3D FACE REGISTRATION USING MULTIPLE AVERAGE MODELS

by

Neşe Alyüz

B.S, in Computer Engineering, İstanbul Technical University, 2005

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computer Engineering

Boğaziçi University

2008

3D FACE REGISTRATION USING MULTIPLE AVERAGE MODELS

APPROVED BY:

Prof. Lale Akarun .....
(Thesis Supervisor)

Prof. Bülent Sankur .....

Dr. Ali Vahit Şahiner .....

DATE OF APPROVAL: 21.01.2008

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis supervisor Lale Akarun for her insight, guidance, and support throughout the preparation of this thesis.

I would like to thank the members of Satellite Laboratory, Birkan Yılmaz, Şükrü Kuran, Evren Önem, Fernaz Alimoğlu, Suzan Bayhan, Bora Zeytinci, Gürkan Gür, Onur Türkyılmaz, Kaan Bür for providing such an amusing environment and for their endless supports through difficult times. I would like to thank the members of Perceptual Intelligence Laboratory, Oya Aran for being my idol with her great personality, Cem Keskin for setting an example of diligence, and especially my spiritual twin, Pınar Santemiz for being such a cheerful friend in my most demoralized moments. I thank all the members of the CMPE basketball team, for providing all those energetic hours that comforted me in my most tense times. I also thank all other friends whose support cannot be forgotten: Rabun Koşar, Hamdi Dibeklioğlu, Oya Çeliktutan, Elif Sürer, Sıla Girgin, Erinç Dikici, Abuzer Yakaryılmaz.

I would especially like to present my gratitude to my mentors, to Koray Balcı for counseling me in any matter I get confused, to Berk Gökberk for being an encouraging co-advisor, and especially to Albert Ali Salah for letting me discover my "potential" with his hardworking and sincere character.

I would like to thank my best friends, Ayşe Gaye Soykök, Demet Nar, Pınar Karagülle for their endurance and tenderness. I am specially indebted to my family for everything they have done for me. Last but not least, I am specifically grateful to Emre Vardar, who bear the actual burden with boundless patience throughout this thesis.

ABSTRACT

3D FACE REGISTRATION USING MULTIPLE AVERAGE MODELS

Three dimensional (3D) face recognition is a frequently used biometric method and its performance is substantially dependent on the accuracy of registration. In this work, we explore registration techniques. Registration aligns two faces and make a comparison possible between the two surfaces. In the literature, best results have been achieved by a one-to-all approach, where a test face is aligned to each gallery face separately. Unfortunately, the computational cost of this approach is high. To overcome the computational bottleneck, we examine registration based on an Average Face Model (AFM). We propose a better method for the construction of an AFM. To improve the registration, we propose to group faces and register with category-specific AFMs. We compare the groups formed by clustering in the face space with the groups based on morphology and gender. We see that gender and morphology classes exist, when faces are categorized with the clustering approach. As a result of registering via an AFM, it is possible to apply regular re-sampling on the depth values. With regular re-sampling, improvements in recognition performance and comparison time were obtained. As another factor causing diversity in the face space, we explore expression variations. To reduce the negative effect of expression in registration and recognition, we propose a region-based registration method. We divide the facial surface into several logical segments, and for each segment we create an Average Region Model (ARM). Registering via each ARM separately, we examine regional recognition performance. We see that even though some regions such as nose or eye area are less affected by expression variations, no single region is sufficient by itself and the use of all regions is beneficial in recognition. We experiment with several fusion techniques to combine results from individual regions and obtain performance increase.

ÖZET

ÇOKLU ORTALAMA MODELLERLE ÜÇ BOYUTLU YÜZ KAYITLAMA

Sıkça kullanılan biyometrik yöntemlerden biri olan 3B yüz tanımanın başarımı, büyük ölçüde kayıtlamanın doğruluğuna bağlıdır. Bu çalışmada kayıtlama yöntemlerini inceledik. Kayıtlama, yüz yüzeylerinin hizalanmasını sağlayarak karşılaştırma yapmayı mümkün kılar. Literatürde en iyi sonuçlar, bir test yüzünün tüm galeri yüzleriyle ayrı ayrı hizalanmasına dayanan birden-tüme kayıtlama yaklaşımıyla elde edilmektedir. Ancak bu yöntemin zaman karmaşıklığı çok yüksektir. Bu sorunu aşmak için, ortalama yüz modeline dayalı kayıtlama yöntemlerini inceledik. Ortalama yüz modeli oluşturmak için daha başarılı bir yöntem önerdik. Ortalama yüz modeli tabanlı kayıtlamayı geliştirmek için, yüzleri gruplayarak kategoriye ait ortalama modeller ile kayıtlama yapmayı önerdik. Şekil uzayında topaklama sonucu oluşan gruplarla, morfoloji ve cinsiyet bilgisine dayanan grupları karşılaştırdık. Topaklama sonucu, morfoloji ve cinsiyete dayalı gruplaşmanın da varolduğunu gördük. Ortalama yüz modeline dayalı kayıtlama sayesinde, kayıtlama sonrasında derinlik değerlerinin eşit aralıklı örneklenmesi mümkün oldu. Bu yöntem, hem tanıma başarımında hem de karşılaştırma hızında artış sağladı. Yüz uzayında çeşitliliğe neden olan diğer bir etken olarak ifadeyi inceledik. Kayıtlama ve tanıma başarımlarında düşüşe neden olan ifadenin etkisini azaltmak için, bölgesel modellerle kayıtlama yapmayı önerdik. Bu nedenle yüzü anlamlı bölgelere bölerek, her bir bölge için ayrı bir ortalama bölge modeli elde ettik. Her bir bölge modeliyle ayrı ayrı kayıtlama yaparak, bölgesel tanıma sonuçlarını inceledik. İfade değişimlerinden daha az etkilenmesine rağmen, burun ya da göz çevresi bölgesinin tek başına yeterli olmadığını ve tüm bölgelerin kullanımının tanıma başarımını arttırdığını gördük. Bu nedenle bölgesel kayıtlama sonuçlarını tümleştirme yöntemleriyle birleştirerek, tanıma performansını arttırdık.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

d	Sorted distances to N gallery classes
d'	Normalized distances
d_{MIN}	Minimum distance
d_{MAX}	Maximum distance
H	Rotation matrix
I	Identity matrix
l	Number of landmarks
L_1	A distance metric
M	Centroid of shapes
N	Number of gallery classes
p	A three dimensional point
P	A shape matrix, made up of landmarks
$P_{i,k}$	k^{th} landmark vector of shape P_i
r	Euclidean distance
$\vec{q_H}$	Quaternion
$\vec{q_T}$	Translation vector
U	Basis function
$\mathbf{v}$	Quaternion vector
$\hat{\mathbf{v}}$	Normalized quaternion vector
Y	Consensus shape
$\mathbf{z}$	A unit quaternion
α	Scaling coefficient
γ	Translation vector
Γ	Rotation matrix
μ	Center of mass
$\mathbf{\Sigma}$	Covariance matrix
τ	Threshold value

2D	Two Dimensional
3D	Three Dimensional
3D-RMA	A face database
AFM	Average Face Model
ARM	Average Region Model
AU	Action Unit
BERC	A face database
BORDA	Borda count
CMC	Cumulative Match Characteristic
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
DP	Dynamic Programming
DWHD	Depth-Weighted Hausdorff Distance
EER	Equal Error Rate
EGI	Extended Gaussian Image
EnterFACE	A face database
FACS	Facial Action Coding System
FAR	False Accept Rate
FRGC	Face Recognition Grand Challenge
FRR	False Reject Rate
FSU	A face database
HC	Highest Confidence
HMM	Hidden Markov Model
ICA	Independent Component Analysis
ICP	Iterative Closest Point
LDA	Linear Discriminant Analysis
LGT	Log-Gabor Templates
MOD-PLUR	Modified plurality voting
ND-2006	A face database
NMF	Non-negative Matrix Factorization
NRCC	National Research Council of Canada

PCA	Principal Component Analysis
PLUR	Plurality voting
PROD	Product rule
PSD	Point Set Difference
R-1	Rank-1
ROC	Receiver Operator Characteristics
ROI	Region Of Interest
SUM	Sum rule
SVD	Singular Value Decomposition
SVM	Support Vector Machines
TPS	Thin-plate Spline
v.1	Version 1
v.2	Version 2

1. INTRODUCTION

Recognition of humans by computers, has attracted increasing interest especially over the last 20 years. The main application areas are security systems that are required for a wide variety of applications ranging from international border crossing to database security. Automatic recognition of identities also plays an important role in Human-computer interaction systems.

In identity management systems, the task of determining the correct identity of a person is critical. Identity representation systems utilizing a password of an identification card are not reliable, since these representations can easily be forgotten or lost, shared with unauthorized acquaintances, or stolen by malignant parties. To overcome these difficulties, identity management system studies have moved towards the use of biometrics.

In biometric studies, methods are developed to recognize human beings from their intrinsic or behavioral characteristics. Commonly used biometrics include fingerprint, face, iris, hand geometry, voice, palm-print, handwritten signatures and gait. Iris and fingerprint biometrics offer great accuracy when cooperative subjects who do not willingly mislead the acquisition process are present. Human recognition systems that are based on face information have several advantages over other biometrics: The main advantage of face recognition emerges in the acquisition process, where the acquisition of faces is a non-intrusive process requiring no contact with the subject. The recent studies [1] also showed that the performance of face recognition have reached a level where it is comparable with the high accuracies obtained from iris or fingerprint recognition systems. It is also stated that the computer-aided recognition of faces has overtaken the accuracy of recognition of faces by humans.

1.1. Terminology

Before going on, some terminology used widely in the face recognition studies is defined in this section for clarification. Recognition can signify different scenarios. The term "recognition" is often used to denote "identification" scenarios. Another scenario can be "authentication" or sometimes referred to as "verification". In both scenarios, a set of faces with known identities are present and this set is referred to as the "gallery" set. The faces that are to be identified against the gallery are referred to as the "probe" or "test" set. In recognition or identification, a probe face is tested against the whole gallery set to determine the identity of the person. In authentication or verification, a probe face is tested against only the gallery images with the claimed identity and the face is accepted or rejected. In recognition tests, Cumulative Match Characteristic (CMC) curves may be consulted. A CMC curve plots the percentage of correctly matched probe faces as a function of the rank that is considered as a correct match. In recognition studies, rank-1 recognition rates are reported. In verification tests, Receiver Operator Characteristics (ROC) curves are studied. A ROC curve summarizes the percentage of falsely rejected probes as a function of the falsely accepted probes. Conceived from the curve, Equal Error Rate (EER), which is the percentage of probes where the False Reject Rate (FRR) is equal to the False Accept Rate (FAR), is often reported.

In face recognition, different representations can be used. In most of the studies, a face is represented in two dimensional space and the representation is named "intensity image" or "texture map". In recent studies, three dimensional surface information has been used. The 3D information can have various representations such as point clouds giving 3D coordinates of facial points, range or depth images holding the distance or depth values of a surface from the sensor, or wire-frame meshes where the 3D facial surface is represented as a graph with 3D points as vertices with edges and faces representing the surface.

1.2. Motivation

We are interested in face recognition that exploits 3D facial data, because it is believed that the use of 3D information of facial surface leads to more accurate recognition performances than the results obtained from the traditional two-dimensional intensity images of faces. Medioni and Waupotitsch [2] state that by working in 3D space, they surmount the constraints caused by the changes in the viewpoint and lighting conditions. In [3], Hesher *et al.* express that by using 3D information, the changes in the shape space can be captured without being effected by the lighting variations. Gordon declares in [4], that depth and curvature features extracted from 3D space are more advantageous over traditional 2D images because by 3D features surface information can be more accurately described, especially for areas such as cheeks and forehead and also because these features are not dependent on the viewpoint. In [5], Nagamine *et al.* state that using 3D information instead of 2D gives the advantage of being irrelative of changes caused by lighting and makeup. Chua *et al.* state in [6], that even though the appearance of a face can be altered with makeup visually, the biometric information still exists in 3D facial surface. As stated in [7], use of 3D geometric information allows better handling of alterations caused by expression.

We are motivated to work on 3D data, because face is a 3D surface and the examination of the geometry can lead to more accurate registration and recognition of faces. By using 3D information, the effect of illumination differences can be avoided and small pose changes can be rectified. The expression variations deform the facial surface, thus the faces under presence of expressions can be analyzed comprehensively when they are considered as 3D surfaces.

This work focuses mainly on registration of faces prior to recognition. It is necessary to align two faces, before they can be compared. If the registration process is executed on 3D information, more realistic and accurate alignments can be achieved.

1.3. 3D Face Recognition Literature Review

This literature review mainly focuses on the recognition studies that have performed matching on the three-dimensional shape information of the face. Recent surveys of earlier works in 3D face recognition area are given in [8] by Bowyer *et al.* and in [9] by Scheenstra *et al.*.

The earliest work done in 3D face recognition area dates back to the early 90's [4],[5],[10],[11]. Studies in this area has increased in number in the recent years. Most of the earlier studies were on small data sets with a few number of subjects not exceeding 100 and on neutral and frontal faces. Therefore, in most of these researches, the reported performances reached 100 percent recognition rates. Only a small number of studies cover the variations in pose and expression [12],[13],[14],[15]. The algorithms developed to cope with challenges caused by alterations in pose and expression space are important if 3D recognition is to be performed with uncontrolled acquisition of faces. The most recent works focus on these challenges [16],[17],[18].

1.3.1. Holistic Approaches

Some algorithms developed for 3D face recognition take face into account as a whole. The face can be expressed in different representations in holistic methods.

In [19], Achermann *et al.* extend the use of two 2D face recognition approaches to be used on range images: the Eigenface approach and the Hidden Markov Model (HMM) approach. In the Eigenface method, the range images are transformed into face space by Principal Component Analysis (PCA) and the base vectors formed from the eigenvalues are used for similarity calculations. In the HMM method, feature vectors are formed by using a sliding window that scans the image from top to bottom and records the pixel value. The dataset used consists of 120 training and 120 test images of 24 subjects each with ten images. The Eigenface approach with 100 per cent performance outperforms the HMM method which has 90 per cent recognition accuracy.

In [6], Chua *et al.* propose to represent facial surfaces using point signatures. In the registration phase, the facial parts with low registration error are identified as rigid regions. These portions of the face that are less affected by expression are used in the recognition phase. The tests run on a data set of six subjects with multiple scans of four different facial expressions give 100 per cent performance.

In [20], Achermann and Bunke represent the facial range data in two different ways, as point sets and voxel arrays. The Hausdorff distance, which is a similarity measure of two point sets in 2D, is extended to be applied in 3D space. Registration is carried out by fitting a plane to the set of points and transforming this plane to bring the facial surfaces in alignment. A data set of 24 subjects, each with ten images of different head orientations are used to test the methods and they achieve 100 per cent recognition performance.

Hesher *et al.* [3] use PCA and Independent Component Analysis (ICA) to reduce the dimensionality of range images. Prior to dimensionality reduction, images are registered via some feature points extracted heuristically. The nearest neighbor rule is applied on the reduced dimensions for identification. Experiments are run on the FSU 3D face database with 37 subjects, each with six different expression scans. The experiments include tests with a gallery containing one scan of each subject and a gallery with multiple images of subjects.

Medioni and Waupotitsch [2] align 3D facial surfaces with the Iterative Closest Point (ICP) algorithm. After the alignment, distances between the surfaces are computed and classification is based on the distance map. A database of 100 subjects, each with seven different head poses, is used and an EER better than two per cent is obtained.

Pan *et al.* [21] explore verification approaches based on the Hausdorff distance and Eigenfaces. The 3D facial surfaces are registered before comparison and the registration phase is based on partial directed Hausdorff distance. Verification tests based on the Hausdorff distance and PCA approaches are run on the 3D-RMA database including

pose and expression variations and results of approximately four per cent and six per cent EER values are reported on the two approaches, respectively.

In [22], Lee and Shim implement a recognition system based on the Depth-Weighted Hausdorff Distance (DWHD) that encapsulates both the depth and the curvature information. A dataset of 42 subjects each with two images is used for the experiments and performances of 92.8 per cent, 97.6 per cent, and 92.6 per cent are achieved with the DWHD method respectively on maximum, minimum and gaussian curvatures.

Russ *et al.* [23] use Hausdorff metric, which is extended to be used for alignment and matching of range images. Verification experiments are run on a dataset of 200 subjects with an additional 68 impostors in the probe set. A verification rate of 98 per cent with a FAR of zero per cent is achieved. In recognition experiments, a subset of 30 subjects from this data set was used to obtain a performance of 93 per cent.

