

University of Tennessee, Knoxville Trace: Tennessee Research and Creative Exchange

Masters Theses

Graduate School

12-2005

Pose Invariant 3D Face Authentication based on Gaussian Fields Approach

Venkat Rao Ayyagari University of Tennessee - Knoxville

Recommended Citation

Ayyagari, Venkat Rao, "Pose Invariant 3D Face Authentication based on Gaussian Fields Approach." Master's Thesis, University of Tennessee, 2005. https://trace.tennessee.edu/utk_gradthes/1593

This Thesis is brought to you for free and open access by the Graduate School at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Masters Theses by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a thesis written by Venkat Rao Ayyagari entitled "Pose Invariant 3D Face Authentication based on Gaussian Fields Approach." I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Electrical Engineering.

Mongi A. Abidi, Major Professor

We have read this thesis and recommend its acceptance:

Andreas Koschan, Seong G. Kong

Accepted for the Council: <u>Carolyn R. Hodges</u>

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

To the Graduate Council:

I am submitting herewith a thesis written by Venkat Rao Ayyagari entitled "Pose Invariant 3D Face Authentication based on Gaussian Fields Approach". I have examined the final electronic copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Electrical Engineering.

Mongi A. Abidi

Major Professor

We have read this thesis and recommend its acceptance:

Andreas Koschan

Seong G. Kong

Accepted for the Council:

Anne Mayhew

Vice Chancellor and Dean of Graduate Studies

(Original signatures are on file with official student records.)

Pose Invariant 3D Face Authentication based on Gaussian Fields Approach

A Thesis

Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Venkat Rao Ayyagari

December 2005

Dedicated to my parents, Vishwanath Ayyagari and Lakshmi Ayyagari

Acknowledgments

With a deep sense of gratitude, I wish to express my sincere thanks to my advisor Dr. Mongi Abidi for his immense help in planning and executing the works in time. The confidence and dynamism with which Dr. Abidi guided the work requires no elaboration. I owe him lots of gratitude for having me shown this way of research. His company and financial support would be remembered lifelong. On the other hand, profound knowledge and timely wit came as a boon under the guidance of Dr. Andreas Koschan. His valuable suggestions as final words during the course of the research are greatly acknowledged. What I know today about the process of research, I learned from Dr. Abidi and Dr. Koschan.

My sincere thanks are due to Dr. Seong Kong, for serving on my thesis committee and providing me valuable suggestions. I specially thank Dr. Andrei Gribok for the help extended to me when I approached him and the valuable discussion that I had with him during the course of research would be cherished. Special thanks are due to Dr. Besma Abidi and Dr. David Page for critiquing me during my presentations and showing directions to solve those problems. The cooperation I received from other faculty members of this department is gratefully acknowledged. I will be failing in my duty if I do not mention the laboratory staff and administrative staff of this department for their timely help, especially Justin who was always ready with his magic touch to resolve any computer issues.

I wish I would never forget the company I had from my fellow lab mates of IRIS Lab. In particular, I am thankful to Faysal for his initial push in my research. It was a pleasure to work along with Priya, Anis, Sophie, Shafik, Chris, Hong, Jessica, Nikhil, Rangan, Vivek, Brad, Hariharan, Sarat, and Hari.

The episode of acknowledgement would not be complete without the mention of my roommates Prasad and Anil as they have helped to lighten many a dark day with their banter and jokes, thank you guys.....you rock. I couldn't have asked for a more supportive Gult gang (Spidermen) in Golf Range who have mastered the art of elevators and ladders.

I also want to thank my parents, who taught me the value of hard work by their own example. I would like to share this moment of happiness with my mother Lakshmi and sister Gayathri. They rendered me enormous support during the whole tenure of my research. Without their loving support and understanding, I would never have completed my present work. Finally, I would like to thank all whose direct and indirect support helped me completing my thesis in time.

Abstract

This thesis presents a novel illuminant invariant approach to recognize the identity of an individual from his 3D facial scan in any pose, by matching it with a set of frontal models stored in the gallery. In view of today's security concerns, 3D face reconstruction and recognition has gained a significant position in computer vision research. The non intrusive nature of facial data acquisition makes face recognition one of the most popular approaches for biometrics-based identity recognition. Depth information of a 3D face can be used to solve the problems of illumination and pose variation associated with face recognition.

The proposed method makes use of 3D geometric (point sets) face representations for recognizing faces. The use of 3D point sets to represent human faces in lieu of 2D texture makes this method robust to changes in illumination and pose. The method first automatically registers facial point-sets of the probe with the gallery models through a criterion based on Gaussian force fields. The registration method defines a simple energy function, which is always differentiable and convex in a large neighborhood of the alignment parameters; allowing for the use of powerful standard optimization techniques. The new method overcomes the necessity of close initialization and converges in much less iterations as compared to the Iterative Closest Point algorithm. The use of an optimization method, the Fast Gauss Transform, allows a considerable reduction in the computational complexity of the registration algorithm. Recognition is then performed by using the robust similarity score generated by registering 3D point sets of faces. Our approach has been tested on a large database of 85 individuals with 521 scans at different poses, where the gallery and the probe images have been acquired at significantly different times. The results show the potential of our approach toward a fully pose and illumination invariant system. Our method can be successfully used as a potential biometric system in various applications such as mug shot matching, user verification and access control, and enhanced human computer interaction.

Contents

1	INT	TRODUCTION	1
	1.1	Problem Statement	3
	1.2	Motivation	5
	1.3	Thesis Contributions	6
	1.4	Thesis Outline	7
2	RE	LATED WORK	8
	2.1	3D Free Form Registration	8
	2.1.	1 The Iterative Closest Point Algorithm	9
	2.1.	2 Modifications to the Iterative Closest Point Algorithm	9
	2.1.	3 Registration with Invariant Features	10
	2.2	3D Face Recognition	11
	2.2.	1 Profiles/Sections/Contours based Approaches	11
	2.2.	2 Curvature based Approaches	12
	2.2.	3 PCA based Approaches	14
	2.2.	4 Point Signature based Approaches	15
	2.2.	5 Template Matching/ Brute Force/ Distance Map	16
	2.2.	6 Global Features and Local Shape Variation based Approaches	16
	2.2.	7 ICP based Methods	17
	2.2.	8 Model Fitting Methods	17
	2.2.	8 Other Methods	18
3	GA	USSIAN FIELDS FOR 3D FACE REGISTRATION AND	
R	ECOG	NITION	19
	3.1	Gaussian Fields and Energy Function	19
	3.2	Attributes	22
	3.3	Optimization	23
	3.4	Fast Gauss Transform	25
	3.5	Extension of the Gaussian Criterion for Recognition	26
4	RE	SULTS FOR 3D FACE REGISTRATION	28
	4.1	Introduction and Objectives	
	4.2	Data Acquisition	29
	4.3	Effect of Varying the Parameter σ	

		vi
4.4	Noise Analysis	
4.5	Resolution Analysis	
4.6	Effect of Overlap	
4.7	Comparison with ICP	
5 RI	ESULTS FOR 3D FACE RECOGNITION	49
5.1	3D Face Database	
5.2	Implementation Details	
5.3	Recognition Scores	
5.3	3.1 Frontal Faces	
5.3	B.2 Effect of Pose Variations on Recognition	
5.4	Recognition with Glasses	67
5.5	Recognition with Complete 3D Head Models	
5.6	Recognition with Expression Variations	
5.7	Recognition System Performance	
5.8	Accuracy based on Database Size	
5.9	Computational Time	
6 CC	DNCLUSIONS AND FUTURE WORK	100
6.1	Thesis Summary	
6.2	Future Research	
BIBLI	OGRAPHY	103
VITA		

Tables

5.9. The normalized similarity scores computed on a gallery of complete head models.

The horizontal row represents the gallery and vertical column represents probe dataset consisting of frontal face models, (b) a graphical representation of the scores.

5.12. The normalized similarity scores computed on a gallery of complete head models The faces in the horizontal row are the complete frontal faces from the gallery whereas the vertical columns contain probe faces with glasses and pose of 30° , (b) a graphical representation of the scores. 82

5.15. The scores computed based on Gaussian criterion with expressions. The faces in the horizontal row are the frontal faces from the gallery whereas the vertical columns contain frontal faces with shocked expression, (b) a graphical representation of the scores. 89

5.16. The scores computed based on Gaussian criterion with expressions. The faces in the horizontal row are the frontal faces from the gallery whereas the vertical columns contain frontal faces with sad expression, (b) a graphical representation of the scores.

Figures

1.1. Block diagram of a typical face recognition system 4
3.1. Framework of the automatic face recognition based on 3D facial data
4.1. The data used in the experiments (a) Mannequin Head Data (b) Original Face Data
4.2. The two range scanning systems used in our registration experiments. (a) Genex3D FaceCam, (b) IVP Ranger 31
4.3. Triangulation principle for 3D imaging
4.4. Rainbow principle: θ_i is calculated by solving one-to-one correspondence problem between color λ_i and projection angle $\theta_i \cdot \alpha_{k,l}$ is geometrically calculated using the coordinates of each pixel $(k; l)$ in the image of the sensor. Then, using triangulation principle, each visible point <i>O</i> of the object can be calculated
4.5. 3D model generated from Genex 3D FaceCam 33
4.6. The recommended set up for Genex 3D FaceCam 500
4.7. The Genex 3D FaceCam experimental setup (a) Front view of the experimental setup mounted on a tripod (b) Side view
4.8. Position/ Capture screen of the Genex 3D FaceCam. The head should be positioned in center for a better result
4.9. Processing screen of the Genex 3D FaceCam. The color coded structured light is also observed in the first and third windows
4.10. Plots showing (a) the rotation and (b) translation error of the real face data and the synthetic face as a function of parameter σ . The parameter sigma and translation error are in terms of fraction of the length of the face model
4.11. Registration error versus uniform noise level; (a) rotation error in degrees and (b) translation error as a fraction of the length of the face model. We show plots for three

values of the confidence parameter	39
------------------------------------	----

Х

4.16. Effect of noise on the registration accuracy for models with different amount of overlap; (a) rotation error and (b) translation error for different values of overlap..... 45

4.17. Effect of sampling on the registration accuracy for different overlap models; (a) rotation error and (b) translation error in terms of length of the model... 47

5.5. Graphical representation of the effect of noise on the similarity scores using grayscale coding for frontal faces. The horizontal row represents the gallery and vertical column represents the uniform noise (%) added to the probe model A_P . The

similarity between the probe and the gallery decreases as the intensity increases...... 59

5.15. Graphical representation of the similarity score using grayscale coding for a probe dataset which consists of faces with glasses and pose of 50° . The faces in the

5.17. Graphical representation of the similarity score using grayscale coding on a gallery of complete head models. The horizontal row represents the gallery and vertical column represents probe dataset consisting of frontal face models. The similarity between the probe and the gallery decreases as the intensity increases...... 77

5.18. Graphical representation of the similarity score using grayscale coding on a gallery of complete head models. The horizontal row represents the gallery and vertical column represents probe dataset consisting of face models with a pose of 30° . The similarity between the probe and the gallery decreases as the intensity increases.....

xii

5.25. Graphical representation of the similarity score using grayscale coding for faces with expressions. The faces in the horizontal row are the frontal faces from the gallery whereas the vertical columns contain frontal faces with shocked expression. The similarity between the probe and the gallery decreases as the intensity increases...... 90

5.27. The receiver operating characteristic curve for 2.5D face gallery. The EER is around 3.7%
5.28. The cumulative match characteristic curve for 2.5D face gallery
5.29. The receiver operating characteristic curve for the complete head gallery. The EER is around 3.6%
5.30. The cumulative match characteristic curve for complete head models
5.31. Effect of database size on the recognition accuracy
5.32. Computational time for our recognition system. (a) The X axis represents the number of points and the Y axis represents the time taken. (b) The enlarged version

1 INTRODUCTION

Vision is indeed the paragon of the human senses. With it we can distinguish millions of shades of colors, recognize thousands of faces instantly, and even detect the presence of a single photon of light. The popular saying "Seeing is Believing" ascribes an enormous power to the image as the embodiment of truth. The vast amount of information that sighted individuals acquire comes through the eyes and this reiterates the belief that we tend to use visual medium for communicating and recording information. Infact, almost one third the brain is involved in processing visual information.

The effortless ease and accuracy with which the eyes present the world to us tends to foster an attitude that this process can be easily replicated by machines. Our ability to construct and process visual images is clearly demonstrated in our pattern recognition abilities that are absolutely crucial for dealing with and understanding the world. We are pattern-seeking animals because recognizing patterns in nature allowed our ancestors to survive. However, it is not possible now nor will it be in near future to create a computing machine that actually understands what it sees and matches our abilities. The main difficulty in vision problems is that almost all of them are ill-defined or ill-posed, mainly because the information is lost in the transformation from the 3D world to a 2D image. Therefore, we cannot uniquely reconstruct the 3D representation from the 2D image and multiple solutions are often 'correct'.

Within the field of computer vision, a considerable amount of research has been performed since time immemorial, particularly in the areas of biometrics. The September 11 and the July 7 terrorist attacks have changed the way, the world looks towards security. Hence, the need for a robust and effective biometric system for security application has been highlighted by security agencies all over the world. Biometrics was traditionally defined as the study of measurable biological characteristics. However, in computer vision, biometrics refers to a measurable physical or behavioral characteristic used to recognize the identity, or verify the claimed identity of a person through automated means. Biometrics represents a more secure way to identify individuals because instead of verifying identity and granting access based on the possession or knowledge of cards, passwords, tokens, or keys, an individual is recognized based on his unique physical and biometric characteristic. There are several types of biometric identification schemes:

- Face: the analysis of facial characteristics.
- **Fingerprint**: the analysis of an individual's unique fingerprints.
- **Hand geometry**: the analysis of the shape of the hand and the length of the fingers.
- **Retina**: the analysis of the capillary vessels located at the back of the eye.
- Iris: the analysis of the colored ring that surrounds the eye's pupil.
- Signature: the analysis of the way a person signs his name.
- Vein: the analysis of pattern of veins in the back of the hand and the wrist.
- Voice: the analysis of the tone, pitch, cadence, and frequency of a person's voice.
- Gait: the analysis of the individual's walking pattern.
- Ear: the analysis of the human ear characteristics.

The human face remains one of the most popular and irreplaceable cues for identity recognition in biometrics, despite the existence of alternative technologies such as fingerprint or iris recognition. This is majorly attributed to the non-intrusive nature of face recognition methods, which makes them especially suitable for surveillance purposes. Other biometric methods do not possess these advantages as they require some voluntary action. For instance, retinal recognition method requires an individual to look into the eyepiece while some light is being reflected off the back of the eye to capture the vein patterns. Similarly, fingerprint recognition methods require the users to make explicit physical contact with the surface of a sensor. Iris scans can provide very high accuracy rates for personal identification but scanning the iris is an expensive and a motion sensitive process. Voice recognition is not robust in noisy environments like public places and is sensitive to throat conditions when people are sick with colds. It can be easily fooled by using a recorded correct person's voice on a tape. In a similar fashion, signature as a biometric cue suffers from reliability because people tend to vary their signatures from time to time and from mood to mood. Furthermore, people physiologically tend to associate fingerprints with crime which is not the case with faces. Fingerprinting and iris recognition are viewed distrustfully by the general public as these system are assumed to have an element of 'Big Brother' about them. The easy availability of face database along with their inherent nature of human readable media justifies the need to use faces as a potential biometric source. Nevertheless, despite the above mentioned advantages of face recognition as a potential source of biometric system, there are some other issues which cast a gloom over its dominance.

In practical scenario, automated face recognition system operates in three modes which are described below.

• Identification/ Recognition: "Who are you?"

In this mode an image of an unknown individual (probe) is collected and the identity is found by searching a large number of images in the database known

as gallery. The nearest match is reported as the identity of the probe and if requested, top N similar matches are also reported.

- Verification/ Authentication: "Are you the same person, you claim to be?" In this mode, rather than identifying the person, the system takes the probe and matches against the claimed image from the gallery after the person submits an ID. The system provides the result in the form of YES/NO.
- **Surveillance**: "*Are we currently looking for you?*" This mode is similar to the recognition mode, but uses an additional threshold to identify a hit. The gallery size is small database of the intruders and the system triggers an alarm only if the probe matches with any image from the gallery.

Face recognition systems are no longer limited to identity verification and surveillance tasks. It has myriad applications in the areas such as medicine, law enforcement, and entertainment. Growing numbers of applications are starting to use face-recognition as the initial step towards interpreting human actions, intention, and behavior, as a central part of next-generation smart environments. Many of the actions and behaviors humans display can only be interpreted if you also know the person's identity, and the identity of the people around them. Examples are a valued repeat customer entering a store, or behavior monitoring in an eldercare or childcare facility, and command-and-control interfaces in a military or industrial setting. In each of these applications identity information is crucial in order to provide machines with the background knowledge needed to interpret measurements and observations of human actions. Some more areas where face recognition is utilized are:

- Physical access control in areas like networks, ATMs, mobile services
- Time and attendance
- Border control, Drug trafficking
- Virtual reality, Human computer interaction
- Security applications like check-in at airports
- Secure financial transactions, Internet banking
- Detection and tracking of people in asylums and prison facilities
- Human flow analysis in shopping centers

1.1 Problem Statement

"When humans are generally very good at recognizing faces, why do we need an automatic face recognition system?"

People are generally very good at recognizing faces that they know. However, they experience difficulties when they perform facial recognition in surveillance or watch

post scenario. This may be attributed to different factors such as short attention spans, difficulty in recognizing unfamiliar faces. In addition to unfamiliar face recognition problems, the ability of human beings to detect critical signals drops rapidly from the start of a task, and their ability to focus their attention drops significantly after just half an hour. Thus, there is an indispensable need for a system that can automatically recognize faces without any manual intervention.

Automated face recognition can be defined as a system that looks through a stored set of signatures in the gallery and picks the one that best matches the features of the unknown individual. Our problem statement is shown in Fig. 1.1 in the form of a pipeline depicting a typical face recognition system. The face image (probe) is captured by a sensor and is then subjected to some preprocessing operations to reduce noise. However, the lighting, background, scale, pose, and parameters of acquisition are all variables in facial images acquired under real-world scenarios. Suitable features are extracted from the image and the obtained signature is normalized so that they are in the same form as the signatures in the gallery. The obtained signature is matched with the signatures of the other images in the gallery and a score is computed. The best match is chosen based on the score and the degree of match is compared to the threshold. If the match is close enough, the probe image is identified as belonging to the individual whose signature produced the best match.



Fig. 1.1: Block diagram of a typical face recognition system.

1.2 Motivation

Increasing demands from security applications (e.g., surveillance, secure access, human/computer interface) and the availability of cheap and powerful hardware led to the development of many commercial face recognition systems. Most of the commercially available face recognition systems have used 2D images of human faces, the reason being the cost effectiveness and easy availability of 2D sensors. However, 2D face recognition techniques are known to suffer from the inherent problems of illumination, pose, and are sensitive to factors such as occlusion, change in human expression, and aging. The appearance of human faces is subject to several different factors mentioned above. As stated by Moses et al. [Moses94] "The variations between the images of the same face due to illumination and viewing directions are almost larger than image variations due to the change in the face identity".

Utilizing 3D face information was shown to improve face recognition performance, especially with respect to pose variations [Blanz03, FRVT02]. Range images captured by 3D sensor provide much more information than a conventional 2D sensor. These models are more accurate because the range sensor captures absolute measurements invariant to camera distance. Since the complete geometry of a person's face is available instead of just color and texture, the models are invariant to illumination change. Pose normalization in 3D space turns out to be a significant advantage of such a technology. This is in contrast to the pose normalization from 2D images, which is a significant challenge considering that information is lost in the transformation from the 3D world to a 2D image. Also, enough invariant information is present to cope with change in expressions and other occlusions such as glasses and beard.

Face recognition based on 3D data has been addressed in many different ways [Bowyer04]. Morphing was tested in the latest FRVT 2002 but the method requires human intervention to align a 2D image to a generic 3D model and introduces an additional algorithm to deal with illumination. The ICP approach also requires an additional step for close initialization, which is mostly done manually and suffers from slowness and problems with local minima. The method proposed by us is fully automatic, does not require an initialization step, and converges rapidly to a global maximum.

We propose a technique that uses 3D geometric (point sets) for face representation. The use of 3D point sets to represent human faces in lieu of 2D texture makes this method robust to changes in illumination and pose. The method first automatically registers facial point-sets of the probe and the gallery through a criterion based on Gaussian force fields. The registration method defines a simple energy function, which is always differentiable and convex in a large neighborhood of the alignment parameters; allowing for the use of powerful standard optimization techniques. The new method overcomes the necessity of close initialization, and converges in much

less iterations as compared to the Iterative Closest Point algorithm (ICP) [Besl92]. The use of an optimization method, the Fast Gauss Transform, allows a considerable reduction in the computational complexity of the registration algorithm. Recognition is then performed by using the robust similarity score generated by registering 3D point sets of faces.

