

3D FACE RECOGNITION BASED ON POSE AND EXPRESSION INVARIANT ALIGNMENT



by

Naeem Iqbal Ratyal
PE101002

A thesis submitted to the
Department of Electrical Engineering
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY IN ELECTRONIC ENGINEERING

Faculty of Engineering
Capital University of Science & Technology
Islamabad
November 2016

Copyright © 2016 by Naeem Iqbal Ratyal

All rights reserved. Reproduction in whole or in part in any form requires the prior written permission of Naeem Iqbal Ratyal or designated representative.

Dedicated to my Family

ACKNOWLEDGMENT

I am thankful to my supervisor Dr. Imtiaz Ahmad Taj for his precious guidance and valuable suggestions about this research.

ABSTRACT

3D face recognition has made considerable progress during the last decade as an emerging biometric modality. In order to ensure reliable 3D face recognition, novel 3D alignment and recognition algorithms are proposed in this research work. The principal objective of this dissertation is to investigate and introduce novel techniques to construct a fully automatic 3D facial recognition system.

The first study presents a novel, pose and expression invariant approach for 3D face alignment based on intrinsic coordinate system (ICS) characterized by nose tip, horizontal nose plane and vertical symmetry plane of the face. It is observed that distance of nose tip from 3D scanner is reduced after pose correction which is presented as a quantifying heuristic for the proposed alignment scheme. In addition, motivated by the fact that a single classifier cannot be generally efficient against all face regions, a two tier ensemble classifier based 3D face recognition approach is presented which employs Principal Component Analysis (PCA) for feature extraction. The individual regions are classified using Mahalanobis Cosine (MahCos) distance, Euclidean distance, Mahalanobis (Mah) distance, and Manhattan distance in separate experiments. The resulting matching scores are combined using weighted Borda Count (WBC) based combination and a re-ranking stage. The performance of the proposed approach is corroborated by extensive experiments performed on two databases, namely, FRGC v2.0 and GavabDB, confirming effectiveness of fusion strategies to improve performance.

In the second study, a novel and fully automatic pose and expression invariant 3D face recognition algorithm is proposed using two-pass 3D face alignment based on minimum distance and two-pass 3D face alignment based on classification approach. The proposed alignment approaches are capable of aligning neutral and expressive 3D faces acquired at frontal and non-frontal poses whereas the former is capable of aligning profile face images as well. For the face recognition framework, multi-view 3D faces are synthesized to exploit real 3D facial information. The matching scores are computed between multi-view face images using Mahalanobis Cosine (MahCos)

distance, Euclidean distance, Mahalanobis (Mah) distance and Manhattan distance in separate experiments. Inspired by the effectiveness of fusion approaches, Support Vector Machine (SVM) is employed using scores obtained from multi-view face pairs for face verification. In addition, a three stage unified classifier based face identification algorithm is employed which combines results from seven base classifiers at first stage, two parallel face recognition algorithms at second stage and an exponential rank combiner at third stage in a hierarchical manner.

For profile face images, the face identification algorithm combines results using four base classifiers, two parallel face recognition algorithms and the rank combiner stage. The performance of the proposed methodology is demonstrated by extensive experiments performed on two databases: FRGC v2.0 and GavabDB. The results show that the proposed methodology can be efficiently used to construct a pose and expression invariant facial recognition system.

LIST OF PUBLICATIONS

- [1] N. I. Ratyal, I. A. Taj, U. I. Bajwa and M. Sajid, "3D face recognition based on pose and expression invariant alignment," *Computers & Electrical Engineering*, vol. 46, pp. 241-255, 2015.
- [2] N. I. Ratyal, I. A. Taj, U. I. Bajwa and M. Sajid, "Automatic multi-view 3D face recognition based on two-pass pose and expression invariant alignment," *Image and Vision Computing*, (submitted, under second revision).
- [3] N. I. Ratyal, I.A. Taj, U. I. Bajwa, M.Sajid, M. J. A. Baig and F. M. Butt, "3D face recognition based on region ensemble and hybrid features", in *Proceedings of the IEEE International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, April 11-12, 2016.
- [4] M. Sajid, I. A. Taj, U. I. Bajwa and N. I. Ratyal, "The role of facial asymmetry in recognizing age-separated face images," *Computers & Electrical Engineering*, (2016), doi: <http://dx.doi.org/10.1016/j.compeleceng.2016.01.001>
- [5] M. Sajid, I. A. Taj, U. I. Bajwa and N. I. Ratyal, "Facial asymmetry based age-group estimation: role in recognizing age-separated face images," *Neural Computing and Applications*, (submitted, under review).
- [6] M. Sajid, I. A. Taj, U. I. Bajwa and N. I. Ratyal, "The Role of Demographic Estimation in Recognizing Age-Separated Face Images," *Image and Vision Computing*, (submitted, under review).

TABLE OF CONTENTS

Acknowledgment	v
Declaration	vi
Abstract	vii
List of Publications	ix
Table of Contents	x
List of Figures	xiv
List of Tables	xvii
List of Acronyms	xviii

Chapter 1

INTRODUCTION	1
1.1 Overview	1
1.2 Face Recognition Challenges	2
1.3 Motivation	3
1.4 Disadvantages of the Existing Approaches	3
1.5 Aims and Objectives	4
1.6 Hypothesis	5
1.7 List of Specific Contributions	5
1.8 Dissertation Organization	6

Chapter 2

LITERATURE REVIEW	8
2.1 3D Image Alignment Approaches	8
2.1.1 Iterative Closest Point based Alignment	8
2.1.2 Simulated Annealing based Alignment	9
2.1.3 Average Face Model based Alignment	10
2.1.4 Intrinsic Coordinate System based Alignment	10
2.2 Face Recognition Approaches	12
2.2.1 Holistic Approaches	12
2.2.2 Local Feature based Approaches	15

2.2.3 Region based Approaches	20
2.3 Discussion.....	21
2.4 Summary	22

Chapter 3

3D FACE ALIGNMENT FRAMEWORK	23
3.1 3D Face Alignment.....	23
3.2 Proposed 3D Face Alignment Algorithms	25
3.2.1 3D Face Alignment based on Intrinsic Coordinate System	25
3.2.2 Two-Pass 3D Face Alignment based on Minimum Distance.....	34
3.2.3 Two-Pass 3D Face Alignment based on Classification pproach.....	40
3.3 Experiments and Results	43
3.3.1 Computational Complexity Analysis	47
3.4 Summary	48

Chapter 4

RECOGNITION BASED ON REGIONAL SEGMENTS OF DEPTH IMAGES	50
4.1 3D Face Recognition	50
4.2. Proposed 3D Face Recognition Algorithm	51
4.2.1 Region Creation	53
4.2.2 Regional classifiers.....	53
4.2.3 Fusion Techniques.....	56
4.3 Experiments and Results	58
4.3.1 Experiments on FRGC v2.0 Database	58
4.3.2 Experiments on GavabDB Database	63
4.3.3 Comparison with other Algorithms.....	64
4.3.4 Computational Complexity Analysis	65
4.4 Summary	68

Chapter 5

RECOGNITION BASED ON MULTI-VIEW DEPTH IMAGES	69
5.1 Proposed Methodology	70
5.1.1 Multi-view Synthesis.....	71
5.1.2 Handling Expression Variations	72
5.1.3 Classifier Fusion	72
5.1.4 Face Verification Algorithm	73
5.1.5 Face Identification Algorithm.....	73
5.2 Experiments and Results	75
5.2.1 Experiments on FRGC v2.0 Database.....	76
5.2.2 Experiments on GavabDB Database	87
5.2.3 Computational Complexity Analysis	89
5.3 Summary	94

Chapter 6

DISCUSSION.....	95
6.1 Properties of the Proposed Approaches.....	95
6.1.1 Alignment Algorithms.....	95
6.1.2 Face Recognition Algorithms.....	96
6.2 Handling Disadvantages of the Existing Techniques.....	97
6.3 Differences from Existing Methods	98
6.4 Results Related Discussion	99
6.4.1 Alignment Approaches	99
6.4.2 Recognition based on Regional Segments of Depth Images.....	101
6.4.3 Recognition based on Multi-View Depth Images	102
6.5 Summary	104

Chapter 7

CONCLUSION AND FUTURE WORK.....	105
7.1 Contributions.....	105
7.1.1 Comparison among the Proposed Techniques.....	108
7.2 Future Work.....	108

REFERENCES.....	111
Appendix- 3D Face Databases	120
A.1 FRGC v2.0 database	120
A.2 GavabDB database	122

LIST OF FIGURES

Fig. 2.1 Steps involved in aligning one point cloud to the other using Iterative Closest Point based alignment approach	9
Fig. 2.2 An example Average Face Model used in the Average Face Model based alignment	11
Fig. 2.3 Steps involved in aligning the point cloud using Intrinsic Coordinate System based Alignment	11
Fig. 2.4 Some examples of the extended Local Binary Pattern operator using circular neighborhoods	17
Fig. 2.5 Extraction of spin image using triangular mesh	18
Fig. 2.6 Illustration of (a) tensor based face representation (b) spherical face representation	20
Fig. 3.1 World and intrinsic coordinate systems shown with angles α , β and γ in yz, xz and xy planes respectively for 3D face alignment	25
Fig. 3.2 Example puffy face (a) xy view (b) yz view (c) yz view at angle 15° showing nearest point from 3D scanner on lips instead of nose	26
Fig. 3.3 Example range image showing horizontal nose plane	27
Fig. 3.4 Frontal view geometry of a face in xz plane	28
Fig. 3.5 Frontal and rotated view face geometry in xz plane presenting a face rotated at angle β	30
Fig. 3.6 Example range image in yz plane showing vertical symmetry plane to determine angle α	31
Fig. 3.7 Example range image in xy plane presenting determination of angle γ	32
Fig. 3.8 Intrinsic and world coordinate systems along with example image	35
Fig. 3.9 Minimum and maximum nose-tip scanner distances for example subject from GavabDB database (a) profile images (b) frontal image	36
Fig. 3.10 Example depth image in xz plane showing two-pass alignment algorithm	37
Fig. 3.11 Example depth image in yz plane showing two-pass alignment algorithm	38
Fig. 3.12 Distance of nose tip from scanner for first 20 scans from FRGC v2.0	44
Fig. 3.13 Distance of nose tip from scanner for first 20 scans from GavabDB	45
Fig. 3.14 Example 3D scans aligned using 3D face alignment based on intrinsic coordinate system: original (top row) and pose corrected (bottom row) from (a) FRGC v2.0 (b) GavabDB	45
Fig. 3.15 Example 3D scans aligned using two-pass 3D face alignment based on minimum distance: original (row 1, 3, 5) and pose corrected (row 2, 4, 6) from (a) FRGC v2.0 (b) GavabDB	46

Fig. 3.16 Example 3D scans aligned using two-pass 3D face alignment based on classification approach: original (top row) and pose corrected (bottom row) from (a) FRGC v2.0 (b) GavabDB	47
Fig. 4.1 Block diagram of the proposed 3D face recognition algorithm	52
Fig. 4.2 3D scan showing (a) spikes (b) holes	53
Fig. 4.3 Region creation for fusing the results. White areas show the selected face regions used in recognition experiments	54
Fig. 4.4 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and MahCos distance (first experiment).....	59
Fig. 4.5 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and Euclidean distance (first experiment)	59
Fig. 4.6 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and Mah distance (first experiment)	60
Fig. 4.7 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and Manhattan distance (first experiment)	60
Fig. 4.8 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and MahCos distance (second experiment).....	61
Fig. 4.9 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and Euclidean distance (second experiment).....	62
Fig. 4.10 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and Mah distance (second experiment).....	62
Fig. 4.11 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and Manhattan distance (second experiment).....	63
Fig. 5.1 Block diagram of proposed facial recognition system	70
Fig. 5.2 Synthesized multi-view depth images: example subject from FRGC v2.0 (top row) and GavabDB (bottom row)	71
Fig. 5.3 (a) Binary mask for segmentation of expression invariant area (b) expression invariant area of the face for example subject from FRGC v2.0	72
Fig. 5.4 Receiver Operating Characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and MahCos distance	77
Fig. 5.5 Receiver Operating Characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and Euclidean distance	77
Fig. 5.6 Receiver Operating Characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and Mah distance	78

Fig. 5.7 Receiver Operating Characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and Manhattan distance	78
Fig. 5.8 Receiver Operating Characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and MahCos distance	79
Fig. 5.9 Receiver Operating Characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and Euclidean distance	79
Fig. 5.10 Receiver Operating Characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and Mah distance	80
Fig. 5.11 Receiver Operating Characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and Manhattan distance	80
Fig. 5.12 Cumulative match characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and MahCos distance	81
Fig. 5.13 Cumulative match characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and Euclidean distance	82
Fig. 5.14 Cumulative match characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and Mah distance	82
Fig. 5.15 Cumulative match characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and Manhattan distance	83
Fig. 5.16 Cumulative match characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and MahCos distance	83
Fig. 5.17 Cumulative match characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and Euclidean distance	84
Fig. 5.18 Cumulative match characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and Mah distance	84
Fig. 5.19 Cumulative match characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and Manhattan distance	85
Fig. A.1 Example 3D faces from FRGC v2.0 database	122
Fig. A.2 Example 3D faces from GavabDB database	123

LIST OF TABLES

Table 3.1 Computational complexity analysis of the proposed algorithms	48
Table 4.1 Rank-1 recognition rates using GavabDB database (%)	64
Table 4.2 Rank-1 recognition rates using FRGC v2.0 database	65
Table 4.3 Comparison of the computational complexity of the proposed algorithm with Iterative Closest Point (ICP) algorithm	66
Table 5.1 Verification and identification rates using two-pass 3D face alignment based on minimum distance on FRGC v2.0.....	86
Table 5.2 Verification and identification rates using two-pass 3D face alignment based on minimum distance on GavabDB	88
Table 5.3 Computational complexity analysis using two-pass 3D face alignment based on classification approach	91
Table 5.4 Computational complexity analysis using two-pass 3D face alignment based on minimum distance	92
Table 5.5 Computational complexity analysis for profile face images using two-pass 3D face alignment based on minimum distance	93
Table A.1 Introduction of public databases	121

LIST OF ACRONYMS

3DWWs	3D Weighted Walkthroughs
AFM	Average Face Model
CMC	Cumulative Match Characteristic
DC	Direct Current
DNN	Deep Neural Network
DT-CWT	Dual Tree Complex Wavelet Transform
FAR	False Accept Rate
FRS	Facial Recognition System
ICA	Independent Component Analysis
ICP	Iterative Closest Point
ICS	Intrinsic Coordinate System
LBP	Local Binary Pattern
LDA	Linear Discriminant Analysis
LPP	Locality Preserving Projections
Mah	Mahalanobis
MahCos	Mahalanobis Cosine
MSAC	M-estimator SAmple Consensus
MSE	Mean Squared Error
N vs. E	Neutral vs. Expressions
N vs. N	Neutral vs. Neutral
NCC	Normalized Cross Correlation
OSH	Optimal Sparating Hyper plane
PCA	Principal Component Analysis
ROC	Receiver Operating Characteristic
ROI	Region of Interest
SA	Simulated Annealing
SFR	Spherical Face Representation
SIFT	Scale Invariant Feature Transform
SIM	Surface Interpenetration Measure
SVM	Support Vector Machine
WBC	Weighted Borda Count

Chapter 1

INTRODUCTION

1.1 Overview

Biometrics refer to physiological and behavioral attributes of human beings which are used for their automatic recognition. Over the last few decades, systems based on biometric modalities like face, iris, gait, fingerprint, and palmprint have been extensively industrialized. Among these, human face is widely employed as a biometric modality of choice because of its contactless acquisition, social acceptance and applicability to non-cooperative scenarios.

In the past decades, machine based face recognition has received substantial attention in the biometric, pattern recognition, and computer vision research community. This common interest among researchers is motivated not only by fundamental and challenging problems in this domain, but also its practical applications such as financial, forensic, access control, and video surveillance etc. Many commercial face recognition systems are already available and able to meet special requirements and contribute to the society.

The general term “face recognition” refers to two main scenarios, namely, verification or authentication, and recognition or identification. Face recognition in terms of verification task means that a person’s biometric template is matched against the claimed identity only, whereas in identification task it is matched with every template enrolled in the gallery. The results for verification experiments are evaluated in terms of Receiver Operating Characteristic (ROC) Curve and it depicts Verification or True Acceptance Rate (TAR) as a trade-off against the False Acceptance Rate (FAR). Verification Rate represents percentage of a set of probe face images that is correctly accepted and FAR represents the percentage that is falsely accepted. The Verification Rate (TAR) of 0.1% is the most commonly stated single number from ROC curve. The performance of face recognition experiments is evaluated using Cumulative Match Characteristic (CMC) curve and it summarizes the percentage of a set of probe images that is considered to be correctly matched as a function of the match rank that

is counted as a right match. Rank-1 recognition rate is the most commonly stated single number from the CMC curve.

Numerous applications of face recognition include access control, smart cards, surveillance systems, public security and criminal investigations.

1.2 Face Recognition Challenges

Over the past three decades, numerous approaches have been proposed and a great deal of efforts has been devoted to intensity based face recognition. The performances of these approaches have achieved a satisfactory level in controlled environment. However, face recognition in an unconstrained daily life environment without the user's cooperation is currently a far from solved problem and a very challenging task [1]. There are still many challenging issues (e.g. pose, illumination and expression variations) to be addressed [2] [3]. Generally, a 3D face recognition system is expected to be robust to the variations of pose, illumination, facial expression and occlusions. In this section, some main challenges of face recognition are summarized.

(i) Pose variations: Pose variations convey rich and interpersonal information. Unfortunately, different head poses result in distinct changes in facial appearances in face recognition degrading system performance dramatically. Head pose has three degrees of freedom: roll, yaw, and pitch, which can be either in-plane or out-plane. In-plane rotation (i.e. along roll) is a pure 2D problem which can be resolved trivially. On the other hand it is a very challenging task to handle large, out-plane rotations (i.e. along pitch or yaw) in 2D. Using multiple multi-view facial images for training might be a feasible strategy to handle this problem [4].

(ii) Illumination variations: Facial illumination is another most challenging issue in 2D face recognition. Illumination variations have extremely complex effects on the image of a face appearance, because varying illumination directions, leads to shifts in the shape and location of shadows, changes in highlights, and reversal of contrast gradients. In general, 3D face recognition is independent of illumination variations.

(iii) Facial expression variations: Facial expressions are caused by movement of facial muscles in response to a person's internal emotion states, intentions, or social communications [1]. Intuitively, facial expression variations can change both the

topology structure of the whole facial appearance and the geometry positions of some local facial features such as eyes, lips, etc. Because severe movements of facial muscles influence the face appearance, facial expression is an adverse factor to face recognition and expression based recognition is a challenging task.

(iv) Occlusions: Occlusions represent another challenge into face recognition, because in this case only partial faces are available as the input for classification. System accuracies are likely to deteriorate when the percentage of occluded facial part becomes larger. Common occlusions can be caused by hair, beard, glasses, hands, hats, scarfs, or even cosmetics.

1.3 Motivation

Over the past two decades, numerous attempts have been made to address the problem of face recognition and a voluminous literature has been produced. Current face recognition systems perform very well in controlled acquisition scenarios using face images captured under frontal pose with strict constraints as defined in related face recognition literature. However, in unconstrained circumstances where a face may be acquired under large pose variations, these systems fail to work. Although efforts have been made to address the pose problem in 3D, but existing techniques possess many drawbacks. Nonetheless, due to the very demanding nature of the problem, there is a need to overcome such constraints to reliably recognize face images captured under unconstrained scenarios of large pose variation including profile face images. This dissertation presents a fully automatic face recognition system that is able to recognize faces under large pose variations including profile images.

1.4 Disadvantages of the Existing Approaches

The main algorithms evolved for face recognition are based on holistic and local features of the face images. The holistic face recognition approaches utilize global facial features and preserve configural (inter relations between facial parts) information of the face. These approaches capture the most prominent features of the face images. The main disadvantages of holistic approaches are

- Their recognition performance could be significantly affected by a probe set

deviating from the average face of a gallery set because of lighting, orientation and scale.

- They are sensitive to face alignment and their performance can be impaired due to outliers and occlusions [5].
- In order to reliably learn the facial subspace, they require a large number of training images of several subjects under diverse imaging conditions [6].

On the other hand, local feature based face recognition approaches employ local descriptive features and compare them in a certain feature space (e.g. original coordinates, distances, angles and areas etc.). These approaches represent face images by a set of low dimensional local feature vectors at low computational cost and memory requirements. The disadvantages of the local feature based approaches are given as under.

- They require an additional step of locating features manually or automatically and the results are reliant on accuracy of feature localization.
- They are not capable of face recognition where images are noisy and two different subjects possess identical facial features [7].
- They are sensitive to changes in facial expressions because changes in facial expressions introduce local motion and change local shape of the face.
- They are less effective for age based face recognition [8].

1.5 Aims and Objectives

This dissertation aims to design 3D face alignment and recognition algorithms. Following are the objectives of the proposed research work.

- To develop accurate and computationally inexpensive 3D face alignment approaches to handle
 - frontal faces
 - profile faces
 - neutral and non-neutral faces
- To develop automatic 3D face recognition approaches based on

- regional segments of depth images
- multi-view depth images

1.6 Hypothesis

(i) The world and intrinsic coordinate systems become concurrent by minimizing the three angles between them whereas each of the three angles gives estimation of face rotation around x , y and z -axis respectively. The two coordinate systems can be defined for 3D face scans which are aligned at a position where the angles between the two coordinate systems attain their minimal.

(ii) For a frontal face image, the nose-tip scanner distance is the minimum. By rotating a non-frontal face and finding the minimum nose-tip scanner distance, it is aligned to a frontal position.

(iii) The variance of a frontal face is the maximum and nose-tip scanner distance is the minimum. Based on classification of the maximum variance and minimum nose-tip scanner distance, the 3D face scans are aligned.

(iv) Regional segments of depth images are capable of handling facial hair, artifacts, wrinkles and local shape deformations caused by facial expression variations. The region based face recognition results are fused for improved performance using ensemble classifier.

(v) Employing more input facial information results into improved face recognition accuracy. The 3D facial information obtained from multi-view synthesized depth images increases the classification accuracy using ensemble classifier.

1.7 List of Specific Contributions

The main contributions of this dissertation are summarized as follows.

- A 3D face alignment approach is proposed based on Intrinsic Coordinate System (ICS) characterized by nose tip, slope of the nose bridge and vertical symmetry plane of the face and is capable to align frontal and non-frontal face images [9] [10].
- A two-Pass 3D face alignment algorithm is proposed based on minimum distance capable of aligning frontal and non-frontal face images including

profile face images [11].

- A two-pass 3D face alignment algorithm is proposed based on classification approach capable of aligning frontal and non-frontal face images.
- A distance reduction measure is proposed as a quantifying heuristic for the proposed alignment approaches, based on the fact that distance of nose tip from 3D scanner is reduced after pose correction [9].
- A pose invariant 3D face recognition approach is proposed using Intrinsic Coordinate System based alignment. The proposed approach employs a two-tier ensemble classifier that combines results from multiple face regions [9].
- An automatic, pose and expression invariant face verification method is proposed employing multi-view synthesized face images and Support Vector Machine (SVM) along with two-pass 3D face alignment based on minimum distance and two-pass 3D face alignment based on classification approach in separate experiments [11].
- An automatic, pose and expression invariant face identification method is proposed based on multi-view synthesized face images and three stage, hierarchical unified classifier using the two aforementioned alignment approaches [11].
- An automatic, 3D face identification approach is proposed to handle classification of profile face images using two-pass 3D face alignment based on minimum distance [11].
- All face recognition experiments are conducted using Mahalanobis Cosine (MahCos) distance, Euclidean distance, Mahalanobis (Mah) distance and Manhattan distance based classifiers separately.
- A comparative analysis is performed with state-of-the-art methods in terms of ROC curves and rank-1 recognition rates for face verification and identification scenarios respectively, along with computational complexity analysis.

1.8 Dissertation Organization

The dissertation is organized as follows:

In chapter 2, first of all 3D face alignment approaches are reviewed. Then a comprehensive review of holistic and local feature based 3D face recognition approaches is given.

Chapter 3 explains limitations of existing 3D face alignment approaches. The proposed 3D face alignment algorithms are also presented in this chapter. The performance evaluation of the proposed alignment algorithms is presented using FRGC v2.0 and GavabDB databases along with computational complexity analysis.

Chapter 4 presents a novel 3D face recognition algorithm based on regional segments of depth images and a two stage ensemble classification approach. The performance of the proposed algorithm is evaluated using FRGC v2.0 and GavabDB databases along with computational complexity analysis.

Chapter 5 describes the proposed 3D face verification and identification algorithms based on multi-view depth images. The face verification is realized using Support Vector Machine based classification approach and face identification is performed using a three stage unified classifier. The performance of the proposed methodology is evaluated on FRGC v2.0 and GavabDB databases along with computational complexity analysis.

The findings of the proposed methodology are reported in Chapter 6.

Chapter 7 concludes this dissertation and suggests the related future work.

Chapter 2

LITERATURE REVIEW

For the past few decades, face recognition has been one of the most dynamic research areas in pattern recognition, computer vision, and biometrics communities, and a great deal of progress has been made for its development. The past years have witnessed the research attention on face recognition from 2D intensity images to 3D shape models. In this chapter, the insight into the 3D face alignment and recognition approaches is highlighted. Similar to the taxonomy of 2D face recognition, 3D face recognition approaches can be broadly classified into two categories, namely, holistic approaches and local feature based approaches. A detailed review of representative holistic and local feature based approaches is presented.

2.1 3D Image Alignment Approaches

Alignment is the process of establishing correspondence between images for their reliable comparison. Establishment of one-to-one correspondence between facial feature points is a very significant and compulsory step for classification of faces. The famous approaches existing in literature for alignment of 3D facial data are based on Iterative Closest Point (ICP) [12] [13], Simulated Annealing (SA) [14], Average Face Model (AFM) [15], and Intrinsic Coordinate System (ICS) [16].

2.1.1 Iterative Closest Point based Alignment

The aim of Iterative Closest Point [12] [13] based alignment approach is to determine rotation and translation parameters iteratively in order to transform one point cloud such that it lies as close as possible to the other point cloud (please see Fig. 2.1). For this purpose, a distance metric must be defined between two point clouds such as the Mean Squared Error which gets minimal when the two point clouds are aligned. The distance between the point clouds is minimized by rotating and translating one of the point clouds relative to the other and is determined by finding distance from each point of the first point cloud to the second one and then averaging all distances. A

significant drawback of the Iterative Closest Point based alignment approach is that it needs initial course alignment for convergence. Another disadvantage of this approach is that it is computationally expensive.

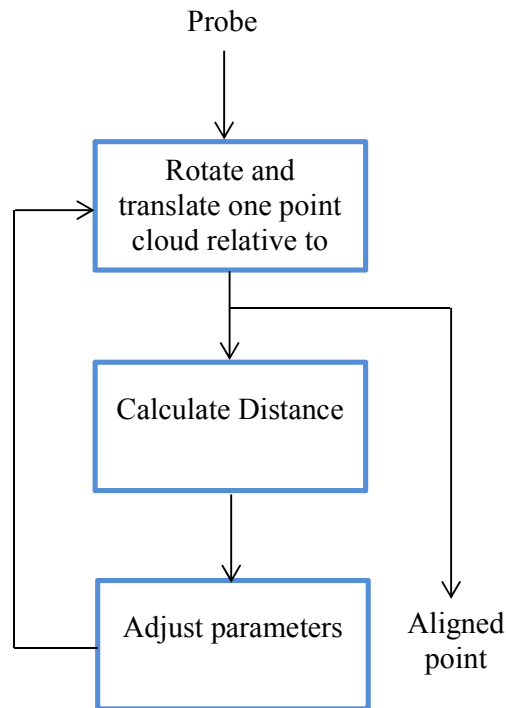


Fig. 2.1 Steps involved in aligning one point cloud to the other using Iterative Closest Point based alignment

2.1.2 Simulated Annealing based Alignment

Simulated Annealing [14] is a stochastic process based algorithm that is used for local search. The difference between Simulated Annealing and other local search methods such as Hill Climbing is that Simulated Annealing may admit a worse solution compared to current possible in the iteration process. Because of this fact Simulated Annealing does not remain “tied” to the local minima. Therefore, it has higher probability to find a solution which is close to the global one.

Simulated Annealing algorithm needs six parameters (three for each of rotation and translation with reference to a 3D coordinate system) to define a transformation

matrix for alignment of two 3D faces. The Simulated Annealing approach aligns two face images in three steps (i) alignment at initial level (ii) alignment at coarse level and (iii) alignment at fine level. Initially centers of the masses of two faces are aligned. Then a course alignment is accomplished with a searching method based on Simulated Annealing. This method works to reduce an estimation measure which uses M-estimator Sample Consensus (MSAC) along with Mean Squared Error of corresponding points of two faces to be compared. After that, a precise alignment is achieved by a searching algorithm based on Simulated Annealing which uses Surface Interpenetration Measure (SIM) as estimation criterion.

The drawback of Simulated Annealing based alignment is its excessive computational time which is comparable to Iterative Closest Point based alignment. Alignment of 3D face images based on Simulated Annealing algorithm was presented in the study Queirolo et al. [17] where it was reported by the authors that they performed an experiment with a dataset from FRGC v2.0 database for evaluation of time used in Simulated Annealing based alignment. The calculated average time for comparing two facial regions was 1.3s for nose, 2.0s for upper head region, and 3.1s for the entire face region, which is comparable to Iterative Closest Point based alignment.

2.1.3 Average Face Model based Alignment

In Average Face Model [15] based alignment, first of all landmarks are located on the face either manually or automatically. Subsequently, average of landmark coordinates is calculated followed by the Procrustes analysis [18]. The transformed landmark points are averaged again to obtain an Average Face Model [19]. An Average Face Model and corresponding landmark points are shown in Fig. 2.2 [19]. In this approach the probe face image is aligned to the Average Face Model using Iterative Closest Point based alignment. A considerable weakness of the Average Face Model based alignment is its low accuracy rates because some of the spatial information is lost while creating the Average Face Model.

2.1.4 Intrinsic Coordinate System based Alignment

The Intrinsic Coordinate System based alignment requires the accurate localization of facial landmarks on face image. These landmarks are then mapped on corresponding



Fig. 2.2 An example Average Face Model used in the Average Face Model based alignment

points in the Intrinsic Coordinate System to find transformation parameters which are then used to transform whole point cloud to obtain the required facial alignment (please see Fig. 2.3). A problem with Intrinsic Coordinate System based alignment is that 3D landmarks are not stable under facial expression variations and/or can be occluded by hair or other facial parts. Some detail of landmark based alignment procedure is given in the study Papatheodorou and Rueckert [16].

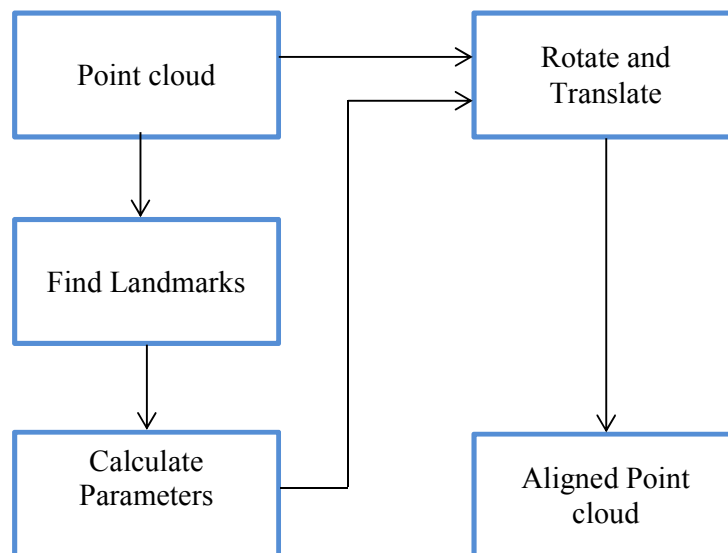


Fig. 2.3 Steps involved in aligning the point cloud using Intrinsic Coordinate System based Alignment

2.2 Face Recognition Approaches

With the development of 3D scanning and capturing techniques, the task of recognizing 3D face scans has been discussed in many ways, leading to numerous face recognition approaches. In this section, the representative and most successful approaches for 3D face recognition are reviewed. Similar to the taxonomy of 2D face recognition, 3D face recognition approaches can be generally classified into two categories: holistic and local feature-based.