Bronstein *et al.* [12] develop an expression-invariant face recognition system, assuming that facial expressions can be defined as isometries of the neutral facial surface. Isometric embedding of calculating geodesic distances between surface points is used and multi-dimensional scaling is applied to model facial isometries. This approach is problematic in cases of topology changes such as mouth opening and the tests were run scans with closed mouths. A database of 30 subjects with 220 scans of varying expression and pose is considered and multiple images of subjects are used in the gallery. They achieve a rank-1 recognition rate of 100 per cent and report that they are also able to recognize identical twins with zero per cent error.

In [24], Russ *et al.* apply the Hausdorff metric for matching on range images. The registration of faces prior to recognition is handled with an iterative procedure similar to ICP. A subset of the Face Recognition Grand Challenge (FRGC) version 1 (v.1) face database is used for experimental purposes, where the gallery consists of 198 subjects that have multiple scans and the probe set contains only one probe image per subject. 98.5 per cent rank-1 recognition performance and 93.5 per cent verification

accuracy with 0.1 per cent FAR is reported.

In [7], Pan *et al.* apply PCA over range images where Region Of Interest (ROI) of facial surfaces is automatically extracted. Prior to ROI extraction, nose tip localization and axis of symmetry detection are used for alignment. Relative depth values, which are pose-invariant, are mapped into a circular range image and PCA is applied. Rank-1 recognition rate of 95 per cent and an EER of 2.83 per cent is achieved on FRGC v.1 face database.

In [25], Lu *et al.* combine surface matching with appearance-based matching. They apply a hybrid ICP algorithm in registering and matching phases of 3D facial surfaces. In the hybrid ICP, two classical ICP algorithms, using point-to-point and point-to-plane distances are the similarity metrics, where the first algorithm is used for alignment and the second for refinement. Coarse alignment prior to ICP is handled by extracting three corresponding feature points. For appearance-based matching, Linear Discriminant Analysis (LDA) on 2D textures is implemented. The weighted sum rule is used to combine the two classifiers. On a database of 200 subjects in the gallery and 598 probe scans with lighting, pose and expression variations, recognition results of 86 per cent, 77 per cent and 90 per cent are obtained respectively for ICP, LDA and ICP-LDA combination.

Lu and Jain [18] propose a method to model expression deformations to deal with expression variations. A small group of subjects, namely the control group, is used to calculate different deformations caused by expressions. When matching a test scan to gallery faces, all deformation models obtained from the control group are applied to the gallery and ICP algorithm is used to find the best fit. A database of ten subjects, each with three different pose and seven different expression scans is used. A recognition rate of 92.1 per cent is achieved when deformable models are used, whereas without deformation modeling, 87.6 per cent accuracy is obtained.

In [26], Faltemier *et al.* explore the effect of using multiple gallery images per subject with expression variations to improve the performance. They used a superset

of FRGC version 2 (v.2), namely ND-2006, data set which consists of 13,450 scans from 888 subjects with six different facial expressions. For multi-instance enrollment, the best results are obtained by putting two neutral and three happiness scans of each subject in the gallery. The matching is done by applying ICP on one single nose region. A recognition performance of 97.2 per cent is reported on ND-2006 database.

1.3.2. Feature-based Approaches

In some face recognition approaches, instead of considering a face as a whole, some features are extracted and matching is performed on these features.

Cartoux *et al.* [10] deal with recognition of faces from range images. Gaussian curvatures of the facial surfaces are extracted from range images and the plane of symmetry is found using these curvature values in an iterative manner. The plane of symmetry is used both for matching purposes and for pose normalization, where the normalized facial surfaces are used for recognition. For a small data set of five subjects, a recognition rate of 100 per cent is reported.

Gordon [4] represents the face in the form of features extracted from range images. The features extracted for a face include both curvature and metric properties such as Gaussian curvatures or distance from the symmetry plane. Matching is done by nearest neighbor clustering on a test set of 24 faces, three views of each of eight subjects. Tests run on different sets of features give recognition results in the range of $80 - 100$ per cent.

Nagamine *et al.* propose to use curves of intersections of the facial surface with the vertical and horizontal planes and with a cylinder [5]. The poses of faces are normalized with five heuristically found feature points on the surfaces. After the pose correction, curves are matched using the Euclidean distance. Various curves and profiles extracted are evaluated and it is stated that the intersections near the vertical profile are the most distinctive facial features for recognition. In a data set of 16 subjects each with ten images, 100 per cent recognition performance is achieved using the vertical profile

curves that pass through the center of the face.

In [27], Tanaka *et al.* approach the face recognition problem by analyzing maximum and minimum principal curvatures and their directions. The faces are represented by Extended Gaussian Imagess (EGIs) and the similarities for recognition are calculated by Fisher's spherical correlation of EGIs of faces. The data set of NRCC (National Research Council of Canada) consisting of 37 range images are used for testing purposes and 100 per cent recognition performance is reported.

Lee *et al.* [28] develop a system that recognizes faces using facial features extracted from surface geometry. Using depth and curvature, generic points especially related to the nose and eyes are extracted and relative features based on distance, ratio and angle values are defined. Two recognition algorithms are implemented. One approach uses depth information and is based on Dynamic Programming (DP). The second approach is based on Support Vector Machines (SVM) and the features extracted are used. For experimental purposes, frontal faces from the BERC face database is used. For depth-based DP, scans of 20 subjects are used and 95 per cent rank-1 accuracy is achieved. For feature-based SVM, 96 per cent rank-1 recognition rate is obtained on a data set of 100 subjects.

In [29], Li and Barreto integrate an expression recognition system with face recognition, where for each expression a different recognition approach is to be used. As an initial work, they only deal with neutral and smiling expressions. Their database consists of 30 subjects, each subject having two neutral and two smiling scans. They extract facial metrics such as distances between points, ratios of distances and angles between facial points. These metrics are used to recognize the expression of a given face. For neutral face recognition, central vertical profiles and contours are used for matching, whereas the smiling face recognition is based on eigen decomposition combined with vertical profile and contour matching. Neutral face recognition achieves an accuracy of 97 per cent on neutral probes and 57 per cent on smiling probes, whereas by the smiling face recognition a performance of 87 per cent is obtained on smiling test scans.

1.3.3. Classifier Fusion Approaches

In 3D face recognition, some approaches combine results obtained from different classifiers. These fused classifiers can be based solely on 3D information or classifiers implemented for 2D and 3D face recognition can be combined together.

In [30], Xu *et al.* use both global and local features to represent a 3D facial surface. The face is globally represented with a regular mesh fitted to the 3D point cloud of the face. Local shape variation information especially near the areas such as mouth, nose and eyes, is extracted using the Gaussian-Hermite moments to define local features. The low dimensional feature vector is obtained with PCA and the matching is done with nearest neighbor classifier. The method is tested on the 3D-RMA database and recognition rates of 72.4 per cent and 96.1 per cent are obtained respectively with 120 and 30 subjects in the test database.

Maurer *et al.* [31] fuse the matching based on 2D and 3D. For alignment and matching in 3D space, ICP is used. The matching scores obtained from 2D and 3D separately are combined with a weighted sum rule. For the experiments, FRGC v.2 database is used and for a FAR of 0.01 per cent recognition rates of 85 per cent and 92 per cent are obtained respectively for 3D only and 3D together with 2D information. For these tests, the gallery is composed of neutral faces and probe set is the whole data set.

In [32], Gökberk *et al.* use different 3D shape representations to define a face, namely point clouds, surface normals, facial profiles and depth images. Classifiers based on these different representations, are implemented and decision-level fusion techniques in both parallel and hierarchical modes are applied to obtain better recognition rates. In the experiments, 3D-RMA face database is used with 571 images from 106 subjects. The best individual performance of 96 per cent is outperformed with an accuracy of 99 per cent using the nonlinear rank-sum fusion technique.

Gökberk and Akarun in [33], use different classifiers implemented for different

facial representations, namely point cloud based, depth image based, profile based and shape index based classifiers. Decision-level fusion techniques that run in parallel are considered; consensus voting, borda count, sum and product rules, weighted sum rule, consensus voting via confidence values, weighted consensus voting, highest confidence. Also serial fusion of classifiers is implemented. On 3D-RMA database, the best result of 97.93 per cent accuracy was obtained by serially fusing the top 20 classes selected by the point cloud based classifier with depth-image based approach at the second stage.

1.3.4. Component-based Approaches

To deal with facial expressions or partial occlusions that degrade the recognition performance, some approaches consider multiple regions segmented from the facial surface and fuse classification scores of individual region recognizers.

In [11], Lee and Milios calculate the sign of mean and Gaussian curvatures at each point of a range image and segment the face into convex regions. Their work is based on the observation that distinct facial features correspond to convex regions. After the convex regions are extracted, EGIs corresponding to each region are created. Regional matching is obtained by the correlation between the EGIs. This approach is reported to be able to cope with expression variations and occlusions.

In [15], Moreno *et al.* segment the 3D facial surface using mean and Gaussian curvatures and various descriptors for the segments are extracted. A database of 420 scans from 60 subjects, each with seven scans of different pose and expression, is used. The database is used for testing different subsets of feature vectors. Experimental results of 78 per cent rank-1 recognition on frontal views and 92 per cent rank-5 recognition are obtained.

Cook *et al.* [34] use Log-Gabor Templates (LGT) on range images to deal with expression variations. A range image is divided into multiple regions both in spatial and frequency domains. Each individual region is classified separately and these classifier are fused at the score level. The facial image is divided into 147 regions and the LGT

responses are reduced in dimension by PCA. For classification, Mahalanobis Cosine distance metric is used and the classifiers are fused by the sum rule. The experiments on FRGC v.2 database with a gallery of neutral expressions, 94.63 per cent recognition performance is achieved.

In [13], Chang *et al.* propose a matching method based on multiple regions selected around the nose area which has the least shape variation due to facial expressions. Facial surfaces are registered via ICP and similarity measures computed from individual alignments are fused using sum, min or product rules. FRGC v.2 database is used, where the gallery consists of one neutral image per subject. A rank-1 recognition result of 92 per cent was reported using the product rule.

In [35], Faltemier *et al.* extend the use of multiple regions. Seven overlapping regions around nose are extracted and ICP is used to separately align individual segments. Threshold values are determined for regions and the identities determined for each individual region and its threshold value are combined with committee voting. On the FRGC v.2 database, a rank-1 accuracy of 94.9 per cent is reported.

Faltemier *et al.* [16] expand the use of multi-regions to segments extracted from the whole facial surface. 38 regions are used in the experiments that run on the FRGC v.2 database. The individual regions are again aligned with ICP and the best performance reported is 97.2 per cent recognition accuracy which uses the best 28 regions and fusing these classifiers with a modified borda count method.

In [17], Kakadiaris *et al.* use a deformable facial model to describe the facial data. After fitting the deformable model to facial surface, a 2D geometry image is calculated via regular re-sampling to a 2D grid and a normal map is constructed for multiple regions of the facial surface. Both representations of each region are analyzed with a wavelet transform and the classifiers where combined using a weighted sum. The experiments run on the FRGC v.2 face database give rank-1 identification result of 97.0 per cent, whereas verification accuracy of 97.3 per cent with a FAR of 0.001 per cent is obtained.

Mian *et al.* [36] develop a multi-modal algorithm which combines 2D and 3D and the matching is handled in the hybrid mode where feature-based and holistic approaches are fused. Automatic extraction of inflection points around the nose tip are used to segment the face into eyes-forehead and nose regions which are less affected by facial expressions. Separate matching of regions are handled with ICP and the similarity measures are fused at the metric level. The FRGC v.2 database is used at the experiments and use of 3D information alone gives 98.82 per cent and 92.36 per cent recognition rates for neutral and non-neutral probe sets respectively. When multi-modal approach is applied recognition accuracies of 98.20 per cent and 93.74 per cent are achieved respectively for neutral and non-neutral probes.

1.4. Approach and Contributions

A face recognition system depends mainly on a good registration of facial surfaces before any comparisons can be made. To compare a test face to the gallery, it needs to be aligned to each of the faces in the gallery set. The accuracy of the registration step can alter recognition performances.

In general, the registration of a test face to any of the gallery faces is handled separately and this greatly increases the computational cost. To overcome this bottleneck of registration, Irfanoğlu *et al.* [37] proposed to construct an Average Face Model (AFM) which can be used as an indexing surface. The gallery faces are registered to this AFM, where for each point on the model, the corresponding point on the gallery face is labeled. When a test face comes, this face is registered only to the AFM and this alignment procedure automatically leads to finding point-to-point correspondences between a test surface to all of the gallery faces. For classification, Point Set Difference (PSD) calculation is used, where the total Euclidean distance between registered points of the test face and a gallery face is computed and the identity of the gallery face giving the minimum distance is assigned as the identity of the test scan. They have experimented on the 3D-RMA face database and obtained 96.66 per cent recognition accuracy.

In [38], we explored and expanded the AFM-based registration. Instead of using only one AFM for the whole data set, the use of multiple AFMs is proposed to increase the registration accuracy. For each facial category a separate average model is constructed and faces that fall in a specific category are registered with the respective AFM. The construction of facial categories is handled in two different ways. One method is based on cognitive cues that are present in facial information such as gender or morphology data. The other one is based on clustering of facial surfaces in an unsupervised manner. Recognition is handled with PSD calculations between gallery and test faces. Prior to PSD calculation, the registered faces are aligned to a regularly re-sampled grid. After regular re-sampling, the registered faces all have the same x and y values, and only z values are used for calculating the PSD. By regularizing the facial point set, the PSD computation is simplified down to one dimensional space instead of using all three dimensions.

Motivated from the previous work on using multiple AFMs, we explored the use of multiple models for registering facial surfaces prior to recognition when expression variation is present. Also in the presence of expression, some regions of a facial surface are deformed, leading to error in PSD calculations. Therefore instead of using whole face models, we proposed to use Average Region Models (ARMs) to register individual regions of faces. The registered regions are again used in PSD calculations and the individual distances are fused in the decision level to augment the affect of regions that are relatively rigid under expression, whereas the deformed regions will be less effective.

1.5. Outline of Thesis

Chapter 2 describes the mathematical details of the registration methods used, including Procrustes Analysis, Thin Plate Spline Warping, and Iterative Closest Point Algorithm.

In Chapter 3, the use of multiple AFMs in registration is explained. Recognition results are given when multiple AFMs for supervised and unsupervised facial categories

are constructed. Also in Chapter 3, the construction method for average models is given in detail.

Chapter 4 includes the method of using multiple average models when expression is present in the face space. The construction of ARMs for different expression classes is explained. The proposed method is applied to a new database containing multiple expressions and recognition results are presented.

Chapter 5 concludes the thesis by giving a summary of the results obtained, followed by a list of future directions.

2. 3D FACE REGISTRATION

The 3D data may have different translation, rotation or scaling due to the controlled environment parameters such as the acquisition setup, device properties or due to uncontrolled conditions such as the pose variations of the acquired subjects. In either case, the 3D shapes need to be aligned to each other and should be brought into a common coordinate frame before a comparison can be made. Registration is the alignment procedure of two similar shapes.

Three dimensional face registration is the process of defining a transformation that will closely align two faces. For the purpose of alignment, some parameters should be defined. First of all, affine transformations such as translation, rotation or scaling should be applied to roughly align the surfaces. Secondly, a similarity measure should be determined that will be maximized as a result of a successful alignment. The transformation used for registration can either be rigid, affine, elastic or liquid [39]. Different similarity measures are used for registration techniques such as the point-to-point or point-to-surface distances.

In general, the registration process is guided by a set of fiducial points called *landmarks* [40], that are used to define the transformation between two surfaces. The registration techniques that are examined in the scope of this thesis make use of a set of landmark points that are labeled for each face surface to be aligned.

The most accurate results in the literature so far have been achieved by one-to-all registration approaches. In one-to-all registration, a test face is registered to each of the training faces in the gallery separately, and a comparison is made for each face couple. This approach despite its accuracy, is not preferred because of its high computational cost. To reduce the computation complexity, Irfanoğlu *et al.* proposed to use an Average Face Model (AFM) for the registration of 3D faces [37]. For the registration process, a generic face model is provided and it is used for the alignment of all the faces. When a test face is examined, it is registered to the average model

for only once, because the gallery faces are already aligned with the AFM in an offline fashion. Therefore the computation time of registering a test face to all of the gallery faces has been greatly reduced.

For accurate and fast registration, we used the AFM based approach, where the average model is used for indexing the face points of the model on any given facial surface. The training faces were aligned to the AFM beforehand, using the ICP method. A one-to-one correspondence is found between the points of the AFM and each of the gallery faces, and the faces in the gallery are cropped according to the AFM point set. The alignment and cropping steps are also applied to a test face in question. In this approach, by only one registration, a test face is brought into alignment with all the faces in the gallery. The cropping process facilitates the computation of Point Set Difference (PSD), which is the similarity measure used for comparison in this work.