1.3 Thesis Contributions

The main contributions of this thesis are as follows:

- Firstly, we built a 3D face database named as IRIS 3D Face Database [IRIS3DFD] for automatic face recognition experiments and other possible applications such as pose correction and 3D face model registration. This 3D face database is one of its kinds due to the variety and variations in the face models. The richness of image quality with pronounced variations such as expressions, glasses, and various pose are some of the most relevant aspects of our database. The database consists of 495 three dimensional facial surfaces corresponding to 25 individuals taken over a period of time. Most of the individuals are aged between 20 years to 35 years, but vary in gender and ethnicity. There are systematic variations over pose, facial expression of each person. Complete ear to ear face models (25 individuals) are built by registering these different views of each individual. However, the striking feature of our database is the 3D facial surfaces of people with glasses.
- We then present a set of experiments to test the robustness of the 3D registration method [Boughorbel04] to various factors using 3D facial datasets. These factors include the effect of noise, resolution accuracy, and amount of overlap between the two datasets. A comparison of the region of convergence with the standard Iterative Closest Point algorithm is also undertaken.
- Finally, we propose a 3D face recognition strategy which is invariant to light and is capable of recognizing faces of individuals over a wide variety of poses. The strategy involves extending the 3D registration algorithm [Boughorbel04] and utilizing it for the purpose of recognizing faces. However, the major contribution of this thesis is the evolution of a similarity score for faces based on the registration algorithm. The recognition results generated using the registration of the facial datasets and the corresponding similarity scores demonstrate the effectives of our method.

1.4 Thesis Outline

The remainder of this thesis is organized as follows:

- Chapter 2 presents a literature review of the topics most relevant to our research. Namely: 3D free-form registration and 3D face recognition.
- Chapter 3 describes the theory utilized in our work including the development of Gaussian Fields framework using mollification and relaxation approaches. The attributes used in the criterion are also described along with the optimization strategy used. The Fast Gauss Transform along with its usefulness is also discussed. Finally, the recognition pipeline and the similarity score is evolved.
- Chapter 4 contains a thorough experimental analysis on 3D face datasets of the 3D registration method. Included are studies of robustness to noise, overlap, resolution, as well as convergence properties.
- Chapter 5 describes the relevant experiments for the face recognition approach and the recognition results obtained with the 3D face database we have used.
- Chapter 6 will present a short summary of the advantages of our recognition method along with the concluding remarks, and opportunities for future research.

2 RELATED WORK

Automatic reconstruction of 3D face models typically involves three stages: a data acquisition stage, wherein the samples of the face are collected from different views using sensors; a data registration stage, which aligns the different 3D views into a common coordinate system; and an integration stage, which simplifies the aligned views into parametric models. Generally, some parts of the face will be unobservable from any given position, either due to occlusion or limitations in the sensor's field of view. When seen from a slightly different viewpoint, the missing data in unobserved regions is readily apparent. However, these different views will be in their local coordinate system and some transformations have to be employed to align these views in a common coordinate system. It is in this capacity that registration becomes an integral part of the reconstruction process.

3D face reconstruction techniques can be broadly classified into active and passive methods, based on their imaging modalities [Bronstein03]. Active reconstruction techniques such as laser scan and structured light use external sources of illumination for reconstruction. Passive techniques such as stereo vision, morphing, structure from motion, etc. do not depend on external sources of illumination. Most of the above mentioned methods make use of registration techniques in the process of building a complete face model. The bulk of this chapter is devoted to discuss in detail about the related literature relevant to our work. The state of the art in 3D free form registration would be discussed first followed by the recent advances in the field of 3D face recognition.

2.1 3D Free Form Registration

The majority of the registration algorithms attempt to solve the classic problem of absolute orientation: finding a set of transformation matrices that will align all the data sets into a world coordinate system [Horn87]. In the literature, a common distinction is found between fine and coarse registration methods [Campbell01]], which are often used in a two stage fashion: a coarse registration followed by fine registration using the ICP and its variants.

2.1.1 The Iterative Closest Point Algorithm

The original ICP algorithm developed by Besl and MacKay [Besl92] aligns the two point sets by minimizing the sum of squared distances between them. It is a locally convergent scheme that requires parameter initialization close to the aligned position. First described by Besl and McKay [Besl92], ICP is the standard solution to register two roughly aligned 3D point sets D_1 and D_2 . At each ICP iteration, each point of D_1 is paired with the closest point in D_2 and a transformation is computed that minimizes the mean squared error (MSE) between the paired points. The new transformation is applied to D_1 and MSE is updated. The above steps are iterated until the MSE falls between a certain threshold or a maximum number of iterations is reached. Without *apriori* approximate estimate of the transformation, the ICP often ends in a local minimum instead of the global minimum which represents the best transformation. Hence, a good estimate of the initial transformation between point sets is required.

2.1.2 Modifications to the Iterative Closest Point Algorithm

Modifications to the original ICP algorithm have been made to improve the convergence and register partially overlapping datasets. Chen and Medioni [Chen92] used an iterative refinement of initial coarse registration between views to perform registration utilizing the orientation information. They devised a new least square problem where the energy function being minimized is the sum of the distances from points on one view surface to the tangent plane of another views surface. Zhang [Zhang94] proposed a method based on heuristics to remove inconsistent matches by limiting the maximum distance between closed points allowing registration of partially overlapping data. While the basic ICP algorithm was used in the context of registration of cloud of points, Turk and Levoy [Turk94] devised a modified registration the set of mesh vertices to achieve efficient local search.

In order to improve the robustness of ICP, Masuda and Yokoya [Masuda95] used a Least Mean Square (LMS) error measure that is robust to partial overlap. The integration of ICP algorithm with random sampling and the LMS estimator has an added advantage of reduced computational complexity. The algorithm is carried out in two stages: an initial stage which calculates the motion parameters followed by a stage which evaluates the quality of the estimation. Some other methods involved in the same effort at robustness were the Least Median Squares (LMedS) proposed by Trucco et al. [Trucco99], and Minimum Variance Estimate (MVE) of the registration error proposed by Dorai et al. [Dorai97]. In contrast to the approach adopted by Masuda and Yokoya, Trucco et al utilize a dynamic translation estimate based on outlier free data in the ICP iteration which is facilitated due to the integration of LMedS and ICP algorithm. This leads to a larger basin of convergence and more accurate registrations than ICP. Dorai et al [Dorai97] employed the variance of the

point to plane distance as a measure of uncertainty in the distance resulting from noise and this minimum variance estimator is used to estimate the transformation parameters reliably. Also some other variants were introduced for reducing the computational complexity such as the use of k-D trees to partition datasets [Zhang94], and the use of spatial subdivision to partition mesh vertices [Turk94].

2.1.3 Registration with Invariant Features

Stoddart et al. [Stoddart96] studied the relationship between surface shape complexity and registration accuracy, and devised a force based optimization method to register the datasets. They addressed the registration problem based on an analogy with physical system of rigid bodies connected by springs. The equations of motion considering the friction in play evolve over time to a local minimum in potential energy. Finally, the registration problem is solved by integrating the equations of motion over time. Early work by Arun et al. [Arun87] on estimating 3D rigid body transformations presented a solution using the singular value decomposition (SVD). The method requires a connected set of correspondences and accurately registers the 3D data. Faugeras and Hebert [Faugeras87] employed the quaternion method to solve the registration problem directly. They solved the minimization of the cost function for rigid motion in a quadratic form of a unit quaternion, which is 4D vector that determines a 3D rotation matrix.

Eggert et al. [Eggert98] proposed a method in which data from each view is passed through Gaussian and Median filters, and point position and surface normal orientation are used to establish correspondence between points. They claim a larger radius of convergence of up to 20° while eliminating any need of distance threshold for removing outlier correspondence. Chen et al. [Chen99] proposed a random sample consensus (RANSAC) scheme that is used to check all possible data-alignments of two data sets. They formulate the registration problem as an optimization problem which uses rigidity-constraints in the search space, thus making the process more efficient. The authors claim that their scheme works with featureless data, requires no initial pose estimate, and is not influenced by outliers. Blais and Levine also formulated the 3D registration task as an optimization problem of the error function computed by the sum of Euclidian distances between a set of control points on one of the surfaces and their corresponding points on the other. The resulting optimization problem was approached through a very fast simulated reannealing (VFSR) technique. The non differentiability of the ICP cost function imposes the use of specialized heuristics for optimization. Addressing the registration in the context of gradient-based optimization has attracted some interest recently. In his work, Fitzgibbon [Fritzgibbon03] showed that a Levenberg-Marquardt approach to the point set registration problem offers several advantages over current ICP methods. The proposed method uses Chamfer distance transforms to compute derivatives and Huber kernels to widen the basins of convergence of existing techniques. The method overcome the limitations of the ICP algorithm by introducing a straightforward differentiable cost function, explicitly expressed in terms of point coordinates and registration parameters.

2.2 3D Face Recognition

Although the research in the field of 3D face recognition dates back to many years, much literature has not been published on this topic. This section summarizes and critiques the state of the art related to our work. 3D face recognition can be broadly classified into two categories: appearance based methods and feature based methods. Appearance based methods such as Eigenface method treats the entire face as a global entity whereas the feature based methods use the relationship between the different features of the face as a measure of facial similarity. However, we further categorize the 3D face recognition methods as described in the sections below.

2.2.1 Profiles/ Sections/ Contours based Approaches

Cartoux et al. [Cartoux89] approached 3D face recognition by the use of both 3D face surface and profile. However, these two modes are not combined explicitly; rather the profile is used to aid the overall process for face matching. The range image is segmented based on principal curvature and a plane of bilateral symmetry is found out which aids in pose normalization. This is done by roughly finding the symmetry plane and profile along the nose tip and later refining iteratively to produce the necessary transformation matrix, also used for face matching in later stages. They consider methods of matching the profile from the plane of symmetry and of matching the face surface utilizing correlation coefficient and mean quadratic distances as a similarity measure, and report 100% recognition for either in a small dataset. However, the performance of their algorithm is affected by the quality of the data more in profile than in frontal face.

Nagamine et al. [Nagamine92] approached the human face identification problem by analyzing the 3D facial section obtained by the intersections of vertical plane, horizontal plane, and cylinder on the face surface. Based on heuristics they find five feature points such as inner corner of eyes, the top and the bottom of the nose, and nasion and later utilize the obtained information for pose normalization. A template profile is created by averaging nine images out of the available ten for each subject, and feature vectors consisting of section curves are extracted. For the purpose of matching, the difference between the two patterns is evaluated with the Euclidean distance between feature vectors. The reported recognition accuracy is highest (100%) for both vertical profile and circular profile in the upper part of the face when

compared with the horizontal profile (96.3%). However, the section extraction was not robust enough against locational and rotational variations.

Pan et al. [Pan03] present an approach for automatic 3D face verification comprising of range data registration followed by comparison. They employ a two stage registration process in order to improve the accuracy and speed up the process. The registration process consists of coarse normalization exploiting the a priori knowledge of human facial features; followed by fine registration utilizing the hausdorff distance approach. The similarity measure between the two face models is defined by the hausdorff distance employed between them. Verification results are reported for images from the multimodal verification for teleservices and security applications (M2VTS) database and a best equal error rate (EER) of 3.24% is reported. This approach was later extended by fusing the information obtained with facial profile matching and surface matching. The best EER reported improved significantly to 2.22%.

Lee et al. [Leey03] introduce a novel face recognition algorithm using multiple features for the area in the contour line of face which has depth information. Having detected the exact tip of the nose, the face is geometrically normalized. This is followed by the extraction of the contour areas using iterative selection theory. After reducing the dimensionality, average and variance features are computed and are used as feature vectors. Euclidian distance is used as a similarity metric for matching at a given contour line threshold. The reported results show a rank five recognition rate of 94% and rank ten recognition rate of 100% at the contour threshold of 40. However, the size of the database is very small (70 images of 35 people). Also, the method is very sensitive to discretization in depth values of the contours.

Beumier et al. [Beumier00] propose a face recognition method based on facial surface analysis as well as facial profile analysis. The normalized profiles are extracted from each face with an assumption that the face is almost symmetric along the vertical plane passing through the nose. Having optimized the transformation parameters, the profiles from the test and the reference facial surface are compared using the minimum distance approach. These individual distances are combined into global distance which when optimized is used as a criterion for face similarity. The results reported on multimodal biometric identity verification (BIOMET) 3D database show a best EER of 3.6% considering 6 shots for each person.

2.2.2 Curvature based Approaches

Segmentation and interpretation of general range images using surface curvatures has been given considerable attention in the past by many researchers [Besl86] [Fan85] [Vemuri86]. Most of the earlier work focused on recognition of geometrically simple objects, attempting to classify surfaces into planar regions, spherical regions, or surfaces of revolution. While Besl and Jain [Besl86] considered just regions of zero mean or Gaussian curvature for segmentation, Fan et al. [Fan85] additionally incorporated the information obtained from the local maxima in maximum curvature. Based on the same strategy, Gordon [Gordon91] explored face recognition from a representation based on features extracted from range images. High level surface feature descriptors in terms of points, lines, and regions are extracted along with the low level scalar features in terms of distance or curvature measurements. The sensed surface regions are classified as convex, concave and saddle by calculating the minimum and maximum principal curvature; then the locations of nose, eyes, mouth, and other features are determined. Additionally, umbilic points are calculated to obtain rich information to describe human face. These features assist in normalizing the position of both the source and target facial surfaces and later a simple brute force strategy is used for face recognition. In his work, Gordon demonstrated the face recognition strategy utilizing the plethora of useful surface primitives that cannot be seen from intensity images. However, the recognition results are reported to be in the range of 80% to 100% on a small database of 24 faces. Also, this approach can deal with faces different in size, but needs extension to cope with changes in facial expression.

Lee et al. [Lee90] propose a method to detect corresponding regions in two range images by graph matching based on Extended Gaussian Image (EGI) and perform a region based matching of range images of human faces. They make use of the idea that distinct facial features (nose, cheek, chin, or eyebrows) correspond to convex regions and can be segmented based on the curvature relationships of the range image. Each convex region is represented by an EGI which is a one to one mapping between points on the unit sphere that have the same surface normal. Matching is then performed based on a similarity metric between the two convex regions generated by correlating the Extended Gaussian Images. To find the optimal correspondence, a graph matching algorithm is applied to incorporate additional relational constraints in addition to the correlation co-efficient between pairs of matched regions (convex regions). Their method is expression invariant to a certain degree due to the assumption that convex regions of the face are more insensitive to changes in facial expression than the non-convex regions. However, EGIs are not sensitive to change in object size, and so two similar shape but different size faces will not be distinguishable in this representation. Furthermore, the correlation coefficient used by them was not robust enough as it was tested on range images of only six people.

Tanaka et al. [Tanaka98] presented a correlation based face recognition approach based on the analysis of maximum and minimum principal curvatures and their directions. First, they analyze face structure based on 3D principal curvatures and their directions from range images. The information obtained from principal directions at high curvature is used to calculate the ridge and valley lines. The former are a set of vectors that correspond to local maxima in the values of the minimum principal curvature whereas the latter are a set of vectors that correspond to local minima in the values of the maximum principal curvature. The EGI's of feature vectors are later constructed by mapping the maximum and minimum principal directions on the two unit sphere for face representations. Finally, matching between the input and the model image is performed by a rotation invariant similarity measure known as Fisher's spherical correlation taking into consideration the respective ridge and valley EGI's. Also, it is simple, efficient, and robust to distractions such as glasses and facial hair, but it has not been tested on faces in different sizes and facial expressions. Although this method does not require either face feature extraction or surface segmentation, the reported results are on a small database of 37 face range images. Furthermore, the reported results are not clear and just claim an average similarity of 44% for a correct match and 13% for an incorrect match.

2.2.3 PCA based Approaches

Achermann et al. [Achermann97] extended the two approaches which were well known from face recognition based on grey level images to range images. They made a comparative study of utilizing eigenface and hidden markov model (HMM) methods on range images. Principal component analysis (PCA) technique was used to build an eigenspace out of five poses each of 24 different people and the vectors with the most significant eigen values were taken as base vectors. Test images were projected into the face space, and recognition was performed based on a certain threshold. In the HMM method, the human face was represented by a linear left right model consisting of five states. The parameters were calculated for every person in the database during the training phase. During the testing phase, the probability of producing that test image by every model in the database was calculated which aids in the recognition process. A recognition rate of 100% was reported for eigenface method using five training images per person. However, a smaller recognition rate of 89.17% was reported in the case of HMM method. Also, the methods based on PCA do show disadvantage on a large database due to the deterioration in the performance caused by the effect of outliers.

Hesher et al. [Hesher03] utilized principal component analysis and independent component analysis (ICA) for mathematically representation and analysis of facial surfaces. They examined 222 frontal range images of 37 people with six different expressions. The range images were geometrically normalized for pose changes by first locating the nasal bridge and aligning it with the Y axis for rotational correction. Also, each range image was translated in the image plane so that the tip of the nose corresponds to the center point location followed by depth correction (Z position). After some preprocessing and hole filling, PCA and ICA are implemented on the range images resulting in the projection of the images onto a lower dimension space. The results are reported for different size of training sets but the best results are achieved when the largest training set is used and ICA with first 10 independent

components is used. However, the mode of PCA and ICA utilized by them were not robust to noise in data induced by error in mesh capture, reduction techniques, or background clutter. Also, the effect of expression variation on the recognition accuracy is not reported.

The face recognition method proposed by Tsalakanidou et al. [Tsalakanidou03] is based on the PCA and the extraction of color and depth eigenfaces. The main motivation is to evaluate three different approaches (color, depth, combination of color and depth) for face recognition and quantify the contribution of depth in FR. PCA is performed on each of the components of the YUV and the range image to obtain multiple sets of eigen vectors. They select a range image each of the 40 people in the XMVTS database to build an eigenspace for training. The test dataset consists of artificially rotated range images of all the 295 people present in the database. For a rotation of $\pm 2^{\circ}$ around the Y-axis, the recognition rate claimed is 93%, while the recognition rate falls down to 89% for a rotation of $\pm 5^{\circ}$. Also, for larger rotations the recognition rate further reduces to 85%.

Chang et al. [Chang05] have presented a report on the largest experimental study on 3D face recognition till date on 166 subjects imaged in both 2D and 3D; the probe and gallery datasets taken over different time intervals. Using a PCA based approach separately on both 2D and 3D, the rank one recognition rate obtained was 83.1% for 2D and 83.7% for 3D, which are not statistically different. However, when the 2D and 3D scores were fused using a weighted sum of distance approach, the recognition performance improved to 92.8%. The main drawback in this method is the manual pose normalization employed to geometrically standardize the 3D images.

2.2.4 Point Signature based Approaches

Chua et al. [Chua00] extended the concept of Point Signature – a representation for free form surfaces to 3D face recognition. The main motivation is the identification of faces, despite having different facial expressions. For this purpose the facial surface is treated as a non-rigid surface. Based on certain heuristics the rigid surface is identified and correspondence is established between the rigid surfaces of the two faces by means of correlation of point signature vectors and other criteria such as distance, and direction. Furthermore, the optimal transformation between the surfaces is estimated in an iterative manner using ICP. After registering the two different facial surfaces, the rigid portions are distinguished from the non-rigid regions by an adaptive threshold for the Gaussian distribution and subsequently a model library is built. For identification of each test scene, the models are voted using the index table created from model library. However, the experimental results are reported on four range images each from six people. Also, the use of ICP for iterative correspondence makes the process computationally expensive.

2.2.5 Template Matching/ Brute Force/ Distance Map

Lao et al. [Lao00] proposed a framework for 3D pose invariant face recognition based on template matching. The 3D facial models are acquired by stereo based system and consist of sparse depth map constructed using isoluminance lines for stereo matching. To normalize the pose, the irises are located by searching arcs whose radiuses are of certain range followed by the location of the mouth. Based on the location of these parts, the model is then transformed into a canonical position. Recognition is then performed by using template matching as follows: a) Both the sample and the data to be recognized are adjusted to their front view and in the same co-ordinate system using the pose recognition algorithm; b) diving the matching area into meshes of width 5×5 mm each; c) mean distance between the local regions; d) choosing the sample with smallest mean distance as the answer. They tested their algorithm with a database of 10 people each with nine different poses ranging from $\pm 15^{\circ}$ to $\pm 30^{\circ}$ both in horizontal and vertical direction. They claim to have a stable and robust recognition rate ranging from 87% to 96%.

Medioni and Waupotittsch [Medioni03] demonstrate a automatic face authentication system by analysis of 3D facial shape. The 3D facial models were generated with the help of an acquisition system consisting of two stereo cameras. The recognition process consists of one-to-one comparison of a probe 3D model with an existing model in the database. The two models are automatically aligned and a brute force is used to calculate the distance map between the two facial surfaces. The final classification is based on statistics derived from the distance maps. The framework is validated on a database of 100 subjects, each with seven poses within $\pm 20^{\circ}$ of the frontal view giving an EER of less than 2%.

2.2.6 Global Features and Local Shape Variation based Approaches

Xu et al. [Xu04] developed an automatic face recognition method combining the global geometric features with local shape variation information. A robust universal fitting algorithm is developed to convert the original 3D point cloud to a regular mesh. The nose region being a prominent and robust feature is used to align the basic mesh with the original point cloud. An average mesh model is thus generated by averaging the mesh models from the pre-modeling process followed by remodeling of the mesh models for pose compensation. The local shape variation information is then extracted to represent face feature together with global geometric feature as a vector. To improve the recognition performance and reduce computational complexity, PCA is used for feature space dimensionality reduction and then nearest neighbor is used for classification. Experimental results are reported on the 3D_RMA database which consists of 120 and 30 people in automatic database (ADB) and manual database (MDB) respectively. The best recognition rates reported are 72% and 96% for ADB

and MDB respectively illustrating that the experimental results are highly dependent on database size and quality.