2.2.1 Holistic Approaches

Holistic approaches directly operate on 3D face data to compute similarity scores. The leading algorithms in this category include subspace based methods i.e. Principal Component Analysis, Linear Discriminant Analysis, Independent Component Analysis, Locality Preserving Projections; and Iterative Closest Point based matching.

(i) Principal Component Analysis

Principal Component Analysis (PCA) [20] is a standard dimensionality reduction technique that approximates the original high-dimensional data to lower-dimensional feature vectors. Principal Component Analysis relies on a set of basis vectors corresponding to maximum variance direction of the image data.

Suppose there are N number of images, each of size $m \times n$ where m and n represent rows and columns of an image respectively. Each of the N images is reshaped into a vector of size $mn \times 1$. These vectors are then combined in a matrix \mathbf{X} of dimensions $mn \times N$. The image covariance matrix \mathbf{S} of the size $N \times N$ is computed as under.

$$\mathbf{S} = (\mathbf{x}_i - \boldsymbol{\mu})^T (\mathbf{x}_i - \boldsymbol{\mu}) \quad (2.1)$$

where \mathbf{x}_i is the i th column vector of matrix \mathbf{X} representing i th image and $\boldsymbol{\mu}$ is the average image vector computed from column vectors of \mathbf{X} . Solving for Eigen values of \mathbf{S} produces a matrix with $N \times N$ Eigen vectors. Then images are multiplied with these Eigen vectors to produce a matrix \mathbf{B} of basis vectors as given below.

$$\mathbf{B} = \mathbf{X} f(\mathbf{S}) \quad (2.2)$$

where $f(\cdot)$ is a function that computes Eigen vectors of \mathbf{S} . These basis vectors of size $m \times n$ are then normalized. After that \mathbf{d} vectors out of \mathbf{N} from \mathbf{B} are selected against d largest Eigen values. These selected vectors which are known as Eigen faces, form $mn \times d$ dimensional projection matrix \mathbf{P} . In the projection phase, the required M number of image vectors \mathbf{e} , are projected using the projection matrix \mathbf{P} to obtain a matrix of templates \mathbf{T}_m which has the size $d \times M$ as given below.

$$\mathbf{T}_m = \mathbf{P}^T \mathbf{e} \quad (2.3)$$

Principal Component Analysis is the most descriptive representation in terms of the least square reconstruction error. Moreover, it operates efficiently and is easy to implement.

(ii) Independent Component Analysis

Independent Component Analysis (ICA) [21] is quite similar to Principal Component Analysis and the only difference between them lies in that the distribution of the components is designed to be non-Gaussian. In general, minimizing non-Gaussianity enhances statistical independence [22]. It was reported in the study Bartlett et al. [21] that first and second order statistics retain information only about the amplitude spectrum of an image while discard the phase spectrum. Some experiments brought out that the human capability to recognize objects is mainly driven by the phase spectrum; therefore, Independent Component Analysis was investigated by Bartlett et al. [21] as a powerful method for face recognition. They provided two architectures based on Independent Component Analysis, namely, statistically independent basis images and a factorial code representation, both of which showed similar accuracies. Nevertheless, it was concluded in the study Baek et al. [23] that Principal Component Analysis outperforms Independent Component Analysis in a face recognition task.

(iii) Linear Discriminant Analysis

Principal Component Analysis and Independent Component Analysis are unsupervised techniques since they do not use the face class information to construct the face space. Unlike them, Linear Discriminant Analysis (LDA) [24] aims to represent the face vector in such a way that maximizes the discrimination between

various classes. The recognition performance can be improved by exploring class information [24]. Linear Discriminant Analysis directly discriminates the classes by finding a base of vectors, attempting to maximize the between class differences and minimize the within class differences. It seeks a transformation such that the projection matrix is chosen to maximize the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples.

Unfortunately, Linear Discriminant Analysis is affected by “small sample size problem” specifically common in face recognition due to small number of available training samples compared to the dimensionality of the sample space. Another limitation of Linear Discriminant Analysis algorithm is that in cases where insufficient training data is available, the algorithm can overfit the solution to the data which leads to poor generalization on unseen data. Moreover, it was revealed in the study Beveridge et al. [25] that Principal Component Analysis algorithm performs better than Linear Discriminant Analysis for face recognition.

(iv) Locality Preserving Projections

Locality Preserving Projections (LPP) or Laplacianfaces algorithm [26] is a subspace method that applies dimensionality reduction while preserving the locality information of feature space. Each face image in the image space is mapped to a low dimensional face subspace, which is characterized by a set of feature images, called Laplacianfaces. The face subspace preserves local structure using a graph based approach that is induced from the data points. Locality Preserving Projections algorithm finds a projection that employs this graph structure. The advantages of this algorithm are that it is a linear method and preserves the local information of the face image space. On the other hand, it was reported by Bajwa et al. [27] that the result accuracy of Locality Preserving Projections algorithm is lower than Principal Component Analysis in a face recognition task.

(v) Iterative Closest Point

Another 3D holistic matching algorithm is based on Iterative Closest Point (ICP) [12] approach which iteratively attempts to align two 3D surfaces represented as point

clouds or meshes. As explained in section 2.1.1, the goal of Iterative Closest Point algorithm is to find the transformation parameters, for which the error between the transformed query shape points and the closest points of the reference shape achieves minimal.

The downside of Iterative Closest Point approach is that its accuracy and convergence speed highly depends on the initial coarse estimation. Therefore, probe and gallery centroids of 3D facial surfaces are frequently aligned before Iterative Closest Point algorithm is applied to ensure an accurate final alignment estimate. If the initial estimate is not good enough, this algorithm finds a local minima that corresponds to an improper alignment resulting into a mismatch.

2.2.2 Local Feature based Approaches

Local feature based approaches concentrate on local descriptive features and compare them in a certain feature space (e.g. original coordinates, distances, areas, angle, and curvatures etc.). Gabor Filters, Log-Gabor Filters, Local Binary Pattern, Scale Invariant Feature Transform, Spin Images, Point Signatures, Tensor Based Face Representation, and Spherical Face Representation are the representative algorithms in this category.

(i) Gabor Filters

Gabor filters are commonly cited as sharing many properties with mammalian cortical cells [28] and appear a logical choice for the task of frequency partitioning. Gabor wavelets demonstrate two desirable characteristics: spatial locality and orientation selectivity. Since Gabor filters detect amplitude invariant spatial frequencies of pixel gray values, they are known to be robust to illumination variations in 2D. On the other hand, for each pixel of a facial image, Gabor filters tend to generate a high dimensional feature vector. Moreover, the final feature vector is classified in a, rather high dimensional feature space leading to another difficulty in this domain [29]. In 3D, the depth Gabor images are smoother in comparison with the intensity Gabor images due to the fact that the value of the pixels in the depth images changes less than the value in the intensity images. Although the smoother depth Gabor images can reduce the influence of noise but the facial features cannot be described by them in

detail, and the recognition rate is deteriorated. Therefore, the recognition accuracy is improved by combining both of depth and intensity information [30], [31], [32].

(ii) Log-Gabor Filters

Log-Gabor filters are defined [33] to perform the DC compensation and overcome the bandwidth limitation of a traditional Gabor filter. The Log-Gabor filter has a response that is Gaussian when viewed on a logarithmic frequency scale instead of a linear one. This allows more information to be captured in the high frequency areas with desirable high pass characteristics [34]. A limitation of Log-Gabor filter is its dimensionality explosion. Another limitation is that while Log-Gabor coefficients provide better overall performance they are equally susceptible to the effects of severe expressions [35].

(iii) Local Binary Pattern

Local Binary Pattern (LBP) was initially proposed for texture analysis by Ojala et al. [36] and has been proved as an effective approach for describing local features of the face image. Local features of facial images are described by it efficiently by comparing every pixel with its neighbors. The original Local Binary Pattern operator encodes the local structure around each pixel by labeling the pixels with decimal numbers called Local Binary Pattern codes. Each pixel is compared with its eight neighbors in a 3×3 patch. The resulting negative values are encoded with 0 and the others with 1 to obtain a binary number by concatenating all these binary codes in a clockwise direction starting from the top left of the 3×3 patch. The resulting decimal value is used for labeling.

A limitation of the basic Local Binary Pattern operator is that its small 3×3 neighborhood cannot capture the dominant features with large scale structures. Therefore, the original Local Binary Pattern operator was later generalized to extended Local Binary Pattern (eLBP) to incorporate different combinations of neighborhoods [37]. A local neighborhood is defined as a set of sampling points evenly spaced on a circle which is centered at the pixel to be labeled. The sampling points that do not fall within the pixels are interpolated using bilinear interpolation, thereby allowing for any radius and any number of sampling points in the

neighborhood. Fig. 2.4 [38] shows some examples of the extended Local Binary Pattern operator using (8, 1), (16, 2), (24, 3) circular neighborhoods where the notation P, R denotes a neighborhood of P sampling points on a circle of radius R. The histogram of Local Binary Pattern labels calculated over a region can be exploited as a feature descriptor.

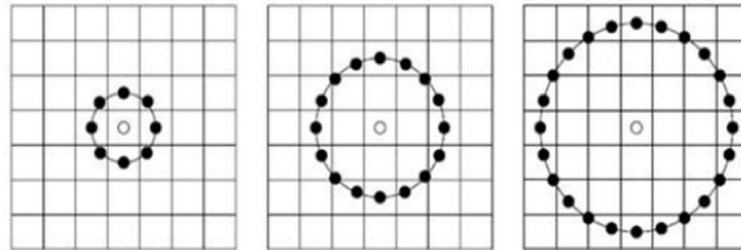


Fig. 2.4 Some examples of the extended Local Binary Pattern operator using circular neighborhoods

The most important properties of Local Binary Pattern are its tolerance to monotonic illumination changes and its low computational cost. Although attempts have been made to employ Local Binary Pattern for 3D face recognition but its coding scheme has following limitations [39]. (i) The feature vector for 3D Local Binary Pattern becomes very large. (ii) The coding principle is very sensitive to the depth variations.

(iv) Scale Invariant Feature Transform

Scale Invariant Feature Transform (SIFT) [40] features are extracted as follows: The first step computes the locations of potential interest points in the image by detecting the maxima and minima of a set of Difference of Gaussian (DoG) filters applied at different scales all over the image. Then, points of low contrast are discarded and orientation is assigned to each key point based on local image features. Finally, a local feature descriptor is computed at each key point based on the local image gradient, transformed according to the orientation of the key point to provide orientation invariance. Scale Invariant Feature Transform based features are invariant to image

rotation and scaling. On the other hand this algorithm does not perform well with noisy images and it is computationally expensive.

(v) Spin Images

Spin images [41] are extracted at each oriented point of the surface by converting the input images into triangular meshes. An oriented point is defined as a point along with its normal. A spin image at an oriented point is a 2D histogram of the cylindrical coordinates of its surrounding points which being a 2D histogram of the vertices, is not invariant to different mesh resolutions. Therefore, the mesh is optimized to a uniform resolution. A triangular mesh with oriented point p and its normal \mathbf{n} is shown in Fig. 2.5 [41]. P represents the tangent plane at point p whereas α and β are the cylindrical coordinates of the point x . Spin images are built for each vertex of the model view by making a 2D histogram of the surrounding points and stored in a spin image stack. Spin images are matched using correlation.

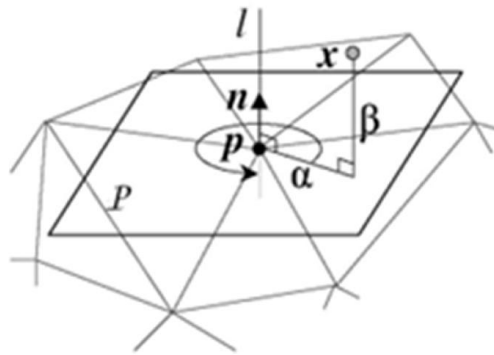


Fig. 2.5 Extraction of spin image using triangular mesh

Spin image matching is accurate and robust for overlapping views. On the other hand the limitations of the spin image representation include: (i) The spin image algorithm is computationally expensive because it involves the search for neighboring points. (ii) Spin image matching is sensitive to noise and variations in the image resolution. (iii) The complexity of the spin image matching increases linearly with the size of the data sets [41].

(vi) Point Signatures

Point signature [42] is a representation scheme developed for 3D object recognition. A point signature is a one dimensional signature that describes the surface surrounding a point. This signature describes the local underlying surface structure in the neighborhood of that point and is obtained by plotting the distance profile of a circle of points to a plane defined by that circle of points.

The point signatures algorithm is invariant to rotations and translations but lacks accuracy. The limitations of the algorithm include: (i) It is affected by noisy points. (ii) It is sensitive to the resolution of the views (iii) Its computational complexity increases linearly with the resolution of the views [41].

(vii) Tensor Based Face Representation

Tensor [41] is a local surface descriptor which corresponds to a local representation of the surface inside a 3D cubic grid. To compute the tensors from point clouds, the point clouds are converted into triangular meshes M_i where $i = 1, \dots, N$ by mapping the 3D points onto the 2D retinal plane of the sensor and performing a 2D Delaunay triangulation over the mapped points. After triangulation, the points are mapped back to the 3D space and the triangles with edges longer than a pre specified threshold are removed. This process separates surfaces which are falsely connected by the Delaunay triangulation. For reasons of efficiency, a mesh reduction algorithm may be applied to each mesh M_i . Normals are then calculated for each vertex of the reduced meshes. Once the normals have been calculated, pairs of vertices along with their normals are selected to define local 3D coordinate basis. This 3D basis is used to define a 3D grid centered at its origin (Fig. 2.6(a)) [43] using two parameters, namely, the number of bins (e.g. $10 \times 10 \times 10$) in the 3D grid and the size of each bin. Once the 3D grid is defined, the surface area of the mesh intersecting each bin of the grid is recorded in a third order tensor. Since more than one triangular facet can intersect a single bin, the calculated areas of intersection in a bin are added up. This process continues until a stage is reached when all the triangular facets are completely included.

Tensor based face representation has following advantages. (i) It is more robust to the

resolution of the views compared to the spin images. (ii) Since a tensor is a more local representation, it gives better results when used for recognition of occluded images. However, on the downside, it is more sensitive to facial expressions [43].

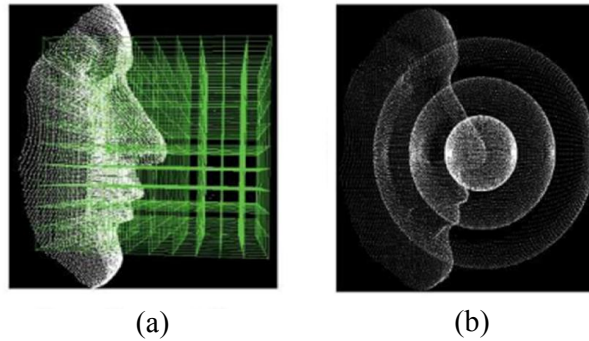


Fig. 2.6 Illustration of (a) tensor based face representation
(b) spherical face representation

(viii) Spherical Face Representation

Intuitively, Spherical Face Representation (SFR) [43] can be imagined as the quantization of the point cloud of a face into spherical bins centered at the nose tip. Fig. 2.6 (b) [43] graphically illustrates Spherical Face Representation of three bins. To compute an n bin Spherical Face Representation, the distance of all points from the origin is calculated. These distances are then quantized into a histogram of $n + 1$ bins. The outer most bin is then discarded since it is prone to errors (for example, due to hairs). The similarity between a probe and gallery face is computed by measuring the point wise Euclidean distance between their Spherical Face Representations. The advantage of Spherical Face Representation is its invariance to facial expressions. On the other hand it is sensitive to the probes with a neutral expression [43].

2.2.3 Region based Approaches

Local region based approaches are generally proposed to handle artifacts, wrinkles, facial hair and local shape deformations caused by facial expression variations. By the selection of multiple small regions on the face, any error caused by a single region can be compensated by fusing the matching scores from multiple regions, thus making the recognition more robust to such factors. This kind of approaches first divides the face

into a number of regions using facial landmarks. Then features from each region are extracted using holistic or local feature based approaches. Finally each region is classified independently and all the region based results are combined using some ensemble classification approach.

2.3 Discussion

Referring to section 2.1, the existing alignment approaches (Iterative Closest Point based alignment, Simulated Annealing based alignment, Average Face Model based alignment and Intrinsic Coordinate System based alignment) have limitations in terms of initial course alignment, excessive computational cost or the accuracy. In this dissertation, three novel 3D face alignment approaches are proposed to overcome the limitations of existing alignment approaches and they exhibit superior alignment results.

For face recognition task, holistic approaches are generally preferred since they preserve the configural information (i.e., the interrelations between facial parts) of the face, which is very important for preserving the identity of an individual as evidenced from psychological [44], neurobiological [45] [46] and computer vision [24] [47] communities.

Because the local feature based approaches concentrate on local facial features, the tendency of preserving the spatial arrangement of different facial parts (configural information) is largely compromised and limited to the local features based information contained in corresponding parts of the two images. These approaches are also sensitive to facial expressions and noise.

Local feature based techniques describe the whole face representation in such a way that a large data is characterized by few local features. From a strict general object recognition stand point, face is one class of object, and thus discriminating within this class requires subtle details of the image that discriminates it among other faces. Therefore, information carried by each pixel of an image is considered valuable which is inherently available in holistic representation.

Referring to section 2.2.1, Principal Component Analysis based holistic approach results in better recognition rates than Independent Component Analysis, Linear

Discriminant Analysis or Locality Preserving Projections; therefore, Principal Component Analysis is employed in this dissertation for face recognition. The performance of the Principal Component Analysis based holistic approach is augmented with (i) a region based fusion strategy as region based approaches are capable to handle facial hair, artifacts, wrinkles and local shape deformations caused by facial expression variations (ii) multi-view synthesized images to exploit real 3D information.

2.4 Summary

In this chapter, 3D face alignment techniques, namely, Iterative Closest Point, Simulated Annealing, Average Face Model and Intrinsic Coordinate System based alignment have been extensively reviewed and their limitations have been analyzed. For face recognition, two broad categories of approaches, namely, holistic and local feature based were elaborately discussed. For each category, the representative and most successful approaches were reviewed and analyzed. Holistic approaches have been investigated to be promising as they use configural information of the face. The leading holistic approaches included Principal Component analysis, Independent Component Analysis, Linear Discriminant Analysis, Locality Preserving Projections and Iterative Closest Point. The local feature based approaches have been studied to involve, accurate feature localization, and sensitivity to facial expressions or noisy face images. These approaches included Gabor Filters, Log-Gabor Filters, Local Binary Pattern, Scale Invariant Feature Transform, Spin Images, Point Signatures, Tensor Based Face Representation and Spherical Face Representation. Region ensemble based approaches were also discussed being capable of improving recognition accuracies of holistic or local feature based approaches in the presence of wrinkles, artifacts, facial hair and local shape deformations caused by facial expression variations.

Chapter 3

3D FACE ALIGNMENT FRAMEWORK

In this chapter, the proposed 3D face alignment algorithms are presented. The objective of this chapter is to describe the nose tip detection and facial alignment procedure with mathematical validation. Three different alignment algorithms are presented, which are capable of aligning neutral and expressive faces acquired at frontal and non-frontal poses whereas one of the algorithms is capable of aligning profile face images. A distance reduction measure is also proposed as a quantifying heuristic for the proposed alignment methods. At the end of the chapter, performance evaluation of the proposed algorithm is given using FRGC v2.0 and GavabDB face databases.

3.1 3D Face Alignment

Facial alignment transforms the facial features in such a way that they can be reliably matched. A few algorithms exist in literature [15], [12], [16] for aligning 3D face scans which can be broadly classified into three categories: (i) One-to-all alignment (ii) Alignment to Average Face Model (AFM) and (iii) Alignment to Intrinsic Coordinate System (ICS).

In one-to-all alignment, 3D point cloud of a probe face image is aligned to every face image in the gallery. Iterative Closest Point (ICP) alignment algorithm [12] falls into this category. In the ICP algorithm, the point clouds are aligned by minimizing distance between them iteratively. The distance between the point clouds is minimized by rotating and translating one of the point clouds relative to the other. The closest point between the surfaces is determined by finding distance from each point of the first point cloud to the second point cloud and averaging all distances. Since, in one-to-all alignment a face is aligned to every other face in the gallery, ICP is not suitable in face identification (one-to-many matches). On the other hand it is suitable in a verification scenario (one-to-one match). A limitation of ICP is that it needs

initial course alignment for convergence. Another disadvantage is the fact that it is relatively slow. ICP based approaches have been used in many leading research papers such as Mahoor and Abdel-Mottaleb [48] and Wang et al. [13].

Simulated Annealing (SA) is another alignment approach in one-to-all category which was implemented in Queirolo et al. [14]. SA is based on stochastic algorithm that is used for local search. Initially, centers of the mass of two faces are aligned. Then a course alignment is accomplished with an SA based searching approach. For course alignment an estimation measure is reduced based on M-estimator Sample and Consensus (MSAC) along with Mean Squared Error of the corresponding points of two faces to be compared. After that, a precise alignment is obtained based on a searching algorithm which uses Surface Interpenetration Measure (SIM) as the evaluation criterion. The drawback of SA based alignment is its computational time which is comparable to ICP.

For alignment to an AFM [15], AFM is constructed by locating landmarks on the face. Average of landmark coordinates is calculated followed by the Procrustes analysis [18]. Then transformed landmark points are re-averaged for getting an AFM. This method is suitable for both of face identification and verification scenarios. A considerable disadvantage of the AFM based alignment is its low accuracy rates because some of the spatial information is lost while creating AFM and a probe may be less accurately aligned to an AFM compared to aligning it directly to a true positive.

The third method, alignment to an ICS requires landmark localization on 3D scans. These landmarks are compared with the corresponding 3D points of the ICS. The resulting transformation is then applied to whole point cloud. This method greatly depends on landmarks which may be less accurately located in case of non-frontal and expressive faces. This issue has been discussed in detail in [16]. A study [49] presents an alignment method based on ICS by finding vertical symmetry plane of the face, slope of the nose bridge and the nose tip. These features can be called the landmark structures instead of landmarks, which mark position only. The choice of landmark structures is beneficial because they remain stable even under pose variations and facial expressions. These landmark structures define an ICS.

3.2 Proposed 3D Face Alignment Algorithms

In this section, three different 3D face alignment algorithms are presented, namely, 3D Face Alignment based on Intrinsic Coordinate System, Two-Pass 3D Face Alignment based on Minimum Distance, and Two-Pass 3D Face Alignment based on Classification Approach, as described in the following subsections.

3.2.1 3D Face Alignment based on Intrinsic Coordinate System

In this algorithm, face image are aligned using Intrinsic Coordinate System (ICS). For 3D face alignment, world coordinate system is represented by x , y and z -axis as shown in Fig. 3.1 with origin defined at nose tip O . The figure shows a model face [50] and ICS defined by a , b and c -axis at the same origin.

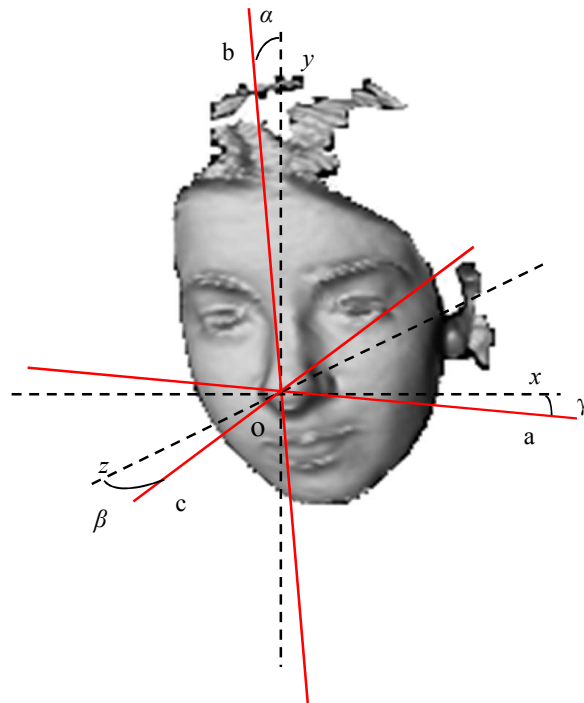


Fig. 3.1 World and intrinsic coordinate systems shown with angles α , β and γ in yz , xz and xy planes respectively for 3D face alignment

Three angles and an origin are necessary to define an ICS. Origin is the nose tip and

three angles defining the ICS are α , β and γ . Angle α is inscribed between nasal bridge slope represented by b-axis of ICS and y -axis of world coordinate system in yz plane. β is the angle between c -axis and z -axis of intrinsic and world coordinate systems respectively describing face rotation in xz plane. Similarly γ is the angle between a -axis and x -axis of intrinsic and world coordinate systems in xy plane. Angle α gives estimation of face rotation around x -axis, whereas angles β and γ contribute towards rotation estimation around y and z -axis respectively. After finding angles α , β and γ , each point cloud is rotated using these parameters and transformed so that its ICS axes become concurrent with the world coordinate system axes. This process brings 3D scans in frontal position. Due to stability of landmark structures, the proposed algorithm is robust for neutral and expressive 3D face images acquired at frontal and non-frontal poses.

(i) Face Localization and Nose Tip Detection

Face of an individual is localized and cropped by using nose tip heuristic. In the proposed methodology, first of all nose tip of 3D scans is found. Nose tip is the point which is nearest to the 3D scanner when scans are captured. While finding nearest point there exist a number of problems due to noise scenarios. In many subjects hairs on forehead come nearer to the scanner instead of nose. Similarly, in case of female subjects, hairs which are spread around neck or ears become nearer to scanner instead of nose.

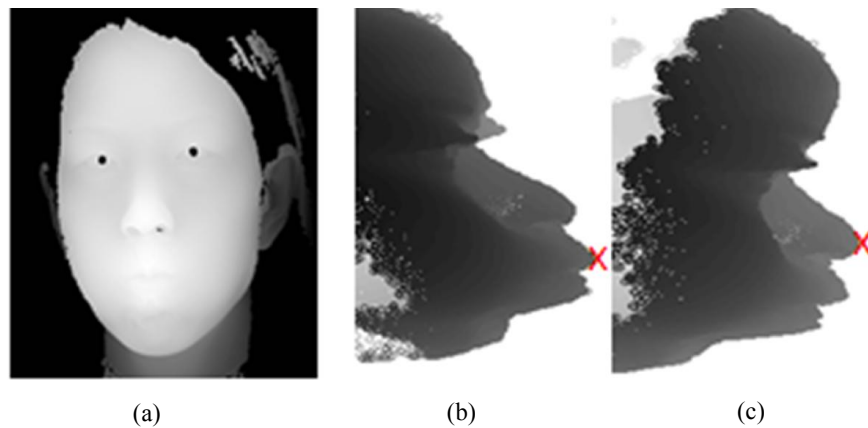


Fig. 3.2 Example puffy face (a) xy view (b) yz view (c) yz view at angle 15° showing nearest point from 3D scanner on lips instead of nose

In order to cope with such noise problems due to hair, the algorithm searches nose tip in an approximate Region of Interest (ROI). Similarly subjects having expression known as “puffy” in literature (as shown in Fig. 3.2(a) for subject 04202d564.abs from FRGC v2.0) present another problem. In such subjects the nearest point may lie on lips instead of nose as shown in Fig. 3.2(b). To deal with this issue, first of all each 3D point cloud is rotated around x-axis using rotation matrix given in equation 3.16 with $\alpha = 15^\circ$, $\beta = 0^\circ$ and $\gamma = 0^\circ$ for all scenarios. This process initially brings the nose tip nearest to the scanner (see Fig. 3.2(c)).

(ii) Correction Parameter Determination in xz Plane

The rotation parameter β is determined using horizontal nose plane of face. For finding this plane, x , y and z coordinates of nose tip at origin, $O(x, y, z)$, are determined. All values of x are found in xy plane at the nose tip by fixing y -coordinate of the 3D point cloud. This process extracts all feature points lying parallel to x -axis passing through the nose tip. These feature points constitute horizontal nose plane which passes through face region at nose tip horizontally as presented in Fig. 3.3 for subject 02463d550.abs from FRGC v2.0.

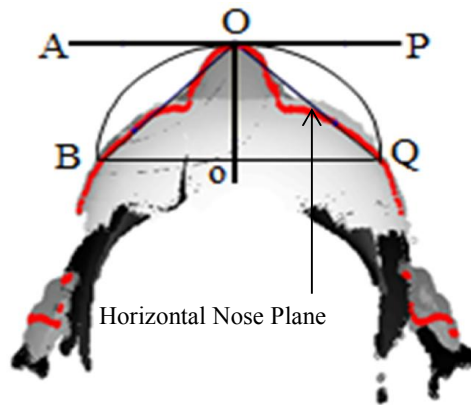


Fig. 3.3 Example range image showing horizontal nose plane

On the horizontal nose plane, where it touches the face, three points are found at equal distances, on both sides of nose. Averages of these points at both sides are taken separately to compensate for outliers and are labeled as B and Q in Fig. 3.3.

To find angle of rotation β for a 3D face around y -axis, $\angle AOB$ and $\angle POQ$ are determined by the algorithm. To describe significance of finding $\angle AOB$ and $\angle POQ$ in determining β , following mathematical reasoning is presented.

In Fig. 3.4, \overline{AP} defines x -axis while \overline{oO} defines z -axis. \overline{OB} and \overline{OQ} represent the lines joining origin and points B and Q which are determined by the proposed algorithm. Arc \widehat{BQ} is the path along which nose tip can move and point o represents center of the arc. \overline{BQ} is the line joining points B and Q through arc center o .

The arc \widehat{BQ} , radius \overline{oO} and angle β are related by the following equation.

$$s = r\beta \tag{3.1}$$

Where s is arc length, r is radius of the circle of which the arc is a part and β is the angle subtended by end points of the arc. Finding β gives

$$\beta = \frac{s}{r} \tag{3.2}$$

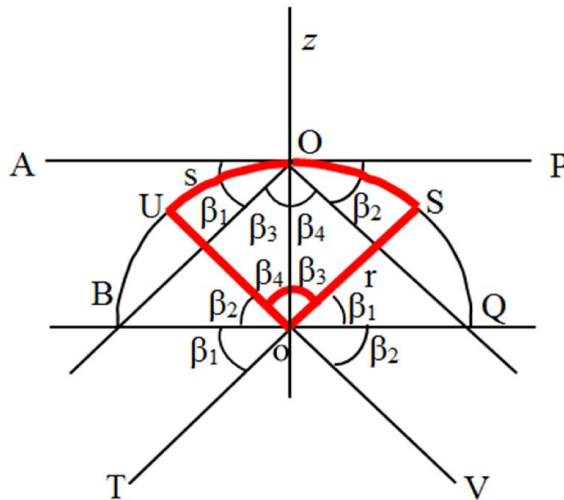


Fig. 3.4 Frontal view geometry of a face in xz plane

The algorithm uses the value of β in xz plane to align the nose tip with the point $x = 0^\circ$ where β becomes 0° .

Referring to Fig. 3.4, it can be observed that

$$\overline{AP} \parallel \overline{BQ} \quad (3.3)$$

where \parallel shows that \overline{AP} is parallel to \overline{BQ} .

$$\overline{OB} \parallel \overline{ST} \quad (3.4)$$

where \parallel that \overline{OB} is parallel to \overline{ST} , Therefore,

$$\angle AOB = \angle BoT = \angle SoQ = \beta_1 \quad (3.5)$$

Referring to equation 3.4

$$\overline{OQ} \parallel \overline{UV} \quad (3.6)$$

where \parallel shows that \overline{OQ} is parallel to \overline{UV} . Therefore,

$$\angle POQ = \angle QoV = \angle UoB = \beta_2 \quad (3.7)$$

As $\angle BoO$ and $\angle QoO$ are right angles, therefore,

$$\angle OoU = 90^\circ - \beta_2 = \beta_4 \quad (3.8)$$

$$\angle OoS = 90^\circ - \beta_1 = \beta_3 \quad (3.9)$$

Subtracting equation 3.9 from equation 3.8

$$\angle OoU - \angle OoS = \beta_1 - \beta_2 = \beta_4 - \beta_3 \quad (3.10)$$

After measuring angles, $\angle AOB$ and $\angle POQ$, their difference (hence difference of angles β_3 and β_4) is calculated and half of the value of this difference angle is the required angle β . The reason of taking half value lies in the fact that decreasing 5° from $\angle AOB$ increases 5° in $\angle POQ$ and vice-versa, e.g. if $\angle AOB = 45^\circ$ and $\angle POQ = 35^\circ$ then half of difference of angles turns out to be 5° . Thus decreasing 5° from $\angle AOB$ sets $\angle AOB = 40^\circ$ and an increase of 5° in $\angle POQ$ adjusts its value equal to 40° . To give the reader a better understanding of angle β , it is drawn in Fig. 3.5 between c-axis and z-axis of intrinsic and world coordinate systems respectively, where c-axis represents a rotated view and z-axis shows a frontal view. It is notable that minimizing the value

of angle β to 0° , folds c-axis of ICS to become concurrent with the z -axis of world coordinate system thus correcting the face pose in xz plane.