The mathematical fundamentals for the registration methods used in this thesis are described in detail in the following sections. Procrustes Analysis, which uses only landmark points defined for facial surfaces, finds the affine transformation between the set of landmarks for alignment. Iterative Closest Point algorithm finds a rigid transformation between the facial surfaces and coarsely aligns two point clouds. The last method, Thin-Plate Spline algorithm, finds an elastic transform in which the two surfaces are deformed for landmark points to be perfectly superimposed. The details and the algebraic formulations for each method are given next.

2.1. Procrustes Analysis

Procrustes analysis proposed by Gower, in [41], analyzes the geometrical shapes in a statistical approach [42]. A *geometrical shape* refers to the characteristics that remain geometrically unaltered even though a translation, rotation or scaling is applied to that figure. If a geometrical shape in $\mathbb{R}^N$ is labeled with l landmark points, it can be represented with a $l \times N$ matrix P, each row giving a significant landmark point in N dimensions.

If two figures are $P : l \times N$ and $P' : l \times N$, they have the same shape if there exists a similarity transformation that relates them. This special similarity transform can be stated as follows:

$$P' = \alpha P\Gamma + \mathbf{1}_l \gamma^T, \tag{2.1}$$

where the parameters of the transformation are defined as $\Gamma : N \times N, |\Gamma| = 1$ standing for the rotation, $\gamma : N \times 1$ standing for the translation, α standing for a positive scaling constant, and $\mathbf{1}_l$ defining a vector of ones. By the triple of these parameters (γ, Γ, α), the similarity transformation consisting of translation, rotation and scaling that maps the shape P to P' is defined.

In [41], the classical Procrustes idea of examining all possible pairs of shapes was generalized. By using generalized Procrustes analysis, a *consensus shape* can be derived from the whole set of shapes, by minimizing the sum-of-squares between each shape and the consensus shape through translating, rotating, reflecting and scaling. The finally obtained consensus can then be used to align a new shape with the whole group of shapes by an affine transformation.

Below, the steps of the generalized Procrustes analysis are listed as given in [41]:

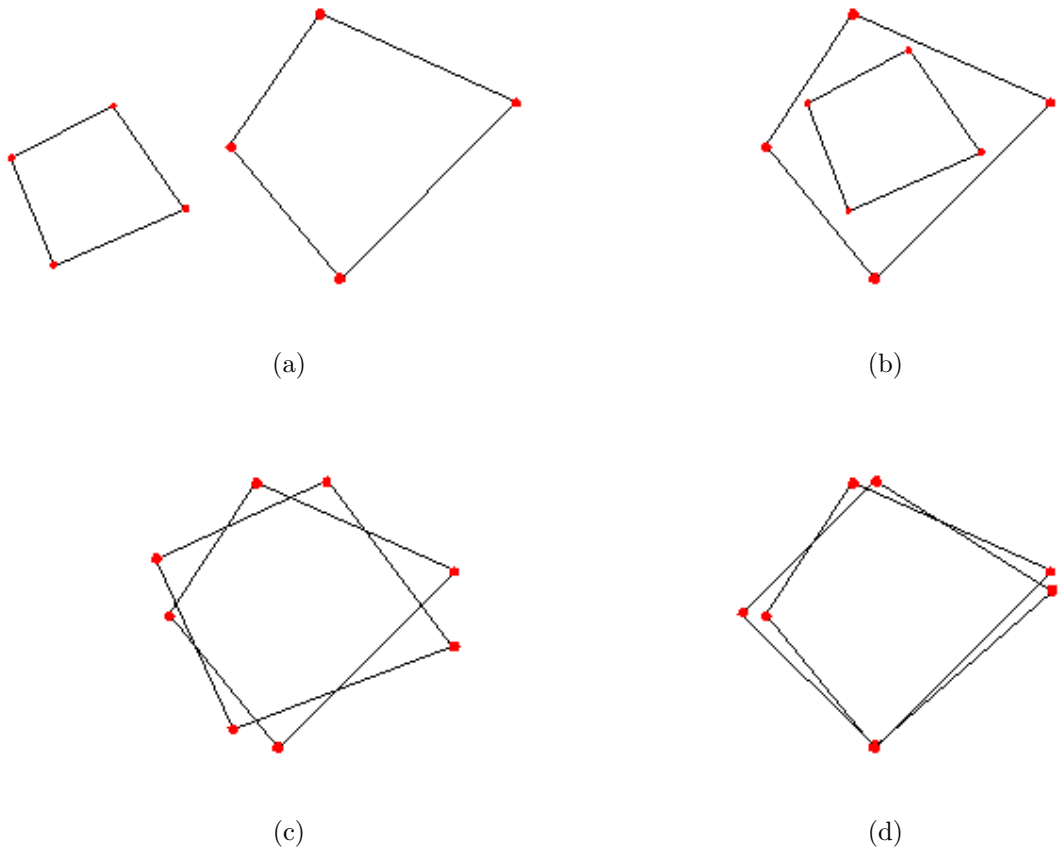
1. First of all, find the centroid of all shapes:

$$M = \frac{1}{N} \sum_{i=1}^{N} P_i \tag{2.2}$$

2. Center all shapes P_i using the centroid M:

$$P_i = P_i - M \tag{2.3}$$

3. Scale shapes so that they all have the average size according to either of these techniques:



(a) (b)

(c) (d)

Figure 2.1. A simple alignment example is given. In (a) the raw landmarks of two shapes are shown. (b), (c) and (d) are the transformed landmarks after the translation, scaling and rotation is applied respectively.

(a) In [41], it was proposed to set the *mean* of the squared landmark distances of each shape to unit value, where $P_{i,k}$ stands for the k^{th} landmark of P_i:

$$P_i = \frac{N * P_i}{\sum_{k=1}^{N} ||P_{i,k}||} \tag{2.4}$$

(b) In [43], bringing the shapes to a common scale is handled by setting the *median* of the squared landmark distances of each shape to unit value. The formulation is given below:

$$P_i = \frac{P_i}{m(D_i)} \tag{2.5}$$

$$where D_i = d_{j,k} = ||P_{i,j} - P_{i,k}|| \quad j,k = 1,\ldots,N \tag{2.6}$$

4. Initialize the consensus shape Y:

$$Y = P_1. \tag{2.7}$$

5. For $i = 2, 3, \ldots, N$, rotate P_i to fit Y.
   (a) in Gower's method as explained in [41] Y is re-evaluated after each update of P_i as

$$Y = \frac{1}{i}\sum_{j=1}^{i} P_j \tag{2.8}$$

   (b) In Rohlf and Slice's method given in [43], Y is updated only once, after the rotation of each P_i.

   The rotation matrix H in two dimensional space can be expressed as:

$$H = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \tag{2.9}$$

   To find the best rotation, singular value decomposition [44] can be used:

$$H = VSU^T \tag{2.10}$$

   where U contains a set of orthonormal output basis vector directions and V contains a set of orthonormal input basis vector directions for H and these two matrices holds for:

$$P_i^T Y = U\Sigma V^T \tag{2.11}$$

   where Σ is a diagonal matrix, containing the singular values. Using S matrix, whose diagonal elements are either $+1$ or -1, instead of Σ, restricts the transform matrix H to be a rotation and not a shear.

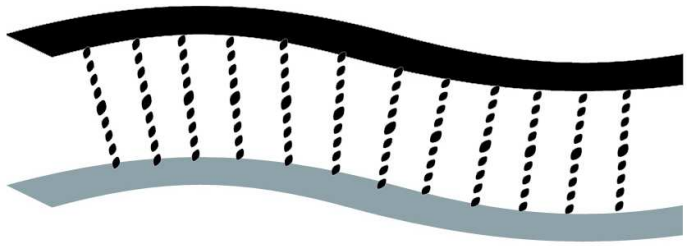


Figure 2.2. A visual example of matching shapes by ICP. A point-to-point correspondence from one surface to the other is found.

6. Update the P_i and Y, while monitoring the residual sum-of-squares:

$$S_r = N(1 - tr(Y_t Y_t^T - Y_{t-1} Y_{t-1}^T)) \tag{2.12}$$

where Y_t stands for the consensus shape at iteration t, and Y_{t-1} is the consensus shape at iteration $t-1$. When S_r is below a threshold value stop iterating, and the consensus shape is hence found.

2.2. Iterative Closest Point Algorithm

In this Section the ICP procedure is summarized as defined in [45]. Since it was introduced by Chen *et al.* [46], it was widely used for matching of three dimensional shapes. A basic visual example is given in Fig. 2.2. Although the ICP algorithm has a high computational cost, its ease of implementation together with its applicability to several geometrical representations of 3D data such as point sets, line segments, parametric curves and surfaces makes this method frequently referred for the registration of 3D objects.

Before going into the details of the algorithm, some mathematical preliminaries about computing the closest point on a model to a given point and finding the correspondence between two surfaces by least-squares registration should be given.

Let p_1 and p_2 be two points such that $p1 = (x_1, y_1, z_1)$ and $p_2 = (x_2, y_2, z_2)$. The Euclidean distance between these two points is formulated as follows:

$$d(p_1, p_2) = ||p_1 - p_2|| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{2.13}$$

If P is a point set of N_p points such that $P = p_i$ where $i = 1, 2, ..., N_p$, the distance between a given point p_g and the point set can be defined as:

$$d(p_g, P) = min_{i\epsilon 1,...,N_p} \quad d(p_g, p_i) \tag{2.14}$$

A closest point p_j in the point set P satisfies the definition below:

$$d(p_g, p_j) = d(p_g, P) \tag{2.15}$$

The closest point computation explained above is in a general form and is applicable to n dimensions. Another method for computing the least-squares rotation and translation is the quaternion-based algorithm which is preferable over the Singular Value Decomposition (SVD) algorithm. SVD approach uses the cross-covariance matrix between the two point sets and it permits reflections which is not desired in the registration of face data. This property makes the quaternion-based approach a preference. Next, the quaternion-based approach will be given in details. Mathematical fundamentals on quaternions are given in Appendix A.

The *unit quaternion* can be defined as a four vector $\vec{q_H} = [q_0 q_1 q_2 q_3]^T$, provided that $q_0 \geq 0$, and $q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1$. The 3×3 rotation matrix H generated by the unit quaternion is given below:

$$H = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & q_0^2 + q_1^2 - q_2^2 - q_3^2 \end{bmatrix} \tag{2.16}$$

If the translation vector is defined as $\vec{q_T} = [q_4 q_5 q_6]^T$ the complete registration state

vector can be given as $\vec{q} = [\vec{q_H}|\vec{q_T}]^T$. Let $P = \{\vec{p_i}\}$ be an obtained point set to be aligned to the model point set $M = \{\vec{m_i}\}$, where both of the point sets have the same number of points such that $N_p = N_m = N$ and that each point p_i is in correspondence with point m_i. The mean-square objective function to be minimized by the ICP procedure is:

$$f(\vec{q}) = \frac{1}{N}\sum_{i=1}^{N} ||\vec{m_i} - H(\vec{q_H})\vec{p_i} - \vec{q_T}||^2. \tag{2.17}$$

The center of mass of the point set P, μ_p, and that of the model set, μ_y can be formulated as follows:

$$\mu_p = \frac{1}{N}\sum_{i=1}^{N} \vec{p_i} \quad , \quad \mu_m = \frac{1}{N}\sum_{i=1}^{N} \vec{m_i} \tag{2.18}$$

and the cross-covariance matrix Σ_{pm} can then be formulated as:

$$\Sigma_{pm} = \frac{1}{N}\sum_{i=1}^{N}[(\vec{p_i} - \vec{\mu_p})(\vec{m_i} - \vec{\mu_m})^T] = \frac{1}{N}\sum_{i=1}^{N}[\vec{p_i}(\vec{m_i})^T]) - (\vec{\mu_p}\vec{\mu_m})^T. \tag{2.19}$$

The cyclic components of the matrix $A_{ij} = (\Sigma_{pm} - \Sigma_{pm}^T)_{ij}$ are used to form a column vector $\Delta = [A_{23} \quad A_{31} \quad A_{12}]^T$, which is then used to form a symmetric 4×4 matrix $Q(\Sigma_{pm})$ as given below:

$$Q(\Sigma_{pm}) = \begin{bmatrix} tr(\Sigma_{pm}) & \Delta^T \\ \Delta & \Sigma_{pm} + \Sigma_{pm}^T - tr(\Sigma_{pm})I \end{bmatrix} \tag{2.20}$$

where I is the 3×3 identity matrix. The unit eigenvector $\vec{q_H} = [q_0 q_1 q_2 q_3]^T$ corresponding to the maximum eigenvalue of the matrix $Q(\Sigma_{pm})$ is chosen, that will give the optimum rotation vector. The optimum translation vector q_T can be computed as follows:

$$\vec{q_T} = \vec{\mu_m} - H(\vec{q_H})\vec{\mu_p} \tag{2.21}$$

The least-squares quaternion operation can be written as:

$$(\vec{q}, d) = \varphi(P, M) \tag{2.22}$$

where $\vec{q}$ denotes the quaternion operation and d the mean square error.

The point set P can be denoted by $\vec{q}(P)$ after the transform represented by $\vec{q}$ is applied.

Now that the mathematical preliminaries are given, the ICP algorithm computing and applying these transformations iteratively can be given. The description of the steps of the algorithm are given in Fig. 2.3.

Usually the number of points for the test surface and the model do not agree. The faces in the gallery or the face model can have fewer points than the test face. In our simulations, the training faces that were aligned to the AFM beforehand consisted of about $30,000$ points, whereas the test faces had a larger number of points varying between $70,000$ and $150,000$. Therefore, the test faces where used as the model in computing the transformation that aligns the cropped gallery faces to it. After one-to-one correspondence between the points of the training and test faces, the test face was cropped discarding the points with no correspondence. The test faces were later transformed back into the coordinate frame of the training faces using the inverse of the transform found by ICP. This way, all the training faces and the test faces were brought into a common coordinate system, making a comparison possible. If the registration is correct, the test faces will be cropped perfectly, possibly leaving out any hair or clutter. After successful registration and hence cropping, the surfaces aligned will have the facial area that is sufficient for face recognition.

2.3. Thin Plate Spline Algorithm

The Thin Plate Spline (TPS) expresses the physical bending energy of a thin metal plate that is fixed at a set of given points [47]. An example of such a deformation

1. Given: the point set P with N_p points and the model point set M with N_m.
2. Initialize the iteration by setting: $P_0 = P$, $\vec{q_0} = [1, 0, 0, 0, 0, 0, 0]^T$, $k = 0$, k standing for the iteration number. The registration vectors should be updated relative to the initial data set P_0 so when convergence is achieved, the obtained final registration will define the complete transformation. The following steps are applied until convergence is achieved.
   (a) Initialize ICP by setting $P_0 = P$, $\vec{q_0} = [1, 0, 0, 0, 0, 0, 0]^T$ and $k = 0$. The registration is defined relative to P_0, which requires a coarse registration. Steps 2-5 are applied iteratively, until convergence is achieved within a tolerance τ.
   (b) Compute the closest points: $Y_k = \mathcal{C}(P_k, Y)$. The computational cost of this step is $O(N_p N_y)$ at the worst case, where N_p is the number of points on the registered point cloud, and N_y is the number of points on the model shape.
   (c) Compute the registration: $(\vec{q_k}, d_k) = \mathcal{Q}(P_0, Y_k)$. The computational cost is $O(N_p)$.
   (d) Apply the registration: $P_{k+1} = \vec{q_k}(P_0)$. The computational cost is $O(N_p)$.
   (e) Terminate the iteration if the change in the mean square error is below pre-set threshold τ. A heuristic value for τ is a multiple of $\sqrt{tr(\Sigma_y)}$, where Σ_x is the covariance matrix of the model shape, and the square root of its trace is a rough indicator of model shape size.

Figure 2.3. The Iterative Closest Point Algorithm

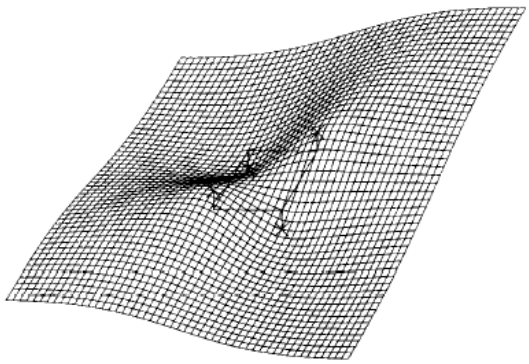


Figure 2.4. Deformation of a thin plate (from [47]).

is shown in Fig. 2.4 (from [47]). Using TPS warping, a shape is deformed while its landmark points are transformed exactly onto the landmarks of another shape. The transformation is one that minimizes the total bending energy over all the points of that surface.

