikuchi, and M. Nakajima, "Face identification using a 3d gray-scale image—a method for lessening restrictions on facial directions," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, 1998, pp. 306–311.

- [64] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Multimodal 2d and 3d biometrics for face recognition," in *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on*, 2003, pp. 187–194.
- [65] K. I. Chang, K. W. Bowyer and P. J. Flynn, "Face recognition using 2d and 3d facial data," in *Multimodal User Authentication, Workshop in*, 2003, pp. 25–32.
- [66] K. I. Chang, K. W. Bowyer, P. J. Flynn, and X. Chen, "Multi-biometrics using facial appearance, shape and temperature," in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, 2004, pp. 43–48.
- [67] F. Tsalakanidou, D. Tzovaras, and M. G. Strintzis, "Use of depth and colour eigenfaces for face recognition," *Pattern Recognition Letters*, vol. 24, no. 9, pp. 1427–1435, 2003.
- [68] A. Godil, S. Ressler, and P. Grother, "Face recognition using 3d facial shape and color map information: comparison and combination," in *Proceedings of the SPIE - The International Society for Optical Engineering*, vol. 5404. Nat. Inst. of Stand. & Technol., Gaithersburg, MD, USA: SPIE-Int. Soc. Opt. Eng, 2004, pp. 351–361.
- [69] C. Heshner, A. Srivastava, and G. Erlebacher, "A novel technique for face recognition using range imaging," in *Signal Processing and Its Applications, 2003. Proceedings. Seventh International Symposium on*, vol. 2, 2003, pp. 201–204 vol.2.
- [70] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. John Wiley and Sons, New York, 2001.

- [71] R. A. Fisher, “The use of multiple measures in taxonomic problems,” *Annals Eugenics*, vol. 7, pp. 179–188, 1936.
- [72] A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley and Sons, New York, 2001.
- [73] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, and W. Worek, “Preliminary face recognition grand challenge results,” in *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, 2006, pp. 15–24.
- [74] S. Malassiotis and M. G. Strintzis, “Pose and illumination compensation for 3d face recognition,” in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, vol. 1, 2004, pp. 91–94 Vol. 1.
- [75] F. Tsalakanidou, S. Malassiotis, and M. G. Strintzis, “Integration of 2d and 3d images for enhanced face authentication,” in *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, 2004, pp. 266–271.
- [76] F. Tsalakanidou, S. Malassiotis and M. G. Strintzis, “Exploitation of 3d images for face authentication under pose and illumination variations,” in *3D Data Processing, Visualization and Transmission, 2004. 3DPVT 2004. Proceedings. 2nd International Symposium on*, 2004, pp. 50–57.
- [77] R. J. Campbell and P. J. Flynn, “A survey of free-form object representation and recognition techniques,” *Computer Vision and Image Understanding*, vol. 81, no. 2, pp. 166–210, 2001.
- [78] B. Achermann and H. Bunke, “Classifying range images of human faces with

- hausdorff distance,” in *Proceedings 15th International Conference on Pattern Recognition*. Dept. of Comput. Sci., Bern Univ., Switzerland: IEEE Comput. Soc, 2000, vol. vol.2, pp. 809–813.
- [79] P. J. Besl and H. D. McKay, “A method for registration of 3-d shapes,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, no. 2, pp. 239–256, 1992.
- [80] X. Lu, D. Colbry, and A. K. Jain, “Three-dimensional model based face recognition,” in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 1, 2004, pp. 362–366 Vol.1.
- [81] M. O. Irfanöglu, B. Gökberk, and L. Akarun, “3d shape-based face recognition using automatically registered facial surfaces,” in *Proceedings of the 17th International Conference on Pattern Recognition*. Dept. of Comput. Eng., Bogazici Univ., Turkey: IEEE Computer Society, 2004, vol. Vol.4, pp. 183–186.
- [82] G. Pan, Z. Wu, and Y. Pan, “Automatic 3d face verification from range data,” in *Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on*, vol. 3, 2003, pp. III–133–6 vol.3.
- [83] G. Pan, Y. Wu, Z. Wu, and W. Liu, “3d face recognition by profile and surface matching,” in *Neural Networks, 2003. Proceedings of the International Joint Conference on*, vol. 3, 2003, pp. 2169–2174 vol.3.
- [84] T. D. Russ, M. W. Koch, and C. Q. Little, “A 2d range hausdorff approach for 3d face recognition,” in *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on*, vol. 3, 2005, pp. 169–169.