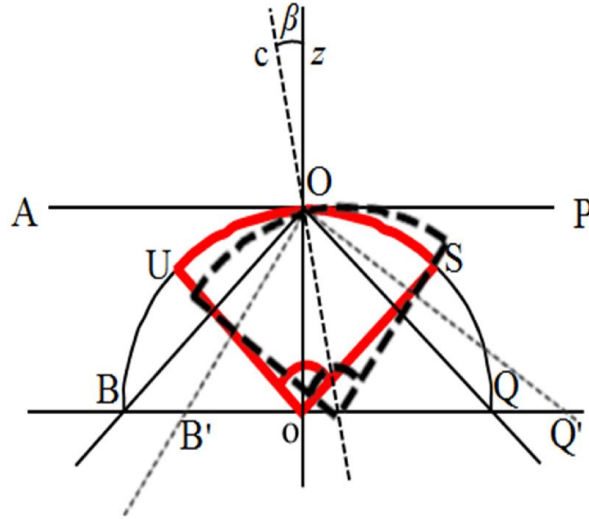


Fig. 3.5 Frontal and rotated view face geometry in xz plane presenting a face rotated at angle β

(iii) Correction Parameter Determination in yz Plane

The vertical symmetry plane is significant because it is used to find angle α . Angle α brings faces which are leaning forward or backward in a position so that they may be treated as frontal. The pivot for finding vertical symmetry plane is nose tip. At the nose tip, all points along y -axis are found by fixing x -coordinate of the 3D point cloud. This way all the points are found which have a fixed value of x but values of y are varying. By finding corresponding z -coordinates against all selected y -coordinates, vertical symmetry plane is constructed as shown in Fig. 3.6 for subject 02463d556.abs from FRGC v2.0.

To find α for 3D faces, 3D point cloud is plotted in yz plane as shown in Fig. 3.6, where $\overline{yy'}$ represents y -axis and $\overline{zz'}$ represents z -axis. From the nose tip a point N is determined on Nose Bridge at the vertical symmetry plane. Thus a central angle α , measured by the algorithm, is subtended by end points of arc l at the nose tip as

shown in Fig. 3.6.

Radius ρ , arc l and angle α are related as:

$$l = \rho\alpha \quad (3.11)$$

where l is arc length, ρ is radius of the arc and α is the angle inscribed by end points of the arc. Solving for α

$$\alpha = \frac{l}{\rho} \quad (3.12)$$

The algorithm determines value of angle α to adjust the head tilt at 30° because this places the face in an upright position, resulting in a frontal view [49].

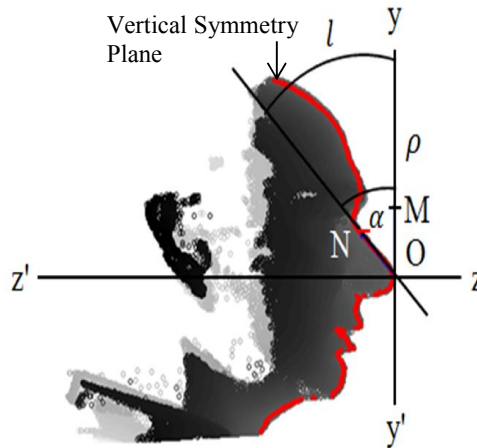


Fig. 3.6 Example range image in yz plane showing vertical symmetry plane to determine angle α

(iv) Correction Parameter Determination in xy Plane

The third angle γ rotates the face in xy plane. This angle is measured by localizing landmarks on inner eye corners of the face as shown in Fig. 3.7 for subject 04201d374.abs from FRGC v2.0. A line $A'A$ is drawn to join landmarks and the line $a'a$ is drawn parallel to $A'A$ which passes through the determined nose tip. This

process defines angle $\gamma = xOa$ with x -axis of world coordinate system to be used as correction parameter in xy plane.

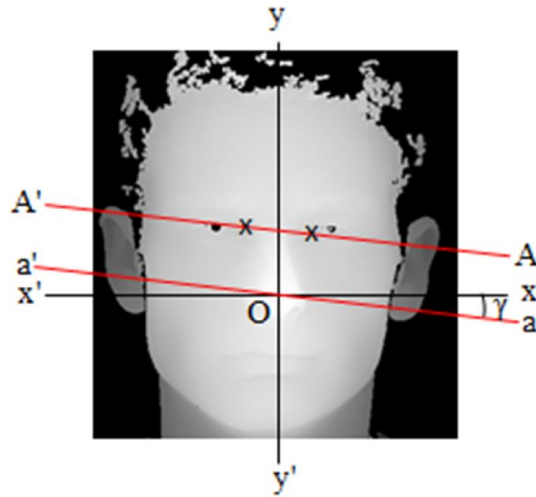


Fig. 3.7 Example range image in xy plane presenting determination of angle γ

The 3D pose of an individual is corrected by substituting values of estimated rotation parameters α , β and γ in rotation matrix [51] given in equation 3.16. The process is repeated until convergence is reached. i.e. angles α becomes 30° while angles β and γ achieve a value equal to zero.

$$A = \begin{bmatrix} \cos \beta \cos \gamma & -\cos \alpha \sin \gamma + \sin \alpha \sin \beta \cos \gamma & \sin \alpha \sin \gamma + \cos \alpha \sin \beta \cos \gamma \\ \cos \beta \sin \gamma & \cos \alpha \cos \gamma + \sin \alpha \sin \beta \sin \gamma & -\sin \alpha \cos \gamma + \cos \alpha \sin \beta \sin \gamma \\ -\sin \beta & \sin \alpha \cos \beta & \cos \alpha \cos \beta \end{bmatrix}$$

(3.13)

The pseudo-code of the alignment approach is given as under.

Pseudo-code 1: 3D Face Alignment based on Intrinsic Coordinate System

BEGIN

I= Image;

Define Nose tip at $\min(z(i, j))$;

FOR xz plane

REPEAT

Compute $Q = \frac{q_1 + q_2 + q_3}{3}$


```

Compute     $B = \frac{b_1 + b_2 + b_3}{3}$ 
Compute     $\beta = \frac{\angle AOB - \angle POQ}{2}$ 
IF  $\beta = 0^\circ$ 
    GOTO END;
ELSE
    I= Rotate I at  $\beta^\circ$  in  $xz$  plane;
ENDIF
UNTIL  $\beta = 0^\circ$ ;
END.
FOR  $yz$  plane
    Search point  $N$  at Nose Bridge on vertical symmetry plane
    REPEAT
        Compute  $\angle \alpha = \text{MON}$ ;
        IF  $\alpha = 30^\circ$ 
            GOTO END;
        ELSE
            I=Rotate I at  $\alpha - 30^\circ$  in  $yz$  plane;
        ENDIF
    UNTIL  $\alpha = 30^\circ$ ;
END.
FOR  $xy$  plane
    Detect landmarks on inner eye corners;
    REPEAT
        WRITE line  $A'A$  joining landmarks and line  $a'a$  ||  $A'A$  passing through
        the nose tip  $O$ ;
        Compute  $\angle \gamma = \text{Xoa}$ ;
        IF  $\gamma = 0^\circ$ ;
            GOTO END;
        ELSE

```

```

        Rotate I  $\gamma^0$  in  $xy$  plane;
    ENDIF;
    UNTIL  $\gamma = 0^\circ$ ;
END.
    Align image I using determined values of  $\alpha$ ,  $\beta$  and  $\gamma$ ;
END.

```

3.2.2 Two-Pass 3D Face Alignment based on Minimum Distance

In this section a novel two-pass 3D alignment algorithm is presented which is capable of aligning frontal and non-frontal face images along with profile images. The two-pass alignment algorithm is based on Intrinsic Coordinate System (ICS) and minimum nose tip-scanner distance. The ICS is shown by u , v and w -axis and world coordinate system is represented by x , y and z -axis in Fig. 3.8 along with a 3D model of subject ‘cara17_frontal1’ from GavabDB, whereas l shows the minimum nose tip-scanner distance.

(i) Nose Tip Detection

The nearest point in a capture from 3D scanner is considered as nose tip by the alignment algorithm which is used to localize and crop a subject’s face. In this section, a novel algorithm is presented to tackle issues related to nose tip detection process for 3D faces with profile views because in such faces, the nearest point from 3D scanner lies on ears or some other facial part as shown in Fig. 3.9 (a) for subject cara6_frontal1.

Referring to Fig. 3.9, the nose tip detection algorithm starts by computing differences $f_x = x_{max} - x_{min}$ and $f_z = z_{max} - z_{min}$ which define the length (L) and width (W) of the face in profile view (Fig. 3.9 (a)) while W and L of the face in frontal view (Fig. 3.9 (b)), along x and z -axis respectively. Then a function $f = f_x / f_z$ is computed which turns out to be a higher numeric value (when compared to an

empirically determined threshold) for profile faces and lower for remaining 3D faces. The lower value of denominator in the former case corresponding to width of half face is regarded as the reason for higher numeric value returned by function f . Based on computed value of function f , the profile faces are separated from rest of the 3D faces. In order to determine whether the profile of probe face is right or left, nose templates of both right and left orientations are correlated with the profile face. The profile face is classified as right or left based on greater value of maximum correlation coefficient using Normalized Cross Correlation (NCC). Because right and left profile faces are not acquired at exactly $+90^\circ$ or -90° , they are rotated in the range of 0° to -90° and $+90^\circ$, respectively, with a step size of 2° to determine the nose tip; and the point on the face at a minimum distance from the 3D scanner is marked as nose tip by the algorithm.

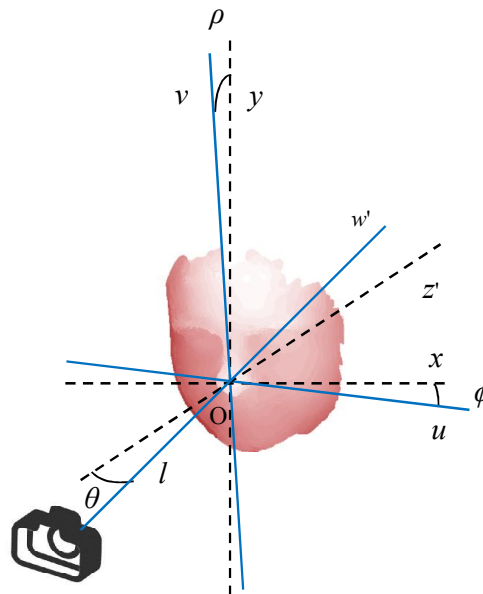


Fig. 3.8 Intrinsic and world coordinate systems along with example image

The robustness of nose tip detection process is estimated by measuring two features which are depth map variance and highest peak in the histogram of depth image. If the face pose is frontal, and the nose tip is correctly detected; the depth map variance is

minimum and the histogram peak is maximum. The minimum depth map variance and maximum peak are determined from nine facial positions separately which are obtained by rotating the face along x -axis at 0° , 60° and -60° while fixing the rotation along y -axis at 0° , 60° and -60° in respective order for each of the aforementioned rotations along x -axis. The face with minimum depth map variance and the maximum histogram peak of all nine faces is used to find the facial landmark at a minimum distance from the scanner and is declared as the nose tip. The above described complete process involves 4 landmarks to classify profile images from rest of the 3D faces.

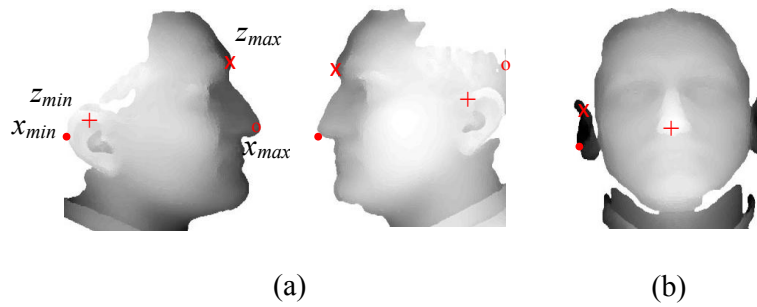


Fig. 3.9 Minimum and maximum nose-tip scanner distances for example subject from GavabDB database (a) profile images (b) frontal image

The proposed alignment approach employs 4 landmarks to find the rotation parameters for 3D faces excluding profile images: nose tip, 2 landmarks in xz plane and 1 in yz plane whereas for profile images, the rotation parameters are determined by employing 2 landmarks: nose tip along with 1 landmark in yz plane as under.

(ii) Alignment in xz Plane

In the first pass, the two-pass alignment algorithm aligns the 3D faces excluding profile images in xz plane by finding the correction parameter θ using the intrinsic coordinate system based alignment algorithm whereas the profile images become in frontal view after the nose tip detection process.

The second pass of the alignment algorithm confirms the alignment accuracy of the proposed two-pass alignment algorithm using the approach depicted in Fig. 3.10 for

subject 02463d550.abs from FRGC v2.0 where it rotates the 3D faces in the range of -2° to $+2^\circ$ with a step size of 0.1° in xz plane and aligns them at a position where the nose tip-scanner distance is the minimum. The rationale behind this strategy is discussed as follows. When a subject's nose tip is not aligned with line joining z -axis of 3D scanner and nose tip of 3D face, it has a larger distance measured from scanner to the nose tip which is reduced when the subject's nose tip is aligned with the z -axis line. The facial pose correction from non-frontal to frontal view causes an essential reduction in the mentioned distance. This statement means that distance l must be smaller than l_1 or l_2 in Fig. 3.10. In order to verify this fact, it can be observed from Fig. 3.10 that the distance l between the point $C(p_0, q_0)$ and the nose tip $O(p_1, q_1)$ is a perpendicular to the line $x'x$ among all nose tip distances such as l_1 and l_2 . The perpendicular distance from a point on a line is always the shortest distance which leads to the conclusion that l is smaller than l_1 or l_2 . Using distance formula between two points, distance between 3D scanner point $C(p_0, q_0)$ and nose tip point $O(p_1, q_1)$ of an aligned face is given as in equation 3.14.

$$l = \sqrt{(p_0 - p_1)^2 + (q_0 - q_1)^2} \quad (3.14)$$

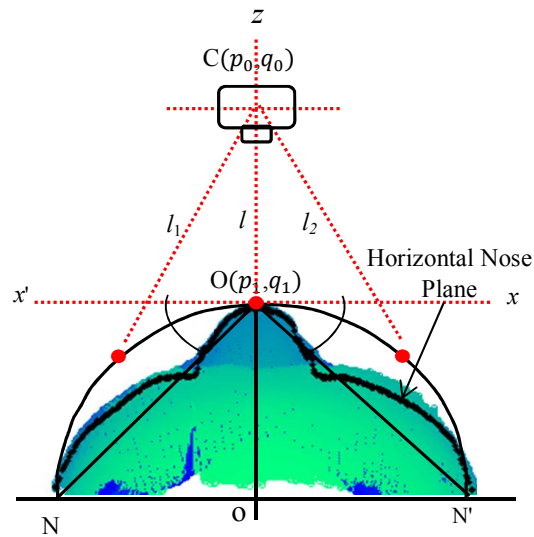


Fig. 3.10 Example depth image in xz plane showing two-pass alignment algorithm

(iii) Alignment in yz Plane

In the first pass, the two-pass alignment algorithm aligns all the 3D faces in yz plane by finding the correction parameter ρ using the intrinsic coordinate system based alignment procedure. The second alignment pass confirms the alignment accuracy of the proposed algorithm by rotating the 3D faces in yz plane in the range of -2° to $+2^\circ$ with a step size of 0.1° and aligns them at a position where the nose tip-scanner distance is the minimum as shown in Fig. 3.11 for subject 02463d556.abs from FRGC v2.0.

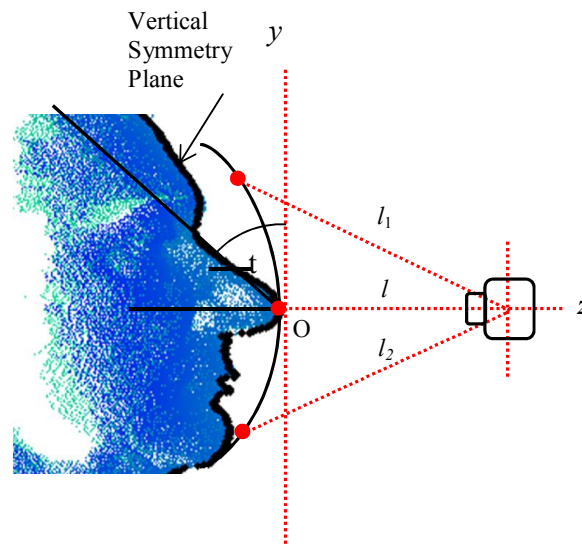


Fig. 3.11 Example depth image in yz plane showing two-pass alignment algorithm

(iv) Alignment in xy Plane

After attaining alignment in xz and yz planes, the algorithm utilizes half face in xy plane using vertical symmetry plane which was determined in yz plane alignment algorithm. The procedure for finding correction parameter ϕ in xy plane for 3D faces excluding profile images is given in the following steps.

1. Rotate right and left halves of face around z -axis in opposite directions up to $\pm 2^\circ$ with a step size of $\pm 0.25^\circ$.
2. For each rotation of step 1 above, mirror any of the half faces in xy plane and shift along the other half so that both halves are fully overlapped.

3. Find sum of differences, $D_{z(i,j)}$, for pixels at same grid position (i,j) of overlapping half faces, considering those values that are less than a threshold T (to rule out outliers introduced in z -direction during face scanning process) such that

$$S = \sum_{i,j} \begin{cases} 0 & D_{z(i,j)} > T \\ D_{z(i,j)} & otherwise \end{cases} \quad (3.15)$$

4. Select the angle at which the value of S in equation 3.15 becomes minimum (representing a good match). This angle is the required correction parameter ϕ in xy plane.
5. Rotate the face using 3D rotation matrix given in equation 3.13 by substituting value of correction parameter ϕ and fixing the values of ρ and θ at 0° . This process aligns the 3D faces in xy plane.

The profile face images, after pose correction in xz and yz planes, are developed into half faces in xy plane. For implementation of proposed pose correction procedure for these images in xy plane, the other half face of a probe image is synthesized by Average Face Model (AFM) using neutral faces of the same identity from the database and the procedure outlined above for 3D faces excluding profile images is adopted for finding correction parameter ϕ .

The pseudo-code of the alignment approach is given as below.

Pseudo-code 2: Two-Pass 3D Face Alignment based on Minimum Distance

BEGIN

I=Image;

Compute $f = \frac{x_{max} - x_{min}}{z_{max} - z_{min}}$

IF $f > threshold$

 Classify profile of I as right or left;

 Define Nose tip at $min(z(i,j))$;

ELSE

 Define Nose tip at $min(z(i,j))$;

ENDIF

```

FOR xz and yz planes
    Coarsely align image I using algorithm 1;
END.
Rotate the image I in the range  $-2^\circ$  to  $2^\circ$  with step size of  $0.1^\circ$ 
    FOR each rotation
        Compute  $\min(z(i, j))$ ;
    END.
Align image I at  $\min(z(i, j))$ ;
FOR xy plane
    IF  $f < \text{threshold}$ 
        Extract half faces
        REPEAT
    MARK:      Rotate half faces in opposite directions up to  $\gamma = \pm 2^\circ$  with a
                step size of  $\pm 0.25^\circ$ ;
                Find  $S$  by subtracting pixel values of first half face from
                mirror image of the second one;
        UNTIL  $S$  becomes minimum;
        Write  $\gamma$ ;
    ELSE
        WRITE missing Average Face Model (AFM);
        GOTO MARK;
    ENDIF
Align image I at determined value of  $\gamma$ ;
END.
END.

```

3.2.3 Two-Pass 3D Face Alignment based on Classification Approach

(i) Alignment in *xz* and *yz* Planes

In this alignment algorithm, a novel Support Vector Machine (SVM) based classification approach is employed to align the facial images in *xz* and *yz* planes where SVM classifier exhibited best experimental results. In order to reduce computational complexity of the algorithm, first of all an initial coarse alignment is obtained by using Intrinsic Coordinate System based alignment approach explained in section 3.2.1. Subsequently SVM classifier is used to classify the best alignment position of the face image as given below.

SVM Classifier

Intuitively, an SVM model is a representation of classes as points in a higher dimensional space, mapped so that their separation could be easier. The SVM classifier defines a decision surface to maximize the distance to the closest points named as support vectors in the training phase. For a binary classification problem, a hyper plane $w_x + b = 0$ having maximum margins, termed as optimal separating hyper plane (OSH) separates training vectors of two classes $(x_1, y_1), \dots, (x_i, y_i)$ where $x_i \in R_n$ and $y_i \in \{1, -1\}$. The objective function of the form given in equation 5.1 is minimized to obtain the OSH with constraints $y_i[(w \cdot x_i) + b] \geq 1 - \xi_i, \xi_i \geq 0$ for $i = 1, \dots, k$.

$$\emptyset(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^k \xi_i \quad (3.16)$$

where ξ_i are slack variables introduced to penalize errors if the data are not linearly separable and C is the regularization constant. Now sign of following OSH decision surface function can be used to classify a test point.

$$f(x) = \sum_{i=1}^k y_i a_i K(x, x_i) + b \quad (3.17)$$

where $a_i \geq 0$ are corresponding support vectors Lagrangian multipliers, b is determined by above mentioned optimization problem, K is the kernel trick used to transform non separable data onto a higher dimensional space where it becomes linearly separable by an hyper plane, x_i is the i th training sample and x is test sample. Radial basis function (RBF) kernel is employed in this study which is of the form as given below.

$$K(x, y) = \exp \left[\frac{-\|x - y\|^2}{\sigma^2} \right] \quad (3.18)$$

where σ^2 is spread of RBF.

The proposed alignment algorithm is employed using a set of four features. Among

these, two features are minimum distance from the nose tip to the 3D scanner in xz and yz planes separately, while the other two are maximum variance of the face image in xz and yz planes, separately. After attaining the initial coarse alignment, face images are rotated in xz and yz planes with a step size of 0.1° and minimum distance from nose tip to 3D scanner and maximum variance of the face images are computed for each rotated position. This entire process results into availability of 4 training vectors which are used to train the SVM classifier in a 4 dimensional space, comprised of genuine and imposter scores. It is worth mentioning to note that genuine scores represent aligned face images, while imposter scores represent misaligned face images.

For a probe face image, similar procedure is adopted to calculate feature vector and the resulting feature vector is used to classify (align) the probe image. The rotated face image corresponding to genuine score is thus regarded as the aligned face image.

(ii) Alignment in xy Plane

In xy plane face image is aligned using the variance of x feature of the 3D data which is maximum when the face is in frontal position. In this method, the face image is rotated in the range of $\pm 2^\circ$ with a step size of 0.1° and variance of the x feature for each of right and left half face is determined. The face image is aligned at a point where the variances of left and right half faces become equal.

The pseudo-code of the alignment approach is given as under.

Pseudo-code 3: Two-Pass 3D Face Alignment based on Classification Approach

BEGIN

I=Image;

FOR xz and yz planes

Coarsely align image I using algorithm 1; $f_x = x_{max} - x_{min}$

END.

Rotate the image I in the range -2° to 2° with step size of 0.1°

FOR each rotation

Compute $min(z(i, j))$;

Compute $\max(\text{variance})$;

END.

Train SVM classifier using minimum distances and maximum variances;

Classify (align) the probe image using classification approach;

FOR xy plane

Compute variance of x dimension for each of half face;

Align the image at equal values of variances;

END.

END.

3.3 Experiments and Results

Referring to section A.1 and A.2 where characteristics of FRGC v2.0 and GavabDB databases are described, experiments were performed to evaluate the performance of proposed alignment algorithms using both databases. Three alignment experiments were conducted in total using the proposed algorithms, namely, 3D face alignment based on intrinsic coordinate system, two-pass 3D face alignment based on minimum distance and two-pass 3D face alignment based on classification approach. For performance evaluation of the proposed alignment algorithms, a novel method is presented to verify accuracy of the alignment, as described below.

Referring to Fig. 3.10 and 3.11 when a subject's nose tip is not aligned with the line joining z -axis of 3D scanner and nose of 3D face scan, it has a larger distance measured from the scanner to the nose tip. On the other hand when the subject's nose tip is aligned with the z -axis line, its distance from the scanner is reduced. In other words, when pose of a face is corrected from non-frontal to frontal position, its nose tip distance from the scanner is essentially reduced. Based on distance reduction measure, the proposed algorithms achieved 99.95% and 99.77% alignment accuracy using FRGC v2.0 and GavabDB databases respectively. The nose tip was not detectable for two subjects in the FRGC v2.0 database and one subject in GavabDB database else the obtained alignment accuracy would have been 100%. The subjects in the face databases carry almost all types of facial variations like non frontal pose,

several types of facial expressions, varying distance of subjects from the scanner and occlusions.

The plots of normalized distances before and after applying the proposed alignment algorithms for first 20 scans from FRGC v2.0 (subjects 02463d546.abs to 02463d672.abs) and GavabDB database (subjects cara1_abajo to cara3_arriba) are shown in Fig. 3.12 and Fig. 3.13 respectively. Please note in Fig. 3.13 that subject 3, 7, 12 and 16 are profile face images aligned using ‘two-pass 3D face alignment based on minimum distance’ algorithm only. It can be observed from the plots that ‘two-pass 3D face alignment based on classification approach’ achieved best alignment accuracy i.e. the maximum reduction of distance of nose tip from the scanner is achieved whereas ‘two-pass 3D face alignment based on minimum distance’ and ‘3D face alignment based on intrinsic coordinate system’ obtained relatively lower alignment accuracies respectively. Original as well as pose corrected range images of the subjects 04343d427.abs, 04385d435.abs, cara4_abajo and cara9_arriba aligned using ‘3D face alignment based on intrinsic coordinate system’ are given in Fig. 3.14 from left to right. Similarly original as well as pose corrected range images of the subjects 04217d399.abs, 04233d396.abs, 04221d553.abs, 04482d418.abs and 04387d322.abs from FRGC v2.0 and images cara1_abajo, cara1_arriba,

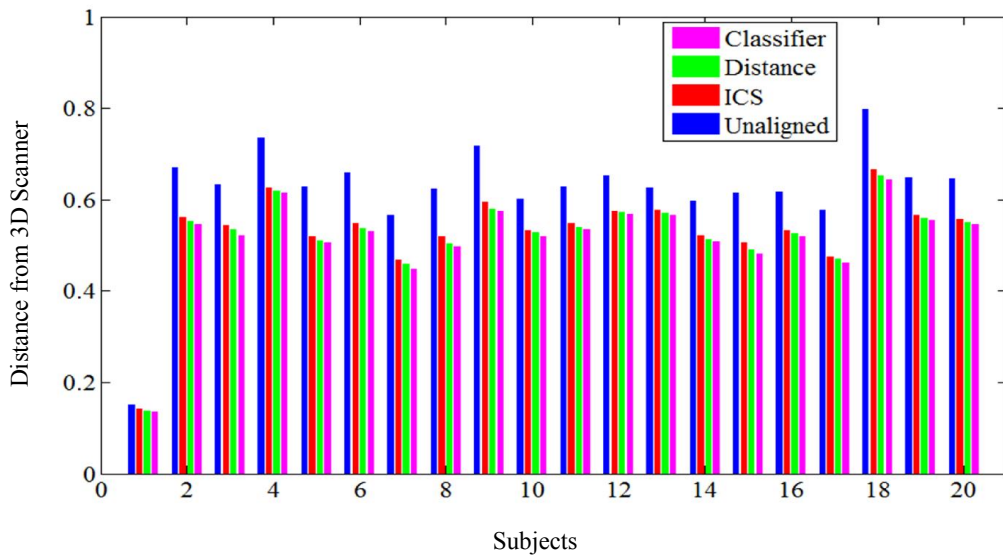


Fig. 3.12 Distance of nose tip from scanner for first 20 scans from FRGC v2.0

caral_frontal1, caral_frontal2, caral_derecha, caral_izquierda, caral_gesto, caral_risa and caral_sonrisa from GavabDB are given in Fig. 3.15 (a) and (b) respectively from top left to right, aligned using ‘two-pass 3D face alignment based on minimum distance’. Fig. 3.16 shows original as well as pose corrected range images of the subjects 04595d149.abs, 04724d146.abs, cara13_frontal1 and cara36_frontal2 from left to right aligned using ‘two-pass 3D face alignment based on classification approach’.