This model algebraically describes the deformations that are specified by given corresponding point sets that are irregularly spaced. At the heart of the analysis is a special surface function stated as follows:

$$z(x, y) = -U(r) = -r^2 \log(r^2) \tag{2.23}$$

where $r = \sqrt{x^2 + y^2}$ is the Euclidean distance from point (x, y) to the Cartesian origin. A circular fragment of the surface defined by this function is given in Fig. 2.5 (from [47]). The origin, included in the surface is marked with X in the figure.

Thin plate splines are used to model deformations while minimizing the bending energy given below:

$$\int_{\mathbb{R}^2} \int ((\frac{\partial^2 z}{\partial x^2})^2 + (\frac{\partial^2 z}{\partial x \partial y})^2 + (\frac{\partial^2 z}{\partial y^2})^2) dx dy \tag{2.24}$$

This minimization equation is called the *integral bending norm*.

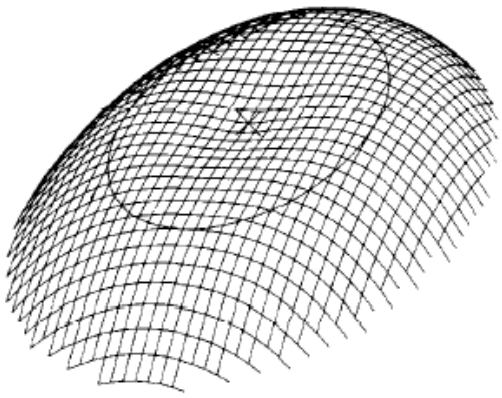


Figure 2.5. A circular fragment of the surface defined by the function $z(x, y) = -r^2 \log(r^2)$ viewed from above (from [47]).

The following derivations show the computation of the weights of several basis functions U, mapping the source landmark points to the target landmark points. When the deformation is expressed as a weighted sum of these basis functions, the total energy of the change will be minimized.

For a set of landmarks given as $P_i = (x_i, y_i), i = 1 \dots n$, the thin-plate spline interpolation is a vector-valued function $f(x, y) = [f_x(x, y), f_y(x, y)]$ that maps the source landmark points to their specified homologues of target landmark points given as $P'_i = (x'_i, y'_i), i = 1 \dots n$. It also specifies a surface with the least possible bending, as measured by the *integral bending norm*. We will give a summarizing mathematical specification of the model here.

Define $r_{ij} = |P_i - P_j|$, giving the distance between the points i and j. Also define the following matrices:

$$K = \begin{bmatrix} 0 & U(r_{12}) & \dots & U(r_{1n}) \\ U(r_{21}) & 0 & \dots & U(r_{2n}) \\ \dots & \dots & \dots & \dots \\ U(r_{n1}) & U(r_{n2}) & \dots & 0 \end{bmatrix} \tag{2.25}$$

$$P = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \cdots & \cdots & \cdots \\ 1 & x_n & y_n \end{bmatrix} \tag{2.26}$$

and

$$L = \left[\begin{array}{c|c} K & P \\ \hline P^T & O \end{array}\right] \tag{2.27}$$

where O is a 3×3 matrix of zeros. Let V be a matrix made up of the homologues of the landmark points:

$$V = \begin{bmatrix} x'_1 & x'_2 & \dots & x'_n \\ y'_1 & y'_2 & \dots & y'_n \end{bmatrix} \tag{2.28}$$

Define w_i and the coefficients a_1, a_x, and a_y as:

$$L^{-1}(V|\mathbf{0}, \mathbf{0}, \mathbf{0}) = (w_1, w_2, \dots, w_n, a_1, a_x, a_y)^T \tag{2.29}$$

The function $f(x, y)$ is defined as:

$$f(x, y) = a_1 + a_x x + a_y y + \sum_{i=1}^{n} w_i U(|P_i - (x, y)|) \tag{2.30}$$

$f(x, y)$ minimizes the nonnegative integral bending norm I_f over all such interpolants:

$$I_f = \iint_{\mathbb{R}^2} \left(\left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 \right) \mathrm{d}x \, \mathrm{d}y \tag{2.31}$$

The thin-plate spline function $f(x, y)$ is invariant under rotations and translations. It maps the landmarks P_i to their homologues P'_i, and defines a smooth interpolation for the rest of the points on the surface. The source and target landmark points,

P_i and P_i', taken together give an exact specification for the function $f(x, y)$, and are therefore crucial to the accuracy of the deformation.

3. RECOGNIZING FACES WITH VARYING MORPHOLOGY USING MULTIPLE FACE MODELS

In this work, we have used the AFM-based approach of [37]. A crucial point for the success of AFM-based registration is the construction of the average face. The average should be as close as possible to all of the faces, so that it can serve as an intermediate surface for one-to-one correspondence between all facial surface points. However, this is not always possible since faces exhibit large variations: Some faces are long and thin, while others are broad. Often, men and women have different facial characteristics; and so do people from different geographical regions. In order to represent all these variations in an average face, we propose to group similar faces together and to use a representative AFM for each. In this chapter, the details of using multiple AFMs for registration are given. The novel AFM construction method, the clustering techniques to obtain multiple models, rigid and non-rigid registration methods and the classification approach are all described and experimental results are reported in the following sections. The details of this work can be found in [48] and also as a preliminary version in [38].

3.1. Methodology

The overall structure of the implemented 3D face recognition system is given in Fig. 3.1. The main steps of the system are the AFM construction, landmarking, registration, post-processing and classification stages.

The gallery images are used in the construction of AFMs. To test the novel AFM construction method, an experiment using a single AFM is also carried out. Two different approaches are implemented for the construction of multiple average models, one based on cognitive cues and the other based on unsupervised clustering of landmark points.

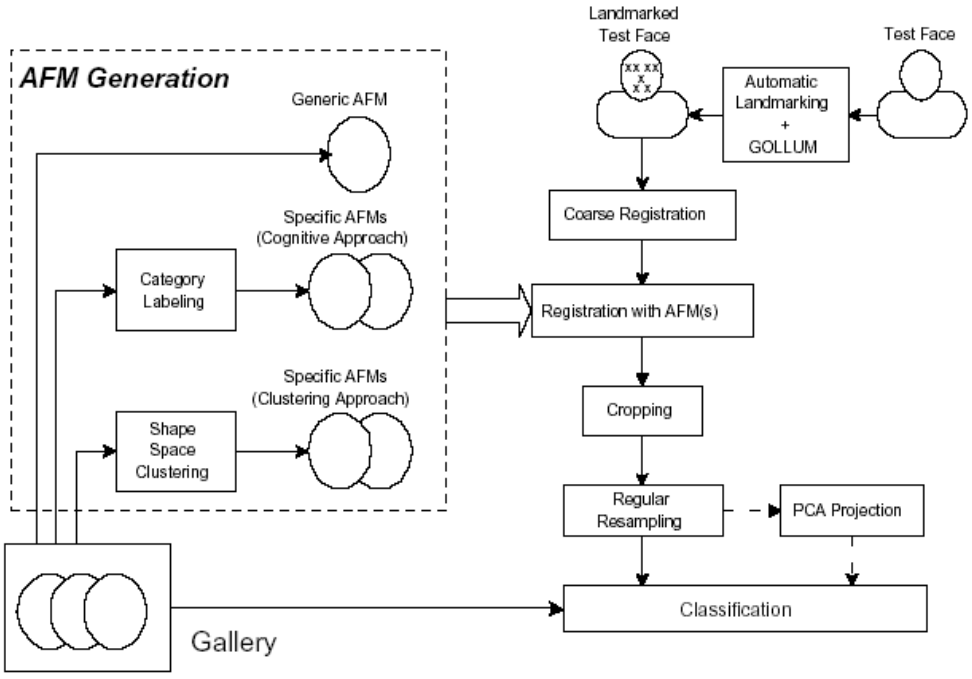


Figure 3.1. The overall 3D face recognition system with registration based on multiple AFMs.

To test the dependence of registration upon the landmarking stage, automatic landmarking is also implemented. The details for the algorithm used can be found in [49], [50]. The automatically located landmark points include the inner and outer eye corners, the nose tip and the mouth corners. These points are also manually labeled as ground truth. Experiments are run using both the automatic and the manual landmarks to make a comparison possible.

The registration procedure consists of two basic stages, coarse initialization and dense alignment. Different approaches for the coarse alignment phase are tested and the affect of coarse initialization on dense registration is reported.

Two different registration approaches are tested to compare rigid and non-rigid registration methods. The rigid registration is based on the ICP algorithm, whereas the non-rigid registration uses TPS warping. Both of these methods are dependent upon landmark localization and for a comparison to be possible, experiments using both manual and automatic landmark points are run for both methods where the registration is handled using a single average model.

After registration, the facial points corresponding to the AFM surface are cropped and regularly re-sampled to avoid any extra error that will arise from small differences in $x - y$ coordinates. After the post-processing, the faces have the same number of points and same $x - y$ coordinates with varying depth values.

The classification is carried out on the aligned and cropped images, and the PSD calculation is adapted to find the nearest gallery image. PSD calculation is tested both for the three-dimensional space where the cropped faces are used and for one-dimensional space where the depth images obtained after regular re-sampling are utilized.

3.2. AFM Generation

For AFM-based registration, construction of an average model is necessary. This model should contain points that are present in nearly all the faces, should exhibit the facial characteristics and should be the mean of the faces in the data set. In [37], a method for generating the AFM was proposed by Irfanoğlu *et al.* which uses the Thin Plate Spline warping algorithm. In this method, the face having the minimum number of points is selected as the candidate base model. The fiducial points are located on all the training faces manually. Irfanoğlu *et al.* uses ten landmark points for the registration. From these manual landmarks, a mean consensus shape is computed using Procrustes analysis. All the training faces are warped onto the computed mean consensus by TPS, using manually located landmarks. The final AFM is constructed by checking the distance between each point on the mesh and its closest match on every warped face and trimming the point if any of these distances is greater than a threshold and averaging the rest. The trimmed points are usually the boundaries of the facial surface. The average model created by this method appears to be a very smooth average face.

We proposed a novel method to construct an average model. For the construction of an AFM, training faces which are manually landmarked are used. The seven landmarks that are manually located for each face are the inner and outer eye corners, the nose tip and the mouth corners. These landmark points can be seen in Fig. 3.2 on an average model.

Using a set of training faces and their manual landmarks, generation of an AFM can be summarized as follows:

- A mean distribution of landmarks (*consensus shape*) from the manual landmarks of the faces is found by Procrustes Analysis.
- The landmarks of the consensus shape are transformed so that they represent a face that is fully frontal. Transformation to a fully frontal space is handled by rotating the face so that the eye and mouth planes are approximately parallel to

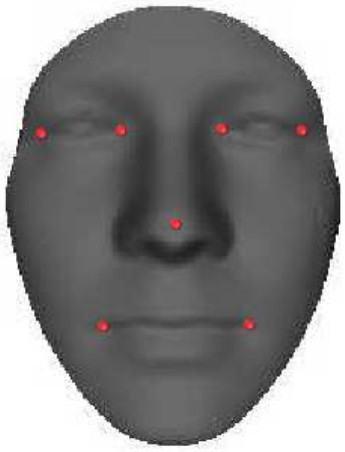


Figure 3.2. The landmarks that are manually located on a face (the inner and outer eye corners, the nose tip, the mouth corners). These landmarks are used for registration purposes.

the x-axis and the z-axis respectively.

- Each training face is warped to the consensus shape using the TPS deformation, where the landmarks of the training face are exactly transformed onto the consensus shape and all the other points of the face are interpolated.
- A regular x-y grid is generated and it is placed on a given face. From this grid, the depth values corresponding to each grid vertex are re-sampled. Using a grid assures that all training faces have points with overlapping x and y values, and the depth values are given for matching points.
- A cropping mask that encloses the facial area is defined. To determine this mask, the distances from nose tip to each landmark are computed. Ten percent of the maximum distance is added to calculate the threshold value. The points that have greater distance to the nose than this threshold value are trimmed off. The remaining points constitute the cropping mask. This cropping mask is used to crop the training faces.
- After all the training faces are cropped according to the cropping mask, the depth values of the cropped training faces are simply averaged.

An example of an AFM generated using this method can be seen in Fig. 3.2 together with the landmarks calculated for the average model. These seven landmarks represent the fiducial points that are used in the coarse registration of faces. Any irregular appearance in the average surface arises from failure in the pre-processing. The database used was collected with a laser sensor which produces holes (especially at the eyes and the mouth). Sometimes the pre-processing is not sufficient to repair some errors that are larger than a limit.

3.3. Cognitive Approach to Multiple AFM Generation

As stated by Gauthier in [51], while objects are mostly recognized at the basic level as being samples from a generic object class such as a cat or a table, humans perceive each face at the most subordinate level and recognize them individually. Humans are *experts* at discriminating faces of individuals by magnifying the differences between the visual items belonging to a class of similar objects.

Although humans become face experts at a young age [52], recognition of faces belonging to another race is a difficult task. This fact is termed as the *the other race effect*. People easily recognize faces belonging to their own race, because they are experienced with their own race rather than others [53],[54]. It is reasonable to conclude that different facial morphologies expose different facial characteristics. Valentine has applied principal component analysis to find different subspace projections and the obtained results emphasized the effect of the other race effect [55].

In the light of the experiments based on the other race effect, we widen the AFM-based registration approach. We propose that better recognition rates can be achieved if the faces are clustered into morphological or gender groups that have greater intra-class similarities than inter-class [38]. For different morphological groups such as Asian, African and Caucasian and different gender groups such as males and females, we construct six different average models. The AFMs constructed for groups having different facial characteristics are given in Fig. 3.5. Note that, the grouping is handled with manually labeling each face with a gender and a morphology information.

3.4. Clustering Approach to Multiple AFM Generation

If morphology and gender are discriminative dimensions of facial surfaces as we have proposed, the intra- and inter-group similarities can be learned for better registration and as a result, for better recognition. Because of ethical concerns, categorization of faces into race groups is not desired. Therefore, based on the suggestion by the other race effect that groups having different morphological or gender characteristics are present in the face space, we propose a method to cluster faces into groups in an unsupervised manner and expect discriminating features to appear automatically.

We propose to use k-means clustering to group faces into discriminative spaces in an unsupervised manner. The clustering of faces is performed on the 3D coordinates of the seven landmarks that are pre-aligned. The number of clusters is predefined manually and it is set to six so that the results can be comparable with the results of the cognitive approach. The cluster consensus shapes are initialized randomly, selected upon landmark points of training faces. At each iteration, training landmark points of training faces are aligned to consensus shapes of clusters with Procrustes analysis and the face is assigned to the cluster whose consensus shape is the closest. The consensus shape is recomputed at each iteration. The multiple AFMs constructed for the shape-space clustered faces are given in Fig. 3.6.

3.5. Experiments

3.5.1. Face Recognition Grand Challenge Data Set

For the experiments carried out in this chapter, the Face Recognition Grand Challenge (FRGC) v.1 face database is used [56]. This database contains 3D surface information represented as a point cloud and the corresponding 2D texture image for a total of 943 frontal and neutral scans collected from 275 subjects. The subjects are either male or female with varying morphology. An example of the data stored for a subject can be seen in Fig. 3.3.

(a) (b)

Figure 3.3. An example of the (a)2D intensity image and (b)the corresponding 3D facial surface information for a subject from the FRGC database.

The methods tested use the three-dimensional facial surface information. Therefore only the 3D part of this database is used. To eliminate irregularities and to fill holes that appear in the facial surface, the 3D information is preprocessed by applying a 9×9 mean and a 9×9 median filter in succession, followed by linear interpolation for filling the gaps.

The gallery is constructed with the first scan of each subject who has multiple 3D images. The remaining images of the gallery subjects appear in the probe set. Therefore an experimental setup with a gallery of 195 subjects and a test set of 659 scans is used.

For each method tested, a recognition experiment and a verification experiment is designed. The rank-1 recognition rate (R-1) and the EER are reported respectively for the recognition and verification experiments. In the recognition experiments, the PSD is used as a distance metric. In the verification experiments, each of the 659 test faces is used for one genuine and 194 false claims.