2.2.7 ICP based Methods

Lu, Colbry, and Jain [Lu04a] employ an approach based on ICP for 3D face recognition. Their recognition pipeline consists of two components; surface matching followed by appearance based matching. The surface matching component is based on a hybrid ICP which dynamically switches between the two ICP algorithms. This strategy results in incorporating the advantages of both the algorithms: the greater speed of the algorithm by Besl and McKay, and the greater accuracy of the method by Chen and Medioni. However, a coarse alignment is performed initially by finding the anchor points based on shape index and transforming the facial surface for pose standardization. The root mean square distance minimized by the ICP algorithm is used as a primary similarity metric. Further, the registered 3D model is utilized to synthesize training samples with facial appearance variations, which are used for discriminant subspace analysis [Lu05b]. Finally, the scores obtained by the two matching components are fused together using the weighted sum rule. Experimental results are reported on a gallery database of 100 3D models and 598 2.5D test scans. The recognition rate is reported to be 87% in the case of surface matching only when compared to an improved performance of 91% in the case of fused components.

2.2.8 Model Fitting Methods

Many attempts were made initially to solve the recognition task by fitting a deformable 3D model to 2D images. For example, Blanz and Vetter [Blanz03] made use of 3D morphable models to perform recognition from 2D images. Their algorithm automatically estimates the 3D shape, texture and other relevant information from a single image of a person. A morphable face model is constructed from a set of laser scanned 3D face models by transforming their shape and texture information into a vector space. This aids in expressing the shape and texture of any face in terms of the linear combination of the shape and texture vectors. For the purpose of recognition these shape and texture vectors of individuals are matched based on a simple nearest neighbor classification rule using a correlation based similarity measure. Verification tests were performed both on the CMU pose, illumination, and expression (CMU-PIE) database and the facial recognition technology (FERET) database. A recognition rate of 77.5% was reported on 4420 probe images of CMU-PIE database. However, the verification rate was better at 89.7% for the FERET database which consisted of 1746 probe images.

In another appearance based approach, Lee et al. [Leem03] consider the synthesis of faces in arbitrary poses for pose invariant 3D face recognition. A generic 3D face model is built using the training images of subjects with faces in arbitrary poses. This

deformable model comprised information from three submodels viz. edge, color region and wireframe models. During the recognition process, the pose of the face to be recognized is estimated, and then all the faces of the people in the database are projected to this view using the 3D deformable representation. This gives an estimate of the corresponding texture points and the intensity values at these locations is stored in a vector. Finally, classification is performed by the least square estimate. The recognition rates reported varied from 56.2% for one training image per subject to 92.3% for 10 training images per subject with 15 subjects in the database.

2.2.9 Other Methods

Bronstein et al. [Bronstein03] proposed a method for invariant 3D face recognition which does not require the facial surface explicitly but utilizes surface gradient field, or the surface metric for constructing the expression invariant face representation. The acquired 3D facial surface is preprocessed by cropping and smoothing operations followed by feature detection. At the last preprocessing stage, the facial contour is extracted using the geodesic mask. The key idea is to map invariant source points on the face and mark an equidistant contour around it. This is then projected onto a three dimensional space using a distance preserving dimensionality reduction technique such as multidimensional scaling. Furthermore, the bending invariant canonical form are aligned and interpolated onto a cartesian grid creating a canonical image. This leads to an efficient, accurate and expression invariant method for representing faces. These images are compared using eigen-decomposition. Experimental results were reported on a database of 220 faces of 30 subjects and a best EER of 1.9% was reported in the case of canonical surface matching.

Eriksson and Weber [Eriksson99] represent each face by sampling the image both spatially and in frequency through the use of Gabor wavelet filters. The faces are stored as image meshes which represent the position and disparity for 40 feature points extracted from input image. Recognition of an unknown image pair is performed by finding the transformation of the template mesh in the real world coordinates that has a projection onto the two image planes, such that the fiducial points on the two meshes are best matched.

Irfanouglu et al. [Irfanoglu04] utilize three dimensional facial information for human identification. They propose an algorithm based on point set distance approach (PSD) that establishes a dense correspondence between faces. The correspondence is performed by first automatically finding landmarks and then these salient facial features are used to find dense correspondence of the points on the facial surface using Thin plate spline (TPS) warping algorithm. In the recognition stage, the similarity between two facial surfaces is estimated using the discrete approximation of the volume difference between the facial surfaces. They report a best recognition rate of 96.66% on 30 people from the 3D_RMA dataset.

3 GAUSSIAN FIELDS FOR 3D FACE REGISTRATION AND RECOGNITION

The registration task of any 3D face datasets consists of the recovery of the transformations that align the partial views. The main parameters which are computed in the case of 3D rigid registration are the rotation and translation parameters provided if the point correspondences are available. To establish point correspondences in 3D face datasets several feature extraction techniques as mentioned in Chapter 2 were proposed but most of them were surface based. The methods used local representations to encode local shape information as well as global descriptors such as spherical attribute images. A typical range scanner returns the 3D model of an object in point sets form and hence we concentrate on point-sets instead of surfaces or meshes.

The method proposed in [Boughorbel04] aims at the design of the point sets registration criterion based on Gaussian fields. This criterion is convex in a large neighborhood of the aligned position (solution) and always differentiable allowing for the use of well proven optimization techniques. This method tries to overcome the problems of ICP which are generally due to the limitations in the differential cost function that imposes local convergence. The proposed method [Boughorbel04] can be used for accurate registration by extending the region of convergence and thus eliminating the need for any close initialization. Also, it doesn't need any additional information about the point correspondences. The main advantage of this method lies in its low computational complexity due to the use of Fast Gauss transform [Greengard91].

3.1 Gaussian Fields and Energy Function

The approach adapted in [Boughorbel04] starts with an assumption that registration between two datasets is a special sub-problem of pattern matching and the registered position is one resulting in the maximum point to point overlap of the two models free from noise. The above definition allows us to work with minimum amount of information about the datasets such as position of the points. However, additional information obtained from local shape similarity between the points can also be used to enhance the quality of the registration.

The main idea used in the 3D registration approach is to make use of the Gaussian fields to measure both the spatial proximity and the visual similarity of the two datasets in the point form. The criterion is introduced on two point sets, $M = \{(P_i, S(P_i))\}$ and $D = \{(Q_j, S(Q_j))\}$ with their associated attribute vectors. As the datasets are considered in point form, 3D moments are utilized as attributes. However, the attributes can also include curvature for smooth surfaces and curves, invariant descriptors, and color attributes when available. At the maximum overlap of the two point set, the transformation Tr^* will lead to a global maximum for the following measure.

$$E(Tr) = \sum_{\substack{i=1...N_{M} \\ j=1...N_{D}}} \delta(d(Tr(P_{i}), Q_{j}))$$
(3.1)
with $\delta(t) = 1$ for $t = 0$
and $\delta(t) = 0$ otherwise

where d(P,Q) is any suitable distance between points such as Euclidean. Although the above measure takes just the position of the points into account, it is an easy task to incorporate local shape similarity in this criterion and requires just using a higher dimensional representation of the datasets where points are defined by both position and a vector of shape attribute: $M = \{(P_i, S(P_i))\}_{i=1...N_M}$ and $D = \{(Q_j, S(Q_j))\}_{i=1...N_D}$.

The criterion derived above (Eq. 3.1) can be visualized by a collection of spikes in the parameter space and is not continuous with respect to the alignment transformations. It would be difficult to apply the standard optimization strategies to this criterion due to the problems associated with finding the global maxima. A smooth approximation of the criterion can be built using an analytical method known as mollification which was introduced by Murio [Murio93]. Mollification is a process of smoothening a non differentiable function by convolving it with the Gaussian kernel. The resulting would be approximation of the original function function an such that $\lim f_{\sigma}(t) = f(t)$. The energy function after the application of mollification is as $\sigma \rightarrow 0$ follows:
The above mollified criterion is a simple sum of Gaussians of distances between all pairs of model and data points and hence the overall profile of the criterion with respect to transformation parameters would have appearance of a Gaussian, with local convexity in the neighborhood of the registered position. Expression (3.2) can be re-interpreted as the integration of a potential field whose sources are located at points in one of the datasets and targets in the other one. Additional information such as intensity, color, and local shape descriptors can be fused in the above criterion by extending the distance measure between points in the criterion as follows:

$$E_{\sigma,\Sigma_{a}}(Tr) = \sum_{\substack{i=1...N_{m}\\j=1...N_{D}}} \exp\left(-\frac{\left\|Tr(P_{i})-Q_{j}\right\|^{2}}{\sigma^{2}} - (S(Tr(P_{i}))-S(Q_{j}))^{T}\Sigma_{a}^{-1}(S(Tr(P_{i}))-S(Q_{j})))\right)$$
(3.3)

The differentiable criterion obtained above can be optimized using any of the powerful optimization techniques such as Quasi-Newton technique and conjugate gradient algorithms. In the noisy case, the Gaussian criterion accounts for noise effects by equating the parameter σ with the noise variance. The parameter σ mainly controls the size of the convex safe region of convergence. The higher the value of σ , the larger the region of convergence, but smaller the localization accuracy. Hence, the value of σ should be properly chosen to maintain an optimum region of convergence and precision. This tricky situation is mainly caused by the effect of outliers, where the term outlier refers to the areas that are outside the intersection of model and data. The effect of outliers can be compensated by associating much more available information to the points which will lead to a low registration error associated with a large area of

convergence. The parameter Covariance Matrix Σ_a is a diagonal matrix with positive components and aids in proper scaling of the different attributes before the fusion. If Σ_a is tactfully chosen, the effect of outliers is further reduced allowing for good localization of the registered position and reducing the need for close initialization.

3.2 Attributes

Various attributes can be extracted from the 3D face scans including curvature, intensity, and color. However, as the shapes are represented as point sets, 3D moment invariants are used as point attributes. These three moment invariants [Sedjadi80] have been used for object recognition tasks in the past and are employed in registration algorithms such as in the extension of ICP by Sharp et al. [Sharp02]. However, for computational simplicity only the first moment J_1 is utilized out of the three moments J_1 , J_2 , and J_3 . These moments J_1 , J_2 , and J_3 are defined for a local neighborhood N around a point $P(X_P, Y_P, Z_P)$ by:

$$J_{1} = \mu_{200} + \mu_{020} + \mu_{002}$$

$$J_{2} = \mu_{200}\mu_{020} + \mu_{200}\mu_{002} + \mu_{020}\mu_{002} - \mu_{110}^{2} - \mu_{101}^{2} - \mu_{011}^{2}$$

$$J_{3} = \mu_{200}\mu_{020}\mu_{002} + 2\mu_{110}\mu_{101}\mu_{011} - \mu_{002}\mu_{110}^{2} - \mu_{020}\mu_{101}^{2} - \mu_{200}\mu_{011}^{2}$$
(3.4)

with

$$\mu_{pqr} = \sum_{(X,Y,Z)\in N} (X - X_P)^p (Y - Y_P)^q (Z - Z_P)^r$$
(3.5)

The concept of Tensor voting introduced by Medioni et al. [Medioni00] is utilized in estimating a local measure of visual saliency. Saliency is similar to the other moments and is analogous to Gaussian curvature in the case of smooth surfaces. The measure is robust to noise and can be estimated even when information from surfaces and curves is difficult to extract. The first pass of tensor voting scheme is used in the computation of saliency. Saliency is evaluated at a site $P_i = (x_i, y_i, z_i)^T$ by collecting votes from neighboring site $P_j = (x_j, y_j, z_j)^T$, which cast the stick tensor at P_i in the case of 2D voting, and plate tensor for 3D. The plate tensor encodes the uncertainty of normals at the voting site in the direction of the unit vector $t_{ij} = \frac{P_i - P_j}{\|P_i - P_j\|} = (t_{ij}^x, t_{ij}^y, t_{ij}^z)^T$. It is methanelise the uncertainty of provide the unit vector $t_{ij} = \frac{P_i - P_j}{\|P_i - P_j\|} = (t_{ij}^x, t_{ij}^y, t_{ij}^z)^T$.

mathematically expressed as:

_

$$T_{ij} = \begin{bmatrix} 1 - (t_{ij}^{x})^{2} & -t_{ij}^{x}t_{ij}^{y} & -t_{ij}^{x}t_{ij}^{z} \\ -t_{ij}^{x}t_{ij}^{y} & 1 - (t_{ij}^{y})^{2} & -t_{ij}^{y}t_{ij}^{z} \\ -t_{ij}^{x}t_{ij}^{z} & -t_{ij}^{y}t_{ij}^{z} & 1 - (t_{ij}^{z})^{2} \end{bmatrix}$$
(3.6)

_

$$= \frac{1}{\left\|P_{i} - P_{j}\right\|^{2}} \begin{bmatrix} (P_{i}^{y} - P_{j}^{y})^{2} + (P_{i}^{z} - P_{j}^{z})^{2} & -(P_{i}^{x} - P_{j}^{x})(P_{i}^{y} - P_{j}^{y}) & -(P_{i}^{x} - P_{j}^{x})(P_{i}^{z} - P_{j}^{z}) \\ -(P_{i}^{x} - P_{j}^{x})(P_{i}^{y} - P_{j}^{y}) & (P_{i}^{x} - P_{j}^{x})^{2} + (P_{i}^{z} - P_{j}^{z})^{2} & -(P_{i}^{y} - P_{j}^{y})(P_{i}^{z} - P_{j}^{z}) \\ -(P_{i}^{x} - P_{j}^{x})(P_{i}^{z} - P_{j}^{z}) & -(P_{i}^{y} - P_{j}^{y})(P_{i}^{z} - P_{j}^{z}) & (P_{i}^{x} - P_{j}^{z})^{2} + (P_{i}^{y} - P_{j}^{y})^{2} + (P_{i}^{y} - P_{j}^{y})^{2} \end{bmatrix}$$

These tensors are then collected from the sites in a small neighborhood around P_i using summation:

$$T_i = \sum_{j \neq i} T_{ij} \tag{3.7}$$

Finally, the scalar measure of saliency is given by the determinant of the tensor T_i which can be interpreted as the square of the volume of the bounding box of the uncertainty ellipsoid.

$$D_i = \det T_i = (\lambda_1 \lambda_2 \lambda_3)^2 = V_i^2$$
(3.8)

The computation of global saliency relies on the fact that at the unregistered position the point sets will have little interaction, due to the local nature of the saliency inference. However, when the two point-sets are aligned, there will be a local increase in the number of votes at the common region resulting in the increase in the saliency measure.

3.3 Optimization

The Gaussian criterion derived is continuous and is always differentiable allowing for the use of well proven optimization techniques. Since the criterion is a mixture of Gaussians closely located in parameter space, the overall profile has a Gaussian shape and hence convexity can be assumed around the registered position. This can be proved from the argument below. For a small value of σ and small rigid displacements near the registered position (i.e. a ball of radius ε around the rotation angle and translation vector (φ , t) the Gaussian criterion (3.3) can be approximated as follows:

$$\sum_{\substack{i=1...N_{m}\\j=1...N_{p}}} \exp(-\frac{d^{2}(Tr(P_{i}),Q_{j})}{\sigma^{2}}) = \sum_{\substack{i=1...N_{m}\\j=1...N_{p}}} \exp(-\frac{(\cos\varphi P_{i}^{x} - \sin\varphi P_{i}^{y} + t_{x} - Q_{j}^{x})^{2} + (\sin\varphi P_{i}^{x} + \cos\varphi P_{i}^{y} + t_{y} - Q_{j}^{y})^{2}}{\sigma^{2}})$$
(3.9)

It can be further simplified into (3.10) using the approximation for small rotation $\cos \varphi \approx 1$ and $\sin \varphi \approx \varphi$, in addition to the first order approximation resulting from the small displacement compared with σ :

$$\exp(-\frac{d^{2}(Tr(P_{i}),Q_{j})}{\sigma^{2}}) \approx 1 - \frac{d^{2}(Tr(P_{i}),Q_{j})}{\sigma^{2}}$$
$$\approx \sum_{\substack{i=1...N_{y}\\j=1...N_{p}}} 1 - \frac{(P_{i}^{x} - \varphi P_{i}^{y} + t_{x} - Q_{j}^{x})^{2} + (\varphi P_{i}^{x} + P_{i}^{y} + t_{y} - Q_{j}^{y})^{2}}{\sigma^{2}}$$
(3.10)

The quadratic nature of the rigid parameters in the expression (3.10) demonstrates the convexity of the criterion. Optimization is performed by using a standard gradient based optimization scheme by the name Quasi-Newton algorithm. *Quasi-Newton* or *variable metric* methods can be used when the Hessian matrix is difficult or time-consuming to evaluate. Instead of obtaining an estimate of the Hessian matrix at a single point, these methods gradually build up an approximate Hessian matrix by using gradient information from some or all of the previous iterates, visited by the algorithm. Using the current iterate, and the approximate Hessian matrix, the decent direction is found out. A line search routine is finally used along the descent direction to find the optimum solution.

There exists a tradeoff between the accurate localization with a small value of σ and a larger region of convergence for a larger σ at the expense of registration accuracy. To solve this tricky situation, a scheme consisting of two or more runs of Quasi-Newton routine with decreasing values of sigma is adopted. However, if the value of sigma is decreased too much, we may get trapped at the local maximum. The local maximum is avoided by studying the rate at which the global maximum is drifting with the change of force range parameter. By ensuring that the drift does not result the next run to start from outside the dominant mode, the problem of getting trapped in local maximum can be avoided.

3.4 Fast Gauss Transform

The registration criterion has a computational cost of $O(N_M \times N_D)$, being a mixture of N_D Gaussians evaluated at N_M points then summed together, which is very high for large datasets. This problem, which is also encountered in other computer vision applications, can be solved by a new numerical technique called as the Fast Gauss Transform. The method, introduced by Greengard and Strain [Greengard91], is derived from a new class of fast evaluation algorithms known as "fast multipole" methods and can reduce the computational complexity of the Gaussian mixture evaluation to $O(N_M + N_D)$. The basic idea is to exploit the fact that all calculations are required only up to certain accuracy. In this framework the sources and targets of potential fields were clustered using suitable data structures, and the sums were replaced by smaller summations that are equivalent to a given level of precision.

$$S(t_i) = \sum_{j=1}^{N} f_j \exp(-(\frac{s_j - t_i}{\sigma})^2), \ i = 1, ..., M$$
(3.11)

where $\{s_j\}_{j=1,\dots,N}$ are the centers of the Gaussians known as sources and $\{t_i\}_{i=1,\dots,N}$ the targets. The following shifting identity and expansion in terms of Hermite series are used:

$$\exp(\frac{-(t-s)^{2}}{\sigma^{2}}) = \exp(\frac{-(t-s_{0}-(s-s_{0}))^{2}}{\sigma^{2}})$$
$$= \exp(\frac{-(t-s_{0})^{2}}{\sigma^{2}})\sum_{n=0}^{\infty} \frac{1}{n!} (\frac{s-s_{0}}{\sigma})^{n} H_{n}(\frac{t-s_{0}}{\sigma})$$
(3.12)

where H_n are the Hermite polynomials. Given that these series converge rapidly, and that only few terms are needed for a given precision, this expression can be used to replace several sources by s_0 with a linear cost at the desired precision. These clustered sources can then be evaluated at the targets. For a large number of targets, the Taylor series (3.13) can similarly be used to group targets together at a cluster center t_0 , further reducing the number of computations:

$$\exp(-(\frac{t-s}{\sigma})^{2}) = \exp(\frac{-(t-t_{0}-(s-t_{0}))^{2}}{\sigma^{2}})$$

$$\approx \sum_{n=0}^{p} \frac{1}{n!} h_{n} (\frac{s-t_{0}}{\sigma}) (\frac{t-t_{0}}{\sigma})^{n}$$
(3.13)

where the Hermite functions $h_n(t)$ are defined by $h_n(t) = e^{-t^2} H_n(t)$. The method was shown to converge asymptotically to a linear behavior as the number of sources and targets increases.

3.5 Extension of the Gaussian Criterion for Recognition

One of the major contributions of this thesis is the extension of Gaussian Fields framework to applications such as 3D face recognition. The overall recognition pipeline is shown in Fig.3.1. The first stage consists of the data acquisition stage, wherein a 3D facial scan is captured by a 3D sensor. This 3D sensor may be stereo based, laser based, or structured light based sensor. However, we make use of a sensor which works on the principle of structured light. The next stage consists of creating a 3D face gallery to be used in conjunction with the existing publicly available 3D face galleries. In this phase, a more complete ear to ear model is also built by registering



Fig. 3.1: Framework of the automatic face recognition based on 3D facial data.

the different views of an individual. For the purpose of recognition, the probe model is registered with each and every facial model in the gallery. This is done using the registration algorithm based on Gaussian Fields. The expression for the Gaussian criterion (3.14) is recalled here to provide a brief insight about the similarity score generated during the registration phase.

$$E(Tr) = \sum_{\substack{i=1...N_{m} \\ j=1...N_{p}}} \exp(-\frac{d^{2}(P_{i}, Tr(Q_{j}))}{\sigma^{2}} - \frac{(S(P_{i}) - S(Tr(Q_{j})))^{T} \Sigma_{a}^{-1} (S(P_{i}) - S(Tr(Q_{j}))))}{C_{a}^{2}})$$
(3.14)

The absolute value of the above Gaussian criterion function is used as a metric to measure the similarity between two faces in our 3D face recognition method. The higher the value of the function, the more similar the two faces are. However, the raw scores obtained from the Gaussian criterion should be normalized to make it suitable for recognition applications. The normalization procedure is explained in detail in Chapter 5 of this thesis. In the final stage, the task of recognition is carried out by fixing a threshold for the above generated normalized scores. A face is considered to be a match if the normalized score crosses the threshold.