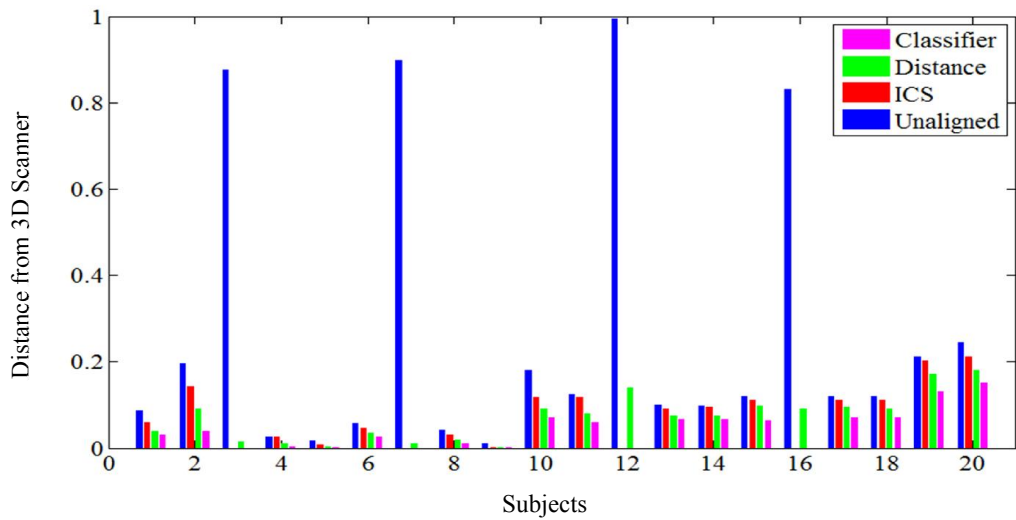


Fig. 3.13 Distance of nose tip from scanner for first 20 scans from GavabDB

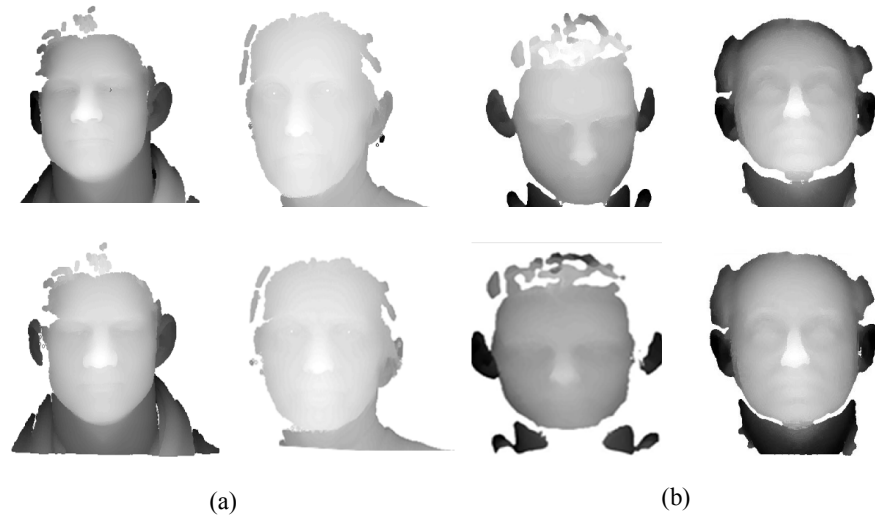
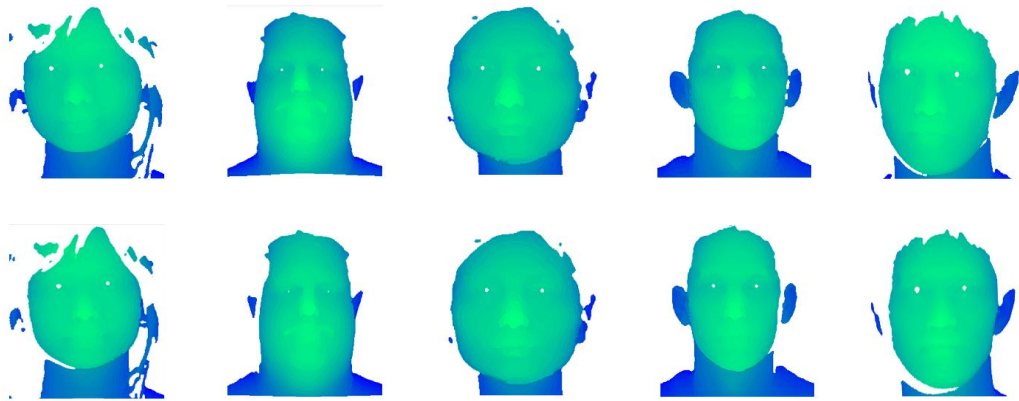
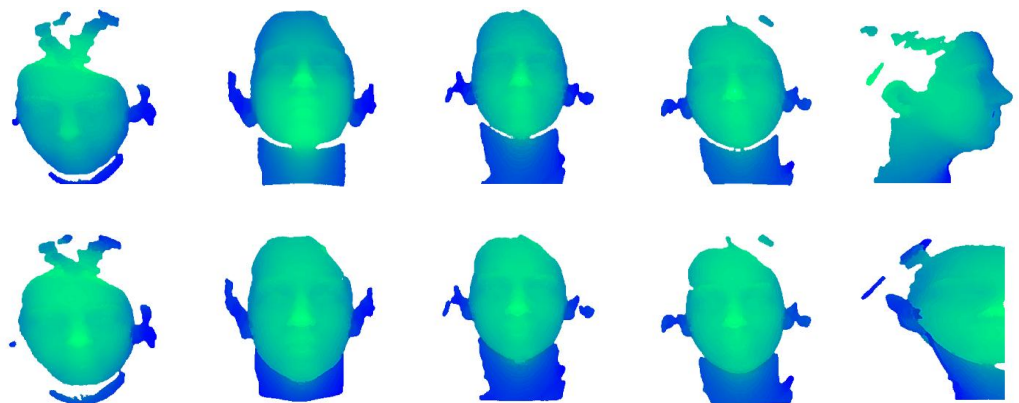


Fig. 3.14 Example 3D scans aligned using 3D face alignment based on intrinsic coordinate system: original (top row) and pose corrected (bottom row) from (a) FRGC v2.0 (b) GavabDB



(a)



(b)

Fig. 3.15 Example 3D scans aligned using two-pass 3D face alignment based on minimum distance: original (row 1, 3, 5) and pose corrected (row 2, 4, 6) from (a) FRGC v2.0 (b) GavabDB

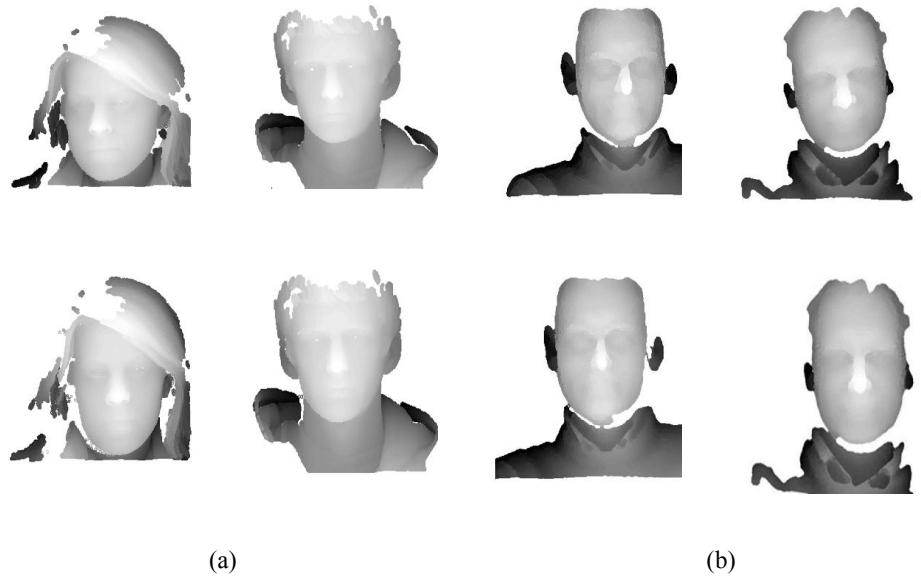


Fig. 3.16 Example 3D scans aligned using two-pass 3D face alignment based on classification approach: original (top row) and pose corrected (bottom row) from (a) FRGC v2.0 (b) GavabDB

3.3.1 Computational Complexity Analysis

Computational complexity analysis of the proposed algorithms considering computationally intensive operations is given in Table 3.1 for point cloud of size $M = 300000$ points. In order to rotate a single 3D point, 9 multiplication operations and 6 addition operations are involved. The multiplication factor of 82 in the table accounts for 82 rotations in the range of -2° to $+2^\circ$ (with a step size of 0.1°) in xz and yz planes whereas multiplication factor of 46 accounts for 46 rotations in the range of 0° to $+90^\circ$ or -90° (with a step size of 2°) in xz plane. The multiplication factor of 147 accounts for 147 support vectors whereas that of 4 and 7 represents multiplication and addition operations for 4 dimensional feature vectors respectively. In terms of big O notation, the computational complexity is of the order of $O(P)$, where P represents total number of points in the point cloud.

Table 3.1 Computational complexity analysis of the proposed algorithms

Operation	Alignment Steps	3D Face Alignment based On Intrinsic Coordinate System	Two-Pass 3D Face Alignment based On Minimum Distance		Two-Pass 3D Face Alignment based On Classification Approach
			Frontal and Non-Frontal Face Images	Profile Face Images	
	Nose tip Detection			$M \times (46 \times 9)$	
Multiplications	Puffy face	$M \times 9$	$M \times 9$		$M \times 9$
	Alignment	$M \times 9$	$M \times 9$		$M \times 9$
			$M \times (82 \times 9)$	$M \times (82 \times 9)$	$M \times (82 \times 9)$
					147×4
	Total	=5400000	=226800000	=345600000	=226800588
Additions	Nose tip Detection	$M \times 1$	$M \times 1$	$M \times (46 \times 6)$	$M \times 1$
	Puffy face	$M \times 6$	$M \times 6$		$M \times 6$
	Alignment	$M \times 6$	$M \times 6$		$M \times 6$
			$M \times (82 \times 6)$	$M \times (82 \times 6)$	$M \times (82 \times 6)$
					147×7
Total	=3900000	=151500000	=230400000	=151501029	

3.4 Summary

In this chapter, three novel 3D alignment algorithms have been proposed to align neutral and expressive faces acquired at frontal and non-frontal poses. ‘3D face alignment based on intrinsic coordinate system’ was based on intrinsic coordinate

system which employed nose tip, vertical symmetry plane and the slope of the nose bridge in the alignment process. The ‘two-pass 3D face alignment based on minimum distance’ aligned the face images using two alignment passes. The first pass coarsely aligned the face images including profile images to reduce computational cost and second pass finely aligned them using a minimum distance algorithm. The ‘two-pass 3D face alignment based on classification approach’ also employed a coarse to fine approach where first pass coarsely aligned face images to reduce computational cost and second pass aligned them finely using a classification based approach. The proposed algorithms successfully aligned neutral and expressive faces acquired at frontal and non-frontal poses whereas ‘two-pass 3D face alignment based on minimum distance’ also aligned profile faces successfully. For performance evaluation of the proposed algorithms, experiments were conducted using FRGC v2.0 and GavabDB databases. Based on a distance reduction measure, ‘two-pass 3D face alignment based on classification approach’ obtained the best alignment accuracy than ‘two-pass 3D face alignment based on minimum distance’ and ‘3D face alignment based on intrinsic coordinate system’ in the respective order. On the other hand ‘3D face alignment based on intrinsic coordinate system’ was computationally inexpensive than ‘two-pass 3D face alignment based on minimum distance’ and ‘two-pass 3D face alignment based on classification approach’.

Chapter 4

RECOGNITION BASED ON REGIONAL SEGMENTS OF DEPTH IMAGES

The goal of this chapter is to present 3D face recognition methodology using intrinsic coordinate system based alignment to deal with varying poses between probe and gallery images. In the beginning, properties of the proposed approach are pointed out. Subsequently, the details of the proposed face recognition algorithm are presented which employs a region based methodology using two-tier ensemble classification approach based on weighted Borda Count method and re-ranking strategy. The performance of the proposed algorithm is evaluated using FRGC v2.0 and GavabDB databases in terms of recognition rates and computational complexity analysis.

4.1 3D Face Recognition

The main algorithms evolved for 3D face recognition employ holistic and local feature based approaches. In general, the advantages and limitations of these approaches are discussed in section 1.4. In order to overcome the limitations of the existing approaches, the proposed region based face recognition algorithm, which is robust to artifacts, wrinkles, facial hair or expression variations [25], employs a two-tier ensemble classification approach because ensemble approaches greatly improve the performance of individual classifiers [52]. The proposed approach employs Principal Component Analysis (PCA) (given in section 2.2.1) for feature extraction and Mahalanobis Cosine (MahCos) distance, Euclidean distance, Mahalanobis (Mah) distance and Manhattan distance for classification of facial regions in different experiments. PCA is a subspace method which provides dimensionality reduction while relying on a set of basis vectors which correspond to maximum variance direction of the image data [27]. The use of PCA is beneficial because it is easy to implement and produces excellent results along with MahCos distance metric [27].

The studies Alyüz et al. [53], Spreeuwers [49] and Wang et al. [13] achieved good

performance in 3D face recognition by using a region based fusion approach considering 40, 60 and 6 face regions respectively. The proposed ensemble classifier based approach takes into account 15 overlapping regions composed of three sets of five regions each and it targets to explore contribution of overlapping regions towards providing complementary information using fusion. It is observed that combining results of multiple overlapping regions using an ensemble classifier produces excellent recognition results.

For performance evaluation of the proposed face alignment and recognition method FRGC v2.0 database [54] comprising of 4007 images from 466 subjects and GavabDB database [50] consisting of 549 images from 61 subjects; both with pose and expression variations are utilized which are the most commonly used databases of 3D face images employed for face recognition. The recognition results of the proposed method are comparable to the best reported results of state-of-the-art studies [55], [48], [13], [56], whereas it is computationally inexpensive.

4.2. Proposed 3D Face Recognition Algorithm

The framework of the proposed 3D face recognition algorithm is presented in Fig. 4.1. The intrinsic coordinate system based alignment step for the 3D images is explained in detail in the previous chapter and aligned range images are used in the face recognition process.

The range images have some artifacts like noise and gaps as shown in Fig. 4.2 for subjects 04217d461.abs and 02463d654.abs from FRGC v2.0, so they are preprocessed before their classification. First of all median filter is applied for removal of sharp spikes which occur during the scanning process. The median filter is a simple nonlinear smoothing filter that can suppress noise while retaining sharp sustained changes (edges) in signal values. It is particularly effective in reducing impulsive-type noise. The output of median filter at a point is the median value of the input data inside the window centered at the point. If $\{x(k)|1 \leq k \leq L\}$ and $\{y(k)|1 \leq k \leq L\}$ respectively represent the input and output of the 1D median filter of window size $2N + 1$, then

$$y(k) = \text{med}\{x(k - N), \dots, x(k - 1), x(k), x(k + 1), \dots, x(k + N)\} \quad (4.1)$$

In order to account for start up and end effect, $x(1)$ and $x(L)$ are repeated N times at the beginning and at the end of the input respectively. After median filtering, interpolation is employed to fill the holes on the face. Interpolation is the estimation of a value of an unknown function f within two points where value of f is known in a sequence of values. For two points $f(x_1, y_1)$ and $f(x_2, y_2)$, an unknown value $f(x, y)$ can be estimated as.

$$f(x, y) = \frac{x_2 - x}{x_2 - x_1} f(x_1, y_1) + \frac{x - x_1}{x_2 - x_1} f(x_2, y_2) \quad (4.2)$$

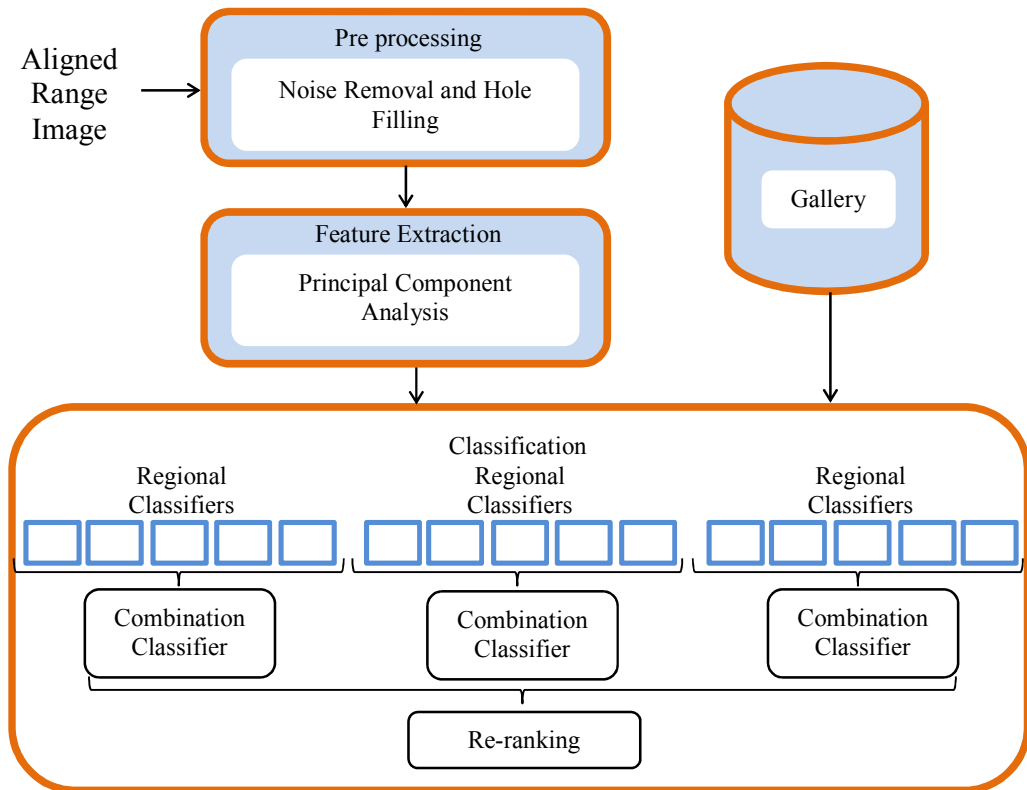


Fig. 4.1 Block diagram of the proposed 3D face recognition algorithm

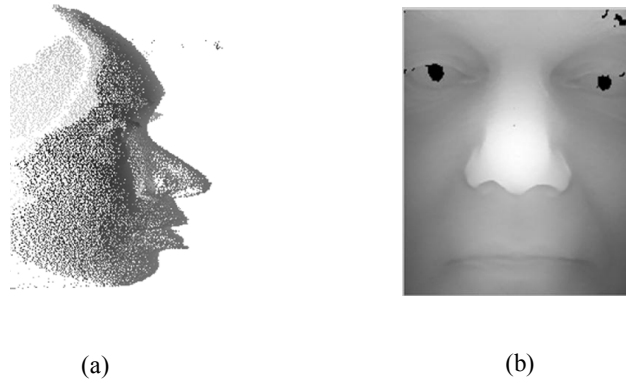


Fig. 4.2 3D scan showing (a) spikes (b) holes

4.2.1 Region Creation

By the selection of multiple small regions on the face, any error caused by a single region can be compensated by fusing the matching scores from multiple regions, thus making the recognition more robust to artifacts, wrinkles, facial hair or expression variations [57]. Similar methodology has been employed by Spreeuwers [49] and Alyüz [15]. It has been deduced in this study that small sized regions produce low recognition rates and large sized overlapping regions yield high recognition rates. Large and overlapping regions have also been used in Spreeuwers [49]; therefore, fusion of many relatively large overlapping regions was explored. The regions were selected in such a way that they would allow stable features to compare different types of local variations. Such regions are cropped by leaving out left or right side of the face, which are less visible under large rotations around the y-axis. Other cropped regions leave out upper or lower part of the face because of variations in caps, hair, or mouth expressions etc. The face regions used in this study are shown in Fig. 4.3 where the five regions in first row are defined as set S_1 while five regions in second and third row are defined as set S_2 and S_3 respectively. White areas show the facial regions that are included in the experiments. The region assemblage densely covers the whole face.

4.2.2 Regional Classifiers

The cropped face regions shown in Fig. 4.3 are classified by using the following four distance metrics.

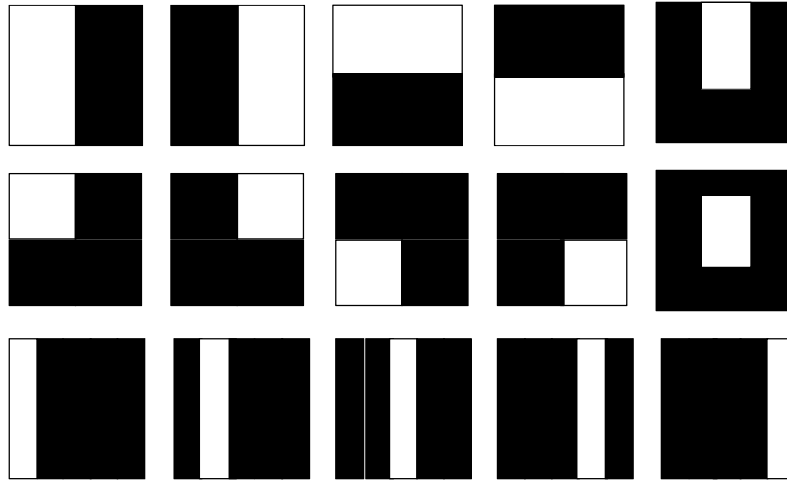


Fig. 4.3 Region creation for fusing the results. White areas show the selected face regions used in recognition experiments

(i) Mahalanobis Cosine Distance

MahCos distance is the Cosine distance computed in the Mahalanobis space. MahCos distance between two vectors \mathbf{u} and \mathbf{v} of image space is explained by equations 4.3 & 4.4 as described in Bajwa et al. [27].

$$S_{\text{MahCos}}(\mathbf{u}, \mathbf{v}) = \frac{\sum_{i=1}^N \left(\frac{u_i v_i}{\sigma_i} \right)}{\sqrt{\sum_{i=1}^N \left(\frac{u_i}{\sigma_i} \right)^2} \sqrt{\sum_{i=1}^N \left(\frac{v_i}{\sigma_i} \right)^2}} \quad (4.3)$$

where σ_i is standard deviation of i th dimension. In this case higher similarity means higher score. Thus, the actual distance is calculated by formula given in equation 4.4.

$$D_{\text{MahCos}}(\mathbf{u}, \mathbf{v}) = 1 - S_{\text{MahCos}}(\mathbf{u}, \mathbf{v}) \quad (4.4)$$

(ii) Euclidean Distance

The Euclidean distance between two vectors \mathbf{u} and \mathbf{v} in image space is calculated by formula given in equation 4.5 [27].

$$D_{EUC}(\mathbf{u}, \mathbf{v}) = \sqrt{\sum_{i=1}^N (\mathbf{u}_i - \mathbf{v}_i)^2} \quad (4.5)$$

(iii) Mahalanobis Distance

The Mahalanobis distance between two vectors \mathbf{u} and \mathbf{v} in image space is calculated by formula given in equation 4.6 [27].

$$D_{Mah}(\mathbf{u}, \mathbf{v}) = \sqrt{\sum_{i=1}^N \frac{1}{\sigma_i^2} (\mathbf{u}_i - \mathbf{v}_i)^2} \quad (4.6)$$

(iv) Manhattan Distance

The Manhattan distance between two vectors \mathbf{u} and \mathbf{v} in image space is calculated as.

$$D_{Manhattan}(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^N |\mathbf{u}_i - \mathbf{v}_i| \quad (4.7)$$

The above explained distance based classifiers are used to classify the individual regions in separate experiments. It was described that this research work presents a two-tier fusion methodology. At the first step, the outputs from 5 regional classifiers in a row (please see Fig. 4.3) are fused with a combination classifier that uses the proposed Weighted Borda Count method. Then results of combination classifiers for all three region sets are fused at second stage with a re-ranking approach into a single

decision. Weighted Borda Count method and re-ranking approach are explained in the following section.

4.2.3 Fusion Techniques

There are many ways to fuse the results of a pool of classifiers. Several approaches are given in the studies Spreeuwens [49] and Bajwa et al. [52] e.g. Dempster-Shafer Theory of Evidence, weighted majority voting and selection of the best combination of classifiers etc. In the study Spreeuwens [49], five levels of fusion are explained as following:

Sensor level fusion: Sensor level fusion fuses raw data from different sensors prior to feature extraction.

- Feature level fusion: Extracted features of a subject are combined in this technique.
- Rank level fusion: It is fusion of results at rank level. e. g. Borda Count method.
- Decision level fusion: It is the fusion of decisions of the different classifiers producing class labels. e.g. majority voting.
- Score level fusion: In this method scores produced by individual classifiers are fused. e.g. sum and product rule.

The Borda Count method [52] is based on the summation of ranks assigned to each of the regions by respective regional classifiers. The image with the lowest rank sum is declared a match. A limitation of the Borda Count method is that, it can be strongly affected by regions which give poor quality contribution towards final result. For example, if each of four regions reports a rank-1 match for a person and fifth region reports a 300th rank due to noise or occlusions, then total rank sums to 304 which will probably result in a mismatch.

To overcome this drawback of Borda Count method, Weighted Borda Count fusion method is proposed (as the first tier of ensemble classifier based matching) which considers only the top ranked matches. For this purpose a generalization of the rank-sum method is presented to transform a set of ranks \mathcal{R} into the set $f(\mathcal{R})$ using a function f which may be any nonlinear monotonically decreasing function such that:

$$\mathcal{R} = \{r, r + 1, r + 2, r + 3, \dots, R\} \quad (4.8)$$

where r and R represents rank 1 and rank N results respectively, and

$$f(\mathcal{R}) = \{f(r), f(r + 1), f(r + 2), f(r + 3), \dots, f(R)\} \quad (4.9)$$

In this study, the function $f(r)$ is proposed as

$$f(r) = (s - r)^w \quad (4.10)$$

where $s = 6$ and $w = 2$. Based on experimental evidences, the values of s and w have been chosen empirically because maximum performance was achieved with this selection. The motivation to use such a function f is to penalize the classes at the bottom of a ranked list. Also, higher weight assignments to top ranked classes returned experimentally higher results than the traditional Borda Count method [52]. When a new probe image is to be classified, the regional classifiers contribute a rank towards the ranked list against each of gallery images. The combination classifiers compute the corresponding ranks using the proposed Weighted Borda Count method for regions shown in Fig. 4.3. Thus, sums of ranked lists against each of gallery images are calculated and rank-1 is assigned to maximum value among all of the sums, rank-2 to next lower value of sum and so on. Re-ranking approach is employed at second ensemble stage, to fuse results produced by combination classifiers for all three region sets. In the re-ranking approach, outputs of combination classifiers are fused by assigning an empirically calculated weight to the combination classifiers and a weight to the output rank of the combination classifier. Finally these ranks are re-ordered to produce a ranked list where each ranked label represents a unique fused result. The mathematical description for such re-ranking [52] is expressed below.

$$W_{Tx} = \sum_{j=1}^c W_{K_i} \cdot W_{R_{x_i}} \quad (4.11)$$

where c is the number of combination classifiers, W_{K_i} is empirically calculated weight for a combination classifier determined by dividing recognition accuracy of classifier K by the total recognition accuracies of both of the classifiers implementing parallel face recognition algorithms, x is the probe image and W_{Tx} is total weight calculated

against the probe image x . $W_{R_{x_i}}$ is the weight assigned to a rank from the output of a combination classifier for the probe image x given as:

$$W_{R_{x_i}} = \frac{1}{e^{R_x}} \quad (4.12)$$

Where R_x is the rank of the corresponding matched label. Hence, the farther the matched label is in rank, lesser is its rank weight towards final weight calculation. Final re-ordered ranks are determined using total weights contributed by each of the combination classifiers computed for each of gallery images against a probe image.

4.3 Experiments and Results

Recognition is a task where the facial recognition system (FRS) is presented a probe image to attempt to match it with a reference image in the gallery. Two closed set face recognition experiments have been performed on each of the FRGC v2.0 and GavabDB databases using PCA based features and MahCos distance, Euclidean distance, Mah distance and Manhattan distance based nearest neighbor classifiers, separately. The dimension of original PCA based feature space is $500 \times 500 = 250000$ and that of reduced feature space is 'g' where 'g' is the size of the gallery employed in an experiment. In closed set recognition, it is guaranteed that for every probe image there is a guaranteed match present in the gallery. The performance of proposed methodology was investigated using the neutral 3D face images. The poses of all subjects were corrected by using the intrinsic coordinate system based alignment algorithm presented in this study.

4.3.1 Experiments on FRGC v2.0 Database

The FRGC v2.0 database [54] (described in section A.1) contains 466 subjects and almost 2410 point clouds with neutral expression. For some of the subjects, more than two neutral images are captured with a time lapse of one week between them. In the first experiment using FRGC v2.0 database, 466 images (one image per subject) were included in the gallery and 1944 were considered as probe. This strategy for splitting gallery and probe sets had been previously employed in the study Al-Osaimi et al. [56]. The experimental results for region sets S_1 , S_2 , S_3 and re-ranking stage (fused) are given by Cumulative Match Characteristic (CMC) plots shown in Fig. 4.4, 4.5, 4.6

and 4.7 using MahCos distance, Euclidean distance, Mah distance and Manhattan distance based classifiers respectively. It is clear from the plots that the experiments using MahCos distance metric obtained best performance whereas the experiments using Euclidean distance, Mah distance and Manhattan distance achieved relatively low performance respectively for region sets S_1 , S_2 and S_3 as well as for fused results at re-ranking stage.

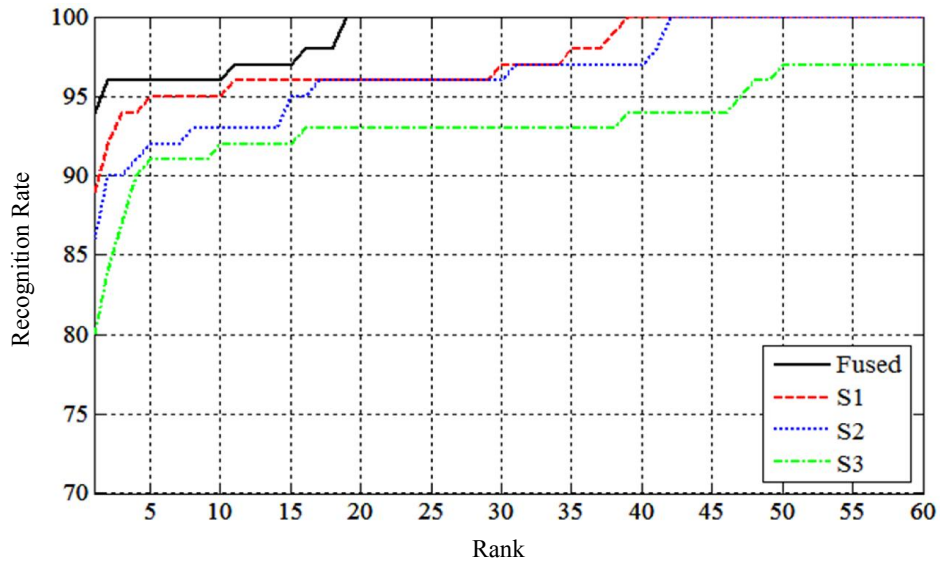


Fig. 4.4 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and MahCos distance (first experiment)

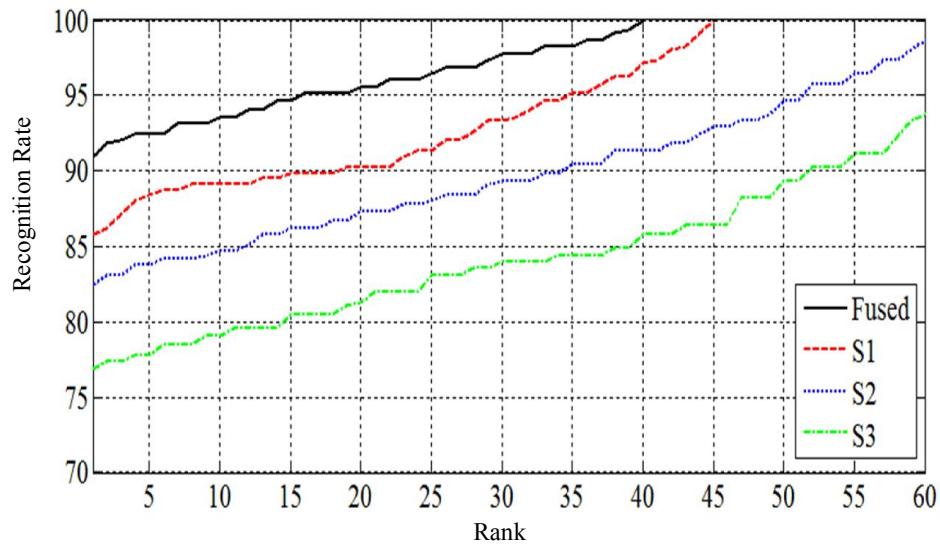


Fig. 4.5 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and Euclidean distance (first experiment)

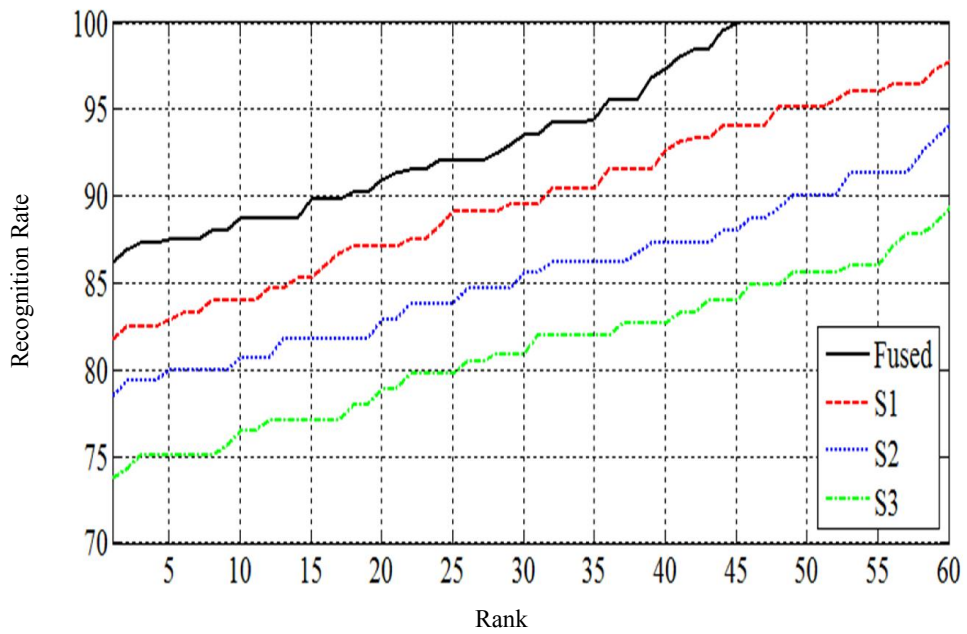


Fig. 4.6 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and Mah distance (first experiment)

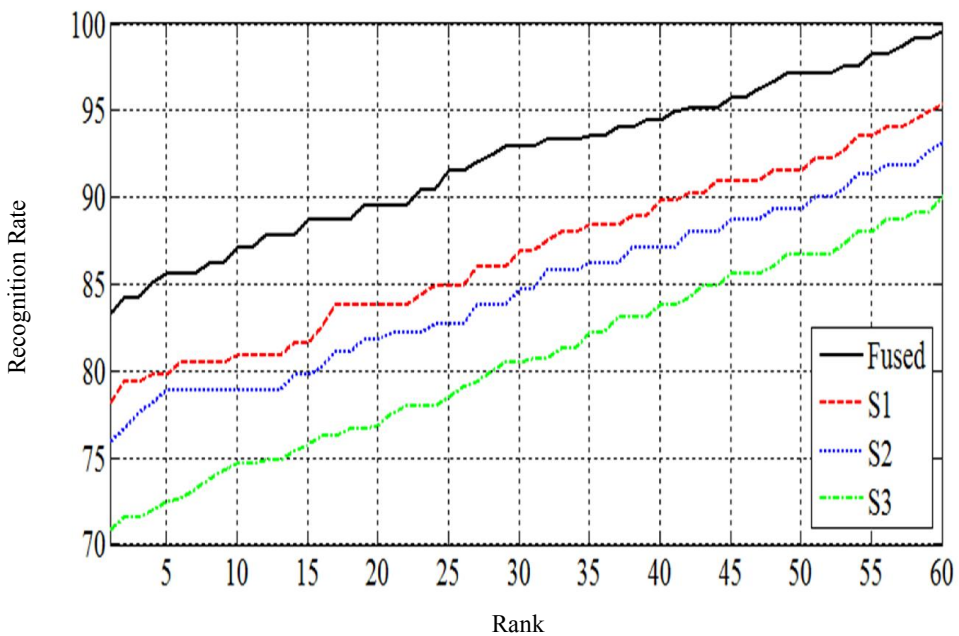


Fig. 4.7 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and Manhattan distance (first experiment)

In the second experiment, experimental protocol of Wang et al. [13] was considered where 943 scans from FRGC v1 were used in the training phase. In this experiment, results produced by combination classifiers for region sets S_1 , S_2 , S_3 ; and re-ranking stage were significantly improved. The improved classification accuracy (in case of GavabDB database as well) comes from the enrollment of multiple training images of known class labels in the gallery as the performance substantially degrades if one gallery image per subject is used in training phase [58]. The CMC plots for region sets S_1 , S_2 , S_3 , and re-ranking stage (fused) of the second experiment are given in Fig. 4.8, 4.9, 4.10 and 4.11 for MahCos distance, Euclidean distance, Mah distance and Manhattan distance respectively to depict the behavior of the considered classifiers. Similar to the results obtained in experiment 1, experiments using MahCos distance metric, achieved highest classification accuracy, whereas Euclidean distance, Mah distance and Manhattan distance classifier based experiments achieved low classification accuracy in the respective order for both of region sets S_1 , S_2 , S_3 , and fused results at re-ranking stage.