3.5.2. Automatic Landmarking

The registration methods used are guided by a few fiducial points (e.g. nose tip, eye and mouth corners). These guiding points greatly influence the registration results. To explore the dependence on landmark points, experimental results are reported for the registration process where the use of manually selected ground truth points and the points extracted with a recent automatic landmark localization algorithm [49, 50] are contrasted. The accuracy of the automatic landmarking procedure is given in Fig. 3.4, where the distance from the ground truth is reported. The x-axis indicates the number of candidates and the y-axis reflects the average distance among that number of candidates.

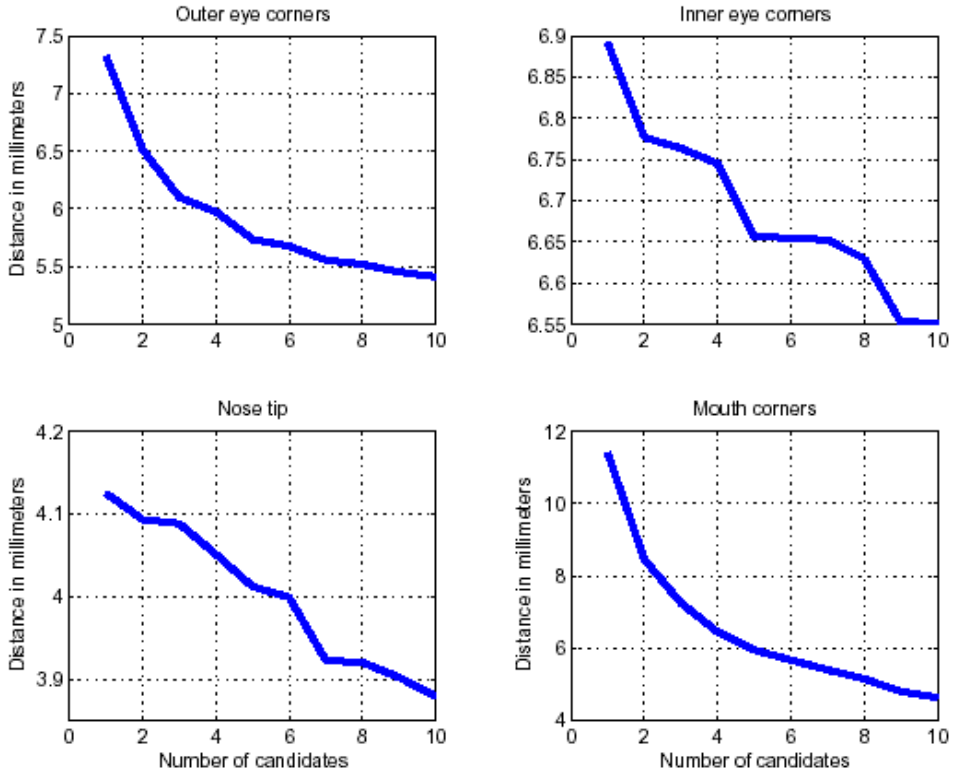


Figure 3.4. Accuracy of the automatical landmarking method, as indicated by average distance to ground truth in millimeters.

3.5.3. Coarse registration

Prior to ICP-based dense registration, coarse alignment is necessary for fast and accurate results. In this subsection, end-results for different coarse alignment approaches are reported as given in Table 3.1. The four different methods use the estimated nose-tip, the ground truth for the nose-tip, seven automatically located landmarks and the ground truth for these seven points. As expected, the worst results were obtained with the nose-tip heuristic, where the nose-tip is accepted as the point with the greatest depth value in the facial surface. Coarse alignment by Procrustes analysis of seven automatically localized landmarks performs better than the nose-tip heuristic. Using only nose-tip ground truth for coarse alignment works well for this dataset, because the faces present have negligible pose differences. The best results are obtained using the Procrustes analysis on ground truth for seven landmark points. For these experiments, only a single AFM is used for ICP-based dense registration.

Table 3.1. Effect of coarse alignment on ICP

	Nose-tip heuristic	Automatic landmarks + Procrustes	Nose ground truth	Manual landmarks + Procrustes
R-1	82.85	87.86	90.60	92.11
EER	14.25	8.12	6.60	6.20

To constrast our novel AFM generation method with the method of Irfanoğlu *et al.* [37], we performed ICP-based registration using both their AFM and the one constructed by the proposed novel method. For coarse alignment, manual landmarks were aligned by Procrustes analysis. With the smoother AFM of Irfanoğlu *et al.*, a rank-1 recognition rate of 86.34 per cent was achieved, as opposed to our 92.11 per cent. The EER was worse by more than two percent when their AFM was used.

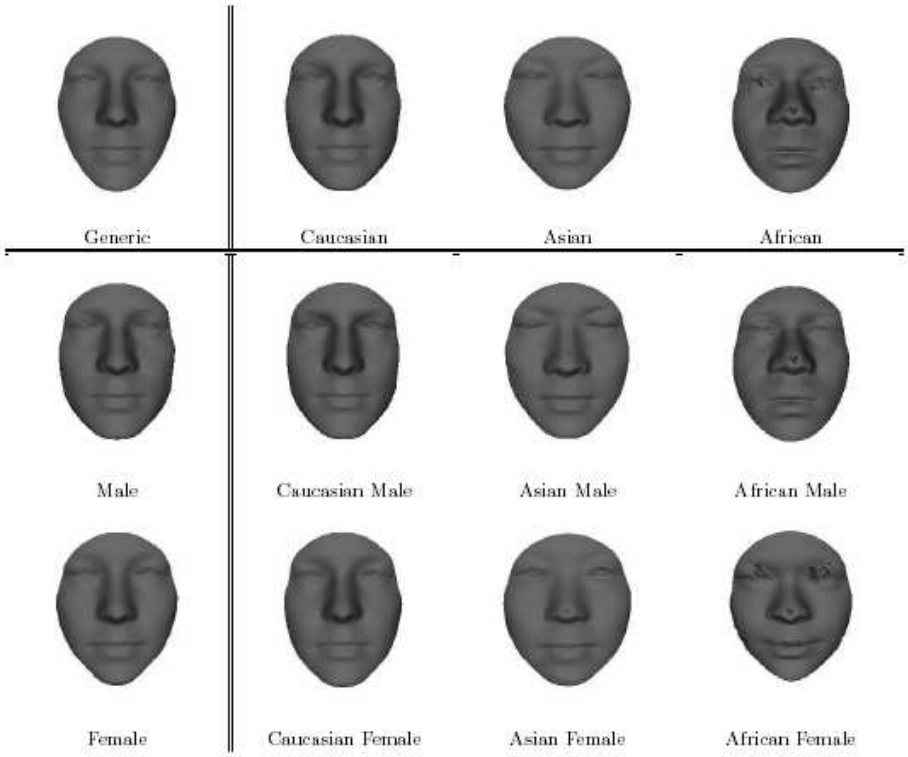


Figure 3.5. Average models for different morphology and gender combinations.

3.5.4. Meta-classification

To test our hypothesis that specialized experts will increase discrimination accuracies, we generated multiple AFMs. For the construction of AFMs, the training faces are grouped manually into gender and morphology classes. The AFMs generated for male and female groups, for three morphological groups (Caucasian, Asian, African) and for all combination groups are given in Fig. 3.5.

In Table 3.2, the verification experiment results with or without specific average faces are given where the registration is handled with the TPS-based method. When the generic-AFM based system and the specific-AFM based system are both supplied with the categorical information, better accuracies are obtained. Improvement in a specific system regarding the corresponding generic system, is a result of better registration using multiple models. The distances between a test face and the gallery faces are computed with an L_1 distance metric, and the worst two percent of corresponding

points are trimmed. If no categorical information is utilized, an EER of 20.10 per cent is obtained. When specific AFMs are used, better results are obtained. Even though the EERs reported are too high because of the deformations in the registration step, the benefit of using multiple models are more obvious.

Table 3.2. Simulation of TPS registration with deformation, EER

	Gender	Morphology	Gender & Morphology
Generic AFM	16.79	18.50	13.87
Specific AFM	14.64	16.97	11.47

It should be noted here that using male and female average models in registration and calculating the point set distance after the aligment can be utilized for gender classification, where a 80 per cent accuracy is achieved.

3.5.5. Shape Space Clustering

Our second hypothesis that discriminating features of face space can be revealed by clustering training faces, is justified if morphology and gender appear as discriminative dimensions of the clusters. We specified the number of clusters manually as six, as in the full morphology-gender combination so that a comparison is possible. The AFMs generated for the clusters are given in Fig. 3.6. The distribution of morphology and gender for each cluster is represented in Fig. 3.7 as pie charts. These charts show that our hypothesis is more or less confirmed where we have clusters dominant in a single gender or a single morphology.

In Table 3.3, recognition rates for ICP and TPS based systems with manual or automatic landmarks are reported. The first row shows the results obtained with a single generic AFM. The next three rows show results with gender-, morphology-, and gender + morphology-based specific AFMs. The results for the last row are obtained with six shape-space derived clusters. For this last case, the registration does not benefit from the injection of categorical information, and each test sample

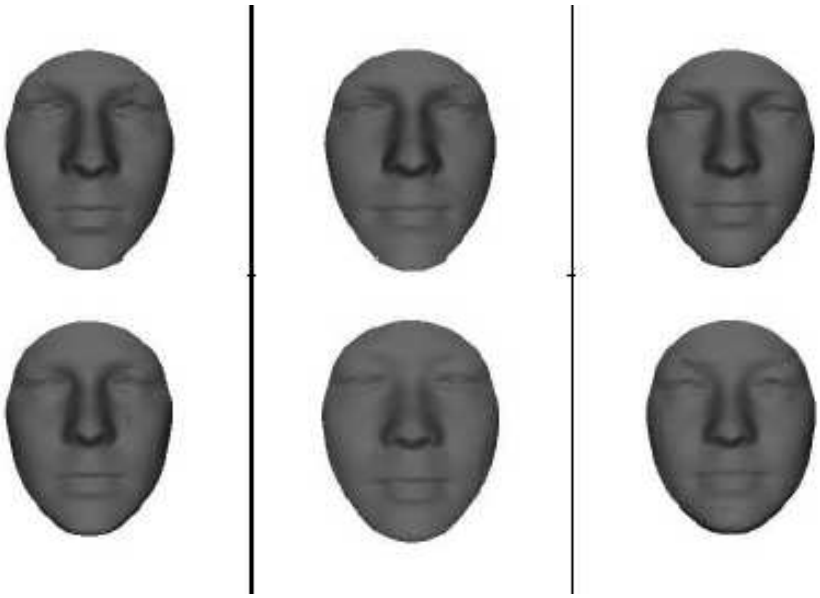


Figure 3.6. Multiple AFMs for Shape space clusters.

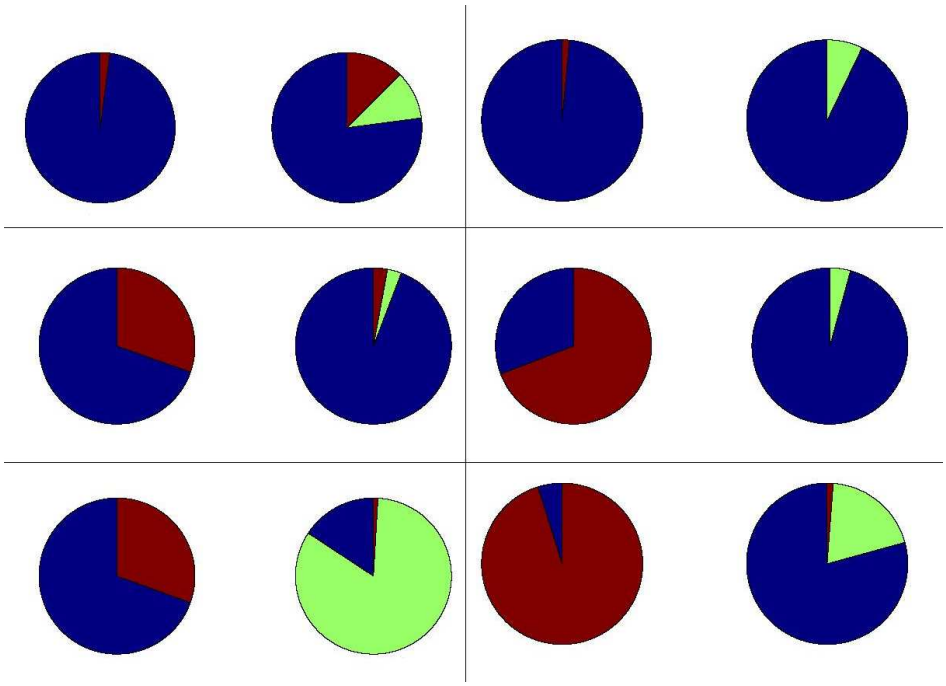


Figure 3.7. Shape space clustering distributions. For each cluster, the gender and morphology distributions are shown in separate pie charts. For six clusters, we have two dominantly male, and two dominantly female clusters, one dominantly Asian cluster, and almost all the males labeled as African are clustered into a single group.

is compared with all the training samples. The best result is obtained when shape space derived specific AFMs are used with ICP-based registration (93.78 per cent). As a baseline experiment, one-to-all ICP is also tested, where each test face is registered to each gallery face separately for distance computation. With manual landmarks, we have obtained 89.07 per cent rank-1 recognition rate, which further demonstrates the usefulness of AFM-based registration.

In [57], Gökberk *et al.*, use the same dataset with the same experimental setup and compare a number of classification methods: Point set difference, non-negative matrix factorization (NMF) and independent component analysis (ICA) coefficients for point clouds, DCT, DFT, PCA, LDA, and ICA projections on depth images, shape indices, mean and principal curvatures, 3D voxel DFT coefficients and 2D Gabor wavelet coefficients. Manual landmarks were used for AFM based registration, where the AFM was generated with the construction method of Irfanoğlu *et al.* [37]. The best classification results are obtained with shape indices (90.06 per cent), principal directions (91.88 per cent) and surface normals (89.07 per cent). When classifiers were fused, best accuracy of 93.63 per cent was achieved using the modified plurality voting method [57].

Table 3.3. Comparison of specific AFMs, rank-1 recognition rates.

	Manual lm. + ICP	Automatic lm. + ICP	Manual lm. + TPS	Automatic lm. + TPS
Generic	92.11	87.86	52.20	42.64
Gender	90.14	86.65	54.63	45.52
Morphology	89.98	86.80	53.87	44.92
Gender & morphology	91.05	86.49	56.90	47.95
Shape space derived	93.78	91.20	47.65	41.58

3.5.6. Regular Re-sampling

In this section, the Eigenface approach over 3D range images is explored. For this purpose, regular re-sampling is applied to acquire the depth images. So far, the classification is done by calculating PSD values over the 3D coordinates of aligned and cropped point clouds and assigning the gallery identity with the smallest total Euclidean distance. Regular re-sampling of faces from a regular $x - y$ grid will make it possible to discard two dimensions and allow a simpler computation in 1D where only the depth values are used. By subspace projection, the gallery and test faces are represented in a much smaller form. Therefore the storage needed for gallery faces will be much smaller and the computation of distances between the gallery and the test faces will be much faster.

For the Eigenface approach, the number of eigenvectors used for subspace projection is determined heuristically so that at least 95 per cent variance is preserved. The subspace projection method is contrasted with the 1D PSD method, where the sum of squared distances of only the depth values obtained after regular re-sampling is used.

Table 3.4 shows results both for PSD approach after regular re-sampling and for Eigenface approach. In the experiments of this section, the African faces are grouped together with Caucasian faces because of the limited number of samples. The drop in accuracy in the subspace projection approach does not exceed one percent if there are sufficient number of training faces. For the gender-morphology combination, the training set is very limited, and the results are reported for only 15 eigenvectors with non-zero eigenvalues in one of the groups and the accuracy decrease is about three percent. The morphology and gender results were obtained using 33 and 49 eigenvectors respectively.

When compared with the results reported in Table 3.3, results obtained by regular re-sampling and by subspace projection are much higher. By discarding the two dimensions, it will also be possible to avoid extra error that will arise from small differences in x and y coordinates. Therefore more accurate distance calculations can be

Table 3.4. Subspace projection after ICP+re-sampling, rank-1 recognition rates

	Manual lm. PSD	Manual lm. Eigenface	Automatic lm. PSD	Automatic lm. Eigenface
Generic	98.18	98.03	98.03	97.88
Gender	96.81	96.21	95.30	95.90
Morphology	96.51	96.05	95.60	94.84
Gender & morphology	96.97	93.78	94.99	91.96
Shape space derived	98.18	97.72	98.03	96.81

carried out. This is illustrated in Fig. 3.8. The facial surface (given in dotted line) initially consists of irregularly spaced surface points (shown with black triangles). In the registration phase, the surfaces are aligned globally by ICP where the total minimum distance between surface points is found. This global alignment does not ensure that the corresponding points are in close alignment when examined locally. After the regular re-sampling step, regularly spaced depth values are obtained (shown in black squares) allowing a more accurate alignment.