4 RESULTS FOR 3D FACE REGISTRATION

In this chapter, we analyze the experimental results obtained from the 3D face registration based on Gaussian Fields. We also discuss about the different sensors used for our data acquisition and the operating principles associated with them.

4.1 Introduction and Objectives

The primary objective of this thesis is to analyze a new method for 3D face recognition. However, as the 3D face registration plays a major role in the recognition pipeline, we also perform an analytic and quantitative study of the Gaussian Fields registration method. The automatic registration problem is addressed at the point level without any explicit point correspondence. Moreover, the Gaussian Field method overcomes the need for close initialization, which is required by Iterative Closest Point algorithm. The expression for Gaussian criterion (4.1) is recalled to provide a brief description about the various parameters associated with it.

$$E(Tr) = \sum_{\substack{i=1...N_{M} \\ j=1...N_{D}}} \exp(-\frac{d^{2}(P_{i}, Tr(Q_{j}))}{\sigma^{2}} - \frac{(S(P_{i}) - S(Tr(Q_{j})))^{T} \Sigma_{a}^{-1}(S(P_{i}) - S(Tr(Q_{j}))))}{C_{a}^{2}}) \quad (4.1)$$

The main advantage of the registration method is the minimum number of free parameters involved. The only parameter which can change and affect the entire registration process is the force range parameter σ . Most of the other parameters are generally computed or derived once at the beginning. The main parameters involved with the Gaussian criterion can be classified as follows:

• σ : Force Range parameter that controls the range of the Gaussian Field. In other terms the width of basin of convergence can be increased by increasing the parameter σ , but this will result in decrease of the localization accuracy of the criterion. Hence, the value of σ should be optimally chosen. If the datasets have sufficient shape complexity, σ can be chosen large for a limited

localization error. However, while choosing σ sufficient care should be taken to avoid being trapped at local minima.

- Σ : Covariance Matrix or the De-Correlation matrix of the feature descriptors. This matrix is computed from the data specifically in the nearly flat regions and used to scale the features to make them independent of dimensions. In short, the main purpose of it is to create the orthogonal features necessary for effective fusion.
- ρ : Radius of the sphere in which the local features are computed. This depends on the resolution and on the information content of the datasets. If the noise is low when compared to the dimensions of the data it would be optimum to consider a smaller neighborhood in which the features are computed. This will spread the values of the descriptors over a larger spectrum allowing for more accurate matching. If the noise level is very high then it would not be appropriate to compute the local features in such a small neighborhood as it could lead to unreliable results.
- C_a : Confidence factor associated with the descriptors. It is added to the criterion to compensate for the noise levels affecting the registration. The confidence level factor is typically chosen to be around 10^{-3} for low noise levels and around unit value for higher noise values.

The Force Range parameter σ being the most significant parameter, we investigate its effect on the registration criterion. The effect of noise on the algorithm is also studied in conjunction with the size of the area over which the descriptors were calculated. Subsequently, we examine the robustness of the criterion to low levels of sampling. Since the amount of overlap between the two face datasets to be registered plays an important role in the registration process, we examine its effects on the registration criterion. Finally, a comparison between the region of convergence with the standard ICP and basin of convergence with the registration technique based on Gaussian fields is undertaken.

4.2 Data Acquisition

In our experiments, we have used a synthetic dataset of a mannequin head and real datasets (Fig 4.1) from our IRIS 3D face database. The 3D faces were scanned using the Genex 3D FaceCam, which operates on the principle of structured light. However, the synthetic face was generated by using a different class of sensor namely Integrated Vision Products (IVP) Ranger 2200, which operates on the principle of triangulation by acquiring several profiles of the face.



Fig. 4.1: The data used in the experiments (a) Mannequin Head Data (b) Original Face Data...

The Genex 3D FaceCam (Fig 4.2) uses three high resolution Charge Coupled Device (CCD) sensors and a color encoded pattern projection system. An accurate 3D surface map is generated using the RGB information from each pixel and multiple 3D views are combined to generate a 3D model having ear to ear coverage. Since the 3D FaceCam uses three CCD cameras to cover the entire face, frame data correspondence and registration is performed to generate the complete 3D face model. It is a powerful three dimensional surface profile measurement system capable of acquiring full frame dynamic 3D images of objects with complex surface geometry at a high speed. The key benefit of the 3D FaceCam is its small image acquisition time (400-500 msec) and its quick processing time (30 sec). However, the 3D FaceCam has a practical limitation in terms of the field of view which is restricted to a volumetric box of 20" width, 16" height, and 12" depth. The minimum and maximum standoff distances are 33" and 45" respectively. The operating principle of Genex 3D FaceCam is based on the triangulation principle which is depicted in Fig 4.3.

The distance R between the CCD sensor and the object can be estimated using the relation:

$$R = B \frac{\sin(\theta)}{\sin(\theta + \alpha)}$$
(4.2)

Since the values of *B* and α can be predetermined, the triangulation method (Fig. 4.4) depends on the computation of the projection angle θ from the image captured by the CCD sensor. This problem is addressed by projecting a light pattern with spatially distributed wavelengths using a linear variable wavelength filter (LVWF). Due to the fixed geometric relationship between the light source, lens, and LVWF, there exists a 1-to-1 correspondence between the projection angle θ of the plane of the light and the wavelength λ of the light ray. The projection angle θ is estimated based on the detected color spectrum in the CCD camera. Angle α can be found out from the pixel



Fig. 4.2: The two range scanning systems used in our registration experiments. (a) Genex 3D FaceCam, (b) IVP Ranger.



Fig. 4.3: Triangulation principle for 3D imaging.



Fig. 4.4: Rainbow principle: θ_i is calculated by solving one-to-one correspondence problem between color λ_i and projection angle $\theta_i \cdot \alpha_{k,l}$ is geometrically calculated using the coordinates of each pixel (k; l) in the image of the sensor. Then, using triangulation principle, each visible point O of the object can be calculated.

information of the CCD camera's image plane. The estimated parameters α , θ , and *B* are used to determine the 3D co-ordinates (*x*, *y*, *z*) of the object.

$$x = \frac{B}{f * \cot \theta - u} * u$$
$$y = \frac{B}{f * \cot \theta - u} * v$$
$$z = \frac{B}{f * \cot \theta - u} * f$$
(4.3)

where *f*: focal length

(u, v): pixel co-ordinates of the sensor image

The final 3D face output model generated by the Genex 3D FaceCam is shown in Fig. 4.5. Also, the experimental setup for the 3D image capture by the Genex 3D FaceCam is shown in Fig. 4.6. As seen in Fig 4.7, the Genex 3D FaceCam is mounted on a top of an adjustable tripod and the person is positioned at a distance of 35" in front of the camera. The person can be properly positioned with the aid of the Positioning/ Capture screen. This screen as shown in Fig. 4.8 allows the user to position the subject optimally within the volume box. i.e. the field of view of the camera. The color coded structured light pattern projected on face can be seen in the processing screen (Fig 4.9).

The IVP Ranger also works on the principle of triangulation. However, it makes use of a sheet-of-light laser which cuts a plane in 3D space and thus projects a single line across the object of interest. This sheet-of-light system allows an increase in scanning speed over a point laser. The sheet-of-light method only requires M images to reconstruct $M \times N$ data points. So, the sheet-of-light approach is much faster and thus the most common method for laser-based triangulation.



Fig. 4.5: 3D model generated from Genex 3D FaceCam.





(a) (b) Fig. 4.7: The Genex 3D FaceCam experimental setup (a) Front view of the experimental setup mounted on a tripod (b) Side view.



Fig. 4.8: Position/ Capture screen of the Genex 3D FaceCam. The head should be positioned in center for a better result.



Fig. 4.9: Processing screen of the Genex 3D FaceCam. The color coded structured light is also observed in the first and third windows.

4.3 Effect of Varying the Parameter σ

The parameter σ controls the region of convergence which should be large for better practical applications. However, increasing the value of σ without any constraints causes a decrease in the localization accuracy. It is with this motivating factor that we analyze the effect of varying σ on the registration accuracy using the synthetic and 3D face dataset from our database. The results of this experiment are shown in Fig. 4.10.

It is interesting to find that both the models exhibit similar trends in the sense that the registration error increases linearly as a function of σ . However, the rate of increase slows down for larger values of σ and tends towards an asymptotic limit. This can be the explained by the fact that as σ exceeds the average distance between the points in the datasets the exponential can be approximated by its first order development:

$$\exp(-\frac{d^{2}(Tr(P_{i}), Q_{j})}{\sigma^{2}}) \approx 1 - \frac{d^{2}(Tr(P_{i}), Q_{j})}{\sigma^{2}}$$
(4.4)

The optimization problem now reduces to minimizing the sum of average distances from one point set to other dataset and doesn't depend anymore on σ . Hence the registration error is bounded. Based on this behavior, we can develop an algorithm that starts with initial rough alignment with a large σ , and then end up with a refinement step where σ is sharply decreased leading to a very low registration error.

4.4 Noise Analysis

Noise may have a significant effect on the 3D registration process, especially in the Gaussian criterion framework, because it influences both the position of the point-sets as well as the descriptors computed from them. In practical applications, noise is more dominant in the radial direction with respect to camera's coordinate frame. However, we focus our experimental analysis on uniform noise to study the worst case scenario. As mentioned in earlier sections, the parameter C_a is added to our criterion to compensate the effect of descriptors which become practically useless at very high levels of noise. This is achieved by forfeiting a part of discriminatory power that the descriptors add at higher levels of noise. For practical applications the confidence level factor is typically chosen to be around 10^{-3} for datasets with low noise levels and around unit value for higher noise values. For the purpose of noise analysis we add uniform noise of amplitude ranging up to 10% of the length of the face to both the models.



Fig. 4.10: Plots showing (a) the rotation and (b) translation error of the real face data and the synthetic face as a function of parameter σ . The parameter sigma and translation error are in terms of fraction of the length of the face model.

The effect of uniform noise on the drift in the maximum of the criterion can be studied from the plots shown in Fig. 4.11. The first conclusion made from the plots is that our algorithm is robust for levels of uniform noise up to $\pm 7\%$, which is very high by any practical standards. The effect of C_a in moderating the effect of registration accuracy at higher levels of noise can also be seen.

4.5 Resolution Analysis

The main criterion of a good registration method is the level of accuracy and the computational complexity involved. There are many optimization techniques which could reduce the computational complexity burden. Although the Fast Gauss Transform was utilized to reduce the computational complexity of the criterion, the sub-sampling of the datasets would lead to a further computational gain. However, the number of points in the datasets (Fig. 4.12) should be sufficient to maintain the accuracy level. Hence, this turns out to be an optimization between the computational complexities and level of accuracy. It was this factor which drove us to experiment on the minimum number of points in space required for an effective 3D registration.

The dataset utilized was taken from our IRIS 3D face database. We start with a relatively low number of 3000 points for each view and then reduced the sampling by half to obtain the next pairs until we reach 350 points. To study the influence of reduction in resolution we sub-sampled our datasets in three different ways: uniform sampling, where the points are sampled at equal intervals; curvature based sampling where points in high curvature regions are retained and points in low curvature region are thinned in order to maintain the accuracy of the curvature line; and random sampling, where the points are randomly sampled throughout the dataset.

Although at higher levels of sampling (lower number of points; Fig. 4.13) the curvature sampling provides a slight edge over others, no particular method can be considered superior to others. The reason that no particular sampling method can be attributed as perfect is due to the following reasons:

- Uniform sampling has better spatial distribution of points but this may lead to coarser description of objects.
- Curvature sampling has better visual description but may sometimes lead to complications due to clustering of points in certain areas.
- Random sampling may create complications due to uneven distribution of points.

Another observation from Fig. 4.13 is that the criterion does not break down even at higher levels of sampling and remains intact even for a few points around 800, thus reducing the computational burden by multi resolution strategy that initializes at



Fig. 4.11: Registration error versus uniform noise level; (a) rotation error in degrees and (b) translation error as a fraction of the length of the face model. We show plots for three values of the confidence parameter.



Original 3D Face Scan - 75,000 points







Fig. 4.13: Effect of sampling on the registration accuracy; (a) rotation error and (b) translation error as a function of number of points for three different sampling methods.

coarser levels.

An experimental analysis was also performed to analyze the drift in the maximum of the Gaussian criterion by performing sampling and adding noise in parallel. For reasons explained earlier, uniform noise was added to the models. The residual error remains small for noise level up to 6% of the length of the head and then increases drastically as seen in Fig 4.14. This general trend is similar for plots with different sampling factor, in the sense that the error increases as the noise increases. However, the error associated with sub sampled points is slightly higher. This puts a limitation on the minimum number of points in space required to register a 3D face.

4.6 Effect of Overlap

The amount of overlap between the different datasets to be registered plays an important role in the accuracy of the registration. In other terms, the lower the relative overlap between the two datasets, the higher the error of registration. The outliers which can be defined as area of datasets not shared by the datasets causes the drift in the maximum of the Gaussian criterion from the correct position, but this can be compensated by a suitable choice of force range parameter.

To study the effects of overlap, partial face models with different levels of overlap ranging from 25% to 80% were generated using our Genex 3D FaceCam scanner. The drift of the criterion maximum caused by the outliers is studied for four different values of the force range parameter σ (20%, 40%, 60%, and 80%). The translation error is computed in terms of the percentage of the largest dimensions of the box bounding the model.

The results are summarized in the plots of Fig 4.15. The above plots show that the algorithm is stable for up to 40% overlap and the registration accuracy decreases rapidly for overlap less than 30%. This is due to the effect of outliers and by the term outliers we mean the area which is not common in both the models. These outliers shift the maximum of the Gaussian away from its true maximum and this effect can be overridden by decrease in the force range parameter or increase in the information content. The slowest drift in the localization error occurs for the curve having low gamma which strengthens the theoretical claim about the effect of force range parameter. Hence, it can be concluded that for practical applications it is suitable to have at least around 40% to 50% overlap.

On similar lines, investigation was performed to analyze the effect of noise on different overlapping models used for registration. Different levels of uniform noise were added to both the face models to be registered and then the Gaussian criterion was applied on them. It is seen from Fig. 4.16 that the localization error increases as



Fig. 4.14: Effect of noise on the registration accuracy for different sampling factors; (a) rotation error and (b) translation error in terms of fraction of the dimensions of the face.



Fig. 4.15: Effect of amount of overlap between two faces on the registration accuracy. Plots of (a) rotation error and (b) translation error for different values of force range parameter.



Fig. 4.16: Effect of noise on the registration accuracy for models with different amount of overlap; (a) rotation error and (b) translation error for different values of overlap.

the level of noise in the models increases. This increase is much higher for the face models having lower amount of overlap. At lower amount of overlap, the localization error shows an oscillating behavior. Also, the criterion is stable to noise levels up to 6% of the length of the model.

A similar kind of experiment was also conducted to study the effect of sampling and different levels of overlap on the localization error. We start with the relatively low number of 3000 points for each view, then sample by two to obtain the next pairs until we reach 300 points. It can be seen from the experimental results shown in Fig 4.17 that the localization error increases as the number of points in the face datasets decreases. The models with lower overlap have higher localization error when compared to models having same points but higher overlap. Furthermore, the criterion is stable for face models up to 700-800 points and the localization error increases drastically below that. Thus for practical purpose it would be suitable to have an overlap more than 40% and number of points more than 800.

4.7 Comparison with ICP

In order to study the effect of σ on the region of convergence and to prove its advantages over the ICP algorithm, we analyzed the basins of convergence of the algorithm for the 3D face dataset. A relationship between the initial value of transformation parameters provided to the algorithm and the residual error at the end of the process with different values of σ can be seen in Fig. 4.18

These plots confirm the tradeoff between a large basin of convergence for a large value of σ associated with a large residual error as well, and a smaller basin of convergence for a small value of σ that comes with better registration accuracy. It can also be seen that the width of the basins grow fast at first but then do not increase much after a certain value of the force range parameter. Also, when these basins are compared with that of ICP, it is found that they are wider even for small values of σ . This can be attributed to the fact that ICP is a locally convergent scheme and needs close initialization. However, the ICP has a small residual error except when compared with algorithm tuned for close Gaussian fields. Thus a balance between residual error and the region of convergence can be obtained by a suitable adaptive optimization scheme.



(b)

Fig. 4.17: Effect of sampling on the registration accuracy for different overlap models; (a) rotation error and (b) translation error in terms of length of the model.



Fig. 4.18: Comparison of our method's basin of convergence to that of ICP; (a) rotation error and (b) translation error for three different values of σ and the ICP.

5 RESULTS FOR 3D FACE RECOGNITION

This chapter describes the methodology implemented in this thesis for the purpose of recognizing faces from 3D face data in point-sets form. The various face database used are described first, followed by the description of the methodology used. This description is followed by a discussion of the experimental results and a study of effect of glasses, and complete head model on the recognition performance.

5.1 3D Face Database

The recognition experiments were performed on a wide variety of database to validate our recognition algorithm on different quality of 3D faces, having considerable variations among them. The main objective was to make use of databases which had a good number of 3D facial models of various individuals, and simultaneously rich in accentuated variations for the same person. The variations offered by a human face can be intrinsically related to his face (e.g. expressions), and extrinsically related to the pose and orientation of his face. In our experiments, the three databases mainly utilized were:

• **IRIS 3D face database:** This database was created by us at the Imaging, Robotics, and Intelligent Systems (IRIS) laboratory, The University of Tennessee, Knoxville. As there were very few 3D face databases which were publicly available, we generated our own 3D face database (IRIS3DFD) using the Genex 3D FaceCam. The operation of the Genex 3D FaceCam is described in Section 4.2 of the previous chapter. The database consists of 495 three dimensional facial surfaces corresponding to 25 individuals (18 males and 7 females) taken over a period of time. Most of the individuals are aged between 20 years to 35 years, but vary in gender and ethnicity. There are seven different images per person and in particular, there is one frontal image and six rotated images with neutral expressions. Each facial scan has around 75,000 points excluding any external background. Complete ear to ear face models (25 individuals) are built by registering these different views of each individual. Apart from variations in the pose, the database includes three views per person

in which there are facial expression viz. happy, sad, and shocked. However, the striking feature of our database is the 3D facial surfaces of people with glasses. The database consists of seven different images per person with glasses (Fig 5.1). These seven images include a frontal image with glasses and six rotated views with glasses.

- GavabDB 3D Face Database: This database [GAVABDB] has been built at the Universidad Rey Juan Carlos for automatic face recognition experiments and other pose and registration experiments. It contains 427 three dimensional facial surface images of 61 individuals. The scanned individuals were aged between 18 years to 40 years and were Caucasians. There were nine different views for each person including two frontal and four rotated images without any facial expression. The other three frontal images of the person consisted of three different facial expressions, out of which two were very pronounced. The four rotated views corresponded to a 90° rotation of head around vertical axis in both the directions and a ±35° rotation around the X axis (Fig 5.2).
- The XM2VTS Database: The complete XM2VTS database [XM2VTS] is a large multimodal database created by the Center for Vision, Speech, and Signal Processing at The University of Surrey, United Kingdom (Fig 5.3). The database contains digital recordings of 295 volunteers taken in four sessions over an interval of one month. All the data was built using the high quality digital video equipment. During the recordings, the subject was either looking towards the camera (Fig 5.3), or was talking and turning his head both in yaw and roll. In the third session, a 3D model was built using an active stereo system provided by the Turning Institute. The 3D database consists of 295 facial surfaces of 295 people and the corresponding models are stored in the VRML format.

5.2 Implementation Details

The images from the databases which were employed to test the robustness of our 3D face recognition algorithm were classified into two categories: a gallery set and a probe set. Our gallery set consisted of 380 frontal 3D facial surfaces in the point form. The gallery set was constructed with the help of 3D models from three different face databases which were explained in detail in Section 5.1. The probe dataset consists of 1195 facial surfaces corresponding to 380 individuals. In fact, the probe datasets incorporated a wide variety of pose variations and expression changes along with facial surfaces with glasses. Each facial scan in the gallery and probe dataset was down-sampled to around 3000 points for reduction in the computational complexity. However, the recognition accuracy is independent of the number of points and is mainly dependent on the accuracy of the registration between the two facial datasets.



Fig. 5.1: Models from IRIS 3D Face Database. 3D face models with texture (a, b, c, d, i,j,k,l) and their corresponding shaded models without texture (e,f,g,h,m,n,o,p). The first three models are of the same person at different time periods. Face models with glasses, expressions are also shown. We use point sets for our experiments.