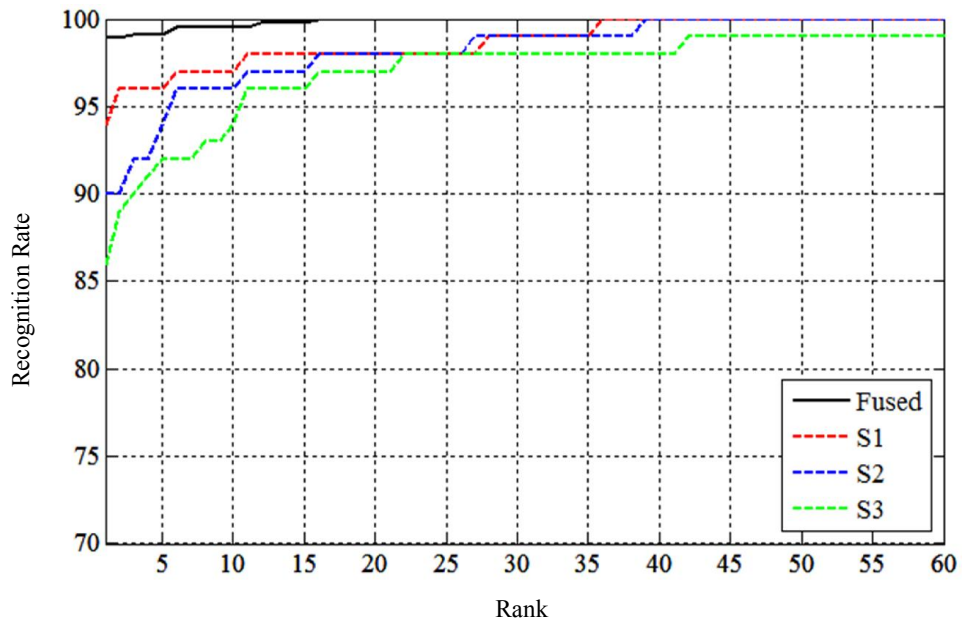


Fig. 4.8 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and MahCos distance (second experiment)

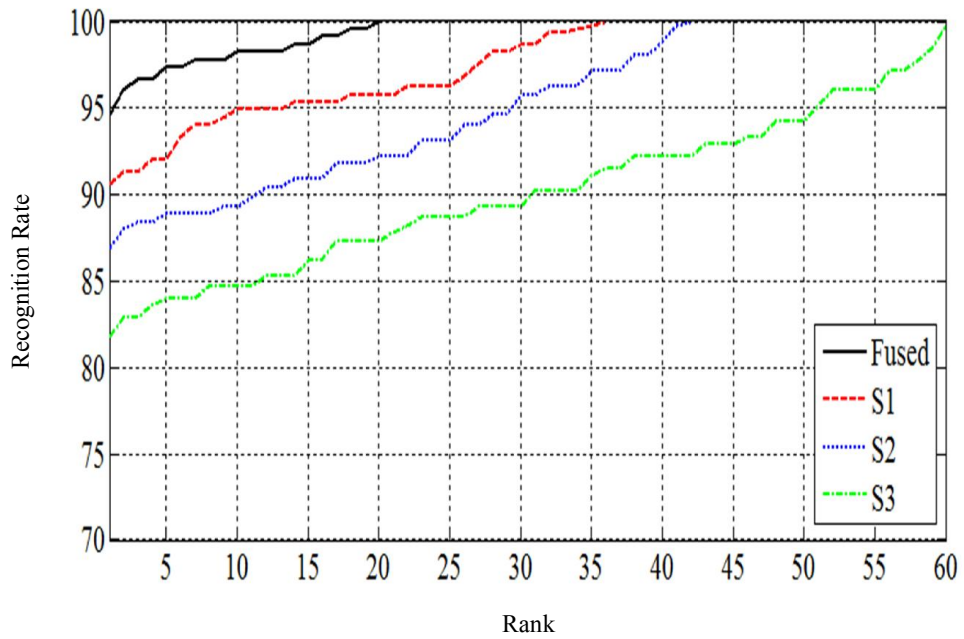


Fig. 4.9 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and Euclidean distance (second experiment)

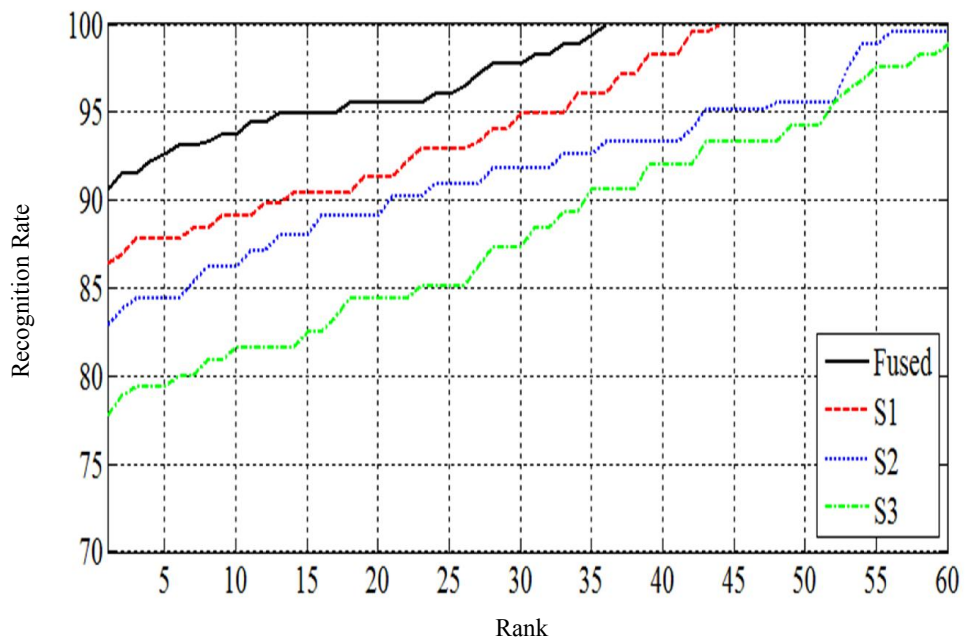


Fig. 4.10 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and Mah distance (second experiment)

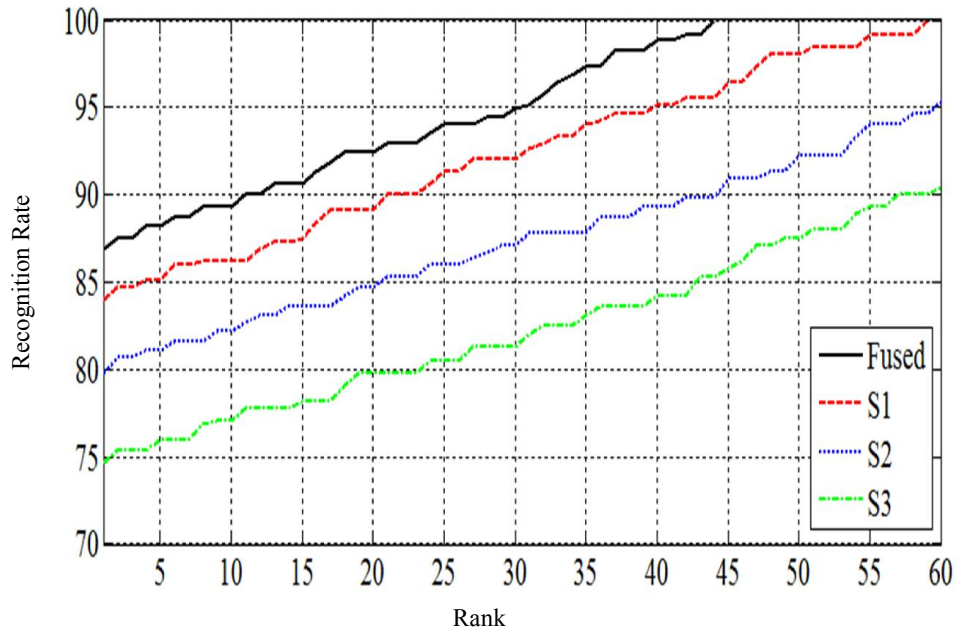


Fig. 4.11 Cumulative match characteristic curves of the proposed method for first rank-60 results using FRGC v2.0 database and Manhattan distance (second experiment)

4.3.2 Experiments on GavabDB Database

Referring to section A.2, which explains the characteristics of the GavabDB database [50], two face recognition experiments were conducted. Experimental protocol of Mahoor and Abdel-Mottaleb [48] was considered for the first experiment where second neutral scan of each subject was enrolled in the gallery and first was used as probe. For the second experiment, experimental protocol of Zhang et al. [55] was employed where three neutral scans of each subject were included in the gallery and remaining neutral scan was used as probe. The rank-1 recognition rates for sets S_1 , S_2 , S_3 , and re-ranking stage for both experiments are shown in Table 4.1. It can be observed from Table 4.1 that MahCos distance based classifier achieved best results for all region sets as well as at re-ranking stage for both experiments while Euclidean distance, Mah distance and Manhattan distance based classifiers obtained relatively

lower recognition rates respectively for both of region set S_1 , S_2 , S_3 , and fused results at re-ranking stage.

Table 4.1 Rank-1 recognition rates using GavabDB database (%)

Author, Year		Gallery Size	Probe Size	Region Set S_1	Region Set S_2	Region Set S_3	Rank-1 Recognition Rate
Mahoor and Abdel-Mottaleb., 2009 [48]		61	61	-	-	-	95
Zhang et al., 2014 [55]		183	61	-	-	-	100
Proposed Methodology (Experiment 1)	MahCos Distance	61	61	90.16	86.89	81.97	95.08
	Euclidean Distance	61	61	86.89	81.97	77.05	90.16
	Mah Distance	61	61	81.97	77.05	73.77	86.89
	Manhattan Distance	61	61	78.69	73.77	70.49	83.61
Proposed Methodology (Experiment 2)	MahCos Distance	183	61	96.72	91.80	85.25	100
	Euclidean Distance	183	61	93.44	88.52	81.97	96.72
	Mah Distance	183	61	90.16	85.25	78.69	93.44
	Manhattan Distance	183	61	86.89	81.97	75.41	90.16

4.3.3 Comparison with other Algorithms

Table 4.1 and 4.2 compare the recognition results of the proposed methodology to

state-of-the-art previous studies which used the GavabDB and FRGC v2.0 databases respectively. Referring to Table 4.1 the algorithm presented in Zhang et al. [55] is based on meshSIFT for facial key points detection with sparse representation based classification and the study Mahoor and Abdel-Mottaleb [48] presented a 3D face recognition approach based on 3D binary ridge images, principal maximum curvature and Iterative Closest Point matching technique. The proposed intrinsic coordinate system based alignment and ensemble classifier based face recognition approach has yielded a recognition rate of 100% using MahCos distance metric.

Referring to Table 4.2, besides the study Mahoor and Abdel-Mottaleb [48], Al-Osaimi et al. [56] integrated local and global geometric cues for 3D face recognition utilizing Principal Component Analysis and 2D histograms of tensors with Euclidean distance matching while the study Wang et al. [13] employed an approach based on Dual Tree Complex Wavelet Transform and Linear Discriminant Analysis with nearest neighbor classifier. The proposed algorithm has realized a better performance with rank-1 recognition result starting at 98.93% using MahCos distance metric.

Table 4.2 Rank-1 recognition rates using FRGC v2.0 database (%)

Author, Year	Gallery Size	Probe Size	Rank-1 Recognition Rate
Mahoor and Abdel-Mottaleb, 2009 [48]	370	370	93.7
Osaimi et al., 2007 [56]	466	1944	93.78
Wang et al., 2014 [13]	-	-	98.71
Proposed Methodology (first Experiment)	466	1944	94.15
Proposed Methodology (Second Experiment)	466	1944	98.93

4.3.4 Computational Complexity Analysis

The complexity of the proposed algorithm considering computationally intensive

operations is given in Table 4.3. Referring to Table 3.1, Intrinsic Coordinate System

Table 4.3 Comparison of the computational complexity of the proposed algorithm with Iterative Closest Point (ICP) algorithm

		Multiplications	Additions
ICP		$(M \times N) \times (1331 \times 3) \times 370$ $= 1.33 \times 10^{17}$	$(M \times N) \times (1331 \times 6) \times 370$ $= 2.66 \times 10^{17}$
ICP (fast)		$M \log(N) \times (1331 \times 3) \times 370$ $= 8.06 \times 10^{12}$	$M \log(N) \times (1331 \times 6) \times 370$ $= 1.61 \times 10^{13}$
ICP (14% points)		$(M \times N) \times (0.14)^2 \times (1331 \times 3) \times 370$ $= 2.61 \times 10^{15}$	$(M \times N) \times (0.14)^2 \times (1331 \times 6)$ $= 5.21 \times 10^{15}$
ICP (fast 14% points)		$0.14M \times \log(0.14 \times N) \times (1331 \times 3) \times 370$ $= 9.53 \times 10^{11}$	$0.14M \times \log(0.14 \times N) \times (1331 \times 6) \times 370$ $= 1.91 \times 10^{12}$
Proposed Methodology	MahCos Distance	$5.40 \times 10^6 + (m \times n) \times 370 \times 5 + 2 \times (m \times n) \times 370 \times 5$ $= 1.39 \times 10^9$	$3.90 \times 10^6 + (m \times n) \times 370 \times 5 + (m \times n) \times 370 \times 5$ $= 9.29 \times 10^8$
	Euclidean Distance	$5.40 \times 10^6 + (m \times n) \times 370 \times 5 + (m \times n) \times 370 \times 5$ $= 9.30 \times 10^8$	$3.90 \times 10^6 + (m \times n) \times 370 \times 5 + 2 \times (m \times n) \times 370 \times 5$ $= 1.39 \times 10^9$
	Mah Distance	$5.40 \times 10^6 + (m \times n) \times 370 \times 5 + 2 \times (m \times n) \times 370 \times 5$ $= 1.39 \times 10^9$	$3.90 \times 10^6 + (m \times n) \times 370 \times 5 + 2 \times (m \times n) \times 370 \times 5$ $= 1.39 \times 10^9$
	Manhattan Distance	$5.40 \times 10^6 + (m \times n) \times 370 \times 5$ $= 4.68 \times 10^8$	$3.90 \times 10^6 + (m \times n) \times 370 \times 5 + 2 \times (m \times n) \times 370 \times 5$ $= 1.39 \times 10^9$

based alignment involves 5.40×10^6 multiplication and 3.90×10^6 addition operations. PCA based face recognition algorithm requires $(m \times n) \times 370 \times 5$ multiplication and addition operations each for a gallery of 370 images, whereas the multiplication factor of 5 compensates for the overlapping regions (4 half face regions in region set S_1 and 11 quarter face regions in region sets S_1 , S_2 and S_3 approximately equal to 5 whole face regions) as given in Fig. 4.3. Each of MahCos distance, Euclidean distance, Mah distance and Manhattan distance is computed for mn dimensional 370 images and 5 overlapping regions. Furthermore, Manhattan distance calculation involves 1 load operation to determine the absolute value.

Among the algorithms [55] [48] [13] [56] which are compared in Tables 4.1 and 4.2, the only available complexity of the study [48] (neglecting initial course alignment complexity) is primarily based on Mean Squared Error (MSE) distance calculation which utilizes 3 multiplication and 6 addition operations for each of 1331 ($11 \times 11 \times 11$) rotations at 0° , $\pm 2^\circ$, $\pm 4^\circ$, $\pm 6^\circ$, $\pm 8^\circ$ and $\pm 10^\circ$ in x , y and z directions assuming that 11 iterations in each of x , y and z directions successfully result in the best geometric alignment of the probe image.

Referring to section 3.1, Iterative Closest Point based approaches align each probe image to every gallery image; whereas the proposed algorithm registers a probe image to the intrinsic coordinate system only once. The complexity calculations for the proposed algorithm and the study Mahoor and Abdel-Mottaleb [48] (considering point cloud of size $M = N = 300000$ points and depth image of size $m \times n = 500 \times 500$) are given in Table 4.3 which clearly reveals that proposed face alignment and recognition algorithm is computationally more efficient than the study Mahoor and Abdel-Mottaleb [48]. In terms of big O notation, the computational complexity of ICP, ICP (fast), ICP (with $\rho = 14\%$ points) and ICP (fast (with $\rho = 14\%$ points)) is of the order of $O(PQ)$, $O(Q \log(P))$, $\rho^2 O(PQ)$ and $O(\rho Q \log(\rho P))$ as given in Mahoor and Abdel-Mottaleb [48]. The computational complexity of the proposed methodology (in big O notation) is analyzed in terms of face alignment and face recognition. For face alignment, it is of the order of $O(P)$, where P represents total number of points in the point cloud. For face recognition, it is of the order of $O(mn)$, where mn represents total number of pixels in the face template.

4.4 Summary

A novel region based 3D face recognition algorithm has been proposed in this chapter employing Intrinsic Coordinate System based alignment. The proposed methodology employed a region based strategy where the face was segmented into 15 overlapping face regions. In order to classify individual face regions, Principal Component Analysis based features were used along with MahCos distance, Euclidean distance, Mah distance and Manhattan distance based classifiers in separate experiments. The region based classification results were combined using a two stage ensemble classifier that employed Borda Count based combination and a re-ranking stage. The experimentation process was carried out using FRGC v2.0 and GavabDB databases. The superior results exhibited by the proposed methodology demonstrated effectiveness of the approach and were compared with the state-of-the-art algorithms in terms of recognition rates and computational complexity.

Chapter 5

RECOGNITION BASED ON MULTI-VIEW DEPTH IMAGES

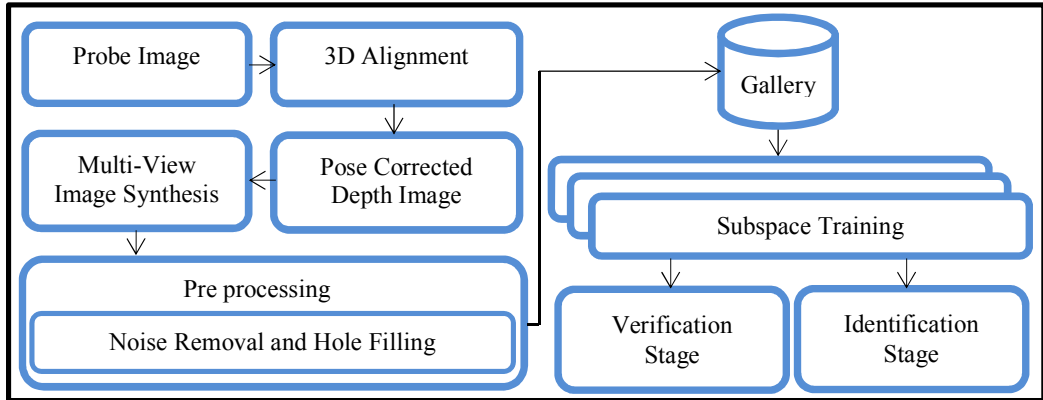
In this chapter, an automatic, pose and expression invariant 3D face recognition approach is presented which is based on facial information extracted from multi-view depth images like other studies [59] [60] [61] [62] [63]. In order to overcome limitations of existing approaches discussed in section 1.4, real 3D facial information is used in the recognition process using multi-view depth images. The proposed approach employs Principal Component Analysis (PCA) (discussed in section 2.2.1) for feature extraction and results in improved recognition rates for both of verification and identification scenarios.

For face verification, a multi-view full face region based recognition approach is proposed where two sets of 3D facial images are synthesized and classified using pairwise scores generated from the synthesized sets. For face identification, a face recognition method is designed that employs a three stage unified classifier which hierarchically fuses the results generated from seven base classifiers at first stage, two parallel face recognition algorithms at the second stage and an exponential rank combiner at the third stage to improve the classification accuracy of neutral and expressive faces. The depth images are aligned using two-pass 3D face alignment based on classification approach and two-pass 3D face alignment based on minimum distance in separate experiments. In order to evaluate the performance of the proposed system, experiments are conducted on FRGC v2.0 and GavabDB databases using Mahalanobis Cosine (MahCos) distance, Euclidean distance, Mahalanobis (Mah) distance, and Manhattan distance based classifiers. In addition, computational complexity analysis of the proposed methodology is also presented. The results are compared with state-of-the-art methods in terms of verification and identification rates.

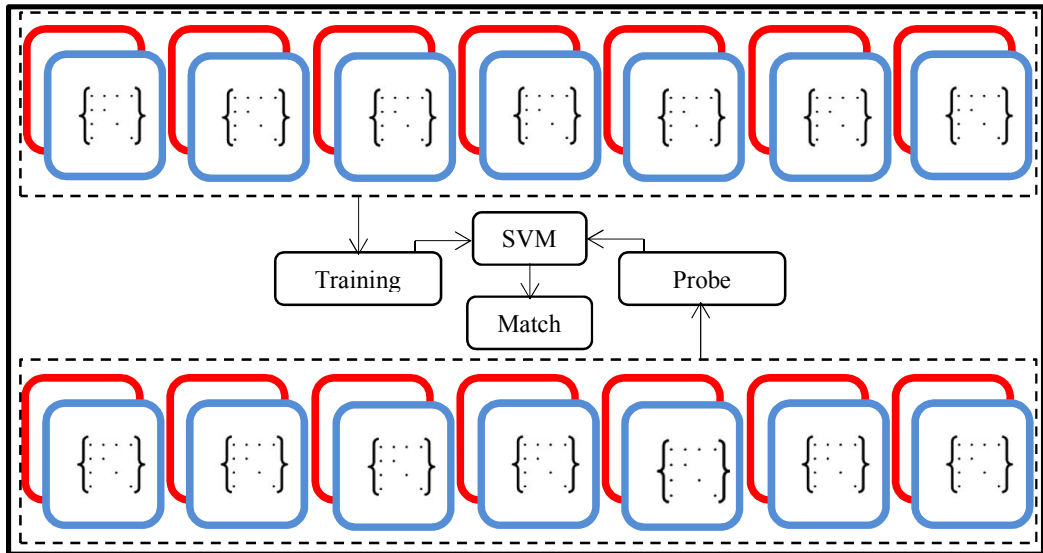
5.1 Proposed Methodology

The proposed face recognition system is comprised of face alignment, verification and identification modules. The face alignment module was discussed in Chapter 3 and an

Alignment and Preprocessing Module



Verification Module



Identification Module

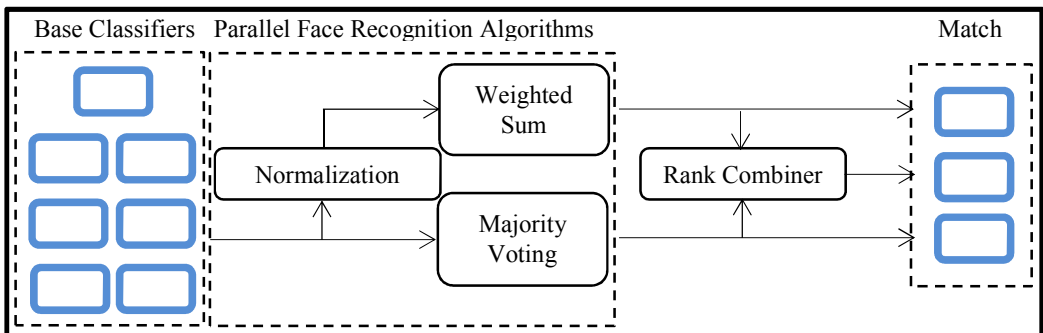


Fig. 5.1 Block diagram of proposed facial recognition system

illustration of the proposed approach is presented in Fig. 5.1.

For classification of depth images, first of all they are passed through a preprocessing step to cope with the artifacts such as noise and gaps. The sharp spikes, which are introduced in the depth images during the face scanning, are removed through median filtering and the holes on the face images are filled using interpolation as described in section 4.2.

5.1.1 Multi-view Synthesis

The images which are aligned at 0° are rotated around y -axis at 0° , $\pm 10^\circ$, $\pm 20^\circ$ and $\pm 30^\circ$ to synthesize novel views, resulting in two sets of seven depth images. The synthesized multi-view face images provide stable features for comparison of various local variations and incorporate occluded regions of the face that are less visible if frontal views are considered only. An example of synthesized multi-view face images employed in this study is shown in Fig. 5.2 (using subject 04203d438.abs from FRGC v2.0 and subject cara33_frontal1 from GavabDB) from which complementary 3D information can be readily observed. This information helps in coming up with a fusion based unified classification approach employed to classify these images for increased classification accuracy.

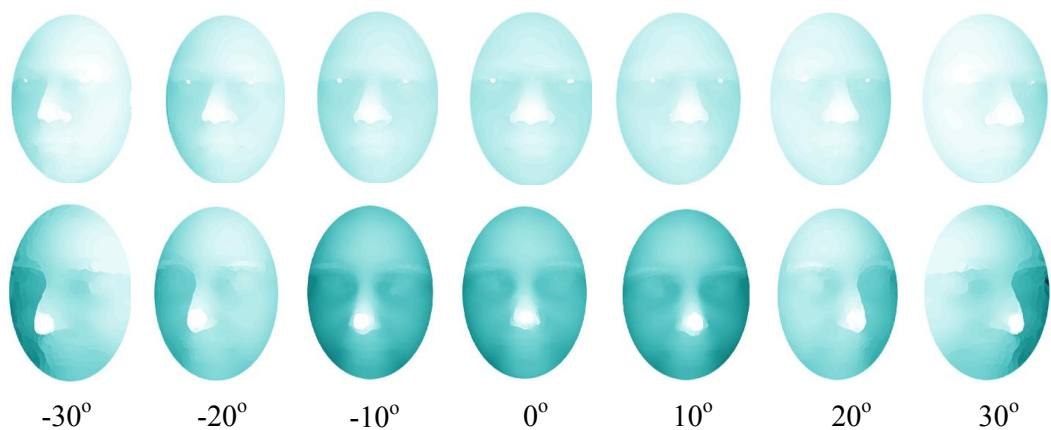


Fig. 5.2 Synthesized multi-view depth images: example subject from FRGC v2.0 (top row) and GavabDB (bottom row)

5.1.2 Handling Expression Variations

In order to handle facial expression variations, a special area called Expression Invariant Area is employed based on rigid facial regions comprised of eyes-forehead and nose. The motivation behind the proposed approach is based on three important findings [43]: (i) Rigid facial regions comprised of eyes-forehead and nose, are less sensitive to expressions compared to mouth and cheeks. (ii) Upper facial part is more significant compared to lower one for face recognition task. (iii) Region based matching is beneficial in case of expressive and neutral faces. By using the binary mask (Fig. 5.3 (a)) and the nose tip heuristic, the proposed approach automatically segments the expression invariant area as shown in Fig. 5.3 (b) for subject 04221d553.abs. After segmenting the expression invariant area multi-view segmented upper faces are synthesized at novel views of 0° , $\pm 10^\circ$, $\pm 20^\circ$ and $\pm 30^\circ$. The novel views carry 3D discriminating facial information as discussed in section 5.1.1.



Fig. 5.3 (a) Binary mask for segmentation of expression invariant area (b) expression invariant area of the face for example subject from FRGC v2.0

5.1.3 Classifier Fusion

Classifier fusion is an effective technique which is beneficial in pattern recognition problems such as fingerprint, speech, character, and face recognition. Facial data can be fused using several methods such as Dempster-Shafer Theory of Evidence, neural networks and majority voting principle [52] etc.

Five distinct levels of fusion are described in the studies Ross et al. [64] [65] namely, sensor level fusion (based on raw facial data), feature level fusion, rank level fusion

(e.g. Borda count), decision level fusion (e.g. majority voting) and matching score level fusion (e.g. weighted sum rule). Score level fusion is commonly chosen because matching scores contain enough information to make genuine and impostor cases separable. Score level fusion methods can be divided into three categories: (i) transformation-based (ii) density-based and (iii) classifier based. Examples of these methods are weighted sum-rule, likelihood ratio test with Gaussian Mixture Model and Support Vector Machine (SVM) based fusion respectively. Referring to section 3.2.3, the classification in verification set up is realized through SVM classifier in this study.

5.1.4 Face Verification Algorithm

Referring to section 5.1.1, for face matching in a verification scenario two sets of seven depth images (which are synthesized using two neutral frontal images of each subject in respective order) are included in the gallery and after Principal Component Analysis (PCA) based subspace learning, the algorithm computes MahCos distance, Euclidean distance, Mah distance or Manhattan distance (in separate experiments) between gallery image pairs for respective angular positions ($0^\circ, 0^\circ$), ($\pm 10^\circ, \pm 10^\circ$), ($\pm 20^\circ, \pm 20^\circ$) and ($\pm 30^\circ, \pm 30^\circ$). In the classification phase, seven synthesized depth images of the neutral probe face image and first gallery image are used to compute aforementioned scores for above mentioned angular positions preceded by PCA based subspace learning of the synthesized probe images.

Referring to section 5.1.2, in order to classify expressive faces, expression invariant area of the neutral images contained in above mentioned gallery and expressive probe image are employed to compute aforementioned scores using above mentioned procedure. The entire process results into availability of seven training and probe vectors of aforementioned scores (in separate experiments). The training vectors are used to train the SVM classifier explained in section 3.2.3 and probe vectors are used to classify the probe images.

5.1.5 Face Identification Algorithm

Face identification algorithm employs a three stage unified classifier based matching that implements MahCos distance, Euclidean distance, Mah distance or Manhattan

distance metric (explained in section 4.2.2) as base classifier at first stage in separate experiments. At the second stage, two parallel face recognition algorithms are employed through transformation based score level fusion using weighted sum-rule, and decision level fusion. The results of these algorithms are fused at the third stage of unified classifier using a rank combiner component.