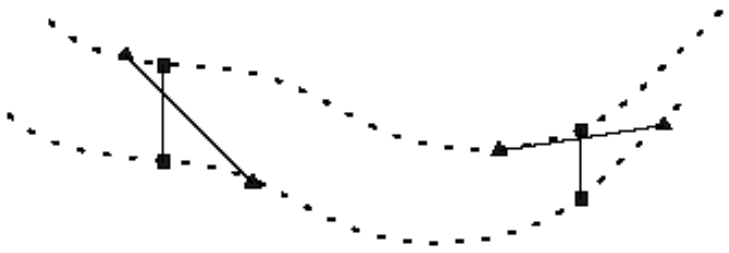


Figure 3.8. Regular re-sampling for ICP is beneficial. Registered surfaces are shown with dotted lines. Surface points are depicted with triangles before re-sampling, and with squares after re-sampling.

We also studied the incorrectly classified samples that were corrected after re-

sampling and we saw that the erroneous classification was caused by irregular sampling of surface points. In Fig. 3.9, the mean distance differences between the correct class and the incorrect class for these samples are shown, where light to dark colors indicate high to low error respectively. When points have irregular distribution, the distance between the corresponding points have a large variance and the error is distributed all over the face. After regular re-sampling, the distribution of error decreases and it is concentrated in areas where the gradient is high, especially for the nose ridge.

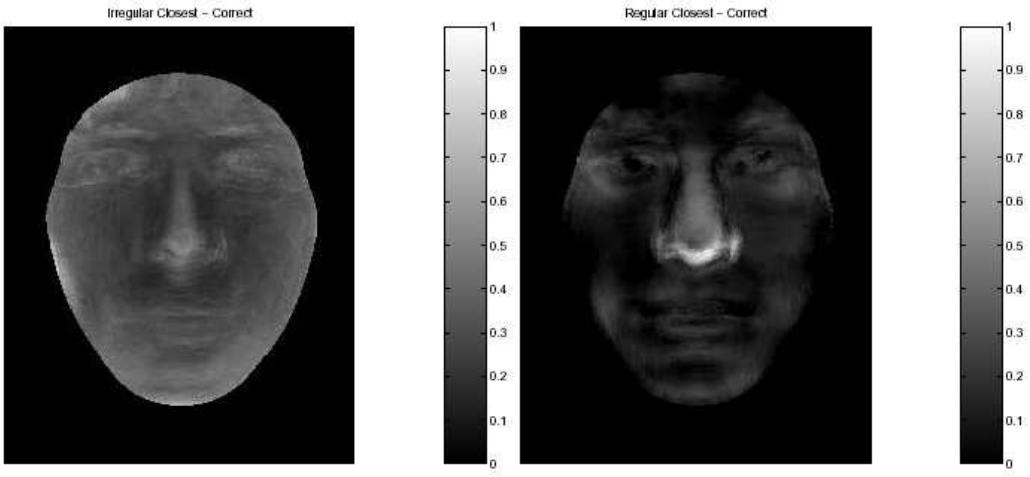


Figure 3.9. Mean point-to-point distance differences in classification. The distances to the correct sample are subtracted from the distances to the closest sample. (a) Irregularly sampled points create errors uniformly on the face surface. (b) Regular sampling reduces the errors on the inner face, close to the nose where the registration is most accurate.

4. RECOGNIZING FACES WITH VARYING EXPRESSION USING MULTIPLE REGION MODELS

Human face is a non-rigid surface which deforms in the presence of expressions initiated by muscle movements. Some of the face recognition systems consider face as rigid and these approaches generally perform well when tested among neutral faces. The accuracy of rigid registration methods decrease when test scans with expressions are introduced. In some studies, for better registration and more accurate recognition region based systems are implemented [13],[17],[35],[36]. In region-based face recognition, face is considered as a composition of facial components. The similarities obtained by registering with individual regions are combined with fusion techniques.

In this work, we introduce a new part-based 3D face recognition method. Motivated by the fast AFM-based registration approach proposed by Irfanoglu *et al.* [37], we propose to generate and utilize Average Region Models (ARMs) for registration. The facial area is divided into several components, and registration of faces encloses separate dense alignments to relative ARMs. The similarity measures between the gallery and test faces obtained for individual regions are fused to improve recognition accuracies.

4.1. Methodology

The overall structure of the implemented part-based 3D face recognition system is given in Fig. 4.1. The main steps of the system are the ARM construction, coarse registration, fine registration, and classification stages.

The gallery images are used in the construction of the AFM. The AFM is subsequently utilized for ARM generation. A test face is aligned to an ARM with two successive alignments: coarse and dense registration. For coarse registration, two approaches are examined: one based on Procrustes analysis of landmark points and the

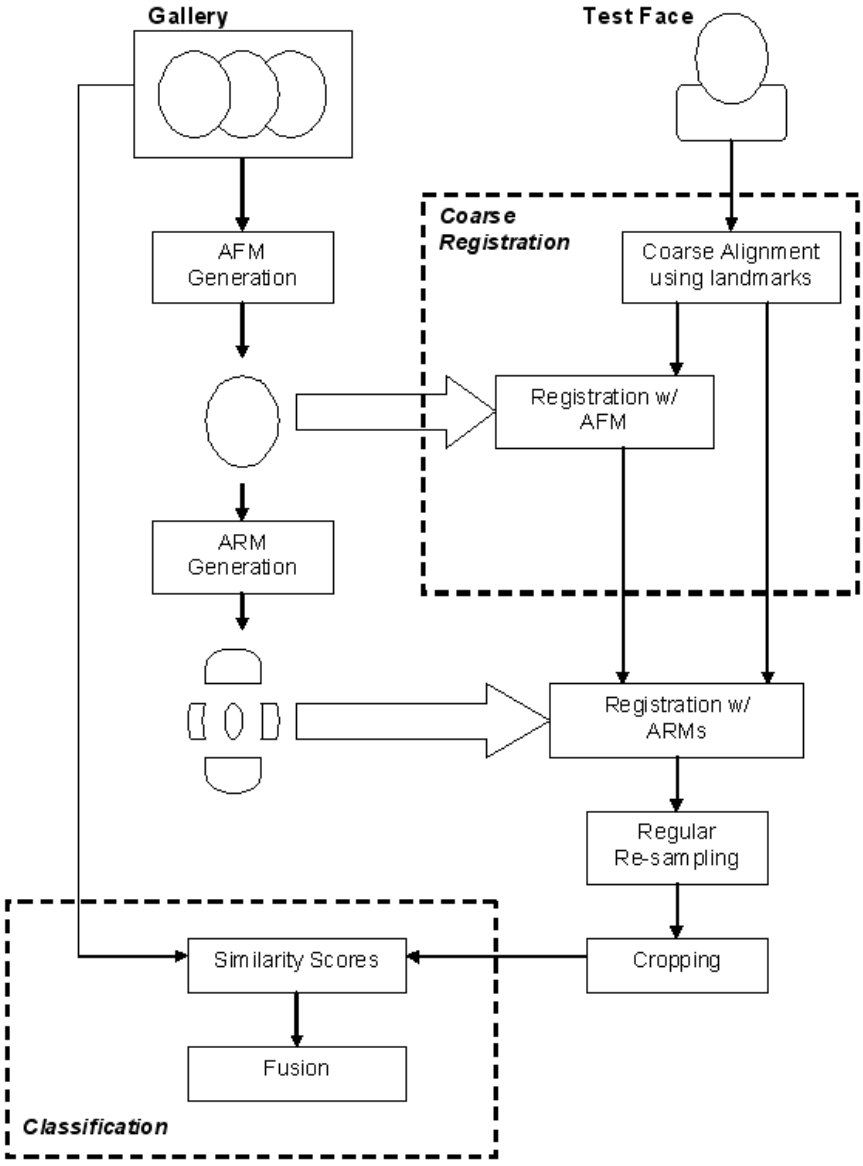


Figure 4.1. The overall 3D face recognition system based on part-based registration.

other based on dense registration with the AFM. The dense registration is carried out by aligning to an average model with the ICP algorithm. To analyze the influence of landmark points on registration, different subsets of landmarks are formed to be used for experiments.

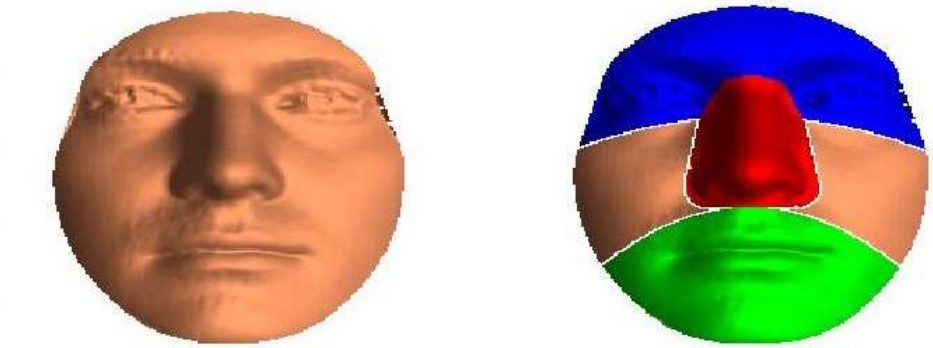
After aligning to an ARM, a test face is regularly re-sampled from an $x-y$ grid, to avoid any extra error that will arise from small differences in the $x-y$ dimensions. Besides decrease in the computed error, the second advantage of regular re-sampling is the reduction of computational complexity. The test face is cropped according to the relative ARM and the PSD is calculated for each of the gallery faces. With regular re-sampling, the PSD calculation is reduced to a simple calculation between the depth vectors. The similarity measures obtained from registering with separate ARMs are combined by fusion techniques to increase the recognition accuracies. As end-results of experiments, rank-1 recognition rates are reported.

4.2. Part-based Registration and Recognition

4.2.1. ARM-based Registration

In registration based on regional models, ARMs obtained for separate facial regions act as index files. ARMs are constructed by manually partitioning an AFM into semantic regions. The AFM is generated from gallery faces using the method explained in Section 3.2. In this work, the whole facial model is divided into four logical parts: forehead-eyes, nose, cheeks, mouth-chin. The AFM generated from EnterFACE v.1 gallery and the subsequent ARMs are given in Fig. 4.2 (a) and (b), respectively.

A test face being examined is registered to each of the regional models separately. After the registration, the related part of the facial surface is labeled and cropped. After the alignment and cropping processes, the region on the test face corresponding to the ARM is obtained. The alignment and cropping processes are already applied to all of the gallery faces. As a gallery set, we have four separate regions for each of the gallery faces. The alignment of a test face to the whole gallery set consists of only



(a) (b)

Figure 4.2. The average models used in registration: (a) neutral AFM constructed from gallery faces, (b) four ARMs for forehead-eyes, nose, cheeks and mouth-chin regions.

four separate registrations, specific for each region. The complexity of registering to separate facial parts is greatly reduced by using ARMs.

4.2.2. Part-based Recognition

In the registration phase, a test face is densely aligned to each regional model separately and the related region of the face is obtained. The registration of gallery faces to ARMs is provided offline, prior to classification. We can use a separate gallery and a separate test set for each facial region. After PSD calculations, four sets of similarity measures between the test and the gallery faces are obtained. These similarities are fused into a single score using several fusion techniques. The fusion techniques can be grouped into three basic categories [57]: abstract-level, rank-level, and score-level methods. A total of six fusion methods from three categories are considered in the scope of this work.

4.2.2.1. Abstract-Level Fusion. In abstract-level fusion, each individual classifier produces one class label. These class labels are combined to provide a single label. For this category plurality voting (PLUR), modified plurality voting (MOD-PLUR), and

highest confidence (HC) fusions are considered.

In PLUR, each expert provides the class label of the nearest gallery subject. Among the set of classifiers, the class label with the highest vote is assigned as the final label. When there are ties, the final label is randomly selected.

Two methods that intend to improve abstract-level fusion are considered: MOD-PLUR and HC. In MOD-PLUR, the approach of plurality voting is improved, where for each classifier, a confidence value is provided together with the class label. When there are ties, the decision is based on the confidence values. HC is another method utilizing confidence values. Among individual experts, the class label having the highest confidence is assigned as the final decision.

The confidence values for MOD-PLUR and HC methods are based on normalized scores. If $d = [d_1, d_2, \ldots, d_N]$ denotes the sorted distances to N gallery classes in an ascending order, the score normalization are normalized as follows:

$$d'_i = \frac{(d_i - d_1)}{median(d) - d_1}, i = 2, \ldots, N \tag{4.1}$$

After score normalization, the classifier confidence can be defined as d'_2. The d'_2 value gives the slope between the normalized scores of the first two top-ranked gallery classes. As the slope increases, the classifier gets more confident in rank-1 class. Therefore this value is selected as the confidence value.

4.2.2.2. Rank-level Fusion. In rank-level fusion, the individual classifiers produce a ranked list of class labels. The ordered set of class codes are fused to provide a single label. We consider Borda count (BORDA) method, which calculates the sum of ranked class labels and returns the class label with the smallest total rank.

4.2.2.3. Score-level Fusion. In score-level fusion, the similarity measures produced by individual experts are combined with simple arithmetic rules. In this work product rule (PROD) and sum rule (SUM) is considered. These rules operate on normalized distance measures. To normalize similarity scores min-max normalization is implemented, where the new distance range is calculated using minimum and maximum distance values, d_{MIN} and d_{MAX}. The normalized distances are computed as follows:

$$d' = \frac{d - d_{MIN}}{d_{MAX} - d_{MIN}} \tag{4.2}$$

4.3. Experiments

4.3.1. EnterFACE Database

We have experimented on the EnterFACE 2D-3D face database, collected during Enterface'07 Workshop held in Boğaziçi University [58]. EnterFACE database consists of several expression, pose and occlusion variations collected from 81 subjects. For each scan the 2D texture image and the corresponding 3D information in the form of a point cloud is acquired. For each 2D and 3D scan, 22 landmark points containing nose-tip, eye corners, mouth corners and tip of the chin are manually labeled. The landmarks are shown in Fig. 4.3 together with the related labels. The age range for subjects is $25 - 35$ and the gender distribution is 51 males to 30 females.

The database consists of two versions. Table 4.1 gives the number of subjects and the number of expression, pose, and occlusion variations for each version. EnterFACE v.1 is acquired for face recognition studies where difficult conditions such as facial expressions, pose variations and partial occlusions are present. Multiple neutral scans for each subject are also available, so that implemented recognition algorithms can be evaluated for neutral scans. EnterFACE v.2 serves both for face recognition and expression understanding studies. The main difference of this version is the variation of expressions. Another difference is the number of neutral scans per person. For most of the subjects only one neutral scans is available and for a few of the subjects a second

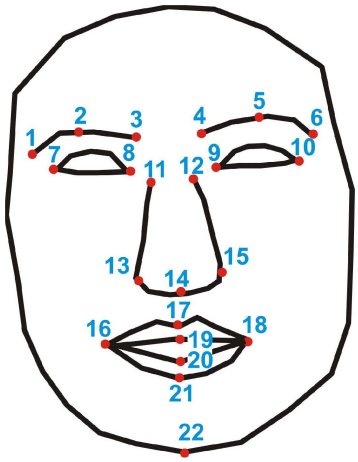


Figure 4.3. The locations and labels of the landmarks manually labeled for EnterFACE database.

neutral scan is also acquired. For EnterFACE v.2, 30 of the 47 subjects are professional actors, contributing more realistic or more evident expressions.

The purpose of the experiments described here is to examine face registration and recognition under expression variations. Therefore, the parts of the database related to the pose variations and occlusions are not utilized.

In the database, there are mainly two groups of facial expressions. The first group consists of Action Units (AU) based on Facial Action Coding System (FACS) which was developed for the taxonomy of plausible facial expressions of humans [59]. For EnterFACE database, more frequently encountered and more easily animated actions are considered. Among the 44 AUs 19 lower face AUs, five upper face AUs, and three upper-lower combination AUs are taken into account. Expressions defined by AUs code the movement of several muscles; thus some AUs are not present for some subjects who cannot control the related muscles. The second group of expressions are related to common emotions: happiness, surprise, fear, sadness, anger, and disgust. The 19 lower face AUs, five upper face AUs, three combination AUs, and six expressions for a sample subject are shown respectively in Fig. 4.4, Fig. 4.5, Fig. 4.6, and Fig. 4.7. All the AUs and expressions are present in v.2, whereas v.1 consists of only the scans

Table 4.1. Number of subjects and number of scan variations for EnterFACE database.

	v.1	v.2
# of subjects	34	47
# of total scans/subject	31	53
# of neutral scans/subject	4	1-2
# of expression variations	10	34
# of pose variations	13	13
# of occlusion variations	4	4

marked with (*).