Fig. 5.3: GavabDB 3D Face Database [GAVABDB].

The overall recognition process is mainly a registration technique that yields a similarity score used for recognition. In the first stage, the probe scan is registered with the different scans in the face gallery. Based on the result of the registration, a similarity score is generated and is used as a similarity metric for 3D face recognition. However, the raw score generated does not convey much information about the recognition and hence normalization is performed on the generated raw scores. For the purpose of normalization, the faces in the gallery were registered, and then matched with the same faces leading to the generated in all the cases was performed for almost 75 three dimensional facial surfaces corresponding to 75 individuals. It was observed that the raw similarity score generated in all the cases was almost similar and had a very small variance among them. This statistical observation led us to take a numerical value which is significantly higher that the highest raw score obtained from the 75 facial matches; and then divide the raw scores with this number. This process of normalization causes the similarity score to be in the range of 0 to 1. More particularly, the higher the normalized similarity score, the more similar the two facial surfaces are.

5.3 Recognition Scores

5.3.1 Frontal Faces

Each of the probe models was first registered with all the models in the database using our registration algorithm and then a similarity score was generated. The obtained similarity score was normalized as explained in Section 5.2. Furthermore, the physical values of the normalized similarity score obtained during experiments are tabulated below. The horizontal row in Table 5.1 and Fig 5.4 represents the frontal face models from the gallery and the vertical column represents probe frontal faces. Henceforth, in all the experimental results reported, the models (gallery models represented by the subscript *g*; probe models represented by subscript *p*) and the similarity scores depicted in the table are a representative of the overall behavior under that category. Each probe face is first registered with each of the models in the gallery and a similarity score is generated. A threshold of 0.7 was heuristically selected as the recognition system that if the normalized score is above the threshold, then the two faces are similar. Also, the score between the two similar faces in the gallery and the probe dataset has a maximum value when compared to scores generated from dissimilar faces.

In practical applications, due to the use of different 3D imaging sensors, the models obtained from different sensors may be affected with different types of noise. Also, the noise is more dominant in the radial direction with respect to camera's coordinate frame. The face models used for inspection (recognition or identification) may have different noise when compared to the models in the database. Keeping these factors in mind we

Table 5.1: The normalized similarity scores computed based on Gaussian criterion. (a) The horizonta	al
row represents the gallery and vertical column represents probe dataset consisting of frontal faces, (b)	a
graphical representation of the scores.	

	Gallery - 2.5D Frontal Face														
Probe - 2.5D Frontal Face		$\mathbf{A}_{\mathbf{G}}$	B _G	C _G	D _G	E _G	F _G	G _G	H _G	I _G	$\mathbf{J}_{\mathbf{G}}$	K _G	L _G	$\mathbf{M}_{\mathbf{G}}$	N _G
	A _P	<mark>0.85</mark>	0.61	0.65	0.68	0.66	0.68	0.62	0.59	0.61	0.69	0.62	0.59	0.6	0.62
	B _P	0.61	<mark>0.91</mark>	0.69	0.66	0.67	0.69	0.67	0.62	0.63	0.68	0.66	0.65	0.68	0.58
	C _P	0.65	0.69	0.86	0.68	0.69	0.67	0.67	0.63	0.68	0.69	0.68	0.66	0.68	0.69
	D _P	0.68	0.66	0.68	0.85	0.63	0.69	0.67	0.64	0.66	0.61	0.69	0.66	0.58	0.69
	E _P	0.66	0.67	0.69	0.63	<mark>0.86</mark>	0.69	0.66	0.68	0.62	0.65	0.64	0.68	0.65	0.67
	$\mathbf{F}_{\mathbf{P}}$	0.68	0.69	0.67	0.69	0.69	0.92	0.64	0.62	0.67	0.69	0.68	0.66	0.63	0.66
	Gp	0.62	0.67	0.67	0.67	0.66	0.64	0.87	0.68	0.62	0.68	0.69	0.63	0.68	0.69
	H _P	0.59	0.62	0.63	0.64	0.68	0.62	0.68	0.86	0.65	0.68	0.68	0.67	0.63	0.62
	I _P	0.61	0.63	0.68	0.66	0.62	0.67	0.62	0.65	0.89	0.64	0.68	0.64	0.65	0.68
	$\mathbf{J}_{\mathbf{P}}$	0.69	0.68	0.69	0.61	0.65	0.69	0.68	0.68	0.64	0.89	0.62	0.67	0.64	0.65
	K _P	0.62	0.66	0.68	0.69	0.64	0.68	0.69	0.68	0.68	0.62	0.88	0.63	0.65	0.68
	L _P	0.59	0.65	0.66	0.66	0.68	0.66	0.63	0.67	0.64	0.67	0.63	<mark>0.86</mark>	0.67	0.68
	M _P	0.6	0.68	0.68	0.58	0.65	0.63	0.68	0.63	0.65	0.64	0.65	0.67	<mark>0.91</mark>	0.68
	N _P	0.62	0.58	0.69	0.69	0.67	0.66	0.69	0.62	0.68	0.65	0.68	0.68	0.68	0.92



(a)



Fig 5.4: Graphical representation of the similarity score using grayscale coding for frontal faces. The horizontal row represents the gallery and vertical column represents probe dataset consisting of frontal faces. The similarity between the probe and the gallery decreases as the intensity increases.
add different levels of uniform noise to models in the probe dataset and the models in the gallery are considered as virtually free from noise. Furthermore, uniform noise is added to the probe models to study the worst case scenario of the effect of noise. The normalized similarity scores computed with noisy probe models is shown in Table 5.2.

It is seen from the Table 5.2 and Fig 5.5 that for a given pair of models, the similarity score decreases as the amount of noise added to the probe models increases. Also, the criterion breaks at levels of noise more than 6% of the dimensions of the face model. In practicality, a level of noise more than 3% of the dimension of the model is considered too high. This leads to the conclusion that even in practical scene, our recognition system is robust to noise.

5.3.2 Effect of Pose Variations on Recognition

Although 3D face recognition methods are more or less invariant to changes in illumination, variation in facial pose still remains a major issue. Most of the face recognition techniques are sensitive to even minor head rotations. To test the robustness of our algorithm to various degree of pose, similarity scores were computed for probe face dataset with various pose variations. As mentioned earlier, seven different views of each individual were taken using the Genex 3D Face-Cam. The seven views and the corresponding direction conventions can be seen in Fig 5.6 and Fig 5.7.

Scores were computed considering the frontal faces in the gallery and the partial data obtained by the different angular views as probe models. The result of this experiment is tabulated in Table 5.3 and Fig 5.8. It can be seen from Table 5.3 that the similarity scores computed for probe faces having a pose of 30° are well above the recognition threshold, and hence the algorithm is robust to pose variations. Furthermore, these scores between similar faces were highest when compared to scores obtained from dissimilar faces.

A similar experiment was conducted using the partial data obtained with the faces having a pose variation of 50° and 90°. The results are shown in Table 5.4, Fig 5.9, Table 5.5, and Fig 5.10. It can be seen from the Table 5.4 that the recognition algorithm is robust to pose variations of 50°, but fails considerably for faces with pose of 90°. The scores for similar faces are well below the recognition threshold and hence are not recognized by the system. Furthermore, the scores between same faces are not always higher than scores between dissimilar faces. This could be due to the fact that the feature information contained in a face with pose of 90° is insufficient for our registration algorithm and hence the similarity scores are not reliable. Based on these experiments, it is concluded that the recognition algorithm is stable for up to pose variations of 50° and fails after that.

Table 5.2: Effect of noise on the similarity scores. (a) The normalized similarity scores computed for noisy probe models. The horizontal row represents the gallery and vertical column represents the uniform noise (%) added to the probe model A_P , (b) a graphical representation of the scores.

						Galle	ery – 2	2.5D I	Front	al Fac	e				
		$\mathbf{A}_{\mathbf{G}}$	B _G	C _G	D _G	E _G	F _G	G _G	H _G	I _G	J_{G}	K _G	L _G	M _G	N _G
noise	1	<mark>0.84</mark>	0.61	0.65	0.67	0.67	0.68	0.62	0.6	0.61	0.69	0.63	0.59	0.6	0.62
orm	2	0.82	0.59	0.64	0.64	0.63	0.67	0.6	0.57	0.6	0.66	0.6	0.57	0.58	0.62
unif	3	0.81	0.57	0.62	0.63	0.62	0.65	0.58	0.58	0.59	0.64	0.58	0.56	0.57	0.6
with	4	<mark>0.78</mark>	0.55	0.61	0.63	0.6	0.64	0.59	0.56	0.59	0.67	0.57	0.56	0.56	0.59
lded	5	<mark>0.76</mark>	0.55	0.61	0.6	0.61	0.63	0.57	0.54	0.59	0.66	0.59	0.55	0.57	0.57
A _P ac	6	<mark>0.75</mark>	0.56	0.59	0.59	0.63	0.61	0.59	0.53	0.57	0.62	0.56	0.56	0.56	0.58
pbe 7	7	0.68	0.53	0.6	0.6	0.6	0.63	0.57	0.53	0.58	0.61	0.57	0.53	0.55	0.54
\Pr	8	0.68	0.55	0.62	0.55	0.58	0.64	0.55	0.5	0.55	0.58	0.55	0.52	0.58	0.55
	9	0.63	0.52	0.64	0.58	0.59	0.61	0.56	0.52	0.56	0.59	0.53	0.52	0.54	0.58
	10	0.6	0.53	0.59	0.56	0.57	0.61	0.54	0.51	0.57	0.57	0.54	0.51	0.53	0.54

(a)

	$\mathbf{A}_{\mathbf{G}}$	B _G	C _G	D _G	EG	$\mathbf{F}_{\mathbf{G}}$	G _G	H _G	I _G	J_{G}	K _G	L _G	M_{G}	N _G	
1															
2															
3															
4															
5															True Accept
6															True Reject
7															False Accept
8															False Reject
9															
10															

(b)



Fig 5.5: Graphical representation of the effect of noise on the similarity scores using grayscale coding for frontal faces. The horizontal row represents the gallery and vertical column represents the uniform noise (%) added to the probe model A_P . The similarity between the probe and the gallery decreases as the intensity increases.



Fig. 5.6: Seven different views of a single individual. (a) Right side 90° view (b) Right side 50° view (c) Right side 30° view (d) Front view (e) Left side 30° view (f) Left side 50° view (g) Left side 90° view.



Fig. 5.7: Diagram showing the direction convention.

Table :	5.3:	The 1	normaliz	zed	similarity	scores	for	faces	with	a p	pose	of	30°.	(a)	The	horiz	ontal	row
represe	ents t	the ga	llery and	d v	ertical col	umn rep	orese	ents pr	obe d	latas	set co	onsi	sting	of	faces	with	a pos	se of
30°, (b)) a gi	raphic	al repres	sent	tation of th	ne score	s.											

					(Galler	y - 2.5	5D Fr	ontal	Face	è				
		$\mathbf{A}_{\mathbf{G}}$	B _G	$\mathbf{C}_{\mathbf{G}}$	$\mathbf{D}_{\mathbf{G}}$	E _G	$\mathbf{F}_{\mathbf{G}}$	$\mathbf{G}_{\mathbf{G}}$	H _G	I _G	$\mathbf{J}_{\mathbf{G}}$	K _G	L_{G}	$\mathbf{M}_{\mathbf{G}}$	N_{G}
	A _P	<mark>0.82</mark>	0.56	0.63	0.62	0.61	0.65	0.6	0.55	0.56	0.64	0.59	0.57	0.57	0.59
Probe - 2.5D Face with Pose of 30°	B _P	0.55	<mark>0.86</mark>	0.64	0.64	0.63	0.67	0.65	0.6	0.57	0.65	0.63	0.61	0.65	0.57
	C _P	0.62	0.63	<mark>0.83</mark>	0.63	0.66	0.64	0.62	0.61	0.64	0.63	0.66	0.63	0.62	0.64
	D _P	0.63	0.65	0.62	0.79	0.6	0.72	0.63	0.59	0.64	0.59	0.58	0.63	0.56	0.64
	E _P	0.62	0.63	0.67	0.6	<mark>0.81</mark>	0.65	0.63	0.65	0.59	0.62	0.61	0.64	0.61	0.62
	F _P	0.65	0.65	0.65	0.66	0.64	<mark>0.86</mark>	0.61	0.59	0.61	0.67	0.65	0.63	0.59	0.61
Fac	G _P	0.61	0.66	0.61	0.61	0.64	0.62	<mark>0.83</mark>	0.65	0.6	0.64	0.65	0.58	0.65	0.64
.5D	H _P	0.55	0.6	0.59	0.6	0.64	0.59	0.61	<mark>0.83</mark>	0.61	0.66	0.64	0.63	0.6	0.59
e - 2	I _P	0.57	0.56	0.62	0.63	0.58	0.6	0.62	0.62	<mark>0.87</mark>	0.62	0.65	0.61	0.62	0.63
Prob	J _P	0.65	0.66	0.63	0.6	0.61	0.65	0.63	0.63	0.59	<mark>0.86</mark>	0.6	0.63	0.6	0.62
I	K _P	0.58	0.62	0.65	0.57	0.6	0.64	0.64	0.62	0.64	0.59	<mark>0.84</mark>	0.6	0.61	0.66
	L _P	0.56	0.61	0.62	0.62	0.63	0.62	0.59	0.63	0.6	0.62	0.59	<mark>0.83</mark>	0.63	0.63
	$\mathbf{M}_{\mathbf{P}}$	0.57	0.63	0.61	0.56	0.55	0.59	0.63	0.61	0.6	0.59	0.6	0.61	<mark>0.88</mark>	0.65
	N _P	0.58	0.56	0.63	0.63	0.61	0.62	0.64	0.57	0.61	0.61	0.64	0.62	0.63	<mark>0.86</mark>





Fig 5.8: Graphical representation of the similarity score using grayscale coding for faces with a pose of 30° . The horizontal row represents the gallery and vertical column represents probe dataset consisting of faces with a pose of 30° . The similarity between the probe and the gallery decreases as the intensity increases.

						Galle	ery - 2	.5D F	'ronta	l Fac	e				
		$\mathbf{A}_{\mathbf{G}}$	B _G	C _G	$\mathbf{D}_{\mathbf{G}}$	E _G	F _G	G_{G}	$\mathbf{H}_{\mathbf{G}}$	I _G	$\mathbf{J}_{\mathbf{G}}$	K _G	L_{G}	$\mathbf{M}_{\mathbf{G}}$	N _G
	A _P	<mark>0.76</mark>	0.54	0.59	0.58	0.55	0.62	0.57	0.54	0.52	0.62	0.56	0.53	0.55	0.61
•	$\mathbf{B}_{\mathbf{P}}$	0.53	0.67	0.61	0.6	0.58	0.71	0.61	0.58	0.54	0.61	0.59	0.57	0.61	0.55
C 10	C _P	0.59	0.6	0.8	0.6	0.61	0.59	0.58	0.57	0.6	0.59	0.62	0.6	0.59	0.59
ose	D _P	0.59	0.61	0.59	0.78	0.56	0.63	0.59	0.55	0.61	0.55	0.55	0.6	0.53	0.6
LN F	E _P	0.54	0.59	0.61	0.56	0.77	0.61	0.59	0.61	0.55	0.59	0.58	0.6	0.57	0.59
ace with	$\mathbf{F}_{\mathbf{P}}$	0.6	0.64	0.58	0.61	0.63	<mark>0.81</mark>	0.57	0.55	0.58	0.64	0.61	0.59	0.55	0.5
rac	Gp	0.56	0.61	0.59	0.58	0.58	0.56	0.7	0.61	0.57	0.6	0.61	0.55	0.61	0.6
Probe - 2.5D Face with Pose or	H _P	0.55	0.56	0.55	0.55	0.61	0.54	0.6	0.8	0.58	0.62	0.59	0.59	0.55	0.5
;	$\mathbf{I}_{\mathbf{P}}$	0.54	0.55	0.59	0.6	0.56	0.59	0.58	0.58	0.83	0.6	0.61	0.58	0.57	0.5
one	$\mathbf{J}_{\mathbf{P}}$	0.71	0.61	0.57	0.56	0.58	0.61	0.58	0.6	0.61	0.73	0.57	0.59	0.57	0.5
Z	K _P	0.57	0.59	0.61	0.56	0.59	0.6	0.61	0.58	0.6	0.57	0.8	0.57	0.58	0.62
	L _P	0.54	0.58	0.6	0.59	0.59	0.59	0.56	0.59	0.56	0.58	0.56	<mark>0.79</mark>	0.59	0.58
	M _P	0.55	0.6	0.57	0.52	0.56	0.54	0.6	0.57	0.57	0.55	0.57	0.57	<mark>0.84</mark>	0.6
	Np	0.59	0.57	0.58	0.61	0.6	0.56	0.61	0.55	0.58	0.57	0.61	0.57	0.6	0.8

Table 5.4: The normalized similarity scores for faces with a pose of 50° . The horizontal row represents the gallery and vertical column represents probe dataset consisting of faces with a pose of 50° , (b) a graphical representation of the scores.

	>
- 1	n
۰.	aı
· · ·	/





Fig 5.9: Graphical representation of the similarity score using grayscale coding for faces with a pose of 50° . The horizontal row represents the gallery and vertical column represents probe dataset consisting of faces with a pose of 50° . The similarity between the probe and the gallery decreases as the intensity increases.

Table 5.5: The normalized similarity scores for faces with a pose of 90° . (a) The horizontal row represents the gallery and vertical column represents probe dataset consisting of faces with a pose of 90° , (b) a graphical representation of the scores.

								G	aller	y - 2.	5D F	ronta	al Fac	ce				
				A	G	B _G	C _G	D _G	E _G	F _G	G _G	H _G	I _G	$\mathbf{J}_{\mathbf{G}}$	K _G	L _G	$\mathbf{M}_{\mathbf{G}}$	N _G
			A _P	0.	33	0.3	0.32	0.34	0.3	0.38	0.33	0.29	0.29	0.38	0.31	0.25	0.31	0.36
	°No	2	B _P	0.	28	0.36	0.36	0.35	0.32	0.4	0.36	0.34	0.3	0.37	0.34	0.33	0.35	0.31
	J.	5	C _P	0	.3 (0.34	0.34	0.37	0.36	0.33	0.33	0.31	0.35	0.34	0.39	0.35	0.34	0.36
	Post		D _P	0.	33 (0.36	0.35	0.52	0.32	0.39	0.38	0.32	0.35	0.3	0.31	0.36	0.29	0.36
	ith	1111	E _P	0.	29 (0.33	0.36	0.31	0.35	0.37	0.35	0.37	0.31	0.33	0.34	0.35	0.32	0.35
	N dJ		F _P	0.	37 (0.39	0.34	0.37	0.36	0.37	0.3	0.31	0.27	0.28	0.28	0.3	0.27	0.29
	Бa	р. т. ст	Gp	0.	34 (0.35	0.32	0.37	0.34	0.31	0.41	0.35	0.32	0.35	0.37	0.31	0.36	0.35
	E C		H _P	0.	28	0.35	0.3	0.33	0.35	0.3	0.35	0.36	0.33	0.37	0.33	0.33	0.3	0.31
	-		Ip	0.	27	0.31	0.33	0.35	0.33	0.26	0.3	0.34	0.45	0.35	0.36	0.31	0.32	0.33
	hor	101	JP	0.	36	0.35	0.34	0.29	0.33	0.27	0.34	0.36	0.34	0.42	0.32	0.31	0.33	0.33
		•	K _P	0.	33 (0.36	0.38	0.3	0.35	0.28	0.35	0.32	0.35	0.31	0.43	0.32	0.33	0.37
			L _P	0.	27 (0.32	0.35	0.34	0.34	0.3	0.36	0.31	0.32	0.32	0.3	0.37	0.35	0.35
			M _P	0.	32 (0.35	0.36	0.29	0.31	0.28	0.34	0.3	0.32	0.32	0.32	0.34	0.39	0.36
			N _P	0.	34	0.3	0.35	0.35	0.34	0.3	0.31	0.32	0.34	0.34	0.35	0.35	0.34	0.43
										(a)							
	•	D		r	D	Б	F	C	T	T	т	V	Т	м	NI	٦		
Aр	AG	D	G	~G	D	<u>;</u> Е	GF	GG	GП	G I	G JG	N G	LG	IVI(; IN(Ť		
B _P																		
2 _P																		
D _P		-																Tana
Ľp Fn		$\left\{ - \right\}$	_										+		+	┥┡		True.
Gp		\vdash											1			╡┢	1	False
H _P																		False
I _P		_										8	-		_			
J _P		$\left - \right $				_							8		-	-		
хр Lp		+				+									+	-		
M _P																		
N _P																		



Fig 5.10: Graphical representation of the similarity score using grayscale coding for faces with a pose of 90° . The horizontal row represents the gallery and vertical column represents probe dataset consisting of faces with a pose of 90° . The similarity between the probe and the gallery decreases as the intensity increases.