In the weighted sum method, an empirically calculated weight for each base classifier is multiplied with distances calculated by the same base classifier for each probe image against the gallery images. The resulting multiplication results from all the base classifiers are added up and arranged in a ranked list. The mathematical description of such a scheme is given in equation 5.1.

$$d_x = \sum_{i=1}^m W_i \cdot d_{ix} \quad (5.1)$$

where W_i is the weight of classifier i , d_{ix} is the computed score vector obtained from base classifier i for probe image x and d_x is the sum of weighted distance scores from all classifiers for probe image x . W_i is determined using the individual recognition accuracy achieved by each of the base classifiers as given in equation 5.2.

$$W_i = \frac{\text{Recognition accuracy of each base classifier}}{\sum \text{Recognition accuracies of all base classifiers}} \quad (5.2)$$

For score normalization, min-max normalization method is utilized in this study which retains original distribution of matching scores except for a scaling factor and maps raw matching scores to interval $[0, 1]$. If maximum and minimum values of the raw matching scores are $\max(X)$ and $\min(X)$ respectively, then normalized score is computed as

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)} \quad (5.3)$$

where $x \in X$ and X is the set of raw matching scores obtained from a base classifier and x' denotes the normalized score of x .

Among the decision-level fusion strategies, majority voting is naïve in nature where each base classifier has right of one vote and the class getting more than half of votes is declared as a match based on the mathematical description, $(n + 1)/2$, if number of votes n is odd. In the context of the proposed application, if four or more base classifiers yield positive identification, then identification process is considered successful. The outputs of parallel face recognition algorithms are fused in a rank combiner element by assigning an empirically calculated weight to them and to the output rank yielded by them as explained in section 4.2.3.

5.2 Experiments and Results

The proposed study targets to develop an efficient, pose and expression invariant; facial recognition system which can handle frontal, non-frontal and profile face images. The images were aligned using two-pass 3D face alignment based on classification approach and two-pass 3D face alignment based on minimum distance in separate experiments. Two databases namely, FRGC v2.0 and GavabDB were employed in the experimentation process of the proposed study. On each of these databases two types of experiments were performed using PCA based features and MahCos distance, Euclidean distance, Mah distance and Manhattan distance based classifiers. The dimension of original PCA based feature space is $500 \times 500 = 250000$ and that of reduced feature space is 'g' where 'g' is the size of the gallery employed in an experiment.

The first type of experiment was face verification experiment where performance of the facial recognition system is reported as verification rate at a given False Acceptance Rate (FAR). The second one was face identification experiment where performance is reported as rank-1 identification rate. For both types, four experiments were performed in total, namely Neutral vs. Neutral (N vs. N) verification, Neutral vs. Non-neutral (represented by N vs. E) verification, N vs. N identification and N vs. E identification as explained in the following.

5.2.1 Experiments on FRGC v2.0 Database

(i) Face Verification Experiments

Referring to section A.1 where characteristics of FRGC v2.0 3D database [54] are given, the performance of proposed face verification algorithm was investigated by creating N vs. N and N vs. E verification experiments using the FRGC v2.0 database which contains 2469 neutral scans and 1538 non-neutral scans [5].

Since the classifier was trained in the score space, two classes being considered were genuine scores and imposter scores. For imposter scores, only one template per subject could be used however for genuine scores, at least two templates per subject were needed. Therefore, two neutral images per subject were used to calculate training scores for SVM classifier, whereas expression invariant area of the same two neutral training images was used for training in N vs. E experiments. The FRGC v2.0 database contains 370 such subjects which possess at least two neutral images [48]. Therefore, 740 images (two images per subject) were included in the gallery set and in case of subjects that had more than two neutral images; the first two stored neutral images were included in the gallery. The remaining neutral and all expressive images were used as probe sets for N vs. N and N vs. E experiments respectively.

The performance metrics of the proposed method using two-pass 3D face alignment based on classification approach are depicted in terms of ROC curves at 0.1% False Acceptance Rate (FAR) in Fig. 5.4, 5.5, 5.6 and 5.7 for N vs. N experiments using MahCos, Euclidean, Mah and Manhattan distances respectively. Similarly Fig. 5.8, 5.9, 5.10 and 5.11 represent ROC curves at 0.1% FAR for N vs. E experiments using aforementioned distances respectively.

The performance metrics using two-pass 3D face alignment based on minimum distance are given in Table 5.1.

It is clear from Fig. 5.4 to Fig. 5.11 and Table 5.1 that the experiments using MahCos distance metric obtained best performance whereas the experiments using Euclidean distance, Mah distance and Manhattan distance achieved relatively low performance respectively. The two-pass 3D face alignment based on classification approach

resulted into better verification rates than two-pass 3D face alignment based on minimum distance.

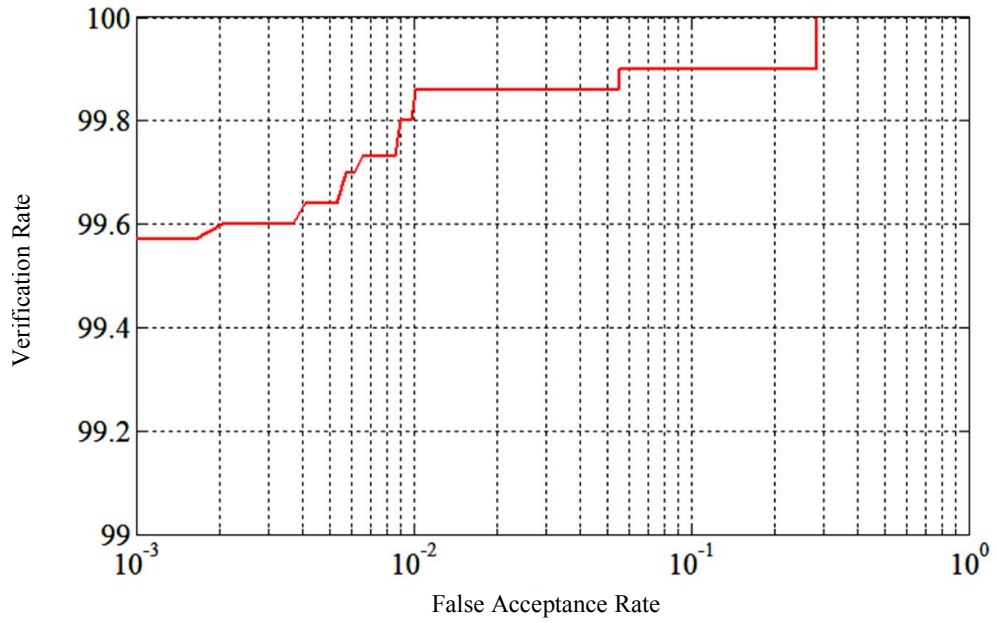


Fig. 5.4 Receiver Operating Characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and MahCos distance

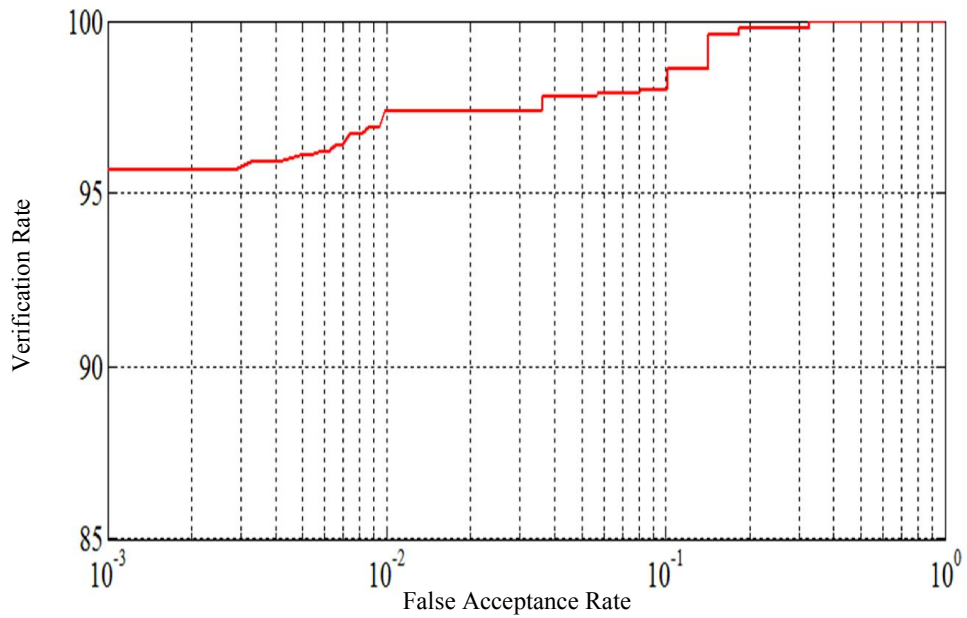


Fig. 5.5 Receiver Operating Characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and Euclidean distance

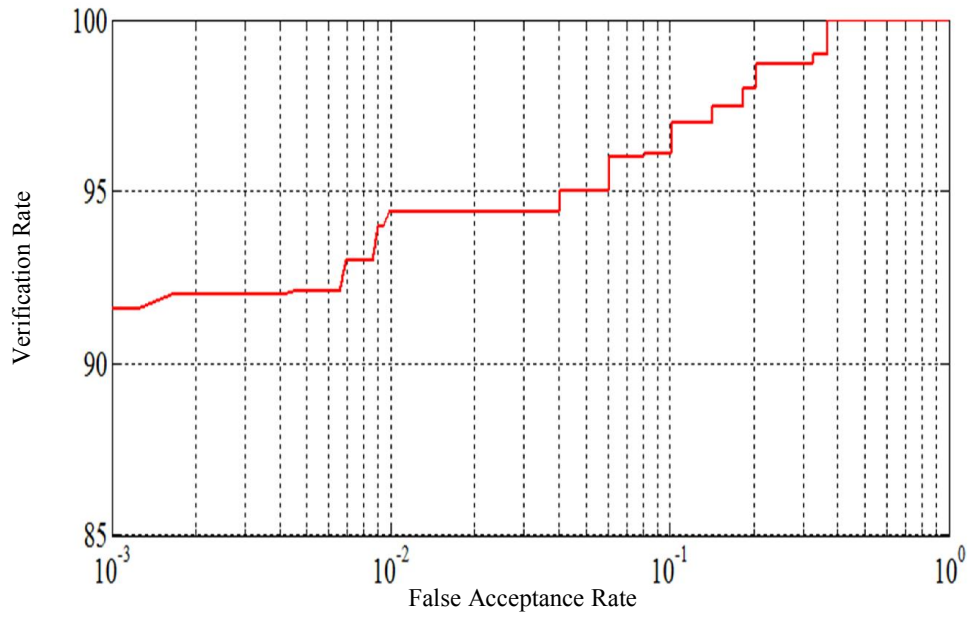


Fig. 5.6 Receiver Operating Characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and Mah distance

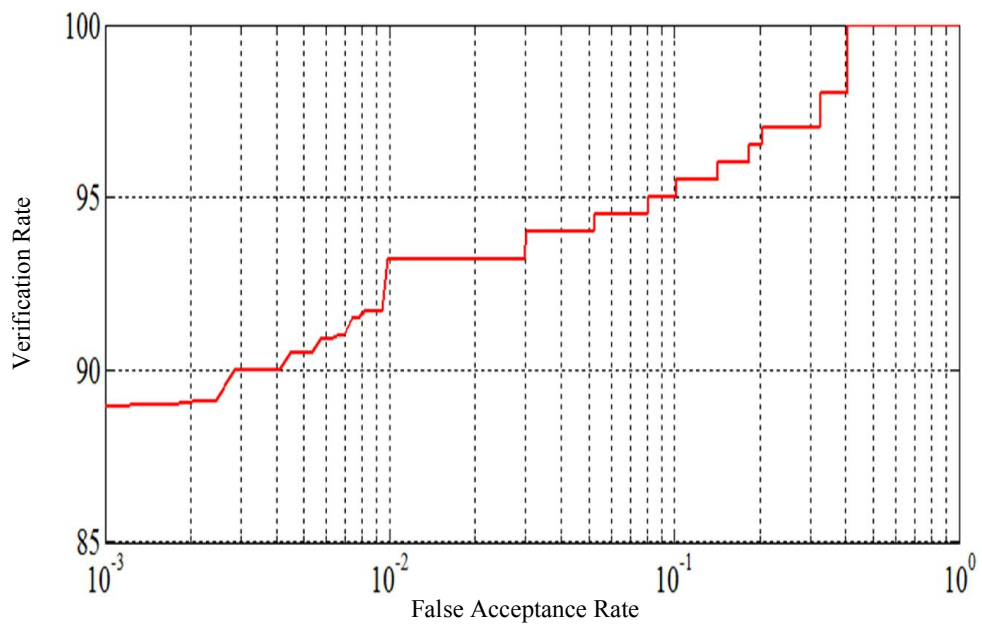


Fig. 5.7 Receiver Operating Characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and Manhattan distance

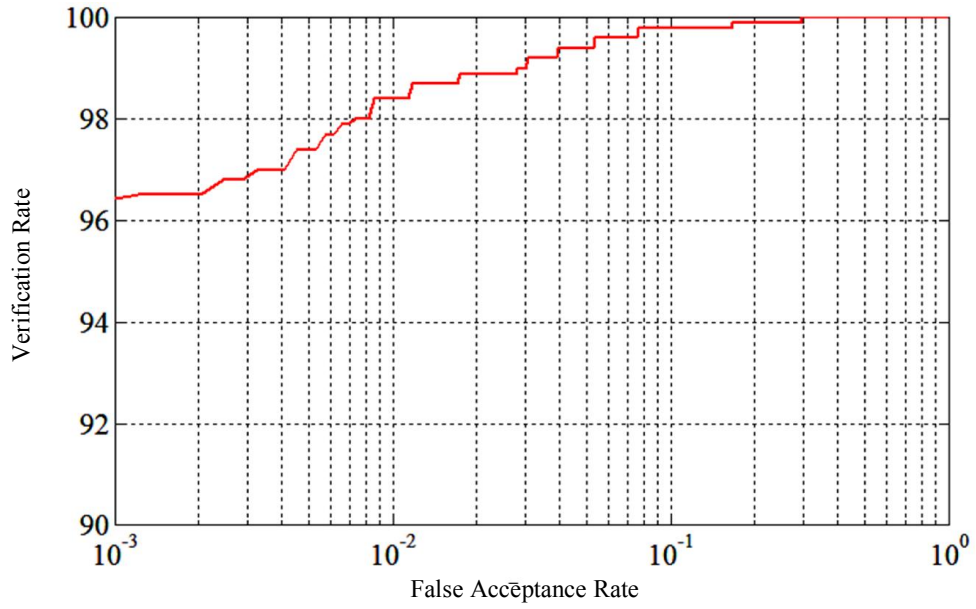


Fig. 5.8 Receiver Operating Characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and MahCos distance

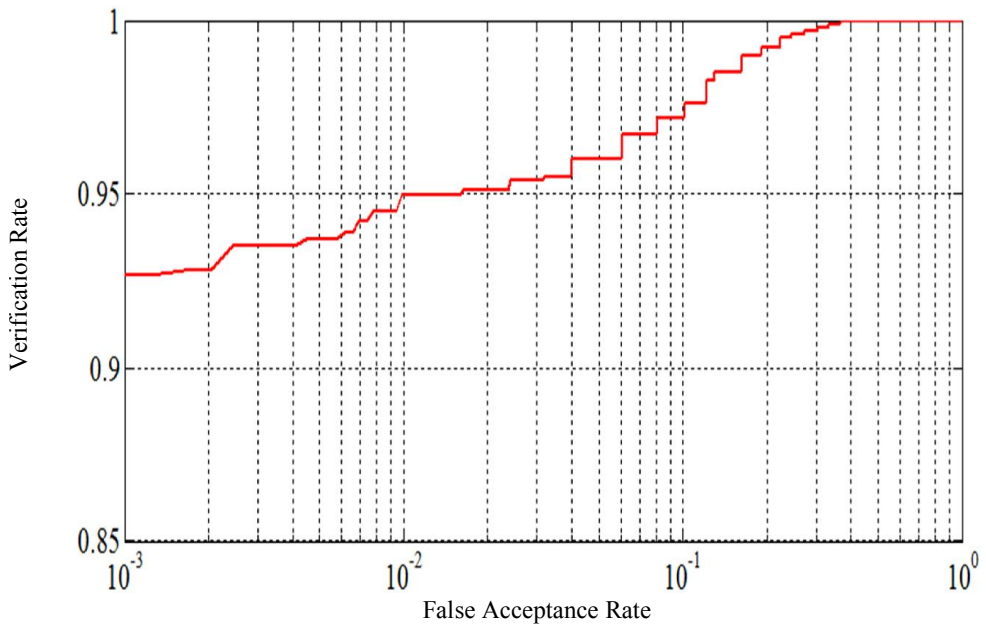


Fig. 5.9 Receiver Operating Characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and Euclidean distance

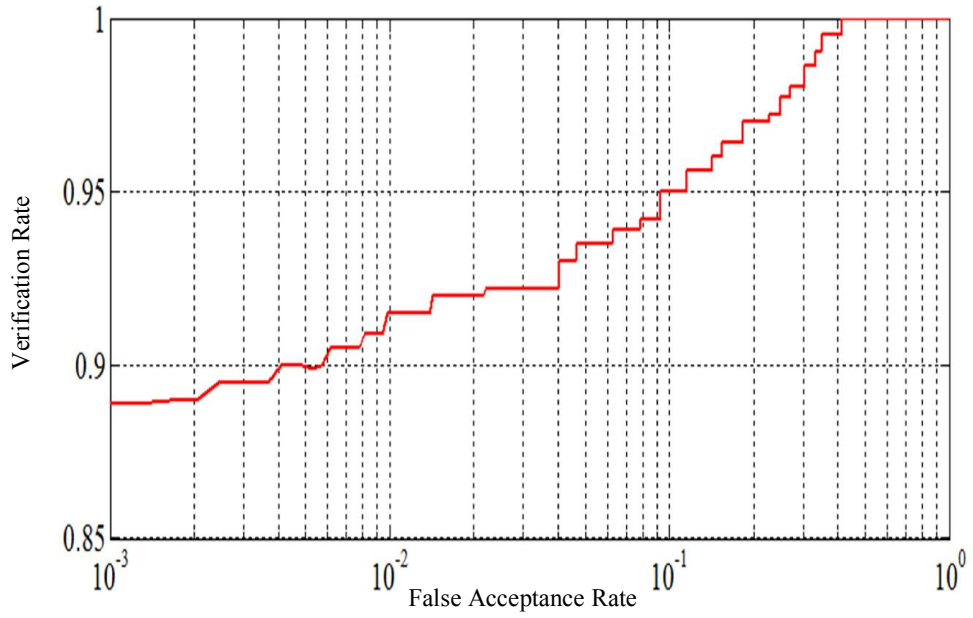


Fig. 5.10 Receiver Operating Characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and Mah distance

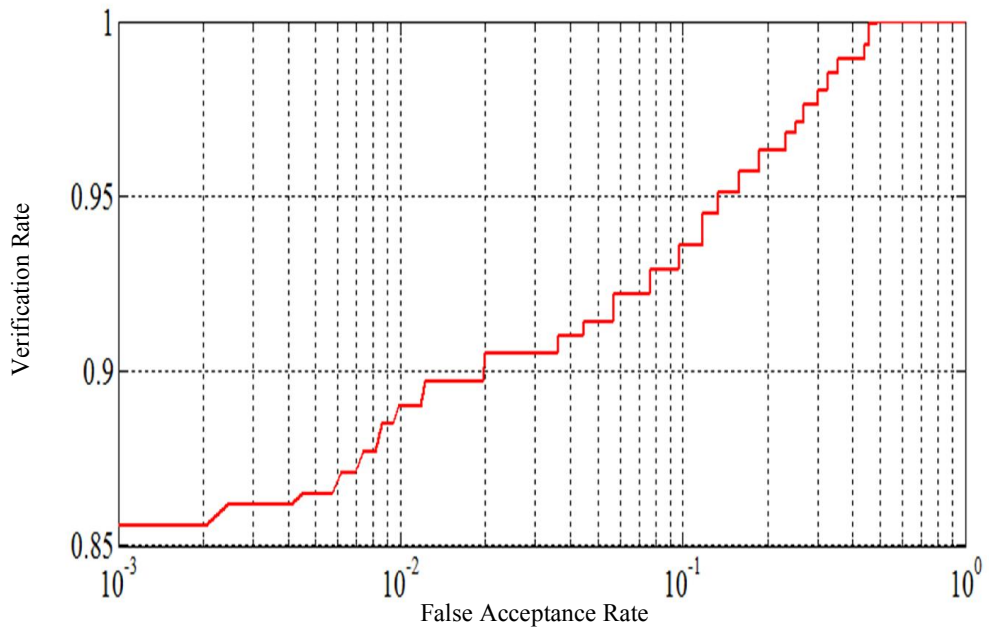


Fig. 5.11 Receiver Operating Characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and Manhattan distance

(ii) Face Identification Experiments:

To evaluate the performance of face identification algorithm proposed in this study, experimental protocol of Berretti et al. [5] was employed where probe sets were created using 2469 neutral scans and 1538 non-neutral scans for N vs. N and N vs. E experiments respectively and images from FRGC v1 were enrolled in the gallery in compliance with guidelines of this database [5]. The cumulative match characteristic (CMC) plots for N vs. N experiments using two-pass 3D face alignment based on classification approach are given in Fig. 5.12, 5.13, 5.14 and 5.15 using MahCos, Euclidean, Mah and Manhattan distances respectively. Similarly Fig. 5.16, 5.17, 5.18 and 5.19 depict CMC plots for N vs. E experiments using the aforementioned distances respectively. The result accuracies using two-pass 3D face alignment based on minimum distance are given in Table 5.1. It is clear from Fig. 5.12 to Fig. 5.19 that best performance was obtained for experiments where MahCos distance metric was used whereas the experiments using Euclidean distance, Mah distance and Manhattan distance obtained relatively low performance respectively. The two-pass 3D face alignment based on classification approach produced better recognition rates than two-pass 3D face alignment based on minimum distance.

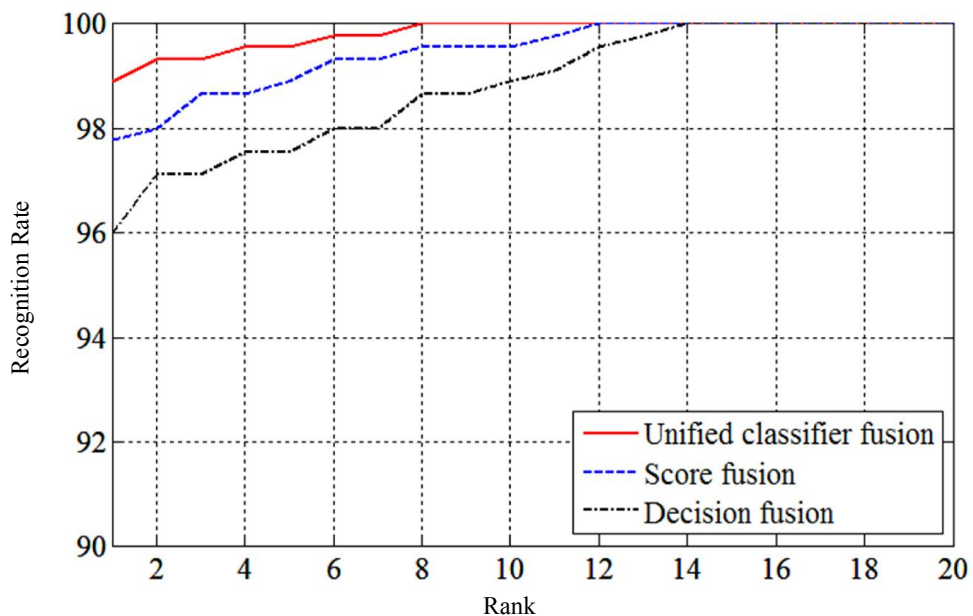


Fig. 5.12 Cumulative match characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and MahCos distance

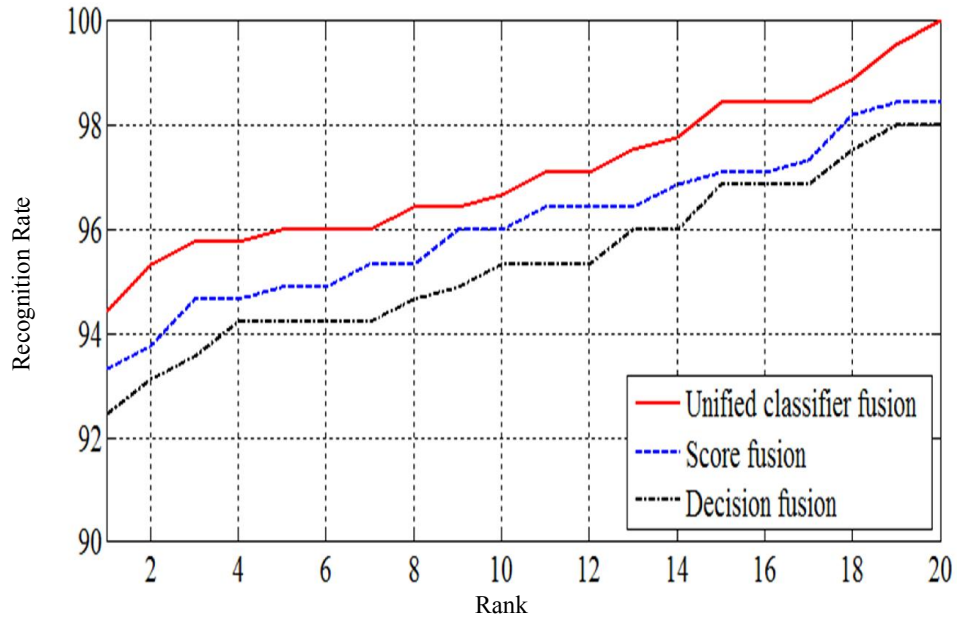


Fig. 5.13 Cumulative match characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and Euclidean distance

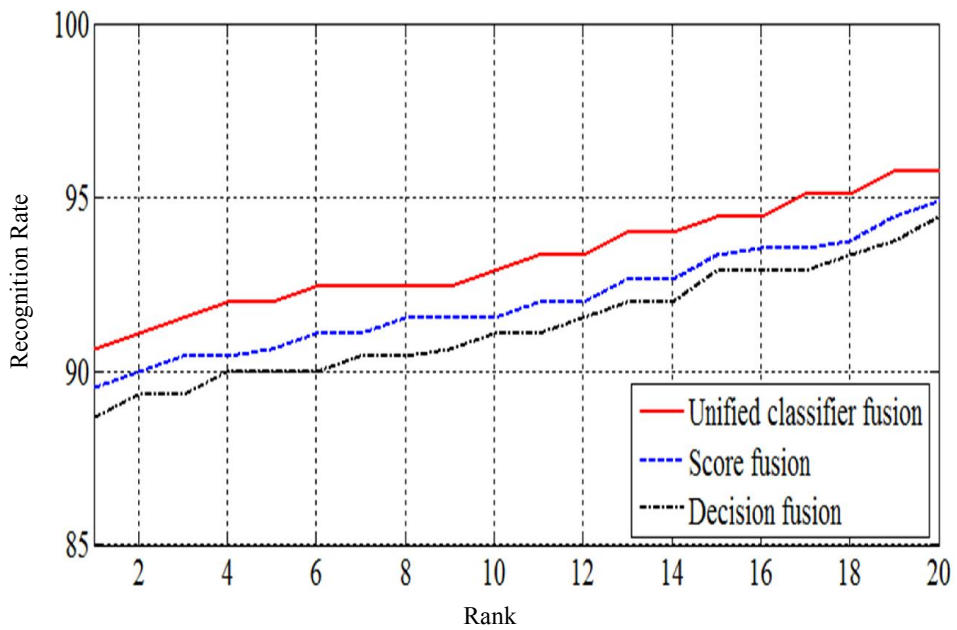


Fig. 5.14 Cumulative match characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and Mah distance

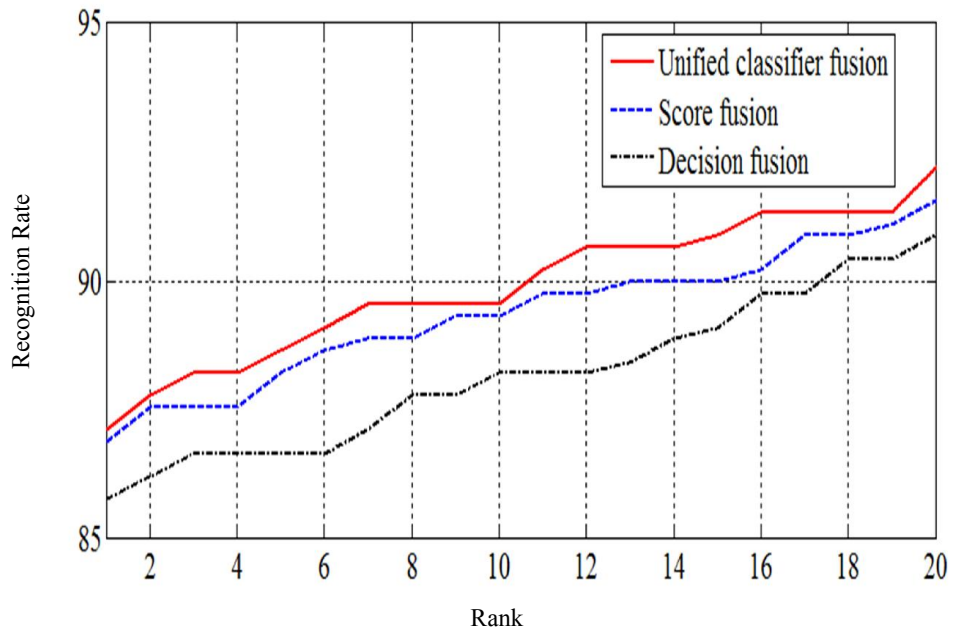


Fig. 5.15 Cumulative match characteristic curve of the proposed method for N vs. N experiment using FRGC v2.0 database and Manhattan distance

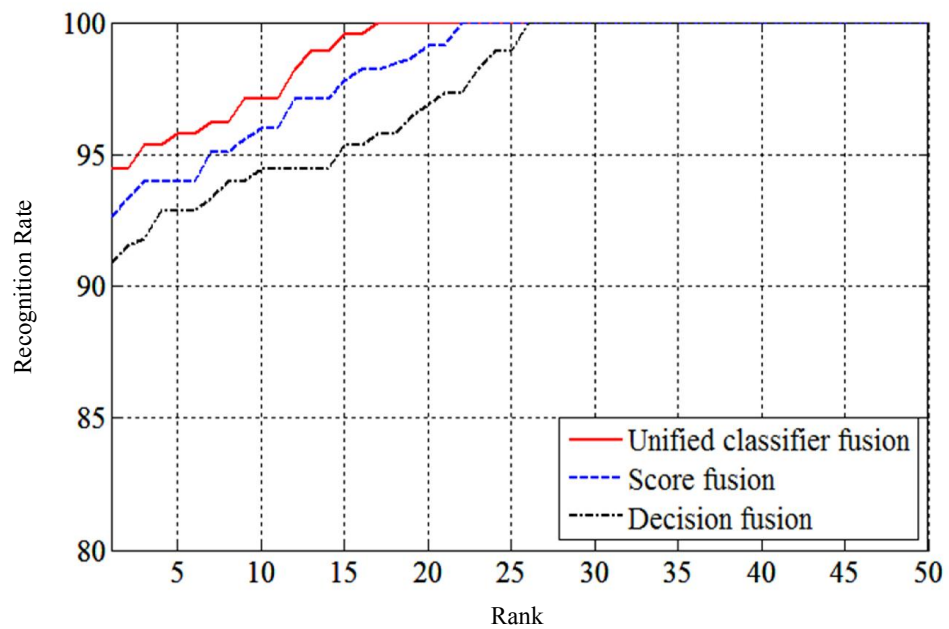


Fig. 5.16 Cumulative match characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and MahCos distance