- [85] X. Lu, D. Colbry, A. K. Jain, and D. Zhang, "Matching 2.5d scans for face recognition," in *Biometric Authentication. First International Conference, ICBA 2004. Proceedings (Lecture Notes in Comput. Sci. Vol.3072)*, D. Zhang and A. Jain, Eds. Dept. of Comput. Sci. & Eng., Michigan State Univ., East Lansing, MI, USA: Croucher Found., 2004, pp. 30–36.
- [86] X. Lu and A. K. Jain, "Integrating range and texture information for 3d face recognition," in *Applications of Computer Vision, 2005. (WACV 2005). Proceedings. Sixth IEEE Workshop on*, 2005.
- [87] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "Adaptive rigid multi-region selection for handling expression variation in 3d face recognition," in *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on*, vol. 3, 2005, pp. 157–157.
- [88] M. L. Koudelka, M. W. Koch, and T. D. Russ, "A prescreener for 3d face recognition using radial symmetry and the hausdorff fraction," in *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on*, vol. 3, 2005, pp. 168–168.
- [89] Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," *Image and Vision Computing*, vol. 10, pp. 145–155, 1992.
- [90] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the hausdorff distance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 15, no. 9, pp. 850–863, 1993.
- [91] J. C. Lee and E. Milios, "Matching range images of human faces," in *Computer Vision, 1990. Proceedings, Third International Conference on*, 1990, pp. 722–726.

- [92] H. T. Tanaka and M. Ikeda, “Curvature-based face surface recognition using spherical correlation-principal directions for curved object recognition,” in *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, vol. 3, 1996, pp. 638–642 vol.3.
- [93] H. T. Tanaka, M. Ikeda, and H. Chiaki, “Curvature-based face surface recognition using spherical correlation. principal directions for curved object recognition,” in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, 1998, pp. 372–377.
- [94] E. Paquet, H. H. Arsenault, and M. Rioux, “Recognition of faces from range images by means of the phase fourier transform,” *Pure and Applied Optics*, vol. 4, no. 6, pp. 709–721, 1995.
- [95] S. Chang, M. Rioux, and C. P. Grover, “Range face recognition based on the phase fourier transform,” *Optics Communications*, vol. 222, no. 1, pp. 143–153, 2003.
- [96] C. Beumier, M. Acheroy, P. H. Lewis, and M. S. Nixon, “Automatic face authentication from 3d surface,” in *BMVC 98. Proceedings of the Ninth British Machine Vision Conference*, P. H. Lewis and M. S. Nixon, Eds. Signal & Image Centre, R. Mil. Acad., Brussels, Belgium: Univ. Southampton, 1998, vol. vol.2, pp. 449–458.
- [97] C. Beumier and M. Acheroy, “Face verification from 3d and grey level clues,” *Pattern Recognition Letters*, vol. 22, no. 12, pp. 1321–1329, 2001.
- [98] J. Y. Cartoux, J. T. Lapreste, and M. Richetin, “Face authentication or recognition by profile extraction from range images,” in *Interpretation of 3D Scenes, 1989. Proceedings., Workshop on*, 1989, pp. 194–199.