5.4 Recognition with Glasses

It is very important for a successful face recognition system to correctly identify/ recognize a person under different physical conditions i.e. glasses, expressions, beard etc. To test the robustness of our method to glasses, we applied our recognition algorithm on different faces with glasses. The 3D scans of 25 different individuals with glasses and different poses were collected with the help of Genex 3D FaceCam as explained earlier. People were voluntarily asked to select glasses of their choice among the four different glasses available. In our experiments, the probe face with glasses was first registered with the frontal faces in the 3D database (no glasses) and then the similarity score was generated. The similarity metric scores obtained when the frontal faces with glasses were matched to the frontal faces in our database with no glasses are shown in Table 5.6 and Fig 5.11.

It can be seen from Table 5.6 that even if the two faces are of the same person, then the similarity scores computed needn't always cross the recognition threshold. On further investigation, it was found that our algorithm worked in the case of faces which had a decent scan with glasses (Fig 5.12). Some of the glasses used by the people were highly reflective (Fig 5.13. (b)) and the 3D reconstruction of their faces was not satisfactory. The noisy spikes near the eyes obtained in such situation caused our algorithm to give erroneous results. However, in case of faces with glasses which didn't have big spikes near the eyes, satisfactory results were obtained. These results are in concurrence with the previous results obtained when probe faces without glasses were matched with the frontal faces in gallery. Hence we conclude that, the algorithm does work in case of faces with glasses if the 3D face reconstruction is satisfactory.

In another set of experiments, faces with glasses obtained at a pose of 30° and 50° were matched with the faces in our gallery and the results are shown below. It is seen from Table 5.7, Fig 5.14, Table 5.8, and Fig 5.15, that when the probe face with glasses and a pose of a 30° or 50° view is matched with a frontal face (without glasses) in the gallery, the trend in the similarity score still holds valid.

5.5 Recognition with Complete 3D Head Models

The recognition results obtained by using 3D face models are encouraging, but the recognition accuracy decreases slightly for faces with pose variations. This is due to the incompleteness of the data in the gallery face models. This motivated us to make use of a more complete face model in the gallery. A more complete face gallery (Fig.5.16) was constructed by registering different views of an individual using our registration algorithm. The gallery consists of 25 complete 3D face models of 25 individuals.

Table 5.6: The normalized similarity scores computed for faces with glasses. (a) The faces in the horizontal row are the frontal faces from gallery whereas the vertical column represents the probe datasets consisting of faces with glasses, (b) a graphical representation of the scores.

					0	Galler	ry - 2.	5D F	ronta	l Face	e				
		$\mathbf{A}_{\mathbf{G}}$	B _G	C _G	D _G	E _G	F _G	G _G	$\mathbf{H}_{\mathbf{G}}$	I _G	$\mathbf{J}_{\mathbf{G}}$	K _G	$\mathbf{L}_{\mathbf{G}}$	$\mathbf{M}_{\mathbf{G}}$	$\mathbf{N}_{\mathbf{G}}$
es	A _P	<mark>0.81</mark>	0.55	0.61	0.63	0.62	0.63	0.57	0.55	0.57	0.64	0.57	0.53	0.56	0.58
lass	B _P	0.54	<mark>0.84</mark>	0.64	0.63	0.62	0.63	0.62	0.58	0.58	0.62	0.61	0.61	0.62	0.54
th G	CP	0.62	0.62	<mark>0.81</mark>	0.63	0.62	0.62	0.61	0.58	0.63	0.62	0.63	0.61	0.6	0.62
s wit	D _P	0.61	0.63	0.62	0.8	0.58	0.63	0.62	0.58	0.61	0.57	0.63	0.61	0.54	0.63
Face	E _P	0.62	0.61	0.62	0.56	0.82	0.63	0.61	0.62	0.57	0.59	0.61	0.62	0.58	0.61
tal]	F _P	0.58	0.62	0.63	0.61	0.62	<u>0.85</u>	0.58	0.57	0.61	0.62	0.63	0.62	0.57	0.6
ron	Gp	0.54	0.63	0.6	0.6	0.6	0.56	<mark>0.83</mark>	0.62	0.56	0.61	0.61	0.58	0.62	0.62
SD F	H _P	0.57	0.62	0.57	0.59	0.63	0.58	0.61	0.81	0.59	0.61	0.61	0.62	0.58	0.57
- 2.5	I _P	0.62	0.57	0.62	0.61	0.58	0.62	0.55	0.58	0.82	0.58	0.62	0.59	0.6	0.61
obe	J _P	0.57	0.58	0.6	0.59	0.6	0.6	0.6	0.6	0.59	<mark>0.83</mark>	0.57	0.62	0.58	0.59
\Pr	K _P	0.52	0.61	0.64	0.61	0.62	0.64	0.6	0.61	0.61	0.54	0.81	0.58	0.6	0.62
	L _P	0.54	0.6	0.62	0.61	0.62	0.72	0.59	0.63	0.6	0.6	0.58	0.8	0.62	0.61
	$\mathbf{M}_{\mathbf{P}}$	0.56	0.6	0.62	0.52	0.58	0.57	0.6	0.63	0.6	0.58	0.61	0.61	<mark>0.86</mark>	0.62
	N _P	0.57	0.62	0.6	0.61	0.59	0.6	0.61	0.56	0.61	0.6	0.6	0.6	0.62	<mark>0.86</mark>







Fig 5.11: Graphical representation of the similarity score using grayscale coding for frontal faces with glasses. The faces in the horizontal row are the frontal faces from gallery whereas the vertical column represents the probe datasets consisting of faces with glasses. The similarity between the probe and the gallery decreases as the intensity increases.



Fig. 5.12: 3D models with glasses used in our experiments. Models have been shown with and without texture. We use point sets for our experiments.



Fig. 5.13: 3D face reconstruction with glasses. (a) Proper reconstruction, (b) big spike near eye due to highly reflecting glasses.

Table 5.7: The normalized scores computed for a probe dataset which consists of faces with glasses and pose of 30° . The faces in the horizontal row are the frontal faces from the gallery whereas the vertical columns contain probe faces with glasses and pose of 30° , (b) a graphical representation of the scores.

					(Galle	ry - 2	.5D I	Front	al Fa	ce				
°0°		$\mathbf{A}_{\mathbf{G}}$	B _G	$\mathbf{C}_{\mathbf{G}}$	$\mathbf{D}_{\mathbf{G}}$	E _G	F _G	$\mathbf{G}_{\mathbf{G}}$	H _G	I _G	$\mathbf{J}_{\mathbf{G}}$	K _G	L_{G}	$\mathbf{M}_{\mathbf{G}}$	N _G
of	A _P	<mark>0.76</mark>	0.54	0.56	0.57	0.59	0.57	0.52	0.51	0.52	0.6	0.54	0.49	0.53	0.53
ose	$\mathbf{B}_{\mathbf{P}}$	0.55	<mark>0.78</mark>	0.59	0.59	0.58	0.77	0.58	0.53	0.54	0.58	0.55	0.56	0.56	0.5
d P	C _P	0.57	0.6	0.67	0.6	0.58	0.57	0.57	0.53	0.59	0.57	0.58	0.55	0.56	0.58
s an	D _P	0.56	0.59	0.61	<mark>0.77</mark>	0.53	0.59	0.58	0.53	0.57	0.54	0.59	0.57	0.5	0.59
ssee	$\mathbf{E}_{\mathbf{P}}$	0.6	0.59	0.57	0.52	<mark>0.78</mark>	0.59	0.57	0.58	0.53	0.55	0.58	0.58	0.55	0.57
Gla	$\boldsymbol{F_{P}}$	0.55	0.56	0.58	0.6	0.58	0.8	0.55	0.52	0.57	0.58	0.58	0.57	0.53	0.56
ith	Gp	0.53	0.57	0.57	0.6	0.56	0.56	<mark>0.79</mark>	0.59	0.52	0.58	0.57	0.55	0.58	0.57
ce w	H _P	0.5	0.52	0.53	0.54	0.59	0.52	0.58	<mark>0.76</mark>	0.55	0.54	0.56	0.58	0.55	0.52
Fac	$\mathbf{I}_{\mathbf{P}}$	0.53	0.55	0.58	0.58	0.54	0.56	0.51	0.54	<mark>0.76</mark>	0.54	0.59	0.55	0.56	0.57
5D	$\mathbf{J}_{\mathbf{P}}$	0.62	0.58	0.76	0.55	0.53	0.57	0.57	0.57	0.53	<mark>0.78</mark>	0.52	0.57	0.53	0.54
- 2	K _P	0.55	0.57	0.57	0.59	0.58	0.58	0.57	0.55	0.57	0.53	<mark>0.76</mark>	0.55	0.54	0.58
obe	L _P	0.5	0.54	0.54	0.55	0.56	0.55	0.56	0.57	0.56	0.56	0.55	<u>0.75</u>	0.58	0.57
\mathbf{Pr}	M _P	0.55	0.56	0.57	0.51	0.56	0.54	0.56	0.55	0.55	0.55	0.52	0.57	<mark>0.81</mark>	0.59
	N _P	0.53	0.5	0.58	0.58	0.57	0.55	0.55	0.53	0.56	0.54	0.56	0.56	0.58	0.8





Fig 5.14: Graphical representation of the similarity score using grayscale coding for a probe dataset which consists of faces with glasses and pose of 30° . The faces in the horizontal row are the frontal faces from the gallery whereas the vertical columns contain probe faces with glasses and pose of 30° . The similarity between the probe and the gallery decreases as the intensity increases.

Table 5.8: The normalized scores computed for a probe dataset which consists of faces with glasses and having a pose of 50° . The faces in the horizontal row are the frontal faces from gallery whereas the vertical columns contain faces with glasses and pose of 50° , (b) a graphical representation of the scores.

						Galle	ery - 2	.5D F	ronta	ıl Fac	e				
°0°	Г	$\mathbf{A}_{\mathbf{G}}$	B _G	C _G	D _G	E _G	F _G	G _G	H _G	I _G	$\mathbf{J}_{\mathbf{G}}$	K _G	L _G	M_{G}	$\mathbf{N}_{\mathbf{G}}$
of ?	$\mathbf{A}_{\mathbf{P}}$	0.72	0.51	0.53	0.53	0.55	0.53	0.49	0.48	0.49	0.57	0.52	0.46	0.51	0.5
ose	B _P	0.5	0.63	0.55	0.56	0.55	0.52	0.54	0.5	0.51	0.55	0.51	0.51	0.52	0.48
d P	$\mathbf{C}_{\mathbf{P}}$	0.52	0.55	0.73	0.58	0.54	0.53	0.54	0.5	0.55	0.53	0.53	0.52	0.51	0.52
s ar	$\mathbf{D}_{\mathbf{P}}$	0.53	0.54	0.58	0.74	0.5	0.55	0.53	0.5	0.53	0.5	0.55	0.53	0.48	0.56
asse	E _P	0.56	0.56	0.53	0.51	0.75	0.55	0.53	0.53	0.5	0.51	0.54	0.53	0.51	0.52
Ъ С	$\mathbf{F}_{\mathbf{P}}$	0.52	0.53	0.54	0.56	0.55	0.76	0.52	0.49	0.52	0.54	0.53	0.53	0.49	0.52
witł	$\mathbf{G}_{\mathbf{P}}$	0.48	0.53	0.55	0.54	0.52	0.51	0.65	0.55	0.49	0.53	0.72	0.51	0.52	0.53
ace	H _P	0.49	0.5	0.7	0.5	0.54	0.48	0.56	0.67	0.51	0.51	0.53	0.52	0.52	0.48
DF	Ip	0.49	0.52	0.54	0.55	0.53	0.5	0.5	0.51	0.72	0.51	0.54	0.52	0.53	0.52
2.5	$\mathbf{J}_{\mathbf{P}}$	0.55	0.53	0.54	0.5	0.5	0.54	0.52	0.5	0.52	0.73	0.5	0.53	0.5	0.51
- əq	K _P	0.54	0.51	0.53	0.53	0.53	0.53	0.52	0.52	0.53	0.51	0.72	0.52	0.51	0.54
\Pr	L _P	0.47	0.52	0.53	0.54	0.53	0.5	0.51	0.54	0.51	0.52	0.52	0.7	0.55	0.53
	M _P	0.51	0.49	0.5	0.49	0.52	0.53	0.5	0.52	0.54	0.5	0.5	0.54	0.75	0.55
	$\mathbf{N}_{\mathbf{P}}$	0.49	0.53	0.52	0.55	0.5	0.51	0.51	0.49	0.52	0.52	0.52	0.52	0.54	<mark>0.74</mark>





Fig 5.15: Graphical representation of the similarity score using grayscale coding for a probe dataset which consists of faces with glasses and pose of 50° . The faces in the horizontal row are the frontal faces from the gallery whereas the vertical columns contain probe faces with glasses and pose of 50° . The similarity between the probe and the gallery decreases as the intensity increases.



Fig. 5.16: Different views of a more complete 3D face model. (a)(c) Side views, (b) Front view. However, we use point sets for our experiments.

The recognition scores were computed using the gallery of complete face models and a probe dataset consisting of frontal face models. From the recognition results shown in Table 5.9 and Fig 5.17, there was no significant improvement in the results for probe datasets consisting of frontal faces. However, when the probe datasets consisted of face models with a pose variation of 30° and 50° , there was a slight improvement in the recognition performance (Table 5.10-5.13, Fig 5.18-5.21). This leads to the conclusion that a more complete face model in the database would improve the recognition results significantly. Similar improvement in the recognition performance was observed for probe datasets consisting of faces with glasses and pose variations of 30° and 50° .

5.6 Recognition with Expression Variations

Although much development has taken in the field of 3D face recognition, still the problem of expression variation is weakly addressed. Most of the existing algorithms fail to recognize faces with expressions. Due to the assumptions and the limitations in the formulation of our criterion, it was expected that our algorithm would fail in case of faces with expressions. However, for the sake of completeness of our investigation and to see where the criterion fails, we investigated the robustness of our algorithm to expression variations by capturing three different expressions of an individual viz. happy, shock, and sad, and a probe dataset of 75 scans from 25 individuals was generated. A sample of three different expressions from the probe dataset is shown in Fig 5.22 and Fig.5.23.

Recognition experiments were conducted using the probe set with expression variations and a gallery with frontal faces models. The similarity scores obtained were normalized as described in Section 5.2. These scores are shown in Table 5.14 -5.16. The horizontal row in Fig 5.24-5.26 represents the gallery containing the frontal face models and the vertical column represents the probe dataset consisting of faces with expression.

Table 5.9: The normalized similarity scores computed on a gallery of complete head models. The horizontal row represents the gallery and vertical column represents probe dataset consisting of frontal face models, (b) a graphical representation of the scores.

					G	aller	y - 31) Ear	to E	ar Fa	ice				
		$\mathbf{A}_{\mathbf{G}}$	B _G	C _G	D _G	E _G	F _G	$\mathbf{G}_{\mathbf{G}}$	H _G	I _G	$\mathbf{J}_{\mathbf{G}}$	K _G	$\mathbf{L}_{\mathbf{G}}$	M _G	N _G
	A _P	<mark>0.84</mark>	0.61	0.65	0.67	0.67	0.68	0.62	0.6	0.61	0.69	0.63	0.59	0.6	0.62
e	B _P	0.61	<mark>0.91</mark>	0.69	0.66	0.68	0.69	0.67	0.64	0.63	0.67	0.66	0.66	0.68	0.58
Fac	C _P	0.65	0.69	<mark>0.86</mark>	0.68	0.68	0.67	0.67	0.64	0.68	0.69	0.68	0.67	0.68	0.69
ıtal	D _P	0.68	0.67	0.68	<mark>0.84</mark>	0.63	0.68	0.67	0.64	0.67	0.61	0.69	0.66	0.59	0.69
l'roi	E _P	0.65	0.67	0.69	0.64	<mark>0.86</mark>	0.69	0.66	0.68	0.64	0.65	0.64	0.68	0.66	0.67
DE	$\mathbf{F}_{\mathbf{P}}$	0.68	0.69	0.67	0.69	0.69	<mark>0.91</mark>	0.64	0.64	0.67	0.69	0.68	0.66	0.63	0.66
2.5	Gp	0.62	0.67	0.67	0.67	0.66	0.64	<mark>0.88</mark>	0.68	0.63	0.68	0.69	0.65	0.68	0.69
be -	H _P	0.62	0.62	0.63	0.64	0.68	0.62	0.68	<mark>0.86</mark>	0.65	0.68	0.68	0.66	0.63	0.64
Pro	Ip	0.63	0.63	0.68	0.66	0.63	0.67	0.62	0.66	<mark>0.88</mark>	0.64	0.68	0.64	0.65	0.68
	$\mathbf{J}_{\mathbf{P}}$	0.69	0.68	0.69	0.61	0.65	0.69	0.68	0.68	0.65	<mark>0.89</mark>	0.62	0.67	0.64	0.65
	K _P	0.64	0.66	0.68	0.69	0.64	0.68	0.69	0.68	0.68	0.63	<mark>0.88</mark>	0.63	0.65	0.68
	L _P	0.6	0.65	0.66	0.66	0.68	0.66	0.64	0.67	0.64	0.67	0.63	<mark>0.86</mark>	0.65	0.68
	M _P	0.6	0.68	0.68	0.58	0.66	0.63	0.68	0.63	0.65	0.64	0.65	0.67	<mark>0.91</mark>	0.68
	N _P	0.62	0.58	0.69	0.69	0.67	0.66	0.67	0.62	0.68	0.65	0.68	0.68	0.68	0.9







Fig 5.17: Graphical representation of the similarity score using grayscale coding on a gallery of complete head models. The horizontal row represents the gallery and vertical column represents probe dataset consisting of frontal face models. The similarity between the probe and the gallery decreases as the intensity increases.

Table 5.10: The normalized similarity scores computed on a gallery of complete head models. The horizontal row represents the gallery and vertical column represents probe dataset consisting of face models with a pose of 30° , (b) a graphical representation of the scores.

					Ga	llery	- 3 E) Ear	to F	Car F	ace				
		$\mathbf{A}_{\mathbf{G}}$	B _G	C _G	D _G	E _G	$\mathbf{F}_{\mathbf{G}}$	G _G	H _G	I _G	$\mathbf{J}_{\mathbf{G}}$	K _G	L _G	M _G	N _G
°0°	A _P	<mark>0.83</mark>	0.56	0.64	0.63	0.61	0.66	0.62	0.58	0.55	0.64	0.58	0.59	0.58	0.6
of 3	B _P	0.57	<mark>0.87</mark>	0.63	0.64	0.65	0.67	0.66	0.62	0.58	0.64	0.63	0.62	0.66	0.59
ose	C _P	0.63	0.64	<mark>0.85</mark>	0.64	0.65	0.66	0.63	0.63	0.64	0.62	0.65	0.63	0.61	0.66
h P	D _P	0.62	0.65	0.62	<mark>0.81</mark>	0.6	0.77	0.63	0.59	0.64	0.59	0.58	0.63	0.57	0.64
wit	E _P	0.63	0.64	0.67	0.61	<mark>0.83</mark>	0.64	0.64	0.64	0.61	0.62	0.61	0.62	0.62	0.63
ace	$\mathbf{F}_{\mathbf{P}}$	0.65	0.65	0.65	0.66	0.63	<mark>0.87</mark>	0.62	0.6	0.61	0.67	0.65	0.63	0.6	0.61
ΟF	G _P	0.61	0.66	0.63	0.61	0.64	0.61	<mark>0.85</mark>	0.65	0.62	0.64	0.63	0.58	0.65	0.64
2.51	H _P	0.57	0.62	0.61	0.59	0.64	0.59	0.63	<mark>0.84</mark>	0.61	0.66	0.64	0.62	0.61	0.6
- ə	$\mathbf{I}_{\mathbf{P}}$	0.58	0.57	0.62	0.63	0.6	0.62	0.63	0.62	<mark>0.87</mark>	0.62	0.62	0.63	0.62	0.63
rob	$\mathbf{J}_{\mathbf{P}}$	0.66	0.65	0.62	0.62	0.61	0.65	0.65	0.62	0.6	<mark>0.88</mark>	0.6	0.63	0.59	0.61
Р	K _P	0.58	0.63	0.65	0.58	0.59	0.64	0.64	0.62	0.62	0.59	<mark>0.85</mark>	0.61	0.61	0.66
	L _P	0.59	0.62	0.63	0.62	0.63	0.61	0.6	0.62	0.6	0.62	0.61	<mark>0.84</mark>	0.62	0.63
	M _P	0.56	0.64	0.61	0.57	0.57	0.6	0.63	0.61	0.6	0.59	0.62	0.62	<mark>0.88</mark>	0.66
	N _P	0.58	0.58	0.63	0.62	0.62	0.62	0.63	0.59	0.61	0.59	0.66	0.62	0.62	<mark>0.86</mark>







Fig 5.18: Graphical representation of the similarity score using grayscale coding on a gallery of complete head models. The horizontal row represents the gallery and vertical column represents probe dataset consisting of face models with a pose of 30° . The similarity between the probe and the gallery decreases as the intensity increases.

Table 5.11: The normalized similarity scores computed on a gallery of complete head models. The horizontal row represents the gallery and vertical column represents probe dataset consisting of face models with a pose of 50° , (b) a graphical representation of the scores.