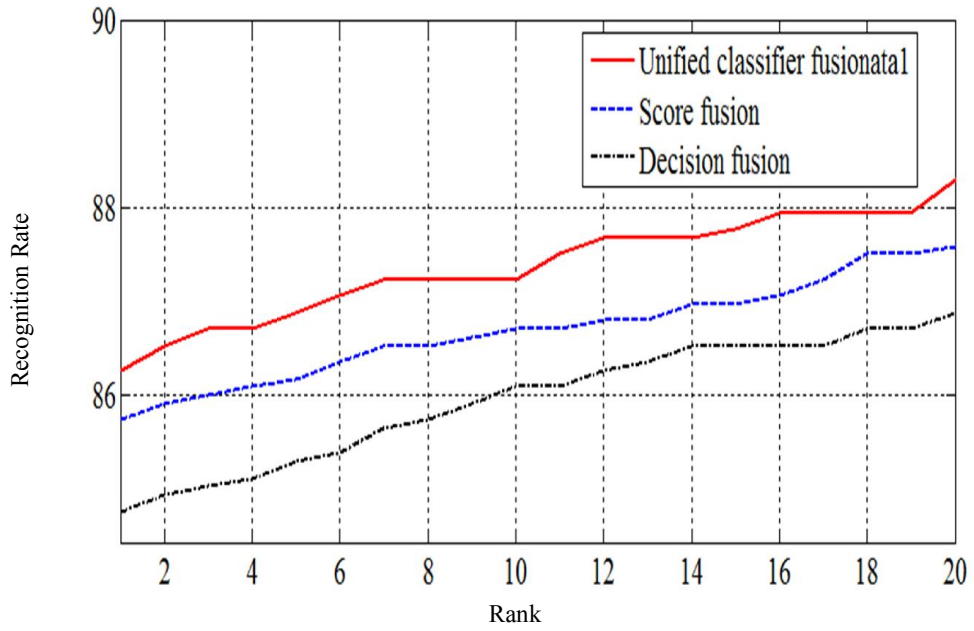


Fig. 5.17 Cumulative match characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and Euclidean distance

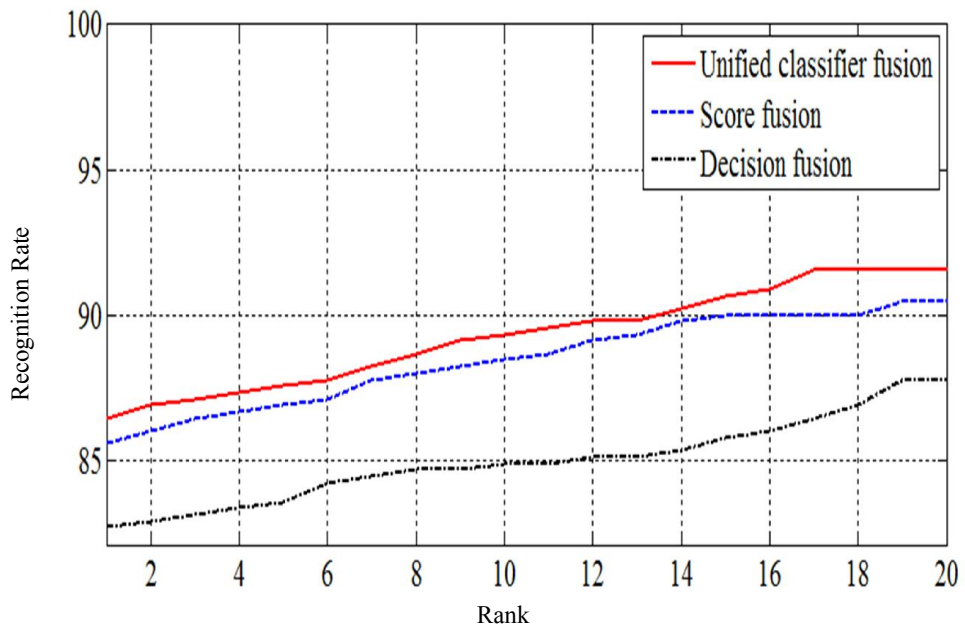


Fig. 5.18 Cumulative match characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and Mah distance

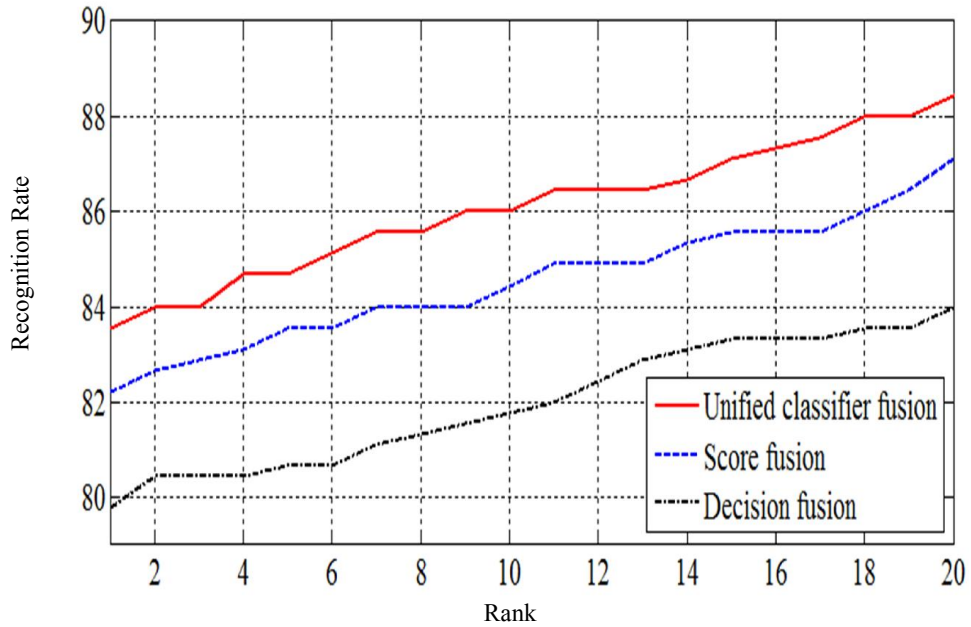


Fig. 5.19 Cumulative match characteristic curve of the proposed method for N vs. E experiment using FRGC v2.0 database and Manhattan distance

(iii) Comparison with Existing Algorithms

The face verification and identification results of the proposed methodology were compared to the existing state-of-the-art methods in Table 5.1 for FRGC v2.0 database. The technique presented in the study Al-Osaimi et al. [56] proposed integration of global and local geometric cues employing Principal Component Analysis and 2D histograms of tensors using Euclidean distance based classification.

The study Berretti et al. [5] is focused on a graph based approach in which recognition is accomplished by partitioning each face into a fixed number of iso-geodesic stripes and measuring displacement for each stripes pair by computing 3D weighted walkthroughs between all point pairs of two stripes. The study Chang and Wang [66] presents a resolution invariant local feature based 3D face recognition method in which scale space extrema on shape index images were detected and matched using score level fusion based SVM classifier, while a Dual Tree Complex Wavelet Transform, Linear Discriminant Analysis and nearest neighbor based face recognition approach is implemented in the study Wang et al. [13].

Table 5.1 Verification and identification rates using two-pass 3D face alignment based on minimum distance on FRGC v2.0

Author, Year		Verification rates (%)				Rank-1 identification rates (%)			
		N vs. N ⁽¹⁾	N vs. N	N vs. E ⁽¹⁾	N vs. E	N vs. N ⁽¹⁾	N vs. N	N vs. E ⁽¹⁾	N vs. E
Proposed methodology	MahCos Distance	99.57	99.14	96.57	96.14	98.93	98.50	94.42	93.99
	Euclidean Distance	95.71	95.28	92.70	92.27	94.42	93.99	90.77	90.34
	Mah Distance	91.63	91.20	88.84	88.41	90.77	90.34	86.4 ¹⁾	86.05
	Manhattan Distance	88.84	88.41	85.62	85.19	87.12	86.70	83.69	83.26
Osaimi et al., 2007 [56]		95.37		-		93.78		-	
Berretti et al., 2010 [5]		97.7		91.4		96.1		90.77	
G. Zhang and Y. Wang., 2011 [66]		98.3		89.5		98		89.1	
Wang et al., 2014 [13]		99.53		96.25		98.71		94.21	

⁽¹⁾ Results using two-pass 3D face alignment based on classification approach.

The proposed SVM classifier based face verification and unified classifier based face identification method using two-pass 3D face alignment based on classification approach and MahCos distance metric has yielded a better performance than state-of-the-art studies presented in Table 5.1, with verification rate of 99.57% and 96.57% at 0.1% FAR and rank-1 identification rate of 98.93% and 94.42% for N vs. N and N vs.

E experiments respectively.

5.2.2 Experiments on GavabDB Database

(i) Face Verification Experiments

Referring to section A.2 which explains GavabDB [50] database features, same protocol as used by Berretti et al. [67], was followed for creation of probe set for the verification experiment, In the training phase, the neutral image ‘abajo’ was included as second image along with ‘frontal1’ in the gallery for pairwise training score calculation, whereas ‘frontal2’ and ‘frontal1’ were used for pairwise test score calculation for N vs. N verification experiment. For the N vs. E verification experiment, 183 expressive images were used for pairwise test score calculation. The performance metrics of the proposed method employing two-pass 3D face alignment based on classification approach and two-pass 3D face alignment based on minimum distance are given in Table 5.2.

(ii) Face Identification Experiments

For the identification experiments, experimental protocol of Berretti et al. [67] was considered for the N vs. N and N vs. E experiments where for each of 61 subjects from GavabDB database, the image ‘frontal1’ was enrolled in the gallery in compliance with the experimental protocol of this database and the image ‘frontal2’ was used as probe. For the N vs. E experiment, three expressive images per subject (183 images in total) were used as probe set. The rank-1 identification rates for N vs. N and N vs. E experiments using two-pass 3D face alignment based on classification approach and two-pass 3D face alignment based on minimum distance are given in Table 5.2.

For classification of profile faces, the unified classifier considered four base classifiers for faces synthesized at 0° , 10° , 20° , 30° and 0° , -10° , -20° , -30° for right and left profiles respectively. The score based parallel face recognition algorithm took into account scores from above mentioned base classifiers and decision based parallel face recognition algorithm considered three votes to decide for a true positive. The rank-1 identification rates for left and right profile faces using two-pass 3D face alignment based on minimum distance are given in Table 5.2.

Table 5.2 Verification and identification rates using two-pass 3D face alignment based on minimum distance on GavabDB

Author, Year		Verification rates (%)				Rank-1 identification rates (%)					
		N vs. N ⁽¹⁾	N vs. N	N vs. E ⁽¹⁾	N vs. E	N vs. N ⁽¹⁾	N vs. N	N vs. E ⁽¹⁾	N vs. E	Left Profile	Right Profile
		Proposed methodology									
	Mah Cos Distance	100	98.36	95.08	93.44	100	98.36	96.72	95.08	95.08	83.61
	Euclidean Distance	96.72	95.08	91.80	90.16	95.08	93.44	93.44	91.80	91.80	80.33
	Mah Distance	91.80	90.16	86.89	85.25	93.44	91.80	91.80	90.16	86.89	77.05
	Manhattan Distance	90.16	88.52	83.61	81.97	88.52	86.89	86.89	85.25	83.61	75.41
Li et al., 2009 [68]		95.08		93.44		-		-		-	-
Berretti et al., 2012 [67]		-		-		100		96.17		93.44	81.97
Zhang et al., 2014 [55] ⁽²⁾		-		-		100		-		-	-
Hariri et al., 2016 [69]						100				83.60	81.96

⁽¹⁾ Results using two-pass 3D face alignment based on classification approach

⁽²⁾ In the study Zhang et al. [55], gallery comprising of 183 subjects is employed.

From Table 5.2 it is revealed that the recognition experiments using MahCos distance metric achieved the best performance whereas those using Euclidean distance, Mah distance and Manhattan distance exhibited relatively low performance respectively for all of the face verification, identification and profile images based experiments. The

two-pass 3D face alignment based on classification approach produced better recognition rates than two-pass 3D face alignment based on minimum distance.

(iii) Comparison with Existing Algorithms

The face verification and identification results of the proposed methodology are compared with the existing state-of-the-art methods in Table 5.2 for GavabDB database. The study Berretti et al. [67] presents a face recognition technique based on 3D surface description by keypoints extraction and measurement of change in face depth along facial curves connecting keypoints pairs, along with sparse comparison based similarity evaluation. The study Zhang et al. [55] employs meshSIFT for detection of facial key points with sparse representation based classification. The study Li et al. [68] presents a sparse representation framework based face recognition method using low level geometric features whereas the study Hariri et. al [69] employs covariance descriptors and geodesic distance matching. The proposed approach has obtained 100% and 95.08% verification rate along with 100% and 96.72% rank-1 identification accuracy for N vs. N and N vs. E experiments respectively. The identification rates yielded by the proposed method using two-pass 3D face alignment based on classification approach (in case of profile face images two-pass 3D face alignment based on minimum distance) and MahCos distance metric are better than state-of-the-art studies presented in Table 5.2 for N vs. E, left and right profile experiments and equal for N vs. N experiment, whereas verification rates are better for both of N vs. N and N vs. E experiments.

5.2.3 Computational Complexity Analysis

(i) Analysis using Two-Pass 3D Face Alignment based on Classification

Approach

Referring to Table 5.3, the values 226800588 and 151501029 represent multiplication and addition operations required to align the point cloud of size $M = 300000$ points using ‘two-pass 3D alignment based on classification approach’. For face identification, the terms $M \times 9 \times 6$ and $M \times 6 \times 6$ represent 9 multiplication and 6 addition operations involved to rotate each point in multi-view synthesis at 6 views (whereas seventh view at 0° is obtained in alignment phase) for frontal and non-frontal

face images respectively. The term $m \times n \times 370 \times 7$ represents additions and multiplications involved in PCA based subspace projection for a gallery of size 370 images synthesized at 7 views for $m = n = 500$. Each of MahCos distance, Euclidean distance, Mah distance and Manhattan distance is computed for mn dimensional 370 images synthesized at 7 views. In case of SVM classifier based face verification, additional multiplications (7), additions (13) and load operations (1) are involved for each of 163 support vectors resulting in 1141, 2119 and 163 multiplications, additions and load operations respectively.

(ii) Analysis using Two-Pass 3D Face Alignment based on Minimum Distance

The computational complexity of the proposed face identification methodology using two-pass 3D face alignment based on minimum distance is given in Table 5.4. Referring to Table 5.4 and section 5.2.3 (i), the values 226800000 and 151500000 represent multiplication and addition operations required to align the point cloud of size $M = 300000$ points. In case of SVM classifier based face verification, additional multiplications (7), additions (13) and load operations (1) are involved for each of 163 support vectors resulting in 1141, 2119 and 163 multiplications, additions and load operations respectively.

(iii) Analysis for Profile Face Images using Two-Pass 3D Face Alignment based on Minimum Distance

The computational complexity of the proposed face identification methodology using two-pass 3D face alignment based on minimum distance is given in Table 5.5. Referring to Table 5.5 the values 345600000 and 230400000 represent multiplication and addition operations required to align the point cloud of size $M = 300000$ points. The terms $M \times 9 \times 3$ and $M \times 6 \times 3$ represent 9 multiplication and 6 addition operations involved to rotate each point in multi-view synthesis at 3 views (whereas fourth view at 0° is obtained in alignment phase) respectively. The term $m \times n \times 370 \times 4$ represents additions and multiplications involved in PCA based subspace projection for a gallery of size 370 images synthesized at 4 views for $m \times n = 500 \times 250$. Each of MahCos distance, Euclidean distance, Mah distance

and Manhattan distance is computed for mn dimensional 370 images synthesized at 4 views. The computational complexity of the proposed methodology (in big O notation) as analyzed in terms of face alignment and face recognition is of the order of $O(P)$ and $O(mn)$ as discussed in section 4.3.4. Comparing the computational complexity of the proposed methodology given in Tables 5.3, 5.4 and 5.5 to the complexity of the Iterative Closest Point given in Table 4.3, it is evident that the proposed methodology is computationally inexpensive than Iterative Closest Point based algorithms.

Table 5.3 Computational complexity analysis using two-pass 3D face alignment based on classification approach

		Multiplications	Additions
Proposed Methodology	MahCos Distance	$226800588 +$ $M \times 9 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $2 \times (m \times n) \times 370 \times 7$ $=2185500588$	$151501029 +$ $M \times 6 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $(m \times n) \times 370 \times 7$ $=1457301029$
	Euclidean Distance	$226800588 +$ $M \times 9 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $(m \times n) \times 370 \times 7$ $=1538000588$	$151501029 +$ $M \times 6 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $2 \times (m \times n) \times 370 \times 7$ $=2104801029$
	Mah Distance	$226800588 +$ $M \times 9 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $2 \times (m \times n) \times 370 \times 7$ $=2185500588$	$151501029 +$ $M \times 6 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $2 \times (m \times n) \times 370 \times 7$ $=2104801029$
	Manhattan Distance	$226800588 +$ $M \times 9 \times 6 +$ $(m \times n) \times 370 \times 7$ $=890500588$	$151501029 +$ $M \times 6 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $2 \times (m \times n) \times 370 \times 7$ $=2104801029$

Table 5.4 Computational complexity analysis using two-pass 3D face alignment based on minimum distance

		Multiplications	Additions
Proposed Methodology	MahCos Distance	$226800000 +$ $M \times 9 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $2 \times (m \times n) \times 370 \times 7$ $=2185500000$	$151500000 +$ $M \times 6 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $(m \times n) \times 370 \times 7$ $=1457300000$
	Euclidean Distance	$226800000 +$ $M \times 9 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $(m \times n) \times 370 \times 7$ $=1538000000$	$151500000 +$ $M \times 6 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $2 \times (m \times n) \times 370 \times 7$ $=2104800000$
	Mah Distance	$226800000 +$ $M \times 9 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $2 \times (m \times n) \times 370 \times 7$ $=2185500000$	$151500000 +$ $M \times 6 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $2 \times (m \times n) \times 370 \times 7$ $=2104800000$
	Manhattan Distance	$226800000 +$ $M \times 9 \times 6 +$ $(m \times n) \times 370 \times 7$ $=890500000$	$151500000 +$ $M \times 6 \times 6 +$ $(m \times n) \times 370 \times 7 +$ $2 \times (m \times n) \times 370 \times 7$ $=2104800000$

Table 5.5 Computational complexity analysis for profile face images using two-pass 3D face alignment based on minimum distance

		Multiplications	Additions
Proposed Methodology	MahCos Distance	$345600000 +$ $M \times 9 \times 3 +$ $(m \times n) \times 370 \times 4 +$ $2 \times (m \times n) \times 370 \times 4$ $=908700000$	$230400000 +$ $M \times 6 \times 3 +$ $(m \times n) \times 370 \times 4 +$ $(m \times n) \times 370 \times 4$ $=605800000$
	Euclidean Distance	$345600000 +$ $M \times 9 \times 3 +$ $(m \times n) \times 370 \times 4 +$ $(m \times n) \times 370 \times 4$ $=723700000$	$230400000+$ $M \times 6 \times 3 +$ $(m \times n) \times 370 \times 4 +$ $2 \times (m \times n) \times 370 \times 4$ $=790800000$
	Mah Distance	$345600000+$ $M \times 9 \times 3 +$ $(m \times n) \times 370 \times 4 +$ $2 \times (m \times n) \times 370 \times 4$ $=908700000$	$230400000+$ $M \times 6 \times 3 +$ $(m \times n) \times 370 \times 4 +$ $2 \times (m \times n) \times 370 \times 4$ $=790800000$
	Manhattan Distance	$345600000 +$ $M \times 9 \times 3 +$ $(m \times n) \times 370 \times 4$ $=538700000$	$230400000 +$ $M \times 6 \times 3 +$ $(m \times n) \times 370 \times 4 +$ $2 \times (m \times n) \times 370 \times 4$ $=790800000$

5.3 Summary

In this chapter, an automatic, pose and expression invariant approach for 3D face recognition has been presented using multi-view face images synthesized at 0° , $\pm 10^\circ$, $\pm 20^\circ$ and $\pm 30^\circ$. The face images used in the study were aligned using two different alignment approaches, namely, two-pass 3D face alignment based on classification approach and two-pass 3D face alignment based on minimum distance. The study also addressed the face identification problem of profile face images aligned using two-pass 3D face alignment based on minimum distance. For face verification set-up, SVM classifier based fusion was employed using pairwise scores calculated from multi-view face images using MahCos, Euclidean, Mah and Manhattan distance metrics. In the face identification scenario, a three stage unified classifier based fusion approach was presented using aforementioned distance metrics as base classifiers along with score and decision based fusion algorithms. The computational complexity analysis of the proposed methodology was also given using MahCos distance, Euclidean distance, Mah distance and Manhattan distance based classifiers. The results obtained from experiments performed on FRGC v2.0 and GavabDB databases demonstrated that the proposed approach outperformed other state-of-the-art methods [56], [5], [66], [13], [67], [55].

Chapter 6

DISCUSSION

In this chapter, properties of the proposed face alignment and recognition approaches, their ability to handle disadvantages of existing techniques and their differences from the existing methods are given. The results related discussion and limitations of the proposed alignment and face recognition approaches are also presented.

6.1 Properties of the Proposed Approaches

The proposed 3D face recognition system is comprised of alignment and recognition modules. In order to align the face images three alignment algorithms and two face recognition algorithms were proposed in this dissertation.

6.1.1 Alignment Algorithms

The first alignment module is based on intrinsic coordinate system and it does not register two face images to each other using the conventional registration process, rather it transforms each 3D scan to an intrinsic coordinate system using vertical symmetry plane of the face, slope of the nose bridge and nose tip which define landmark structures instead of landmarks, which mark position only. The choice of landmark structures is beneficial because they remain stable even under pose variations and facial expressions. The proposed algorithm is able to align neutral and expressive faces acquired at frontal and non-frontal poses.

The second alignment algorithm employs a two-pass alignment approach where first pass is based on intrinsic coordinate system to align the face images and greatly reduces the computational cost. The second pass uses minimum distance between nose tip and the 3D scanner in xz and yz planes to align the face image. Because this algorithm uses minimum distance feature along with intrinsic coordinate system, it gives further improvement in the quality of the alignment and is capable of aligning neutral and expressive faces acquired at frontal and non-frontal poses including profile face images.

The third two-pass alignment algorithm employs a classification based approach where first pass aligns the face image using intrinsic coordinate system and greatly reduces the computational cost. The second alignment pass employs Support Vector Machine (SVM) classifier to align a face using four features: minimum distance of nose tip to the 3D scanner and maximum variance of the face image in xz and yz planes. Because the proposed alignment module employs four features along with intrinsic coordinate system, it further improves the alignment quality. The proposed algorithm is capable of aligning neutral and expressive faces acquired at frontal and non-frontal poses.

6.1.2 Face Recognition Algorithms

The first region based face recognition algorithm employs a two-tier ensemble classification approach, because ensemble classifiers greatly improve the performance of individual classifiers [52]. The proposed approach employs Principal Component Analysis (PCA) for feature extraction and Mahalanobis Cosine (MahCos) matching score for classification of facial regions. The use of Principal Component Analysis is beneficial because it is easy to implement and produces superior results along with MahCos distance metric [27]. By the selection of multiple small regions on the face, any error caused by a single region can be compensated by fusing the matching scores from multiple regions, thus making the recognition more robust to artifacts, wrinkles, facial hair or expression variations [25].

In the second face recognition approach, multi-view faces are synthesized to exploit real 3D information. For face verification, the multi-view faces are classified using Principal Component Analysis based features with MahCos as distance metric and Support Vector Machine as a classifier. Due to use of real 3D information, Principal Component Analysis along with MahCos distance metric and Support Vector Machine based ensemble classification in the score space; the proposed face recognition approach is effective and produces excellent results. The proposed approach also performs effectively by combining results from seven multi-view faces based on Principal Component Analysis; MahCos distance metric and ensemble classifier in an identification scenario and is also capable of recognizing profile face images.

6.2 Handling Disadvantages of the Existing Techniques

In order to overcome the disadvantages of existing techniques, the proposed system employs three different face alignment algorithms along with two face recognition approaches, namely, region based approach and multi-view synthesis based approach.

In 2D face recognition, face images can be aligned in xy plane only whereas in 3D they can be aligned in xz and yz planes as well. A few 3D alignment techniques [12] [14] [70] [16] existing in literature are based on Iterative Closest Point (ICP) [12], Simulated Annealing (SA) [14], Average Face Model (AFM) [70] and Intrinsic Coordinate System (ICS) [16].

Iterative Closest Point based procedure aligns two 3D facial surfaces by iteratively minimizing distance between them; whereas Simulated Annealing based approach employs a local search based stochastic algorithm. Drawbacks of Iterative Closest Point include initial course alignment and slow convergence while Simulated Annealing's limitation is that it suffers from excessive time consumption which is comparable to Iterative Closest Point.

In Average Face Model based alignment, the Average Face Model is constructed by localizing and averaging landmark points on the facial images and probe image is aligned to the Average Face Model. A significant weakness of the Average Face Model based method is less accurate alignment of a probe image to an Average Face Model due to loss of spatial information during averaging process.

The fourth method, alignment to an Intrinsic Coordinate System; mainly involves landmark localization on 3D faces, comparison to corresponding points on Intrinsic Coordinate System and a transformation step to finish alignment. The drawback of this alignment method is low accuracy of landmark localization process especially in case of non-frontal and expressive faces.

Because the proposed alignment algorithms are 3D and can align face images in xz , yz and xy planes, the xy plane alignment limitation of 2D is overcome. Similarly, the proposed alignment algorithms neither require initial course alignment step nor converge slowly; therefore, disadvantages of Iterative Closest Point and Simulated Annealing based alignment approaches are overcome.

Like Average Face Model based alignment, the proposed algorithms do not construct average face model in the alignment process, therefore, they are more accurate than Average Face Model based alignment.

The Intrinsic Coordinate System based alignment, which employs facial landmarks, may be less accurate specially under facial expressions. To overcome the low accuracy limitation of Intrinsic Coordinate System based alignment, the proposed algorithms align the face images to Intrinsic Coordinate System based landmark structures defined by nose tip, vertical symmetry plane of the face and the slope of the nose bridge which define stable landmarks even under facial expressions.

For face recognition, the proposed region based and multi-view synthesis based algorithms employ Principal Component Analysis based features along with Mahalanobis Cosine (MahCos) distance metric. The use of Principal Component Analysis along with MahCos distance metric produces excellent face recognition results [27]. Besides improving alignment of face images, the proposed face recognition approaches handle the disadvantages of existing techniques [55], [48], [56], [13], [5], [66] using an ensemble classifier based approach that combines results from several regions or synthesized multi-view face images.

6.3 Differences from Existing Methods

The proposed alignment and recognition techniques are different from existing methods in various aspects as under.

- Both of Iterative Closest Point and Simulated Annealing methods align face images to every other image. Unlike them, the proposed alignment algorithms align the probe face image to an Intrinsic Coordinate System in a single alignment event without the need of aligning each face to every other face.
- The drawbacks of Iterative Closest Point and Simulated Annealing based alignment techniques limit their applicability to only verification scenario where the probe image is to be aligned to claimed identity only but they are an issue in identification scenario where a probe is to be aligned to the whole gallery. On the other hand, the proposed alignment algorithms align a face image to intrinsic coordinate system only once; therefore, they can be used in

identification as well as verification scenarios.

- Existing methods do not handle alignment of profile face images, whereas the proposed two-pass distance based alignment algorithm is capable of aligning profile face images.
- None of the existing systems employs a classification based approach to face alignment. In contrast, the proposed two-pass 3D face alignment based on classification approach utilizes support vector machine classifier for alignment of the face images.
- The Iterative Closest Point and Simulated Annealing approaches directly employ 3D point clouds in alignment and recognition process. Although the proposed algorithms align the faces by using 3D point clouds but the matching process uses depth images instead of point clouds.
- The first proposed face recognition algorithm employs a region based approach where each region is classified using Principal Component Analysis based features and MahCos distance metric using two tier ensemble classifier whereas the studies [55] [48] [56] are not region based and use different feature extraction and classification approaches.
- The second proposed face recognition study uses an ensemble classifier based approach where multi-view faces are synthesized to exploit real 3D information and classified using Principal Component Analysis based features and MahCos distance metric whereas the studies [13] [5] [66] do not use a multi-view synthesis approach to face recognition and employ different approaches for feature extraction and classification.

6.4 Results Related Discussion

6.4.1 Alignment Approaches

The proposed study addresses the alignment problem of 3D face images. The important findings of the proposed alignment algorithms are.

- The algorithms achieved 99.95% and 99.77% alignment accuracy using FRGC v2.0 and GavabDB databases respectively. The nose tip was not detectable for two images of FRGC v2.0 and one image of GavabDB database; else the accuracy of

the proposed algorithms would have been 100%.

- The algorithms successfully aligned face images having several types of facial expressions and acquired at varying distances from the scanner.
- The ‘3D face alignment based on intrinsic coordinate system’ and ‘two-pass 3D face alignment based on classification approach’ are capable of aligning face images acquired at frontal and non-frontal face images.
- The ‘two-pass 3D face alignment based on minimum distance’ is capable to align face images captured at frontal and non-frontal poses including profile images.
- The algorithms align the face images in a single alignment event; therefore, they can be used in verification as well as identification scenarios.
- ‘Two-pass 3D face alignment based on classification approach’ obtained best alignment accuracy in terms of distance reduction measure between nose tip and 3D scanner than ‘two-pass 3D face alignment based on minimum distance’ and ‘3D face alignment based on intrinsic coordinate system’ respectively.
- ‘3D face alignment based on intrinsic coordinate system’ is computationally inexpensive than ‘two-pass 3D face alignment based on minimum distance’ and ‘two-pass 3D face alignment based on classification approach’ respectively.

Limitations: The proposed alignment approaches have the following limitations.

- The algorithms are based on nose tip heuristic and are not capable to align face images where nose tip is not detectable.
- The nose tip detection and alignment algorithms are capable to align face images captured in the range of $\pm 45^\circ$ in xz plane and $\pm 15^\circ$ in yz plane.
- ‘3D face alignment based on intrinsic coordinate system’ is not capable of aligning profile face images in xz and xy planes because in such cases either right or left half of the face is not available. The ‘two-pass 3D face alignment based on classification approach’ is not able to align profile face images as well. This algorithm employs maximum variance of the full face as a measure of frontal facial position. Because either right or left half face is not available

in case of profile face images, therefore, face is not in a frontal position when variance is the maximum. As a result maximum variance feature cannot be exploited in case of profile face images.

6.4.2 Recognition based on Regional Segments of Depth Images

The proposed study addresses the problem of 3D face recognition based on regional segments of depth images using two-tier ensemble classifier. The proposed methodology employs MahCos distance, Euclidean distance, Mah distance and Manhattan distance based classifiers. The findings of the methodology are summarized as under.

- The proposed approach successfully recognized the images acquired at varying poses.
- The experiments using MahCos distance metric obtained best classification accuracy for individual face regions, combination classifiers and re-ranking stage whereas Euclidean distance, Mah distance and Manhattan distance based experiments achieved relatively lower performance in the respective order.
- Ensemble classifier has clearly outperformed the constituent regional classifiers.
- Fusion methods enhance the performance of regional classifiers using rank based approaches.
- Regions of large size result into better recognition performance than the small cropped regions.
- Increasing the number of regional classifiers might not improve the result accuracy of the ensemble classifier. Rather the results of ensemble classifier may get worse by increasing the number of constituent regional classifiers.
- The proposed methodology employing MahCos distance is computationally inexpensive than Iterative Closest Point based algorithms.
- Results produced by the techniques presented in this study are comparable in both databases i.e. FRGC v2.0 and GavabDB database.

Limitations: The limitations of the proposed methodology are summarized as under.

- The images are sensitive to facial alignment and it is hard to detect useful face regions automatically.
- The performance is dependent on local features and resolution differences.
- The results of individual regions are to be combined using ensemble classification approaches and may not improve more than an extent if region or classifier selection is not optimal.

6.4.3 Recognition based on Multi-View Depth Images

The proposed study addresses the problem of multi-view pose and expression invariant 3D face recognition of frontal, non-frontal, and profile face images. The frontal and non-frontal face images are aligned using two-pass 3D face alignment based on classification approach and two-pass 3D face alignment based on minimum distance separately, whereas profile face images are aligned using two-pass 3D face alignment based on minimum distance. The important findings of the methodology are summarized as under.

- The two-pass 3D face alignment based on classification approach resulted in better verification and identification rates than the two-pass 3D face alignment based on minimum distance.
- Face images at angular position 0° achieved the best individual face recognition accuracy which gradually decreased for synthesized images as the facial view is deviated from 0° for both of neutral and expressive faces.
- The expression invariant area contains the maximum discriminating features resulting in improved expression invariant face recognition performance.
- MahCos distance metric performed best of Euclidean distance, Mah distance and Manhattan distance metrics respectively.
- SVM classifier based face verification method outperformed the best individual multi-view classification accuracy achieving a verification rate of 100% and 95.08% on GavabDB database while 99.57% and 96.57% at 0.1% FAR on FRGC v2.0 database for N vs. N and N vs. E experiments

respectively.