- [99] G. G. Gordon, “Face recognition based on depth and curvature features,” in *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on*, 1992, pp. 808–810.
- [100] M. Hüsken, M. Brauckmann, S. Gehlen, and C. Von der Malsburg, “Strategies and benefits of fusion of 2d and 3d face recognition,” in *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on*, vol. 3, 2005, pp. 174–174.
- [101] Y. Lee, H. Song, U. Yang, H. Shin, and K. Sohn, “Local feature based 3d face recognition,” in *Audio- and Video-based Biometric Person Authentication, 2005 International Conference on, LNCS*, vol. 3546, 2005, pp. 909–918.
- [102] C.-S. Chua, F. Han, and Y.-K. Ho, “3d human face recognition using point signature,” in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, 2000, pp. 233–238.
- [103] Y. Wang, C. Chua, and Y. Ho, “Facial feature detection and face recognition from 2d and 3d images,” *Pattern Recognition Letters*, vol. 23, no. 10, pp. 1191–1202, 2002.
- [104] Y. Wang and C. Chua, “Face recognition from 2d and 3d images using 3d gabor filters,” *Image and Vision Computing*, vol. 23, no. 11, pp. 1018–1028, 2005.
- [105] A. Colombo, C. Cusano, and R. Schettini, “Tri-dimensional face detection and localization,” in *SPIE Internet imaging VI, in Proceedings of*, S. Santini, R. Schettini, and T. Gevers, Eds., vol. 5670, 2005, pp. 68–75.

- [106] A. Colombo, C. Cusano and R. Schettini, “3d face detection using curvature analysis,” *Pattern Recognition*, vol. 39(3), pp. 444–455, 2006.
- [107] I. A. Kakadiaris, G. Passalis, T. Theoharis, G. Toderici, I. Konstantinidis, and N. Murtuza, “Multimodal face recognition: combination of geometry with physiological information,” in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 2, 2005, pp. 1022–1029 vol. 2.
- [108] G. Passalis, I. A. 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 *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on*, vol. 3, 2005, pp. 171–171.
- [109] J. Kittler, M. Hatef, R. Duin, and J. Matas, “On combining classifiers,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 3, pp. 226–239, 1998.
- [110] A. M. Bronstein, M. M. Bronstein, R. Kimmel, J. Kittler, and M. S. Nixon, “Expression-invariant 3d face recognition,” in *Audio- and Video-Based Biometric Person Authentication. 4th International Conference, AVBPA 2003. Proceedings (Lecture Notes in Computer Science Vol.2688)*, J. Kittler and M. S. Nixon, Eds. Dept. of Electr. Eng., Israel Inst. of Technol., Haifa, Israel: Springer-Verlag, 2003, pp. 62–69.
- [111] X. Lu and A. K. Jain, “Deformation analysis for 3d face matching,” in *Applications of Computer Vision, WACV '05 7th IEEE Workshop on*, 2005, pp. 99–104.

- [112] G. Pan, S. Han, Z. Wu, and Y. Wang, “3d face recognition using mapped depth images,” in *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on*, vol. 3, 2005, pp. 175–175.
- [113] A. W. Gruen, *Close Range Photogrammetry and Machine Vision*, K. B. Atkinson, Ed. Whittles, 1987.
- [114] G. Otto and T. Chau, “‘region-growing’ algorithm for matching of terrain images,” *Image and Vision Computing*, vol. 7, no. 2, pp. 83–94, May 1989.
- [115] R. Tsai, “A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses,” *Robotics and Automation, IEEE Journal of [legacy, pre - 1988]*, vol. 3, no. 4, pp. 323–344, 1987.
- [116] S. Gupta, J. K. Aggarwal, M. K. Markey, and A. C. Bovik, “3d face recognition founded on the structural diversity of human faces,” in *Computer Vision and Pattern Recognition, 2007 IEEE Computer Society Conference on*, Minneapolis, MN, June 2007.
- [117] S. Gupta, M. K. Markey, J. K. Aggarwal, and A. C. Bovik, “3d face recognition based on euclidean and geodesic distances,” in *Electronic Imaging, IS&T/SPIE 19th Annual Symposium on*, San Jose, CA, January 2007.
- [118] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and M. Bone, “Face recognition vendor test 2002 evaluation report,” National Institute of Standards and Technology, Tech. Rep., 2003.
- [119] Y. Wang and C. Chua, “Robust face recognition from 2d and 3d images using