		Gallery - 3D Ear to Ear Face														
		$\mathbf{A}_{\mathbf{G}}$	B _G	C _G	D _G	E _G	$\mathbf{F}_{\mathbf{G}}$	G _G	H _G	I _G	$\mathbf{J}_{\mathbf{G}}$	K _G	L _G	M _G	N _G	
°0°	A _P	<mark>0.77</mark>	0.54	0.58	0.59	0.56	0.63	0.59	0.55	0.54	0.63	0.55	0.54	0.57	0.6	
of :	B _P	0.54	0.69	0.62	0.59	0.58	0.64	0.62	0.58	0.55	0.6	0.6	0.58	0.71	0.55	
ose	C _P	0.59	0.61	<mark>0.81</mark>	0.61	0.6	0.62	0.57	0.59	0.6	0.58	0.62	0.61	0.58	0.61	
ЧU	D _P	0.59	0.62	0.58	<mark>0.78</mark>	0.57	0.62	0.58	0.56	0.6	0.54	0.54	0.59	0.54	0.6	
wit	E _P	0.54	0.6	0.61	0.57	<mark>0.75</mark>	0.6	0.58	0.62	0.56	0.6	0.59	0.61	0.58	0.59	
ace	$\mathbf{F}_{\mathbf{P}}$	0.6	0.64	0.6	0.61	0.64	0.8	0.59	0.57	0.58	0.64	0.61	0.59	0.56	0.57	
) F	Gp	0.57	0.6	0.59	0.58	0.6	0.57	<mark>0.71</mark>	0.62	0.59	0.61	0.61	0.56	0.61	0.6	
2.51	H _P	0.55	0.56	0.56	0.56	0.61	0.54	0.62	<mark>0.82</mark>	0.58	0.62	0.6	0.59	0.56	0.56	
e -)	$I_{P} \\$	0.56	0.56	0.59	0.6	0.58	0.59	0.59	0.59	<mark>0.83</mark>	0.62	0.61	0.59	0.57	0.59	
rob	$\mathbf{J}_{\mathbf{P}}$	0.61	0.61	0.57	0.55	0.58	0.61	0.6	0.61	0.63	<mark>0.69</mark>	0.57	0.59	0.58	0.59	
Ъ	Kp	0.57	0.57	0.6	0.57	0.6	0.61	0.63	0.58	0.62	0.57	<mark>0.81</mark>	0.57	0.58	0.62	
	L _P	0.55	0.59	0.6	0.59	0.61	0.59	0.56	0.6	0.56	0.59	0.57	0.8	0.59	0.58	
	M _P	0.55	0.6	0.58	0.54	0.56	0.55	0.6	0.59	0.57	0.55	0.58	0.58	<mark>0.85</mark>	0.62	
	N _P	0.6	0.58	0.58	0.61	0.61	0.56	0.61	0.55	0.57	0.58	0.61	0.57	0.61	<mark>0.82</mark>	







Fig 5.19: Graphical representation of the similarity score using grayscale coding on a gallery of complete head models. The horizontal row represents the gallery and vertical column represents probe dataset consisting of face models with a pose of 50° . The similarity between the probe and the gallery decreases as the intensity increases.

Table 5.12: The normalized similarity scores computed on a gallery of complete head models The faces in the horizontal row are the complete frontal faces from the gallery whereas the vertical columns contain probe faces with glasses and pose of 30° , (b) a graphical representation of the scores.

		Gallery - 3D Ear to Ear Face														
[30°		$\mathbf{A}_{\mathbf{G}}$	B _G	C _G	D _G	E _G	F _G	G _G	H _G	I _G	$\mathbf{J}_{\mathbf{G}}$	K _G	$\mathbf{L}_{\mathbf{G}}$	M _G	N _G	
se of	A _P	<mark>0.78</mark>	0.55	0.57	0.57	0.59	0.58	0.53	0.52	0.54	0.6	0.55	0.51	0.53	0.53	
P_{06}	B _P	0.56	<mark>0.79</mark>	0.6	0.59	0.58	0.57	0.58	0.54	0.54	0.58	0.55	0.56	0.56	0.52	
and	C _P	0.58	0.63	0.69	0.62	0.59	0.58	0.57	0.53	0.6	0.58	0.6	0.56	0.57	0.58	
ses :	D _P	0.57	0.6	0.62	<mark>0.78</mark>	0.53	0.59	0.59	0.54	0.57	0.54	0.59	0.57	0.52	0.6	
las	E _P	0.61	0.58	0.59	0.53	<mark>0.79</mark>	0.59	0.57	0.58	0.55	0.56	0.58	0.58	0.55	0.57	
th G	$\mathbf{F}_{\mathbf{P}}$	0.56	0.55	0.58	0.62	0.59	<mark>0.82</mark>	0.54	0.53	0.57	0.58	0.59	0.56	0.55	0.57	
wi	Gp	0.54	0.57	0.58	0.61	0.56	0.57	<mark>0.81</mark>	0.62	0.54	0.57	0.58	0.55	0.58	0.57	
ace	Hp	0.52	0.54	0.53	0.54	0.6	0.52	0.57	<mark>0.78</mark>	0.56	0.54	0.59	0.58	0.56	0.53	
DF	$\mathbf{I}_{\mathbf{P}}$	0.53	0.55	0.58	0.58	0.55	0.56	0.53	0.55	<mark>0.77</mark>	0.56	0.59	0.56	0.56	0.57	
2.5	$\mathbf{J}_{\mathbf{P}}$	0.62	0.58	0.57	0.55	0.53	0.56	0.58	0.57	0.54	0.8	0.52	0.57	0.54	0.55	
be -	Kp	0.56	0.58	0.57	0.59	0.59	0.58	0.58	0.55	0.57	0.55	<mark>0.76</mark>	0.57	0.54	0.58	
Pro	L _P	0.52	0.54	0.55	0.54	0.56	0.55	0.56	0.59	0.56	0.56	0.56	<mark>0.77</mark>	0.59	0.59	
	M _P	0.55	0.58	0.57	0.51	0.58	0.54	0.58	0.55	0.56	0.56	0.52	0.57	<mark>0.83</mark>	0.59	
	N _P	0.54	0.5	0.59	0.6	0.57	0.55	0.57	0.55	0.56	0.54	0.55	0.58	0.58	<mark>0.81</mark>	





Fig 5.20: Graphical representation of the similarity score using grayscale coding on a gallery of complete head models. The faces in the horizontal row are the complete frontal faces from the gallery whereas the vertical columns contain probe faces with glasses and pose of 30° . The similarity between the probe and the gallery decreases as the intensity increases.

Table 5.13: The normalized similarity scores computed on a gallery of complete head models The faces in the horizontal row are the complete frontal faces from the gallery whereas the vertical columns contain probe faces with glasses and pose of 50° , (b) a graphical representation of the scores.

		Gallery - 3D Ear to Ear Face														
of 50°		$\mathbf{A}_{\mathbf{G}}$	B _G	C _G	D _G	E _G	F _G	G _G	H _G	$\mathbf{I}_{\mathbf{G}}$	$\mathbf{J}_{\mathbf{G}}$	K _G	L _G	M _G	N _G	
se (Ap	<mark>0.73</mark>	0.53	0.54	0.54	0.56	0.54	0.5	0.49	0.5	0.58	0.53	0.47	0.51	0.51	
\mathbf{P}_{0}	B _P	0.51	0.65	0.55	0.56	0.55	0.52	0.54	0.5	0.51	0.55	0.51	0.51	0.56	0.48	
and	CP	0.53	0.56	<mark>0.73</mark>	0.59	0.56	0.54	0.54	0.51	0.55	0.54	0.54	0.53	0.51	0.53	
ses	D _P	0.53	0.55	0.6	<mark>0.75</mark>	0.5	0.55	0.54	0.5	0.54	0.5	0.54	0.53	0.49	0.56	
las	E _P	0.57	0.56	0.55	0.53	<mark>0.77</mark>	0.56	0.53	0.53	0.5	0.51	0.55	0.54	0.51	0.52	
РG	$\mathbf{F}_{\mathbf{P}}$	0.52	0.53	0.54	0.56	0.54	<mark>0.77</mark>	0.54	0.5	0.53	0.53	0.53	0.53	0.5	0.54	
wit	Gp	0.5	0.53	0.56	0.55	0.54	0.52	0.68	0.55	0.5	0.53	0.54	0.52	0.52	0.53	
ace	Hp	0.5	0.52	0.5	0.5	0.55	0.48	0.56	0.7	0.52	0.52	0.53	0.53	0.54	0.49	
) F	$\mathbf{I}_{\mathbf{P}}$	0.51	0.52	0.55	0.56	0.53	0.52	0.52	0.53	<mark>0.73</mark>	0.51	0.55	0.52	0.53	0.52	
2.51	$\mathbf{J}_{\mathbf{P}}$	0.56	0.55	0.56	0.51	0.52	0.54	0.53	0.51	0.54	<mark>0.75</mark>	0.5	0.54	0.52	0.52	
	Kp	0.55	0.51	0.53	0.53	0.53	0.55	0.52	0.53	0.52	0.53	<mark>0.74</mark>	0.52	0.51	0.54	
rob	L _P	0.47	0.54	0.53	0.55	0.54	0.5	0.51	0.54	0.52	0.55	0.53	<mark>0.71</mark>	0.56	0.54	
P	Mp	0.51	0.5	0.51	0.49	0.52	0.54	0.52	0.54	0.55	0.5	0.52	0.54	<mark>0.74</mark>	0.56	
	N _P	0.5	0.53	0.52	0.55	0.52	0.51	0.52	0.51	0.52	0.54	0.52	0.55	0.53	<mark>0.76</mark>	





Fig 5.21: Graphical representation of the similarity score using grayscale coding on a gallery of complete head models. The faces in the horizontal row are the complete frontal faces from the gallery whereas the vertical columns contain probe faces with glasses and pose of 50° . The similarity between the probe and the gallery decreases as the intensity increases.



(a) (b) (c) Fig. 5.22: The 2D images of the three modes of expressions. (a) Happy (b) Shock (c) Sad.



Fig. 5.23: 3D models with different expressions. The models are shown with and without texture. (a)(d) Happy expression, (b)(e) Shock expression, (c)(f) Sad Expression, (g)(h) Neutral expression. The change in the geometry between faces with neutral expression and shocked expression can also be seen.

Table 5.14: The normalized similarity scores computed for faces with expressions. The faces in the horizontal row are the frontal faces from the gallery whereas the vertical columns contain frontal faces with happy expression, (b) a graphical representation of the scores.

					(Galler	ry - 2	.5D F	ront	al Fa	ce				
sion		$\mathbf{A}_{\mathbf{G}}$	B _G	$\mathbf{C}_{\mathbf{G}}$	D _G	E _G	F _G	G _G	H _G	$\mathbf{I}_{\mathbf{G}}$	$\mathbf{J}_{\mathbf{G}}$	K _G	L _G	M _G	N _G
n Happy Expres	A _P	0.69	0.51	0.55	0.6	0.61	0.63	0.55	0.51	0.53	0.62	0.54	0.52	0.54	0.57
	$\mathbf{B}_{\mathbf{P}}$	0.52	<mark>0.79</mark>	0.62	0.61	0.6	0.62	0.61	0.56	0.56	0.61	0.6	0.58	0.59	0.52
	C _P	0.57	0.61	<mark>0.76</mark>	0.58	0.6	0.58	0.62	0.57	0.59	0.63	0.61	0.61	0.6	0.62
	D _P	0.61	0.62	0.59	<mark>0.75</mark>	0.55	0.6	0.59	0.56	0.59	0.53	0.6	0.6	0.51	0.6
with	$\mathbf{E}_{\mathbf{P}}$	0.59	0.6	0.61	0.56	0.7	0.59	0.58	0.59	0.54	0.57	0.54	0.59	0.57	0.61
Ice V	$\mathbf{F}_{\mathbf{P}}$	0.62	0.64	0.58	0.6	0.58	<mark>0.78</mark>	0.58	0.57	0.6	0.61	0.59	0.58	0.55	0.61
ıl Fa	Gp	0.53	0.62	0.61	0.57	0.59	0.57	<mark>0.76</mark>	0.6	0.55	0.59	0.6	0.54	0.58	0.61
onta	H _P	0.49	0.57	0.58	0.57	0.6	0.56	0.6	<mark>0.75</mark>	0.57	0.59	0.56	0.61	0.57	0.55
Fr($\mathbf{I}_{\mathbf{P}}$	0.54	0.55	0.58	0.59	0.53	0.6	0.53	0.56	<mark>0.77</mark>	0.56	0.59	0.57	0.58	0.59
2.5D	$J_{P} \\$	0.63	0.59	0.62	0.52	0.57	0.62	0.56	0.59	0.57	<mark>0.78</mark>	0.54	0.58	0.59	0.54
e - 2	K _P	0.53	0.6	0.6	0.59	0.56	0.56	0.61	0.55	0.58	0.54	<mark>0.77</mark>	0.57	0.59	0.6
rob	L _P	0.51	0.57	0.62	0.61	0.58	0.59	0.53	0.59	0.57	0.59	0.58	<mark>0.78</mark>	0.59	0.61
Р	M _P	0.54	0.61	0.6	0.52	0.58	0.55	0.58	0.59	0.59	0.59	0.57	0.6	<mark>0.79</mark>	0.61
	N _P	0.55	0.52	0.63	0.6	0.58	0.6	0.62	0.55	0.61	0.53	0.62	0.61	0.62	0.8







Fig 5.24: Graphical representation of the similarity score using grayscale coding for faces with expressions. The faces in the horizontal row are the frontal faces from the gallery whereas the vertical columns contain frontal faces with happy expression. The similarity between the probe and the gallery decreases as the intensity increases.

Table 5.15: The scores computed based on Gaussian criterion with expressions. The faces in the horizontal row are the frontal faces from the gallery whereas the vertical columns contain frontal faces with shocked expression, (b) a graphical representation of the scores.

и					G	aller	y - 2	.5D I	Front	al Fa	nce				
essio		$\mathbf{A}_{\mathbf{G}}$	B _G	$\mathbf{C}_{\mathbf{G}}$	$\mathbf{D}_{\mathbf{G}}$	E _G	F _G	G _G	H _G	$\mathbf{I}_{\mathbf{G}}$	J_{G}	K _G	$\mathbf{L}_{\mathbf{G}}$	M _G	N_{G}
xpr	A _P	0.58	0.45	0.51	0.54	0.55	0.56	0.51	0.46	0.47	0.55	0.49	0.47	0.5	0.53
dE	B _P	0.46	0.61	0.58	0.56	0.55	0.57	0.56	0.51	0.5	0.55	0.56	0.53	0.52	0.48
ocke	C _P	0.53	0.57	0.6	0.54	0.55	0.53	0.57	0.52	0.55	0.59	0.55	0.56	0.54	0.57
Shc	D _P	0.55	0.55	0.55	0.59	0.51	0.54	0.55	0.51	0.53	0.48	0.55	0.56	0.48	0.56
ith	E _P	0.56	0.54	0.54	0.5	0.62	0.53	0.54	0.54	0.49	0.54	0.5	0.54	0.52	0.56
ce w	$\mathbf{F}_{\mathbf{P}}$	0.55	0.57	0.52	0.55	0.52	0.58	0.55	0.52	0.55	0.57	0.53	0.54	0.51	0.55
Fac	Gp	0.54	0.58	0.57	0.56	0.54	0.53	0.61	0.56	0.51	0.54	0.54	0.5	0.53	0.55
ntal	H _P	0.47	0.52	0.53	0.52	0.55	0.53	0.55	0.6	0.53	0.55	0.51	0.55	0.52	0.51
roi	$\mathbf{I}_{\mathbf{P}}$	0.48	0.5	0.53	0.55	0.5	0.52	0.5	0.52	0.58	0.51	0.55	0.51	0.53	0.55
D	$\mathbf{J}_{\mathbf{P}}$	0.53	0.54	0.57	0.49	0.53	0.56	0.55	0.51	0.51	0.54	0.51	0.53	0.55	0.49
2.5	K _P	0.51	0.57	0.56	0.54	0.53	0.54	0.56	0.53	0.54	0.52	0.55	0.52	0.54	0.55
pe.	L _P	0.47	0.54	0.55	0.56	0.54	0.54	0.5	0.56	0.51	0.55	0.51	0.56	0.55	0.57
\Pr	M _P	0.49	0.54	0.56	0.47	0.5	0.52	0.56	0.53	0.51	0.55	0.55	0.54	0.59	0.56
	N _P	0.54	0.49	0.55	0.53	0.53	0.54	0.54	0.52	0.53	0.47	0.53	0.58	0.57	0.6







Fig 5.25: Graphical representation of the similarity score using grayscale coding for faces with expressions. The faces in the horizontal row are the frontal faces from the gallery whereas the vertical columns contain frontal faces with shocked expression. The similarity between the probe and the gallery decreases as the intensity increases.

Table 5.16: The scores computed based on Gaussian criterion with expressions. The faces in the horizontal row are the frontal faces from the gallery whereas the vertical columns contain frontal faces with sad expression, (b) a graphical representation of the scores.

						Galle	ry - 2	2.5D H	Front	al Fao	ce				
ion		$\mathbf{A}_{\mathbf{G}}$	B _G	$\mathbf{C}_{\mathbf{G}}$	$\mathbf{D}_{\mathbf{G}}$	E _G	F _G	$\mathbf{G}_{\mathbf{G}}$	H _G	$\mathbf{I}_{\mathbf{G}}$	$\mathbf{J}_{\mathbf{G}}$	K _G	$\mathbf{L}_{\mathbf{G}}$	M _G	N _G
ress	$\mathbf{A}_{\mathbf{P}}$	0.56	0.43	0.51	0.54	0.54	0.54	0.51	0.43	0.45	0.55	0.51	0.47	0.54	0.53
Exp	$\mathbf{B}_{\mathbf{P}}$	0.45	0.63	0.59	0.55	0.55	0.55	0.57	0.5	0.5	0.65	0.52	0.56	0.56	0.49
ad J	$\mathbf{C}_{\mathbf{P}}$	0.54	0.53	0.6	0.55	0.51	0.54	0.56	0.5	0.53	0.58	0.55	0.53	0.54	0.53
ih S	$\mathbf{D}_{\mathbf{P}}$	0.52	0.55	0.54	0.57	0.55	0.55	0.55	0.51	0.52	0.46	0.53	0.54	0.49	0.54
, wit	$E_{P} \\$	0.57	0.56	0.53	0.53	0.63	0.58	0.51	0.52	0.52	0.5	0.5	0.5	0.53	0.56
ace	$F_{P} \\$	0.54	0.57	0.55	0.48	0.5	0.59	0.58	0.55	0.53	0.53	0.56	0.52	0.52	0.55
tal I	G_{P}	0.54	0.55	0.58	0.49	0.56	0.5	0.62	0.56	0.55	0.52	0.54	0.51	0.53	0.5
ron	$\mathbf{H}_{\mathbf{P}}$	0.48	0.54	0.52	0.51	0.52	0.55	0.52	0.64	0.51	0.57	0.53	0.5	0.52	0.56
DF	Ip	0.46	0.52	0.53	0.56	0.51	0.61	0.51	0.53	0.56	0.54	0.55	0.56	0.54	0.55
2.5]	$\mathbf{J}_{\mathbf{P}}$	0.57	0.51	0.56	0.45	0.54	0.54	0.5	0.49	0.54	0.55	0.52	0.48	0.53	0.46
be -	K _P	0.45	0.53	0.57	0.5	0.56	0.53	0.56	0.51	0.5	0.52	0.52	0.48	0.56	0.45
Pro	L _P	0.44	0.51	0.52	0.52	0.54	0.56	0.52	0.52	0.56	0.55	0.51	0.58	0.51	0.57
	M _P	0.49	0.52	0.58	0.49	0.52	0.54	0.5	0.56	0.48	0.54	0.5	0.52	0.61	0.54
	N_P	0.51	0.5	0.53	0.5	0.5	0.55	0.49	0.55	0.55	0.57	0.51	0.57	0.52	0.6







Fig 5.26: Graphical representation of the similarity score using grayscale coding for faces with expressions. The faces in the horizontal row are the frontal faces from the gallery whereas the vertical columns contain frontal faces with sad expression. The similarity between the probe and the gallery decreases as the intensity increases.
In the case of probes with a smile expression (Table 5.14), it is seen that the similarity scores crosses the recognition threshold (0.7) and is higher in the case of similar faces when compared with scores obtained from dissimilar faces. However, the recognition algorithm fails to correctly recognize faces with shock and sad expression as seen from Table 5.15 and Table 5.16. The change in facial geometry is much higher in the case of shock and sad expressions when compared to faces with happy expression.

5.7 Recognition System Performance

A biometric recognition system is usually used in two different modes: identification or authentication. Identification is the process of trying to find out a person's identity by examining a biometric signature calculated from the person's biometric features. Generally, in the identification case, the system is trained with the patterns of several persons. For each of the persons, a biometric template is calculated in this training stage. A pattern that is going to be identified is matched against every known template, yielding either a score or a distance describing the similarity between the pattern and the template. The system assigns the pattern to the person with the most similar biometric template. To prevent impostor patterns (in this case all patterns of persons not known by the system) from being correctly identified, the similarity has to exceed a certain level. If this level is not reached, the pattern is rejected. In the authentication case, a person's identity is claimed a priori. The pattern that is verified only is compared with the person's individual template. Similar to identification, it is checked whether the similarity between pattern and template is sufficient to provide access to the secured system or area.