Both versions are used in the experiments. For each set, the gallery is composed of one neutral scan per subject. The remaining faces, consisting of other neutral scans, expression variations or AUs, form the probe set. The sample numbers of gallery and probe sets for each database version are given in Table 4.2. For EnterFACE v.1, the distribution of test faces into neutral and expression scans is also defined.

Table 4.2. The distribution of faces into gallery and probe sets for both versions of EnterFACE database.

EnterFACE		Gallery	Probe
v.1	neutral scans	34	102
	expression scans	-	339
	total scans	34	441
v.2	total scans	47	1508

4.3.2. AFM-based Registration

Although face is a non-rigid surface, it is considered as non-deformable when rigid registration methods are used for alignment. In the presence of expressions, the 3D facial surface deforms, causing a decrease in the accuracy of rigid registration results.

To support our hypothesis, we tested the AFM-based ICP registration approach on the EnterFACE v.1 database as a baseline algorithm. The coarse alignment is derived by the manually labeled 22 landmark points. The AFM used for fine registration is constructed from the gallery faces. It should also be noted that for each version of the face database, separate average models are generated and the experiments run on each version utilize the respective models.

The PSD classification results for both neutral and non-neutral test sets are given in Table 4.3. The recognition rate for the non-neutral test set is approximately 30 per cent lower than the recognition rate of neutral-test set. The results confirm that rigid-registration approaches considering the whole facial area are less accurate when expression is present. The recognition rate of EnterFACE v.2 test set for the baseline registration approach is also obtained. This result is given in Table 4.3 to allow comparisons with experiments using the proposed part-based approach.

Table 4.3. Recognition rates for AFM-based registration approach.

Recognition Rates	Neutral Test Set	Non-neutral Test Set
v.1	100.00	71.39
v.2	-	67.67

4.3.3. Coarse Alignment Approaches for ARM-based Registration

In ICP-based registration, the end-results are greatly affected by the accuracy of the coarse alignment process. For the proposed ARM-based registration, two different coarse alignment approaches are tested.

In the first method, referred to as the single-pass registration, coarse alignment is handled by Procrustes analysis of 22 manually labeled landmarks. This approach is identical with the coarse alignment phase of the baseline AFM-based registration.

In the second approach, prior to registration with ARMs, a face is densely aligned to the AFM by ICP algorithm. This method, together with ARM-based dense registration, includes two ICP alignments. Therefore it is referred to as the two-pass registration. Utilizing the two-pass approach, we intend to improve the overall registration accuracy.

In Table 4.4, the recognition rates obtained for ARM-based registration with either of the coarse alignment approaches are given. It should be noted that both coarse alignment methods use the whole facial information, either only the landmark locations or the whole facial surface. On the contrary, when densely registering a face with an ARM, only the respective facial region is considered. As the results exhibit, in v.1 no improvement is achieved by the two-pass registration. In v.2, when the expression variety increases, the two-pass approach introduce better results.

4.3.4. Effect of Landmark Selection on Coarse Registration

Both of the proposed coarse registration methods are derived by a coarse alignment of landmark points. Therefore, the accuracy of the registration is highly dependent on these fiducial points. When expression is present, the deformation of the facial surface will alter the landmark locations, thus affecting the overall accuracy. To examine the effect of landmark points on the alignment process, we designed several subsets of landmarks. Results are reported for one-pass ICP registration based on

Table 4.4. Comparison of coarse alignment approaches. (Recognition rates)

ARM	Single-pass		Two-pass	
	v.1	v.2	v.1	v.2
forehead-eyes	82.89	82.16	82.29	83.09
nose	85.55	82.23	85.84	83.95
cheeks	53.39	52.12	54.57	51.72
mouth-chin	42.48	34.55	45.72	34.95

the nose ARM. The landmarks are distributed either over the related area or over the whole surface. In Table 4.5, the labels of landmarks used for each specific experiment are given. The respective locations for the labels are given in Fig. 4.3. In the experiments, initial alignment is handled with Procrustes analysis of selected landmarks. After ARM-based ICP registration for the nose area, classification is done by detecting the nearest neighbor over the PSD values. The PSD-based rank-1 recognition rates for v.1 are given in Table 4.6. The results demonstrate that using points distributed over the face area instead of local landmarks is beneficial. We obtained identical recognition results for 22 and seven (nose-tip, inner and outer eye corners, mouth corners) landmark points. Addition of the tip of the chin to the seven landmarks decreased the accuracy, because the chin is greatly affected by facial expressions.

Table 4.5. Number of landmarks used in different experiments and their labels studying effects of landmark selection to registration.

	Number of Points	Labels
Experiment 1 *(5)*	5	11-15
Experiment 2 *(7)*	7	7-10, 14, 16, 18
Experiment 3 *(8)*	8	7-10, 14, 16, 18, 22
Experiment 4 *(22)*	22	1-22

Table 4.6. Recognition rates for different subsets of landmarks.

	Number of Landmarks			
	(5)	*(7)*	*(8)*	*(22)*
Neutral Test Set	92.16	100.00	100.00	100.00
Non-neutral Test Set	81.12	85.84	81.71	85.84
Total Test Set	83.67	89.12	85.94	89.12

4.3.5. Recognition Performance for Individual Face Regions

The rank-1 recognition rates obtained with registration via individual ARMs are presented in Table 4.4. Nose and forehead-eyes regions appear to be less affected by expressions and they offer better performances. It was proposed by Mian *et al.* [36], that nose and forehead-eyes area can be used for recognition instead of considering the whole facial surface. Some recognition systems were proposed that work on only regions selected from the nose area [13],[35] and they reported efficient results on the FRGC database. Here, we see that the nose area is not sufficient for good recognition performance when diverse expressions are present in the database such as EnterFACE database. In Table 4.7, all possible combinations of regions are generated and for each combination the considered regional classifiers are combined with the product rule. We can conclude from these results that the use of all the regions of the facial surface is beneficial in recognition.

The worst results were obtained with the chin ARM. The chin region encapsulates the mouth, which is the mostly deformed area for some expressions. Especially when the mouth is open, the topology of the facial surface is changed greatly. We proposed to apply TPS warping to the chin area to neutralize the deformation caused by an expression. Even though the deformations were reduced by TPS warping, no improvement was achieved in the recognition phase. While neutralizing the deformation, the facial information was altered by the applied warping, resulting to incorrect

classifications.

Table 4.7. Recognition rates after classifiers are combined with the product rule. R1, R2, R3, and R4 denote nose, eye-forehead, cheek, and mouth-chin regions, respectively.

	One-pass		Two-pass	
	v.1	v.2	v.1	v.2
R1,R2	97.05	94.36	97.05	94.76
R1,R3	84.66	83.22	84.37	83.69
R1,R4	88.50	83.89	88.50	84.55
R2,R3	89.38	89.46	88.79	89.92
R2,R4	94.40	87.86	94.10	89.06
R3,R4	59.88	58.22	58.41	58.36
R1,R2,R3	94.10	93.44	94.40	94.23
R1,R2,R4	97.35	94.83	97.94	95.03
R1,R3,R4	85.55	84.15	86.73	85.21
R2,R3,R4	90.27	89.72	90.86	89.99
R1,R2,R3,R4	95.28	94.76	95.87	95.29

4.3.6. Fusion Methods

Several fusion techniques from three different categories are implemented to improve recognition accuracies of individual classifiers. The different experts are the classifiers based on individual ARMs. The first category includes the abstract-level fusion methods that are used to combine class labels obtained from registering via individual ARMs. The considered methods are plurality voting (PLUR), confidence aided modified plurality voting (MOD-PLUR), and highest confidence (HC). The second category is rank-level fusion, where a ranked list of class labels obtained from registration based on ARMs are fused. From this set, Borda Count (BORDA) is considered. The third category of techniques are in the score-level group, where the similarity scores obtained

from individual experts are fused using basic arithmetic rules. Sum (SUM) and product (PROD) rules are implemented. In Table 4.8, the recognition rates obtained after the fusion of individual classifiers are given. In these experiments, two-pass coarse alignment was applied.

Table 4.8. Recognition rates after individual classifiers are fused.

	Two-pass	
	v.1	v.2
PLUR	89.38	88.53
MOD-PLUR	94.40	94.03
HC	94.40	92.97
BORDA	82.77	76.15
SUM	88.79	91.78
PROD	95.87	95.29

The best recognition results are obtained with the product rule. The improvement caused by fusion is about 10 per cent when compared with individual classifier accuracies. The confidence-aided methods, MOD-PLUR and HC, give better accuracies than BORDA or the basic PLUR. They can be considered as good alternatives for fusion.

Figure 4.4. Lower Face Action Units: (a) lower lip depressor, (b) lips part, (c) jaw drop, (d) mouth stretch*, (e) lip corner puller*, (f) left lip corner puller, (g) right lip corner puller, (h) dimpler, (i) lip stretcher, (j) lip corner depressor, (k) chin raiser, (l) lip funneler, (m) lip puckerer, (n) lip tightener, (o) lip presser, (p) lip suck*, (q) upper lip raiser, (r) nose wrinkler*, (s) cheek puff*, (t) low intensity lower lip depressor.

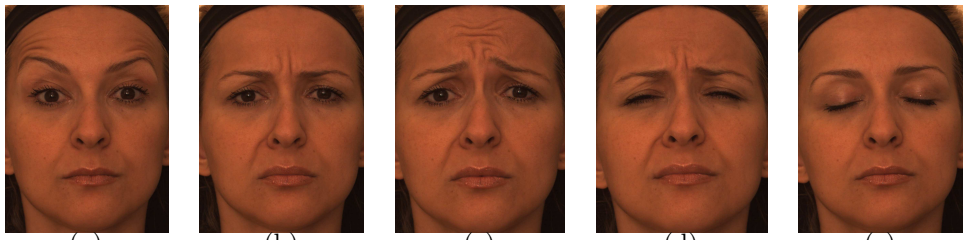


Figure 4.5. Upper Face Action Units: (a) outer brow raiser*, (b) brow lowerer*, (c) inner brow raiser, (d) squint, (e) eyes closed*.

Figure 4.6. Some Action Unit Combinations: (a) lip corner puller + lip corner depressor, (b) jaw drop + lip corner puller, (c) lip funneler + lips part*.

Figure 4.7. Emotional expressions: (a) happiness*, (b) surprise, (c) fear, (d) sadness, (e) angry, (f) disgust.

5. CONCLUSIONS

To recognize a 3D face, one has to compare it with facial surfaces in the gallery. Before any comparisons can be made between two surfaces, the correspondence between the faces must be defined. Therefore, a registration step prior to recognition is necessary.

In real-time 3D face recognition systems, beside the accuracy of the algorithms implemented, the computational requirements must also be considered. Registration greatly influences the end-results of a recognition system and it is usually a costly process. To reduce the computational cost, we considered the AFM-based approach, where an intermediate average model is used for indexing surface points. The AFM-based approach is based on the idea of using a common model to act as an intermediate surface between two surfaces to be registered: Both surfaces are registered to the AFM, which establishes the dense correspondence of points on surface one and two. The establishment of this dense correspondence brings additional advantages such as the benefit of using PCA or LDA. Since the correspondences between gallery faces and the AFM can be done offline, only a single registration with the test face is sufficient at testing, greatly speeding up the registration.

A crucial step for the success of the AFM-based technique is how the AFM is generated: In this work, we have developed a better AFM generation method than the method of Irfanoğlu *et al.*. However, no matter how good the AFM generation method is, it cannot be successful if the underlying faces are too diverse.

This diversity may be due to two factors: different face morphologies or different expressions. In this thesis, we have developed two approaches to deal with these situations.

In the first case, we group faces with similar morphologies together. We construct several morphology/gender groups and construct different models for each.

For AFM-based approach, we investigated two registration methods: a rigid approach based on the ICP algorithm and a non-rigid approach based on TPS warping. The ICP algorithm, which is greatly slower than the TPS method, yields practical results only if the computational complexity is reduced with the use of an average model. As the results show, the TPS based method is much inferior to ICP in accuracy despite its speed. When a face is deformed for registration, the characteristic information concealed in the facial surface disappears, and as a result the identity cannot be correctly specified.

For the construction of category-specific AFMs, we have proposed two approaches to group the training faces: A cognitive based approach and a clustering approach. The use of AFMs constructed with clustering on shape space increases the accuracy of registration and the subsequent recognition. The clusters expose morphology and gender as discriminating dimensions of the face space, verifying the *other race effect*. We obtained the best results with ICP registration via shape space clustered AFMs.

Our experimental results have also confirmed that for the ICP-based approach a good initialization is essential. For an automatic face recognition system, automatic landmark localization is necessary. If the faces are frontal and upright the nose-tip heuristic can be used, nevertheless hair, clothing or chin can sometimes be detected as the nose-tip. When other poses are present, the nose-tip heuristic will perform even worse. When the number of landmarks used is increased, the accuracy of registration is improved.

With regular re-sampling succeeding the alignment, the extra error caused by small changes in x and y coordinates of the corresponding points can be prevented. The results show that the accuracy increases with regular re-sampling. With the Eigenface approach, the dimensionality of the face space is lowered while preserving high recognition accuracies.

For the first part of experiments, exploring registration via multiple AFMs, the FRGC v.1 database is used. The effect of the AFM used is tested and 92.11 per cent

is obtained when the AFM is obtained with the proposed AFM generation method, whereas the accuracy was 86.34 per cent when the AFM of Irfanoğlu *et al.* was used. The effect of landmarks used in coarse registration is examined and it is shown that better registration is obtained with more landmark points. A recognition rate of 92.11 per cent is obtained with seven landmarks as opposed to 90.60 per cent rate when only the nose-tip is used. In generating specific AFMs gender-, morphology-, gender+morphology- and cluster-based groups are formed and the best result of 93.78 per cent achieved with the cluster AFMs. The performance is improved with regular re-sampling of depth values and a recognition rate of 98.18 per cent is obtained. With PCA over regularly re-sampled depth values the performance is 98.08 per cent, where the storage requirements and the computational cost is reduced with no accuracy loss.

In the second part of this work, we proposed a registration method based on ARMs to reduce the effect of surface deformations due to facial expressions. The ARMs are constructed by dividing a face model into meaningful parts. A face is segmented automatically by individual registrations via specific ARMs.

As the results show, using ARMs instead of a whole face model for registration is beneficial in the presence of expression in the face space. The recognition performances show that nose and forehead-eye regions were less affected by surface deformations, but they were not sufficient by themselves for recognition. Combining all of the regions leads to improvement. Fusion techniques were used to combine the results of individual regions, so that the all the regions can be considered. Best results were obtained with the product rule. As an alternative, using the confidence values of classifiers provide to use the decision of regions that are less deformed for a specific expression.

Two different approaches for coarse alignment of faces are examined: In single-pass approach, the faces are coarsely alignment using only the landmark points and fine registration is done via ARM-based ICP. In two-pass approach, the faces are aligned densely to AFM via ICP and a second ARM-based ICP is used to improve the registration. When deformations due to various expressions are present, the two-pass registration offers more accurate results.

The effect of landmark selection on registration is also examined. The results show that the use of landmarks regarding the whole facial surface instead of regional points is beneficial in registration.

In the second set of experiments, where the diversity in the face space due to facial expressions is considered, the EnterFACE database including various expressions is used. As a baseline algorithm, we tested the AFM-based registration approach and with non-neutral test scans we obtained a recognition rate of 71.39 per cent whereas the performance was 100 per cent when only neutral test scans were considered. With region-based registration, a recognition rate of 85.84 per cent was achieved when nose area was used. For all possible combinations of facial regions, we have calculated the performances and concluded that the use of all regions is beneficial in recognition. To consider all of the regions, we experimented with several fusion techniques and achieved a best result of 95.87 per cent with the product rule. Confidence aided methods with a performance of 94.40 per cent appear as good alternatives for fusing classifiers. In general, it is shown that with regional registration a performance improvement of 15 per cent can be achieved and the results can be further improved with the fusion techniques where an extra increase of ten per cent can be obtained.

As future directions, the number of regions for ARM-based registration should be increased, where ARMs better revealing regional face deformations are obtained. Especially the mouth and chin area should be examined exhaustively, where the topology of the surface is greatly altered when mouth is open. By defining more significant regions such as separate upper and lower mouth regions, the deformation neutralization can also be successful.