- structural hausdorff distance,” *Image and Vision Computing*, vol. 24, pp. 176–185, 2006.
- [120] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. von der Malsburg, “Face recognition by elastic bunch graph matching,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 775–779, 1997.
- [121] L. G. Farkas, Ed., *Anthropometry of the Head and Face*. Raven Press, New York, 1994.
- [122] J. Comas, *Manual of Physical Anthropology*. Charles C. Thomas, 1960.
- [123] B. O. Rogers, “The role of physical anthropometry in plastic surgery today,” *Clinical Plastic Surgery*, vol. 1, p. 439, 1974.
- [124] D. DeCarlo, D. Metaxas, and M. Stone, “An anthropometric face model using variational techniques,” in *SIGGRAPH*, 1998, pp. 67–74.
- [125] A. Hrdlička, *Practical Anthropometry*. Wister Institute of Anatomy and Biology, Philadelphia, 1939.
- [126] E. W. Dijkstra, “A note on two problems in connexion with graphs,” *Numerische Mathematik*, vol. 1, p. 269–271, 1959.
- [127] J. B. Tenenbaum, V. de Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [128] A. Hamza and H. Krim, “Geodesic matching of triangulated surfaces,” *Image Processing, IEEE Transactions on*, vol. 15, no. 8, pp. 2249–2258, 2006.

- [129] M. P. Do Carmo, *Differential Geometry of Curves and Surfaces*. Prentice-Hall, Inc. New Jersey, 1976.
- [130] S. Sharma, *Applied Multivariate Techniques*. John Wiley and Sons, Inc., New York, 1996.
- [131] P. J. Besl, *Surfaces in Range Image Understanding*. New York: Springer-Verlag, 1988.
- [132] P. J. Besl and R. C. Jain, "Segmentation through variable-order surface fitting," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 10, no. 2, pp. 167–192, 1988.
- [133] D. Marr and E. Hildreth, "Theory of edge detection." in *Royal Society of London, in Proceedings*, vol. B 207, Feb 1980, pp. 187–217.
- [134] J. J. Rodriguez and J. K. Aggarwal, "Matching aerial images to 3-d terrain maps," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 12, pp. 1138–1149, 1990.
- [135] S. Jahanbin, A. C. Bovik, and H. Choi, "Automated facial feature detection from portrait and range images," in *Image Analysis and Interpretation, 2008. SSIAT 2008. IEEE Southwest Symposium on*, 2008, pp. 25–28.
- [136] A. C. Bovik, M. Clark, and W. S. Geisler, "Multichannel texture analysis using localized spatial filters," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 12, no. 1, pp. 55–73, 1990.
- [137] M. Clark, A. C. Bovik, and W. S. Geisler, "Texture segmentation using gabor modulation and demodulation," *Pattern Recognition Letters*, vol. 6, pp. 261–267, September 1987.

- [138] J. Tanaka and M. Farah, *Perception of faces, objects and scenes: Analytical and holistic processes*. Oxford University Press, New York, 2003, ch. The holistic representation of faces, pp. 53–74.
- [139] J. M. Henderson, C. C. Williams, and R. J. Falk, “Eye movements are functional during face learning,” *Memory and Cognition*, vol. 33, no. 1, pp. 98–106, 2005.
- [140] G. H. Golub and C. F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, 1996.
- [141] K. S. Arun, T. S. Huang, and S. D. Blostein, “Least-squares fitting of two 3-d point sets,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 5, pp. 698–700, 1987.
- [142] S. Gupta, M. P. Sampat, Z. Wang, M. K. Markey, and A. C. Bovik, “Facial range image matching using the complex-wavelet structural similarity metric,” in *Applications of Computer Vision, WACV '07 8th IEEE Workshop on*, Austin, TX, 2007.
- [143] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. M. K., “Complex wavelet structural similarity: A new image similarity index,” *IEEE Transactions on Image Processing*, vol. (In preparation), 2008.
- [144] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *Image Processing, IEEE Transactions on*, vol. 13, no. 4, pp. 600–612, 2004.
- [145] P. Y. Simard, Y. A. Le Cun, J. S. Denker, and B. Victorri, “Transformation invariance in pattern recognition: tangent distance and propagation,” *Interna-*