The higher the score is, the higher is the similarity between them. Access to the system is granted only, if the score for a trained person (identification) or the person that the pattern is verified against (verification) is higher than a certain threshold. In theory, client scores (scores of patterns from persons known by the system) should always be higher than the scores of impostors. If this would be true, a single threshold, that separates the two groups of scores, could be used to differ between clients and impostors. Due to several reasons, this assumption isn't true for real world biometric systems. In some cases impostor patterns generate scores that are higher than the scores of some client patterns. For that reason it is a fact, that however the classification threshold is chosen, some classification errors occur. For example, you can choose the threshold such high, that really no impostor scores will exceed this limit. As a result, no patterns are falsely accepted by the system. On the other hand the client patterns with scores lower than the highest impostor scores are falsely rejected. In opposition to this, you can choose the threshold such low, that no client patterns are falsely rejected. Then, on the other hand, some impostor patterns are falsely accepted. Depending on the choice of the classification threshold, between all and none of the impostor patterns are falsely accepted by the system. The threshold depending fraction of the falsely accepted patterns divided by the number of all impostor patterns is called False Acceptance Rate (FAR). Its value is one, if all impostor patterns are falsely accepted and zero, if none of the impostor patterns is accepted. If a classification threshold that is too high is applied to the classification scores, some of the client patterns are falsely rejected. Depending on the value of the threshold, between none and all of the client patterns will be falsely rejected. The fraction of the number of rejected client patterns divided by the total number of client patterns is called False Rejection Rate (FRR). The value at which the FAR and the FRR of a system are same is called the Equal Error Rate (EER). A good recognition system is expected to have a very low EER.

The performance of a recognition system is characterized by two curves: receiver operating characteristic curve (ROC) and cumulative match characteristic curve (CMC). The ROC curve is a plot between the FAR and FRR of a recognition system. To generate the ROC curve (Fig 5.27), the FAR and FRR of the system is calculated at a particular threshold. This is then marked as a point in the ROC curve. This procedure is repeated for several thresholds and finally a smooth curve is obtained. The CMC curve depicts the relationship between the FAR and the recognition percentage of the system (Fig 5.28, Fig. 5.30). In other words, it shows the recognition percentage of the system for different ranks. The EER of our recognition system (Fig 5.27) is 3.7% when the frontal face models are used in the gallery. However, this EER decreases to 3.6% when the complete 3D face models (Fig 5.29) are used in the gallery.



Fig. 5.27: The receiver operating characteristic curve for 2.5D face gallery. The EER is 3.7%.



Fig. 5.28: The cumulative match characteristic curve for 2.5D face gallery.



Fig. 5.29: The receiver operating characteristic curve for the complete head gallery. The EER is 3.6%.



Fig. 5.30: The cumulative match characteristic curve for complete head models.

5.8 Accuracy based on Database Size

According to the characterization performed by Face Recognition Vendors Test 2002 [FRVT02], the size of the database has a considerable effect on the recognition performance. It was found out that for the best recognition system, the top rank identification rate was 85% on a database of 800 people, 83% on a database of 16,000 people and 73% on a database of 37,437. In terms of mathematics, the identification performance decreases linearly with respect to the logarithm of the database.

To investigate the effect of database size on the registration accuracy of our algorithm, we conducted tests using different size of the gallery. It was found that the recognition rate (Fig. 5.31) was 96.6 % for a database of 150 facial scans. Furthermore, as the database size increased to around 500 facial scans, there was a fall in the recognition rate to 94.7%. However, most of the previous results reported for face recognition based on 3D data have been done on a small database. The use of a large number of probe models (521 models of 85 individuals) for our experiments enhances the credibility of our reported system performance.



Effect of Database size on Recognition Accuracy



5.9 Computational Time

Finally, to observe the computational time of our recognition system, we performed an experiment to calculate the time taken by our algorithm on a Pentium IV machine with 2.8 GHz clock speed and 1GB RAM. Experiments were performed using different number of points and the corresponding time was noted down. It can be seen from Fig. 5.32 and Table 5.17 that as the number of points increases the computational time also increases. We have conducted most of our tests with 3000 points and hence it took around 6 sec for the entire recognition process. The computational time can be further decreased by reducing the number of points used for recognition experiments. The recognition performance is not completely dependent on the number of points used in the datasets. However, the registration accuracy gets affected by the size of the dataset. A large number of points in the datasets to be registered will give a better registration performance, but after a certain threshold the computational burden overcomes the advantage. Keeping these factors in mind, most of the experiments were conducted with 3000 points in the gallery and probe models.



Fig. 5.32: Computational time for our recognition system. (a) The X axis represents the number of points and the Y axis represents the time taken. (b) The enlarged version.

No. of Points	Time (seconds)
600	0.28
1000	0.73
1500	1.64
2000	2.85
2500	4.47
3000	6.57
6000	27.3
12,000	110.3
24,000	408.95
48,000	1586.33

Table 5.17: Computational time for our recognition system.

6 CONCLUSIONS AND FUTURE WORK

6.1 Thesis Summary

This thesis presents a strategy to automatically recognize faces of an individual from their 3D scan. We tried to address the problems of lighting, pose and other factors which influence the accuracy of a face recognition system. This is done by completely ignoring the texture information, in any form and utilizing the geometry information from a 3D scan. However, with not many 3D face databases publicly available, a new 3D face database was built named as IRIS 3D Face Database. This 3D face database is one of its kinds due to the variety and variations in the face models. The database consists of 495 three dimensional facial surfaces corresponding to 25 individuals captured over a period of time. More complete ear to ear models were created using the different views of an individual and registering them with the help of our registration algorithm. The models are rich in texture and contain wide variety of ethnic diversity and pose variations. However, it's the 3D facial surfaces with glasses and expressions which make our database distinct from others.

Registration of the facial datasets plays an important role in our recognition pipeline and hence a comprehensive analysis was performed on the automatic registration method based on Gaussian Fields. This method overcomes the close initialization limitation of the ICP and avoids the two stage registration process employed by the other algorithms. Moreover, it allows us to start from an arbitrary initial position and converge to the registered position. A simple energy function is utilized, and by the application of the Fast Gauss Transform the computational complexity is reduced to linear level. The experiments performed on real, noisy 3D face datasets demonstrate the effectiveness of the method and its robustness to various factors such as noise, resolution, and amount of overlap between the facial datasets.

However, the success of this thesis on face recognition lies in the evolution of a similarity score based on the registration of the two facial datasets. In the first stage, the method first automatically registers facial point-sets through a criterion based on Gaussian force fields. The registration method defines a simple energy function, which is always differentiable and convex in a large neighborhood of the alignment

parameters; allowing for the use of powerful standard optimization techniques. The next phase consists of generating a similarity scores using the energy criterion E. Recognition is then performed by using the robust similarity score generated by registering 3D point sets of faces.

Automatic face recognition can become relevant only if it displays robust performance and can cope up with a considerable number of faces. Experimental results have shown that the proposed face recognition method is invariant to illumination changes and a certain degree of pose $(\pm 50^\circ)$. Our approach has been tested on a large variety of databases in which the gallery and probe images have been acquired at significantly different times. The overall best recognition rate is 94.7% on a database of 521 probe models. Furthermore, the EER of our recognition system is 3.7%. The encouraging experimental results obtained by the use of ear to ear head models in the gallery has justified the need for a more complete 3D face model for recognition applications. According to FRVT2002 many face recognition algorithms require human intervention in order to select the control points. These hand picked points were used for pose normalization and to find correspondences between two facial datasets. Since the registration algorithm does not need any set of external correspondences, our recognition system is designed to be free from anchor point detection, initial alignment, or explicit pose recovery. These features make our recognition system completely automatic removing the need for manual intervention.

6.2 Future Research

Although our recognition method is robust to noise, and invariant to pose and illumination changes, it is not that it's free from any limitations. For practical application scenario, it has some limitations such as considerably high computational time, the problem of expression invariance for emotions which cause a huge change in the facial geometry and the effect of occlusions. One way to solve this problem would be to extend the registration algorithm for non-rigid transformations. The registration algorithm in present form assumes that the two facial surfaces being matched differ only by a rigid transformation. However, human facial expressions are a non-rigid transformation and cause considerable change in geometry and appearance. It is for this reason that the registration algorithm fails to register the two facial surfaces optimally, leading to reduction in recognition rates. Extension of the present registration algorithm to non-rigid transformations would resolve this issue.

Another interesting future research in this direction would be to identify certain regions of the face which are less deformable than others and assign more importance (or "weight") to them. By assigning less importance to regions which are easily deformable, the problem of expression invariance can be addressed. An important extension to our method could be to include texture information along with the

geometry information to obtain better recognition rate. The matching energy function based on Gaussian criterion can be suitably modified to include the color attributes from the facial texture. Finally, though we have tested our method on a database of considerable size, the number of 3D face models in the database could be increased. The future potential of our framework is tremendous and our contribution is just the first step in that direction.

BIBLIOGRAPHY

BIBLIOGRAPHY

[Achermann97] B. Achermann, X. Jiang, and H. Bunke, "Face Recognition using Range Images," Proceedings International Conference on Virtual Systems and Multi Media, Geneva, Switzerland, pp. 129-136, September 1997.

[Arun87] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 9, No. 5, pp. 698-700, 1987.

[Beumier00] C. Beumier and M. Acheroy, "Automatic 3D Face Authentication," Image and Vision Computing, Vol.18, No.4, pp. 315–321, 2000.

[Blanz03] V. Blanz and T. Vetter, "Face Recognition based on Fitting a 3d Morphable Model," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 9, pp.1063-1074, 2003.

[Boughorbel04] F. Boughorbel, A. Koschan, B. Abidi, and M. Abidi, "Gaussian Fields: a New Criterion for 3D Rigid Registration," Pattern Recognition, Vol. 37, No. 7, pp. 1567-1571, July 2004.

[Boywer04] K. W. Bowyer, K. I. Chang, and P. J. Flynn, "A Survey of Approaches to Three-Dimensional Face Recognition," International Conference on Pattern Recognition, England, Vol. 1, pp. 358-361, 2004.

[Besl95] P.J. Besl and R.C. Jain, "Three-dimensional Object Recognition," ACM Computing Surveys, Vol.17, No.1, pp.75-145, 1985.

[Besl92] P. J. Besl and N. D. McKay, "A Method for Registration of 3-D Shapes," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, pp. 239-256, 1992.

[Bronstein03] A. Bronstein, M. Bronstein, and R. Kimmel, "Expression -Invariant 3D Face Recognition," Proceedings Audio and Video-based Biometric Person Authentication, Lecture Notes in Comp. Science 2688, Springer, pp. 62-69, 2003.

[Campbell01] R. Campbell and P. Flynn, "A Survey of Free-form Object Representation and Recognition Techniques," Computer Vision and Image Understanding, Vol. 81, No. 2, pp. 166-210, 2001.

[Cartoux89] J. Cartoux, J. Laprest, and M. Richetin, "Face Authentication by Profile Extraction from Range images," Proceedings of IEEE Workshop on Interpretation of 3D Scenes, pp. 194-199, 1989.

[Chang03] K. Chang, K. Bowyer, and P. Flynn, "Face Recognition Using 2D and 3D Facial Data," 2003 Multimodal User Authentication Workshop, pp. 25-32, 2003.

[Chang05] K. Chang, K. Bowyer, and P. Flynn. "An Evaluation of Multimodal 2D+3D Face Biometrics," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 4, pp. 619-624, April 2005.

[Chen92] Y. Chen and G. Medioni, "Object Modeling by Registration of Multiple Range Images," Image and Vision Computing, Vol.10, No. 3, pp. 145–155, 1992.

[Chen99] C. Chen, Y. Hung, and J. Cheng, "RANSAC-based DARCES: A New Approach for Fast Automatic Registration of Partially Overlapping Range Images," IEEE Transactions on Pattern Analysis and Machine Vision, Vol. 21, No. 11, pp. 1229-1234, 1999.

[Chua00] C. Chua, F. Han, and Y. Ho, "3D human face recognition using point signature," International Conference on Automatic Face and Gesture Recognition, Grenoble, France, pp. 233-238, March 2000.

[Dorai97] C. Dorai, J. Weng, and A. K. Jain, "Optimal Registration of Object Views using Range Data," IEEE Transactions on Pattern Analysis Machine Intelligence, Vol. 19, No. 10, pp. 1131-1138, 1997.

[Eggert98] D.W. Eggert, A.W. Fitzgibbon, and R. B. Fisher, "Simultaneous Registration of Multiple Range Views for Use in Reverse Engineering of CAD Models," Computer Vision and Image Understanding, Vol. 69, No. 3, pp 253–272, 1998.

[Eriksson99] A. Eriksson and D. Weber, "Towards 3-Dimensional Face Recognition," Proceedings of the 5th IEEE AFRICON, Cape town, Africa, pp. 401–406, 1999.

[Fan86] T.J. Fan, G. Medioni, and R. Nevatia, "Description of Surfaces from Range Data using Curvature Properties," Proceedings. IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, pp.86-91, 1986.

[Faugeras87] O. D. Faugeras and M. Hebert, "The Representation, Recognition and Positioning of 3-D Shapes from Range Data," in Takeo Kanade, editor, Three-Dimensional Machine Vision, Kluwer Academic Publishers, Boston, Massachusetts, pp. 301-353, 1987.

[Fritzgibbon03] A. W. Fitzgibbon, "Robust Registration of 2D and 3D Point Sets," Image and Vision Computing, Vol. 21, No. 12-13, pp. 1145-1153, 2003.

[FRVT02] Face Recognition Vendor Test (FRVT 2002), http://www.frvt.org/

[GAVABDB] The GAVABDB 3D Face Database, http://gavab.escet.urjc.es

[Gordon91] G. Gordon, "Face Recognition Based on Depth Maps and Surface Curvature," Proceedings. IEEE Conference on Computer Vision and Pattern Recognition, Champaign, IL, pp. 808-810, 1992.

[Greengard91] L. Greengard and J. Strain, "The Fast Gauss Transform," SIAM J. Scientific Computing, Vol. 12, pp. 79-94, 1991.

[Hesher03] C. Hesher, A. Srivastava, and G. Erlebacher, "A Novel Technique for Face Recognition Using Range Imaging," Seventh International Symposium on Signal Processing and its Applications, Paris, France, 2003.

[Horn87] B. K. P. Horn, "Closed-form Solution of Absolute Orientation using Unit Quaternions," Journal of the Optical Society of America, Vol. 4, No. 4, pp. 629-642, 1987.

[Irfanoglu04] M.O. Irfanoglu, B. Gokberk, and L. Akarun, "3D Shape–Based Face Recognition using Automatically Registered Facial Surfaces", Proceedings. International Conference on Pattern Recognition, Cambridge, England, Vol. 4, pp. 183-186, 2004.

[Lao00] 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 System," Proceedings. International Conference on Pattern Recognition, Vol. 2, pp. 911-916, 2000.

[Lee90] J. Lee and E. Milios, "Matching Range Images of Human Faces," Proceedings. Third International Conference on Computer Vision, Osaka, Japan, pp. 722-726, 1990.

[Leem03] M. W. Lee and S. Ranganath, "Pose-Invariant Face Recognition Using a 3D Deformable Model," Pattern Recognition, Vol. 36, No. 8, pp. 1835–1846, 2003.

[Leey03] Y. Lee, K. Park, J. Shim, and T. Yi, "3D Face Recognition Using Multiple Features for the Local Depth Information," Proceedings. 16th International Conference on Vision Interface, Halifax, Canada, June 2003.

[Lu04a] X. Lu, D. Colbry, and A. Jain, "Three-Dimensional Model based Face Recognition," Proceedings. International Conference on Pattern Recognition, Cambridge, England, Vol. 4, pp. 362-366, 2004.

[Lu05b] X. Lu and A. K. Jain, "Integrating Range and Texture Information for 3D Face Recognition," Proceedings. 7th IEEE Workshop on Applications of Computer Vision (WACV'05), Breckenridge, CO, Vol. 1, pp. 156-163, 2005.

[Masuda95] T. Masuda and N Yokoya, "A Robust Method for Registration and Segmentation of Multiple Range Images," Computer Vision and Image Understanding, Vol. 61, No. 3, pp. 295-307, 1995.

[Medioni00] G. Medioni, M. S. Lee, C. K. Tang, A Computational Framework for Segmentation and Grouping, Elsevier, Amsterdam, 2000.

[Medioni03] G. Medioni and R. Waupotitsch, "Face Modeling and Recognition in 3-D," IEEE International Workshop on Analysis and Modeling of Faces and Gestures, France, pp. 232-233, October 2003.

[Moreno03] A. Moreno, A. Sanchez, J. Velez, and J. Diaz, "Face Recognition using 3d Surface-Extracted Descriptors," Irish Machine Vision and Image Processing Conference, Antrim, N. Ireland, September 2003.

[Moses94] Y. Moses, Y. Adini, and S. Ullman, "Face Recognition: The Problem of Compensating for Changes in Illumination Direction," Proceedings of 3rd European Conference on Computer Vision, Stockholm, Sweden, Vol.1, pp. 286-296, 1994.

[Murio93] D. A. Murio, The Mollification Method and the Numerical Solution of Ill-Posed Problems, J. Wiley & Sons, New York, 1993.

[Nagamine92] T. Nagamine, T. Uemura, and I. Masuda, "3D Facial Image Analysis for Human Identification," Proceedings. International Conference on Pattern Recognition, Netherlands, Vol. 1, pp. 324-327, 1992.

[Pan03] G. Pan, Z. Wu, and Y. Pan, "Automated 3D Face Verification from Range Data," International Conference on Acoustics, Speech and Signal Processing, pp. 192-196, 2003.

[RapidForm02] 3D Reverse Modeler Rapid Form 2002 (Software Manual), Inus Technology Inc., 2002.

[Russ04] T. Russ, K. Koch, and C. Little, "3D Face Recognition: a Quantitative Analysis," 45-th Annual Meeting of the Institute of Nuclear Materials Management, Orlando, FL, 2004.

[Sadjadi80] F. A. Sadjadi and E. L. Hall, "Three-Dimensional Moment Invariants," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 2. No. 2, pp. 127-136, 1980.

[Sharp02] G. C. Sharp, S. W. Lee, and D. K. Wehe, "ICP Registration using Invariant Features," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, pp. 90-102, January 2002.

[Stoddart96] A. J. Stoddart and A. Hilton, "Registration of Multiple Point Sets," Proceedings 13th International Conference on Pattern Recognition, Vienna, Austria, Vol. 2, pp. 40-44, 1996.

[Tanaka98] H. Tanaka, M. Ikeda, and H. Chiaki, "Curvature-Based Face Surface Recognition Using Spherical Correlation - Principal Directions for Curved Object Recognition," 3rd International Conference on Automatic Face and Gesture Recognition, Japan, pp. 372-377, 1998.

[Tsalakanidou03] F. Tsalakanidou, D. Tzovaras, and M. Strintzis, "Use of Depth and Color Eigenfaces for Face Recognition," Pattern Recognition Letters, Vol. 24, pp. 1427-1435, 2003.

[Turk94] G. Turk and M. Levoy, "Zippered Polygon Meshes from Range Images," Proceedings. 21st International Conference on Computer Graphics and Interactive Techniques, Orlando, FL, pp. 311-318, 1994.

[Trucco99] E. Trucco, A. Fusiello, and V. Roberto, "Robust Motion and Correspondence of Noisy 3-D Point Sets with Missing Data," Pattern Recognition Letters, Vol. 20, No. 8, pp. 889-898, 1999.

[XM2VTS] The XM2VTSDB, http://www.ee.surrey.ac.uk/Research/VSSP/xm2vtsdb

[Xu04] C. Xu, Y. Wang, T. Tan and L. Quan, "Automatic 3D Face Recognition Combining Global Geometric Features with Local Shape Variation Information," Sixth International Conference on Automated Face and Gesture Recognition, Seoul, Korea, pp. 308-313, May 2004. [Vemuri86] B.C. Vemuri, Mitiche, and J.K. Agarwal, "Curvature-Based Representation of Objects from Range data," Image and Vision Computing, Vol. 1, No. 2, May 1986.

[Zhang94] Z. Zhang, "Iterative Point Matching for Registration of Free-Form Curves and Surfaces," International Journal of Computer Vision, Vol. 13, No. 2, pp 119-152, 1994.

Vita

Venkat Rao Ayyagari was born on Nov 3, 1981 in Amalapuram, a coastal town in the state of Andhra Pradesh, India. Due to his interest in mathematics, he was motivated to take mathematics and physics as electives in his high school. After completing his schooling in 1999, he attended the engineering program at Muffakham Jah College of Engineering and Technology, a private college affiliated to Osmania University, India. During his undergraduate degree he worked as a summer intern at the Electronics Corporation of India Limited where he was involved in the area of radar simulation. Venkat came to the United States of America in August 2003 for his Master's study in the Department of Electrical and Computer Engineering at The University of Tennessee, Knoxville. Upon his arrival, he was fortunate enough to become a Graduate Research Assistant in the renowned Imaging, Robotics, and Intelligent Systems (IRIS) Laboratory. Under the able supervision of Dr. Abidi, he was actively involved in the research related to the areas of Image Processing, Computer Vision, Pattern Recognition, and Biometrics. He will be awarded the Master of Science degree in Dec 2005.