- The proposed unified classifier clearly achieved enhanced classification accuracy compared to the base classifiers or parallel face recognition algorithms whereas score level fusion based parallel face recognition algorithm performed better than the decision level based algorithm. The proposed approach obtained rank-1 identification rates of 100% and 96.72% on the GavabDB database while 98.93% and 94.42% on FRGC v2.0 database for N vs. N and N vs. E experiments, whereas identification rate of 95.08% and 83.61% was obtained on left and right profile images of GavabDB database respectively.
- The proposed methodology based on MahCos distance is computationally inexpensive than Iterative Closest Point based algorithms.
- The experiments show that face recognition results based on proposed methodology are comparable in both databases and that the proposed methodology achieved better performance for both of face verification and identification experiments than the methods compared in Table 5.1 and 5.2.

Limitations: The proposed face recognition methodology has following limitations.

- Multi-view synthesis of face images at large rotation angles results in self-occlusion, and most of the discriminating facial features around nose and eye area become occluded leading to deterioration of recognition rates.
- Multi-view synthesis, PCA based subspace learning and computation of MahCos distance, Euclidean distance, Mah distance and Manhattan distance of individual views (in separate experiments) increases computational cost of the method.
- A single classifier is not able to classify face images at all synthesized views and recognition results of multi-view face images are to be combined using an ensemble classification approach which is cumbersome to define.

6.5 Summary

In this chapter, properties of the proposed approaches have been presented. The proposed approaches effectively aligned and recognized frontal, profile, neutral and non-neutral face images. The proposed methods successfully overcame limitations of the existing methods and offered various differences from them in several aspects. The key result insights and limitations of the proposed face alignment and recognition techniques have also been discussed.

Chapter 7

CONCLUSION AND FUTURE WORK

This dissertation is focused on the topic of 3D face recognition based on pose and expression invariant alignment. In chapter 1, face recognition challenges, motivation, disadvantages of the existing techniques, aims and objectives, and contributions were given. In chapter 2, 3D alignment and face recognition approaches were reviewed. In chapter 3, three novel face alignment algorithms were presented to handle 3D face images. Chapter 4 was focused on region based face recognition using ensemble classification approach. Chapter 5 presented face verification and identification approaches using multi-view synthesized face images based on ensemble classifier. Properties of the proposed approaches, handling of disadvantages of the existing techniques, differences from existing methods, results related discussion and limitations of the developed algorithms are given in chapter 6. The results of the research work reveal that the aims to develop accurate and computationally inexpensive 3D face alignment and recognition algorithms were achieved. In this chapter, the dissertation is concluded and some suggestions for the future work are given.

7.1 Contributions

The proposed methodology offered the following contributions:

- In order to reliably detect the nose tip in face images under large pose variations, and acquired in noisy scenarios, several novel preprocessing steps were proposed as given in chapter 3.
- The first alignment algorithm was based on Intrinsic Coordinate System using vertical symmetry plane of the face, slope of the nose bridge and nose tip defining landmark structures instead of landmarks, which mark position only. The choice of landmark structures was beneficial because they remain stable even under pose variations and facial expressions. The proposed algorithm

successfully aligned neutral and expressive faces acquired at frontal and non-frontal poses.

- The second alignment algorithm employed a two-pass approach based on minimum distance where first pass used Intrinsic Coordinate System to align the face images while greatly reducing the computational cost. The second pass utilized minimum distance between nose tip and the 3D scanner in xz and yz planes to align the face images. Because this algorithm utilized minimum distance feature along with Intrinsic Coordinate System, it resulted into further improvement in the quality of the alignment and successfully aligned neutral and expressive faces acquired at frontal and non-frontal poses including profile face images.
- The third two-pass alignment algorithm used a classification based approach where first pass aligned the face image employing Intrinsic Coordinate System which greatly reduced the computational cost. The second alignment pass utilized Support Vector Machine (SVM) classifier to align a face using four features: minimum distance between the nose tip and the 3D scanner and maximum variance of the face image in xz and yz planes. Because the proposed alignment approach utilized four features along with Intrinsic Coordinate System, it further improved the quality of the alignment. The proposed algorithm aligned the neutral and expressive faces acquired at frontal and non-frontal poses successfully.
- A novel approach was proposed to quantify the quality of the alignment results based on the fact that when pose of a face is corrected from non-frontal to frontal position, its nose tip distance from the 3D scanner is essentially reduced. Based on the criterion of distance reduction from nose tip to the 3D scanner as a measure of alignment, the proposed algorithms achieved 99.95% and 99.77% alignment accuracy using FRGC v2.0 and GavabDB databases respectively. The nose tip was not detectable for two subjects in the FRGC v2.0 database and one subject in GavabDB database else the alignment accuracy of the proposed algorithms would have been 100%.
- The first face recognition method presented in this dissertation employed a region based, two tier ensemble classification approach. Principal Component

Analysis (PCA) was employed for feature extraction whereas Mahalanobis Cosine (MahCos) distance, Euclidean distance, Mahalanobis (Mah) distance and Manhattan distance based matching scores were used for classification of individual facial regions in separate experiments. The classification results obtained from individual regions were combined using Weighted Borda Count (WBC) based combination and a re-ranking stage. The recognition was based on 15 regional classifiers, used as 3 sets of 5 classifiers. By employing two different fusion techniques, i.e. Weighted Borda Count (WBC) and re-ranking, the results of regional classifiers were fused into 3 combination classifiers which were then combined into a final score. The performance of the proposed approach was corroborated by extensive experiments performed on two databases: FRGC v2.0 and GavabDB and 98.93% and 100% rank-1 recognition rates were obtained using these databases respectively.

- The second face recognition algorithm presented in this dissertation employed a multi-view synthesis based approach to classify 3D faces including profile face images. For the face recognition framework, multi-view 3D faces were synthesized to exploit real 3D facial information. Inspired by the effectiveness of fusion approaches, Support Vector Machine (SVM) was employed using scores obtained from multi-view face pairs for face verification.
- In addition, a three stage unified classifier based face identification algorithm was employed which combined results from seven base classifiers at first stage, two parallel face recognition algorithms at second stage and an exponential rank combiner at third stage in a hierarchical manner. For profile face images, the three stage unified classifier based face identification algorithm combined results from four base classifiers, two parallel face recognition algorithms and an exponential rank combiner in a hierarchical manner.

The performance of the proposed methodology was demonstrated by extensive experiments performed on two databases, namely, FRGC v2.0 and GavabDB. The results exhibited that the proposed methodology can be efficiently used to construct an automatic, pose and expression invariant facial recognition system. The SVM classifier based face verification method achieved 99.57%

and 96.57% verification rate at 0.1% FAR; whereas unified classifier based face identification method obtained 98.93% and 94.42% rank-1 identification rates for Neutral vs. Neutral (N vs. N) and Neutral vs. Non-neutral (represented by N vs. E) experiments respectively using FRGC v2.0 database. Using GavabDB database, the proposed identification and verification methods achieved 100% and 96.72% rank-1 identification rates while 100% and 95.08% verification rates for N vs. N and N vs. E experiments respectively. Similarly identification rate of 95.08% and 83.61% was obtained on left and right profile images of GavabDB database respectively.

7.1.1 Comparison among the Proposed Techniques

The two-pass alignment algorithm based on classification approach produced better alignment results than the two-pass alignment algorithm based on minimum distance and Intrinsic Coordinate System based alignment. The comparative advantages and disadvantages of the proposed alignment approaches have been given in section 6.4.1.

The multi-view synthesis based face recognition approach exhibited superior recognition rates than the region based face recognition approach. On the other hand, the latter approach is computationally inexpensive than the former and is capable to handle facial hair, artifacts, wrinkles and local shape deformations caused by facial expression variations. Therefore, in case of good quality images multi-view synthesis based approach may be used whereas for noisy and relatively poor quality images, region based approach should be preferred. The comparative advantages and disadvantages of region based and multi-view synthesis based face recognition approaches have been enlisted in sections 6.4.2 and 6.4.3 respectively.

7.2 Future Work

Facial recognition systems are commonly employed in real life compared to other biometric modalities (like finger print and palmprint) due to the fact that face image acquisition is non-invasive and the subjects can be captured without their cordial cooperation. Because of promising accuracy of facial recognition systems, future research may be pursued in this field in following directions:

- The proposed alignment approach is based on determining correction parameters in xz , yz and xy planes which are then used to correct the 3D pose of the probe face. As described in Chapter 3, in xy plane the correction parameter γ is determined by using different approaches. The approach employing manual detection of landmarks on inner eye corners can be improved by devising an automatic landmarks detection algorithm. The approach employing Average Face Model (AFM) for missing facial profile synthesis can be improved by devising an algorithm which can better estimate the missing profile.
- In the proposed region based 3D face recognition approach, 15 face regions have been employed in 3 sets of 5 regions using Principal Component Analysis (PCA) based holistic features along with Weighted Borda Count (WBC) and re-ranking approach. The proposed approach may be improved by using other better performing regions and diverse ensemble classification approaches. Another future investigation approach is to study the impact of feature based region level classification techniques for fusing their results to improve recognition performance.
- The proposed multi-view depth images based recognition employs Support Vector Machine (SVM) classifier for verification scenario and a three stage ensemble classifier for identification scenario using images synthesized at seven different angles in xz plane and PCA based holistic features. The SVM based face verification performance may be improved by synthesizing more views in xz and yz planes. In case of face identification, a diverse ensemble classifier may further improve the identification rate, whereas local facial features can also be modeled for improving the recognition performance.
- Given the current trend of using deep learning in every possible application, Deep Neural Networks (DNNs) [71] have established themselves as a dominant technique in machine learning [72]. In contrast to conventional methods, feature learning methods learn a feature extractor based on the statistics of the training data and have been successfully applied to a variety of different domains and modalities. Recently, DNNs have been applied to 2D face recognition but a downside of DNNs is that they require large

amounts of training data, preferably tens of millions of images [72]. DNNs based approaches can be extended to 3D face recognition but currently available 3D datasets are small and do not contain enough images to exploit modern deep architectures. The proposed face recognition approaches can be successfully evaluated using DNNs in future with the availability of larger datasets.

REFERENCES

- [1] S. Z. Li and A. K. Jain, "Handbook of face recognition," Springer-Verlag, New York, USA, 2005.
- [2] W. Zhao, "Face recognition: a literature survey," *ACM computing survey*, vol. 35, pp. 399-458, 2003.
- [3] P. J. Phillips, H. Moon, S. A. Rizvi and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [4] H. Zhang, N. M. Nasrabadi, Y. Zhang and T. S. Huang, "Joint dynamic sparse representation for multi-view face recognition," *Pattern Recognition*, vol. 45, no. 4, pp. 1290–1298, 2012.
- [5] S. Berretti, A. Del Bimbo and P. Pala, "3D face recognition using isogeodesic stripes," *Pattern Analysis and Machine Intelligence*, vol. 32, no. 12, pp. 2162-2177, 2010.
- [6] S. Gupta, M. K. Markey and A. C. Bovik, "Advances and challenges in 3D and 2D+3D human face recognition," in *Pattern Recognition Research Horizons*, pp. 161-200, Nova Science Publishers, New York, 2007.
- [7] Z. Sun, A. A. Paulino, J. Feng, Z. Chai, T. Tan and A. K. Jain, "A study of multi biometric traits of identical twins," in *Proceedings of SPIE, Biometric Technology for Human Identification VII*, vol. 7667, 2010.
- [8] U. Park, Y. Tong and A. Jain, "Age-invariant face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 5, pp. 947–954, May 2010.
- [9] N. I. Ratyal, I. A. Taj, U. I. Bajwa and M. Sajid, "3D face recognition based on pose and expression invariant alignment," *Computers & Electrical Engineering*, vol. 46, pp. 241-255, 2015.
- [10] N. I. Ratyal, I. A. Taj, U. I. Bajwa, M. Sajid, M. J. A. Baig and F. M. Butt, "3D face recognition based on region ensemble and hybrid features," in

Proceedings of the IEEE International Conference on Computing, Electronic and Electrical Engineering, 2016.

- [11] N. I. Ratyal, I. A. Taj, U. I. Bajwa, M. Sajid, “Automatic multi-view 3D face recognition based on two-pass pose and expression invariant alignment,” *Image and Vision Computing*. (Submitted, Under Review)
- [12] 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.
- [13] X. Wang, Q. Ruan, Y. Jin and G. An, “Three-dimensional face recognition under expression variation,” *EURASIP Journal on Image and Video Processing*, 2014:51, 2014.
- [14] C. C. Queirolo, L. Silva, O. R. Bellon and M. P. Segundo, “3D face recognition using simulated annealing and the surface interpenetration measure,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 206–219, 2010.
- [15] N. Alyüz, B. Gökberk and L. Akarun, “Regional registration for expression resistant 3-D face recognition,” *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 425-440, Sep. 2010.
- [16] T. Papatheodorou and D. Rueckert, “3D face recognition,” Vienna: I-Tech Education and Publishing, 2007.
- [17] C. C. Queirolo, L. Silva, O. R. P. Bellon and M. P. Segundo, “3D face recognition using the surface interpenetration measure: a comparative evaluation on the FRGC database,” in *Proceedings of International Conference on Pattern Recognition*, pp. 1-5, 2008.
- [18] C. Goodall, “Procrustes methods in the statistical analysis of shape,” *Journal of the Royal Statistical Society, Series B (Methodological)*, pp. 285–339, 1991.
- [19] B. Gökberk, H. Dutağacı, A. Ulas, L. Akarun and B. Sankur, “Representation plurality and fusion for 3-D face recognition,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 38, no. 1, pp. 155-173,

2008.

- [20] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no.1, pp. 71-86, 1991.
- [21] M. S. Bartlett, J. R. Movellan and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Transactions on Neural Networks*, vol. 13, pp. 1450-1464, 2002.
- [22] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, pp. 411-430, May 2000.
- [23] K. Baek, B. Draper, J. R. Beveridge and K. She, "PCA vs. ICA: A comparison on the FERET data set," in *Proceedings of International Conference on Computer Vision, Pattern Recognition and Image Processing*, March. 2002.
- [24] P. Belhumeur, J. Hespanha and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711-720, 1997.
- [25] J. R. Beveridge, K. She, B. A. Draper and G. H. Givens, "A nonparametric statistical comparison of principal component and linear discriminant subspaces for face recognition," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 2001.
- [26] X. He, S. Yan, Y. Hu, P. Niyogi and H. J. Zhang, "Face recognition using laplacianfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 328-340, 2005.
- [27] U. I. Bajwa, I. A. Taj, M. W. Anwar and X. Wang, "A multifaceted independent performance analysis of facial subspace recognition algorithms," *PLOS ONE*, vol. 8, no. 2:e56510, 2013.
- [28] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vision Research*, vol. 20, no. 10, pp. 847-856, 1980.
- [29] D. Huang, "Robust face recognition based on three dimensional data," Ph. D. Thesis, May, 2012.
- [30] T. S. Lee, "Image representation using 2D Gabor wavelets," *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, vol. 18, no. 10, pp. 959-971, 1996.

- [31] C. J. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," IEEE Transactions on Image Processing, vol. 11, no. 4, pp. 467-476, 2002.
- [32] D. H. Liu, K. M. Lam and L. S. Shen, "Optimal sampling of Gabor features for face recognition," Pattern Recognition Letters, vol. 25, pp. 267-276, 2004.
- [33] D. Fields, "Relations between the statistics of natural images and the response properties of cortical cells," Journal of the Optical Society of America, vol. 4, no. 12, pp. 2379-2394, 1987.
- [34] J. Cook, V. Chandran and S. Sridharan, "Multiscale representation for 3-D face recognition," IEEE Transactions on Information Forensics and Security, vol. 2, no. 3, pp. 529-536, Sept. 2007.
- [35] J. A. Cook, "A decompositional investigation of 3D face recognition," Ph. D. Thesis, Dec. 2007.
- [36] T. Ojala, M. Pietikainen and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," Pattern Recognition, vol. 29, no. 1, pp. 51-59, Jan. 1996.
- [37] T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, 2002.
- [38] T. Ahonen, A. Hadid and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, pp. 2037-2041, 2006.
- [39] A. Aissaoui, J. Martinet, C. Djeraba, "DLBP: A novel descriptor for depth image based face recognition," in Proceedings of IEEE International Conference on Image Processing, Oct. 2014.
- [40] D. G. Lowe, "Distinctive image features from scale-invariant keypoints,"

- International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.
- [41] A. S. Mian, "Representations and matching techniques for 3D free-form object and face recognition," Ph. D. Thesis, 2006.
- [42] C. S. Chua and R. Jarvis, "Point signatures: A new representation for 3D object recognition," *International Journal of Computer Vision*, vol. 25, no. 1, pp. 63–85, 1997.
- [43] A. S. Mian, M. Bennamoun and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 11, pp. 1927-1943, 2007.
- [44] Picozzi Marta, Macchi Cassia Viola and Turati Chiara, "The development of configural face processing: The face inversion effect in preschool-aged children," in Annual meeting of the XVth Biennial International Conference on Infant Studies, June, 2006.
- [45] A. Schwaninger, C. Wallraven, D. W. Cunningham and S. Chiller-Glaus, "Processing of identity and emotion in faces: A psychophysical, physiological and computational perspective," *Progress in Brain Research*, vol. 156, pp. 321-343, 2006.
- [46] W. G. Hayward, G. Rhodes and A. Schwaninger, "An own-race advantage for components as well as configurations in face recognition," *Cognition*, vol. 106, no. 2, pp. 1017-1027, 2008.
- [47] L-F. Chen, H-Y. Liao, J-C. Lin and C-C. Han, "Why recognition in a statistics-based face recognition system should be based on the pure face portion: A probabilistic decision-based proof," *Pattern Recognition*, vol. 34, no. 7, pp. 1393-1403, 2001.
- [48] M. H. Mahoor and M. Abdel-Mottaleb, "Face recognition based on 3D ridge images obtained from range data," *Pattern Recognition*, vol. 42, no. 3, pp. 445–451, 2009.
- [49] L. Spreuwers, "Fast and accurate 3D face recognition using registration to an intrinsic coordinate system and fusion of multiple region classifiers,"

International Journal of Computer Vision, vol. 93, pp. 389-414, 2011.

- [50] A. B. Moreno and A. Sanchez, "GavabDB: a 3D face database," in Proceedings of Second COST Workshop on Biometrics on the Internet: Fundamentals, Advances and Applications, pp. 77-82, 2004.
- [51] G. G. Slabaugh, "Computing Euler angles from a rotation matrix," Technical Report, 1999. Available at: <http://www.gregslabaugh.name/publications/euler.pdf>. (Date of access: 03-Sept.-2014).
- [52] U. I. Bajwa, I. A. Taj and M. W. Anwar, "A unified classifier for robust face recognition based on combining multiple subspace algorithms," Optics Communications, vol. 285, no. 21-22, pp. 4324-4332, Oct. 2012.
- [53] N. Alyüz, B. Gökberk and L. Akarun, "3-D face recognition under occlusion using masked projection," IEEE Transactions on Information Forensics and Security, vol. 8, pp. 789–802, 2013.
- [54] P. J. Phillips, P. J. Flynn, T. Scruggs et al., "Overview of the face recognition grand challenge," in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 947-954, 2005.
- [55] L. Zhang, Z. Ding, H. Li, Y. Shen and J. Lu, "3D face recognition based on multiple keypoint descriptors and sparse representation," PLOS ONE, vol. 9, no. 6: e100120, 2014.
- [56] F. R. Al-Osaimi, M. Bennamoun and A. Mian, "Integration of local and global geometrical cues for 3D face recognition," Pattern Recognition, vol. 41, no. 3, pp. 1030–1040, 2007.
- [57] T. Faltemier, K. Bowyer and P. Flynn, "A region ensemble for 3-D face recognition," IEEE Transactions on Information Forensics and Security, vol. 3, no. 1, pp. 62-73, Mar. 2008.
- [58] H. Mohammadzade and D. Hatzinakos, "An expression transformation for improving the recognition of expression-variant faces from one sample image per person," in Proceedings of IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), pp. 1-6, Sept. 2010.

- [59] K. Niinuma, H. Han and A. K. Jain, “Automatic multi-view face recognition via 3D model based pose regularization,” in Proceedings of IEEE International Conference on Biometrics: Theory Applications and Systems (BTAS), pp. 1-8, Oct. 2013.
- [60] A. Moeini, H. Moeini and K. Faez, “Pose-invariant facial expression recognition based on 3D face reconstruction and synthesis from a single 2D image,” in Proceedings of IEEE International Conference on Pattern Recognition, pp. 1746-1751, Aug. 2014.
- [61] M. Du, A.C. Sankaranarayanan and R. Chellappa, “Robust face recognition from multi-view videos,” IEEE Transactions on Image Processing, vol. 23, no.3, pp.1105-1117, March 2014.
- [62] Y. Guo, X. Ding and J. Xue, “MiLDA: A graph embedding approach to multi-view face recognition,” Neurocomputing, vol. 151, pp. 1255–1261, March 2015.
- [63] M. Kan, S. Shan, H. Zhang, S. Lao and X. Chen, “Multi-view discriminant analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 1, pp. 188-194, 2016.
- [64] A. A. Ross, K. Nandakumar and A. K. Jain, “Levels of fusion in biometrics,” Handbook of Multibiometrics, vol. 6, pp. 59-90, Springer US, 2006.
- [65] A. A. Ross, K. Nandakumar and A. K. Jain, “Score level fusion,” Handbook of Multibiometrics, vol. 6, pp. 91-142, Springer US, 2006.
- [66] G. Chang and Y. Wang, “Robust 3D face recognition based on resolution invariant features,” Pattern Recognition Letters, vol. 32, no. 7, pp. 1009–1019, 2011.
- [67] S. Berretti, A. Del Bimbo, P. Pala, “Sparse matching of salient facial curves for recognition of 3-D faces with missing parts,” IEEE Transactions on Information Forensics and Security, vol. 8, no. 2, pp. 374-389, 2012.
- [68] X. Li, T. Jia and H. Zhang, “Expression-insensitive 3D face recognition using sparse representation,” in IEEE Computer Society Conference on

Computer Vision and Pattern Recognition, pp. 2575–2582, 2009.

- [69] W. Hariri, H. Tabia, N. Farah and A. Benouareth, “3D Face recognition using covariance based descriptors,” *Pattern Recognition Letters*, 2016. doi: 10.1016/j.patrec.2016.03.028
- [70] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis and T. Theoharis, “Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 640-649, 2007.
- [71] Y. Taigman, M. Yang, M. A. Ranzato and L. Wolf, “Deepface: Closing the gap to human-level performance in face verification,” in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, 2014.
- [72] S. Balaban, “Deep learning and face recognition: The state of the art,” in *Proceedings of SPIE International Conference on Biometric and Surveillance Technology for Human and Activity Identification XII*, vol. 9457, pp. 94570B–94570B–8, 2015.
- [73] Y. Baocai, S. Yanfeng, W. Chengzhang and G. Yun, “BJUT-3D large scale 3D face database and information processing,” *Journal of Computer Research and Development*, vol. 46, no. 6, pp. 1009-1018, 2009.
- [74] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur and L. Akarun, “Bosphorus database for 3D face analysis,” *Biometrics and Identity Management*, vol. 5372, pp. 47-56, Springer Berlin Heidelberg, 2008.
- [75] L. Yin, X. Wei, Y. Sun, J. Wang and M. J. Rosato, “A 3D facial expression database for facial behavior research,” in *Proceedings of International Conference on Automatic Face and Gesture Recognition*, pp. 211-216, 2006.
- [76] C. Conde, A. Serrano, L. J. Rodriguez-Aragon and E. Cabello, “An automatic 2D, 2.5D & 3D score-based fusion face verification system,” in *Proceedings of IEEE International Conference on Application-Specific*

Systems, Architectures and Processors, pp. 214-219, 2007.

- [77] N. F. Troje and H. H. Bulthoff, "Face recognition under varying poses: the role of texture and shape," *Vision Research*, vol. 36, no. 12, pp. 1761-1771, 1996.
- [78] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proceedings of Annual Conference on Computer Graphics and Interactive Techniques*, pp. 187-194, 1999.
- [79] K. Messer, J. Kittler, M. Sadeghi et al., "Face verification competition on the XM2VTS database," in *Proceedings of International Conference on Audio and Video Based Biometric Person Authentication*, pp. 964-974, 2003.
- [80] T. Heseltine, "Face recognition: two-dimensional and three-dimensional techniques," thesis (PhD), The University of York, 2005.
- [81] C. Beumier and M. Acheroy, "Face verification from 3D and grey level clues," *Pattern Recognition Letters*, vol. 22, no. 12, pp. 1321-1329, Oct. 2001.
- [82] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer and W. Worek, "Preliminary face recognition grand challenge results," in *Proceedings of International Conference on Automatic Face and Gesture Recognition*, pp. 15-24, 2006.
- [83] K. Bowyer, K. Chang and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1-15, 2006.
- [84] D. Smeets, J. Keustermans, D. Vandermeulen and P. Suetens, "MeshSIFT: local surface features for 3D face recognition under expression variations and partial data," *Computer Vision and Image Understanding*, vol. 117, no. 2, pp. 158-169, 2013.

Appendix- 3D Face Databases

Several 3D databases have been acquired and released for research on face recognition problems. Examples include BJUT-3D [73], Bosphorus [74], BU-3DFE [75], FRAV3D [76], FRGC v2.0 [54], GavabDB [50], MPI [77], USF 3D [78], XM2VTS [79], York [80] and 3D_RMA [81] shown in Table A.1 with corresponding database characteristics. The labels E, I, O and P are acronyms for expression, illumination, occlusion and pose respectively.

The objective of the studies presented in this dissertation was to propose automatic face alignment and recognition algorithms which can tackle neutral and expressive faces acquired at frontal and non-frontal poses including profile face images. Keeping in view these circumstances, two 3D databases namely, FRGC v2.0 and GavabDB were selected to conduct experiments. The selected databases carry 3D images acquired with neutral and expressive facial conditions under frontal and non-frontal poses. The choice of these databases was also motivated by the fact that most of the state-of-the-art face recognition algorithms employ these databases to report the performance, therefore, implementation of these databases in this research facilitated for a direct comparison of results with the state-of-the-art methods. FRGC v2.0 and GavabDB databases are reviewed in the following subsections.

A.1 FRGC v2.0 Database

FRGC v2.0 [54] is a publically available license based database which consists of 50,000 recordings divided into training and validation partitions. It supports 6 experiments, however, the focus was on experiment 3 in which gallery and probe sets consist of both the shape and texture images for each subject. Experiment 3 allows for an assessment of the contribution of the shape and textures to the performance of 3D facial imagery [82]. 3D scans are provided in form of four matrices of size 480 x 640 each, out of which three represent x, y and z coordinates of scans, whereas z is the distance from the scanner.

Table A.1 Introduction of public databases

Database	Total Subjects	Images/ Subject	Texture	Characteristics	Acquisition Device
BJUT-3D	100	-	Yes	E	Cyberware 3030PS
Bosphorus	105	Total:46 66	Yes	P,E,O	-
BU-3DFE	100	4	No	E	3DMD
FRAV3D	106	16	Yes	P ,E	Minolta Vivid 700
FRGC v2.0	466	Total:40 07	Yes	I,E,O	Minolta Vivid 910
GavabDB	61	9	No	P,E	Minolta Vivid 700
MPI	200	7	-	-	Cyberware
USF 3D	100	1	Yes	-	Cyberware 3030PS
XM2VTS	295	1	Yes	-	-
York	350	15	Yes	P,E	-
3D_RMA	120	3	No	P,E,O	-

The fourth matrix denotes binary representation of valid 3D points in x, y and z matrices. Both male and female subjects with age 18 years and above are included in the database. Facial images of individuals have been acquired with frontal view and

pose variations. Almost sixty percent of the subjects have neutral expressions, whereas others have expressions of happiness, disgust, surprise, sadness and inflated cheeks or puffy mouth [83]. Images have been captured at varying lengths from the scanner; therefore, resolution of the scans is variable. Example 3D faces 04217d399.abs, 04233d396.abs, 04221d553.abs, 04482d418.abs, 04343d427.abs, 04385d435.abs, 04387d322.abs and 04595d149.abs from FRGC v2.0 database are shown in Fig. A.1 from top left to right.

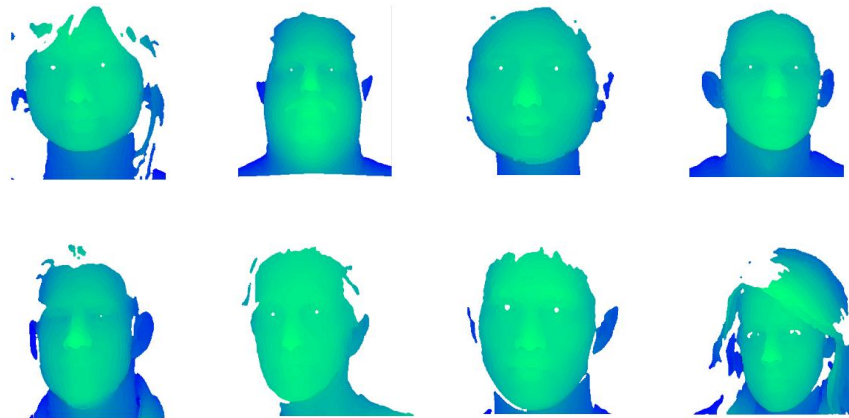


Fig. A.1 Example 3D faces from FRGC v2.0 database

In FRGC v2.0 database some of the subjects have occlusions such as hair on face but none of them is wearing glasses [43]. 2D images are also provided in one-to-one correspondence to the 3D scans and the scans have artifacts such as spikes and holes. The 3D face scans are separated in three sets, namely, Spring 2003 set (comprised of 943 scans of 277 subjects) and Fall 2003 + Spring 2004 set (consisting of 4007 scans of 466 subjects in total).

A.2 GavabDB Database

The GavabDB [50] database is comprised of 427 [48] facial scans from 61 Caucasian subjects among which 45 are male and remaining are female. For each subject there are nine facial scans even with large pose variations and accentuated expressions. The database includes two frontal scans for each subject with neutral expression while

another two scans are captured with neutral facial expression where subjects are looking up or down at nearly $+35^\circ$ and -35° respectively. Among the remaining five, three scans present smile, laugh or a random expression and two are right and left profile scans that are acquired at $+90^\circ$ and -90° rotation respectively. Example 3D faces `caral_abajo`, `caral_arriba`, `caral_frontal1`, `caral_frontal2`, `caral_derecha`, `caral_izquierda`, `caral_gesto`, `caral_risa` and `caral_sonrisa` from GavabDB database are shown in Fig. A.2, from top left to right.

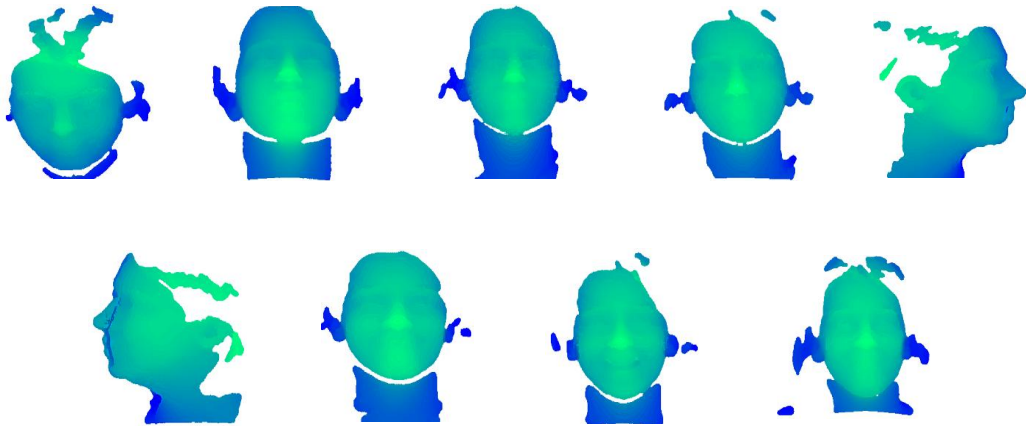


Fig. A.2 Example 3D faces from GavabDB database