- tional Journal of Imaging Systems and Technology*, vol. 11, no. 3, pp. 181–197, 2000.
- [146] M. P. Sampat, Z. Wang, G. J. Whitman, T. Stephens, M. K. Markey, and A. C. Bovik, “Measuring intra- and inter-observer agreement in identifying and localizing structures in medical images,” in *IEEE International Conference on Image Processing*, vol. 08, no. 11, Atlanta, Georgia, October 2006.
- [147] Z. Wang and E. P. Simoncelli, “Local phase coherence and the perception of blur,” in *Adv. Neural Information Processing Systems (NIPS03)*, vol. 16. Cambridge, MA: MIT Press, May 2004.
- [148] A. Oppenheim and J. Lim, “The importance of phase in signals,” *Proceedings of the IEEE*, vol. 69, no. 5, pp. 529–541, 1981.
- [149] D. J. Fleet and A. D. Jepson, “Stability of phase information,” *IEEE Trans. Pattern Analysis Machine Intell.*, vol. 15, no. 12, pp. 1253–1268, 1993.
- [150] J. Daugman, “Statistical richness of visual phase information: update on recognizing persons by iris patterns,” *Int’l J Computer Vision*, no. 45, pp. 25–38, 2001.
- [151] C. Kuglin and D. Hines, “The phase correlaton image alighment method,” in *Proc. IEEE Int. Conf. Cybern. Soc.*, 1975, pp. 163–165.
- [152] M. C. Morrone and R. A. Owens, “Feature detection from local energy,” *Pattern Recognition Letters*, vol. 6, pp. 303–313, 1987.
- [153] P. Kovesi, “Phase congruency: A low-level image invariant,” *Psych. Research*, vol. 64, pp. 136–148, 2000.

- [154] J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *Int’l J Computer Vision*, vol. 40, pp. 49–71, 2000.
- [155] J. A. Solomon and D. G. Pelli, “The visual filter mediating letter identification,” *Nature*, vol. 369, pp. 395–397, 1994.
- [156] B. F. J. Logan, “Information in the zero crossings of bandpass signals,” *The Bell Systems Technical Journal*, vol. 56, pp. 487–510, 1977.
- [157] S. Curtis, A. Oppenheim, and J. Lim, “Signal reconstruction from fourier transform sign information,” *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing]*, *IEEE Transactions on*, vol. 33, no. 3, pp. 643–657, 1985.
- [158] D. Rotem and Y. Zeevi, “Image reconstruction from zero crossings,” *Acoustics, Speech, and Signal Processing [see also IEEE Transactions on Signal Processing]*, *IEEE Transactions on*, vol. 34, no. 5, pp. 1269–1277, 1986.
- [159] J. Portilla and E. P. Simoncelli, “A parametric texture model based on joint statistics of complex wavelet coefficients,” *International Journal of Computer Vision*, vol. 40, no. 1, pp. 49–71, 2000.
- [160] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, “Shiftable multiscale transforms,” *Information Theory, IEEE Transactions on*, vol. 38, no. 2, pp. 587–607, 1992.
- [161] K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*. Academic Press, London, 1979.

- [162] D. Colbry and G. Stockman, "Identity verification via the 3bid face alignment system," in *Applications of Computer Vision, 2007. WACV '07. IEEE Workshop on*, 2007, pp. 2-2.

Vita

Shalini Gupta was born in Chandigarh, India on December 10, 1979. She is the daughter of Usha Gupta and Ram Niwas Gupta. She attended the Carmel Convent School, Chandigarh, India for her primary and secondary education. In June 2002, she received the degree of Bachelor of Electronics and Electrical Communication Engineering from the Punjab Engineering College, Chandigarh, India. She then enrolled for graduate studies at The University of Texas at Austin and in May 2004 she received the degree of Master of Science in Electrical and Computer Engineering. In August 2004, she entered the doctoral program in the department of Electrical and Computer Engineering at the University of Texas at Austin. She is the recipient of the 2001 Summer Research Fellowship from the Jawaharlal Nehru Center for Advanced Scientific Research, Bangalore, India.

Permanent Address: House No. 990, Sector 9,
Panchkula, Haryana 134113,
India

This dissertation was typeset with L^AT_EX 2_ε¹ by the author.

¹L^AT_EX 2_ε is an extension of L^AT_EX. L^AT_EX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuths TEX Program.