To obtain better registration in the presence of expression, the number of ARMs specific for a region can be increased. A region can have various forms under different expressions and multiple ARMs can be achieved to represent the diversity. Better registration can then be obtained via registering to multiple ARMs specific for each individual regions.

The ARM-based approach can be applied for face recognition under occlusion. The faces can be registered with ARMs and the recognition can be achieved by fusion where the occluded regions can be detected by defined threshold values.

APPENDIX A: MATHEMATICAL FUNDAMENTALS FOR QUATERNIONS

In mathematics, quaternions are an extension of complex numbers with a non-commutative property. They were first described by the Irish mathematician Sir William Rowan Hamilton in 1843 and applied to mechanics in three-dimensional space. Quaternions are used in both theoretical and applied mathematics for calculations involving three-dimensional rotations, such as in 3D computer graphics. Basic information about quaternions is given in this chapter. Further details about quaternions can be found in [60].

A.1. Definition

A general rotation in 3D is fully defined by four numbers: one for the angle of the rotation and three for the axis of the rotation. Quaternions are 4-tuples defined by Hamilton and they can be considered as elements of a four-dimensional vector space, whose bases are:

$$\mathbf{i} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}, \quad \mathbf{j} = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix},$$
$$\mathbf{k} = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \quad \mathbf{1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{A.1}$$

Quaternions have four dimensions, one real dimension and 3 imaginary dimensions. Each of these imaginary dimensions has a unit value of the square root of -1, but they are different square roots of -1 all mutually perpendicular to each other, known as i,j and k. A quaternion $\mathbf{q}$ can be defined as:

$$\mathbb{H} = \{w + x\mathbf{i} + y\mathbf{j} + z\mathbf{k} | w, x, y, z \in \mathbb{R}\} \tag{A.2}$$

A.2. Basic Operations

Quaternions satisfy the following identities, known as Hamilton's Rules:

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -\mathbf{1}, \quad \mathbf{ij} = -\mathbf{ji} = \mathbf{k}. \tag{A.3}$$

Unlike multiplication of real or complex numbers, multiplication of quaternions is not commutative: e.g. $\mathbf{ij} = \mathbf{k}$, while $\mathbf{ji} = -\mathbf{k}$. The multiplication rules are shown in Table A.1:

Table A.1. Multiplication rules for quaternions.

	$\mathbf{1}$	$\mathbf{i}$	$\mathbf{j}$	$\mathbf{k}$
$\mathbf{1}$	$\mathbf{1}$	$\mathbf{i}$	$\mathbf{j}$	$\mathbf{k}$
$\mathbf{i}$	$\mathbf{i}$	$-\mathbf{1}$	$\mathbf{k}$	$-\mathbf{j}$
$\mathbf{j}$	$\mathbf{j}$	$-\mathbf{k}$	$-\mathbf{1}$	$\mathbf{i}$
$\mathbf{k}$	$\mathbf{k}$	$\mathbf{j}$	$-\mathbf{i}$	$-\mathbf{1}$

Quaternions can also be interpreted as a combination of a scalar and a vector, where a quaternion can be represented as the sum of the 4-tuple (w, x, y, z): $\mathbf{q} = w + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. A quaternion can also be represented as the sum of the scalar-vector pair $[s, \mathbf{v}]$, where $s = w$ and $\mathbf{v} = (x, y, z)$. The conjugate quaternion is given by

$\mathbf{q}^* = w - x\mathbf{i} - y\mathbf{j} - z\mathbf{k}$.

The sum/difference of two quaternions $\mathbf{q_1}$ and $\mathbf{q_2}$ is:

$$\mathbf{q_1} \pm \mathbf{q_2} = (w_1 \pm w_2) + (x_1 \pm x_2)\mathbf{i} + (y_1 \pm y_2)\mathbf{j} + (z_1 \pm z_2)\mathbf{k} = [(s_1 \pm s_2), (\mathbf{v_1} \pm \mathbf{v_2})]. \quad \text{(A.4)}$$

The product is:

$$\begin{aligned} \mathbf{q_1} \cdot \mathbf{q_2} &= (w_1 w_2 - x_1 x_2 - y_1 y_2 - z_1 z_2) + (w_1 x_2 + x_1 w_2 + y_1 z_2 - z_1 y_2)\mathbf{i} \\ &\quad + (w_1 y_2 + x_1 z_2 + y_1 w_2 + z_1 x_2)\mathbf{j} + (w_1 z_2 + x_1 y_2 + y_1 x_2 + z_1 w_2)\mathbf{k} \quad \text{(A.5)} \\ &= [(s_1 s_2 - \mathbf{v_1} \bullet \mathbf{v_2}), (s_1 \mathbf{v_2} + s_2 \mathbf{v_1} + \mathbf{v_1} \times \mathbf{v_2})]. \end{aligned}$$

The quaternion product has the associative property, i.e., $(\mathbf{q_1 q_2})\mathbf{q_3} = \mathbf{q_1}(\mathbf{q_2 q_3})$, but it does not have the commutative property.
The size of a quaternion can be given by its *norm*, which is defined as:

$$|\mathbf{q}| = \mathbf{q} \cdot \mathbf{q}^* = \mathbf{q}^* \cdot \mathbf{q} = \sqrt{w^2 + x^2 + y^2 + z^2} = \sqrt{s^2 + x^2 + y^2 + z^2}. \quad \text{(A.6)}$$

A unit quaternion is one for which $|\mathbf{q}| = 1$.
The inverse of a quaternion is given by:

$$\mathbf{q^{-1}} = \frac{\mathbf{q}^*}{(\mathbf{q q}^*)} = \frac{\mathbf{q}^*}{|\mathbf{q}|^2} = \frac{\mathbf{q^2}}{w^2 + x^2 + y^2 + z^2}. \quad \text{(A.7)}$$

The multiplication of a quaternion with its inverse can be expressed as follows:

$$\mathbf{q q^{-1}} = [1, (0, 0, 0)] = [1, \mathbf{0}]. \quad \text{(A.8)}$$

A.3. Rotation Operations

Quaternions provide a convenient mathematical notation for representing orientations and rotations of objects.

The goal is to find a formula that expresses rotation in 3D space using quaternion multiplication. Ideally, this formula should be an analog to the complex multiplication used to represent 2D rotations:

$$f(w) = zw, \tag{A.9}$$

where $z = e^{\alpha i}$ is used to define a rotation by an angle α.

The desired 3D rotation formula cannot be simple quaternion multiplication, because rotating a vector (represented as a quaternion with zero real part) should yield a vector, but multiplying a vector with an arbitrary quaternion may result in a non-vector (with non-zero real part).

However, it turns out that we can cancel the real part if we multiply by a quaternion from one side and by the inverse of that quaternion from the other side. Let $\mathbf{z} = a + \mathbf{u}$ be a non-zero quaternion, and consider the function:

$$f(\mathbf{v}) = \mathbf{z}\,\mathbf{v}\,\mathbf{z}^{-1} \tag{A.10}$$

where $\mathbf{z}^{-1}$ is the multiplicative inverse of $\mathbf{z}$, and $\mathbf{v}$ is a vector in quaternion form. The function f is known as *conjugation* by $\mathbf{z}$. Note that the real part of $f(\mathbf{v})$ is zero, because in general $\mathbf{zw}$ and $\mathbf{wz}$ have the same real part for any quaternions $\mathbf{z}$ and $\mathbf{w}$. Therefore:

$$\Re(\mathbf{z}\ \mathbf{v}\ \mathbf{z}^{-1}) = \Re(\mathbf{z}\ (\mathbf{v}\ \mathbf{z}^{-1})) = \Re((\mathbf{v}\ \mathbf{z}^{-1})\ \mathbf{z}) = \Re(\mathbf{v}\ 1) = 0 \tag{A.11}$$

Furthermore, we have $f(\mathbf{v}) = \mathbf{v}$ if and only if $\mathbf{v}$ and the imaginary part $\mathbf{u}$ of $\mathbf{z}$ are

collinear (because $f(\mathbf{v}) = \mathbf{v}$ means $\mathbf{vz} = \mathbf{zv}$). Hence f is a rotation whose axis of rotation passes through the origin and is given by the vector $\mathbf{u}$.

Note that conjugation with $\mathbf{z}$ is the equivalent to conjugation with $r\mathbf{z}$ for any real number r. We can thus restrict our attention to the quaternions of absolute value one, the so-called *unit* quaternions. Note that even then $\mathbf{z}$ and $-\mathbf{z}$ represent the same rotation.

Inverting unit quaternions is especially easy: If $|\mathbf{z}| = 1$, then $\mathbf{z}^{-1} = \mathbf{z}^*$ (the conjugate $\mathbf{z}^*$ of the quaternion $\mathbf{z} = a + \mathbf{v}$ is defined as $\mathbf{z}^* = a - \mathbf{v}$) and this makes our rotation formula even easier.

It turns out that the angle of rotation α is also easy to read off if we are dealing with a unit quaternion $\mathbf{z} = a + \mathbf{v}$:

$$a = \cos\frac{\alpha}{2}. \tag{A.12}$$

To summarize, a counterclockwise rotation through an angle α about an axis $\mathbf{v}$ can be represented via conjugation by the unit quaternion $\mathbf{z}$:

$$\mathbf{z} = \cos\frac{\alpha}{2} + \sin\frac{\alpha}{2}\,\hat{\mathbf{v}} \tag{A.13}$$

where $\hat{\mathbf{v}}$ is the normalized vector:

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{\|\mathbf{v}\|}. \tag{A.14}$$

The composition of two rotations corresponds to quaternion multiplication: if the rotation f is represented by conjugation with the quaternion $\mathbf{z}$ and the rotation g is represented by conjugation with $\mathbf{w}$, then the composition of f with g is represented by conjugation with $\mathbf{zw}$.

If one wishes to rotate about an axis that does not pass through the origin, then one first translates the vectors into the origin, conjugates, and translates back.

REFERENCES

1. Phillips, P. J., W. T. Scruggs, A. J. O. Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott and M. Sharpe, *FRVT 2006 and ICE 2006 Large-Scale Results, Technical Report NISTIR 7408*, Tech. rep., NIST, 2007.

2. Medioni, G. and R. Waupotitsch, "Face modeling and recognition in 3-D", *Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03)*, pp. 232–233, 2003.

3. Hesher, C., A. Srivastava and G. Erlebacher, "A novel technique for face recognition using range imaging", *Seventh International Symposium on Signal Processing and Its Applications*, pp. 201–204, 2003.

4. Gordon, G. G., "Face recognition based on depth and curvature features", *Computer Vision and Pattern Recognition (CVPR)*, pp. 108–110, 1992.

5. Nagamine, T., T. Uemura and I. Masuda, "3D facial image analysis for human identification", *International Conference on Pattern Recognition (ICPR 1992)*, pp. 324–327, 1992.

6. Chua, C., F. Han and Y. K. Ho, "3D human face recognition using point signature", *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 233–238, 2000.

7. Pan, G., S. Han, Z. Wu and Y. Wang, "3D face recognition using mapped depth images", *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005.

8. Bowyer, K. W., 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, pp. 1–15, 2006.

9. Scheenstra, A., A. Ruifrok and R. C. Veltkamp, "A survey of 3D face recognition methods", *International Conference of Audio- and Video-based Biometric Person Authentication (AVBPA 2005)*, Vol. 3546, pp. 891–899, 2005.

10. Cartoux, J. Y., J. T. LaPreste and M. Richetin, "Face authentification or recognition by profile extraction from range images", *Proceedings of the Workshop on Interpretation of 3D Scenes*, pp. 194–199, 1989.

11. Lee, J. C. and E. Milios, "Matching range images of human faces", *International Conference on Computer Vision*, pp. 722–726, 1990.

12. Bronstein, A. M., M. M. Bronstein and R. Kimmel, "Three-dimensional face recognition", *International Journal of Computer Vision*, Vol. 5, No. 30, 2005.

13. Chang, K. I., K. W. Bowyer and P. J. Flynn, "Adaptive rigid multi-region selection for handling expression variation in 3D face recognition", *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pp. 157–164, 2005.

14. Lu, X., D. Colbry and A. K. Jain, "Matching 2.5D scans for face recognition", *International Conference on Pattern Recognition (ICPR 2004)*, pp. 362–366, 2004.

15. Moreno, A. B., A. Sanchez, J. F. Velez and F. J. Diaz, "Face recognition using 3D surface-extracted descriptors", *Irish Machine Vision and Image Processing Conference (IMVIP 2003)*, 2003.

16. Faltemier, T., K. W. Bowyer and P. J. Flynn, "A region ensemble for 3d face recognition", *(submitted)*, 2007.

17. Kakadiaris, I. A., 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, 2004.

18. Lu, X. and A. K. Jain, "Deformation modeling for robust 3D face matching", *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, pp. 1377–1383, 2006.

19. Achermann, B., X. Jiang and H. Bunke, "Face recognition using range images", *International Conference on Virtual Systems and Multimedia*, pp. 129–136, 1997.

20. Achermann, B. and H. Bunke, "Classifying range images of human faces with Hausdorff distance", *15-th International Conference on Pattern Recognition*, pp. 809–813, 2000.

21. Pan, G., Z. Wu and Y. Pan, "Automatic 3D face verification from range data", *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 193–196, 2003.

22. Lee, Y. and J. Shim, "Curvature-based human face recognition using depth-weighted Hausdorff distance", *International Conference on Image Processing (ICIP)*, pp. 1429–1432, 2004.

23. Russ, T. D., M. W. Koch and C. Q. Little, "3D facial recognition: a quantitative analysis", *45-th Annual Meeting of the Institute of Nuclear Materials Management (INMM)*, 2004.

24. Russ, T. D., M. W. Koch and C. Q. Little, "A 2D range hausdorff approach for 3D face recognition", *IEEE Workshop on Face Recognition Grand Challenge Experiments*, 2005.

25. Lu, X., A. K. Jain and D. Colbry, "Matching 2.5D face scans to 3D models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 1, pp. 31–43, 2006.

26. Faltemier, T., K. W. Bowyer and P. J. Flynn, "Using a Multi-Instance Enrollment Representation to Improve 3D Face Recognition", *IEEE International Conference*

on Biometrics: Theory, Applications, and Systems (BTAS'07), pp. 1–6, 2007.

27. Tanaka, H. T., M. Ikeda and H. Chiaki, "Curvature-based face surface recognition using spherical correlation principal directions for curved object recognition", *Third International Conference on Automated Face and Gesture Recognition*, pp. 372–377, 1998.

28. Lee, Y., H. Song, U. Yang and H. S. K. Sohn, "Local feature based 3D face recognition", *International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA 2005)*, pp. 909–918, 2005.

29. Li, C. and A. Barreto, "Intelligent expression-independent face recognition algorithm", *Florida Conference on Recent Advances in Robotics (FCRAR'06)*, 2006.

30. Xu, C., Y. Wang, T. Tan and L. Quan, "Automatic 3D face recognition combining global geometric features with local shape variation information", *Sixth International Conference on Automated Face and Gesture Recognition*, pp. 308–313, 2004.

31. Maurer, T., D. Guigonis, I. Maslov, B. Pesenti, A. Tsaregorodtsev and D. W. G. Medioni, "Performance of geometrix activeID 3D face recognition engine on the FRGC data", *Face Recognition Grand Challenge Workshop*, pp. 154–160, 2005.

32. Gökberk, B., A. A. Salah and L. Akarun, "Rank-based decision fusion for 3D shape-based face recognition", *International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA 2005)*, pp. 1019–1028, 2005.

33. Gökberk, B. and L. Akarun, "Comparative analysis of decision-level fusion algorithms for 3D face recognition", *18th International Conference on Pattern Recognition (ICPR'06)*, pp. 1018–1021, 2006.

34. Cook, J., V. Chandran and C. Fookes, "3D face recognition using log-gabor templates", *Biritish Machine Vision Conference*, pp. 83–92, 2006.

35. Faltemier, T., K. W. Bowyer and P. J. Flynn, "3D face recognition with region

committee voting", *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, pp. 318–325, 2006.

36. Mian, A. S., 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.

37. İrfanoğlu, M. O., B. Gökberk and L. Akarun, "3D Shape-Based Face Recognition Using Automatically Registered Facial Surfaces", *International Conference on Pattern Recognition (ICPR'04)*, Vol. 4, pp. 183–186, 2004.

38. Salah, A. A., N. Alyüz and L. Akarun, "Alternative face models for 3D face registration", *SPIE Conf. on Electronic Imaging, Vision Geometry*, San Jose, 2007.

39. Lester, H. and S. R. Arridge, "A Survey of Hierarchical Non-Linear Medical Image Registration", *Pattern Recognition*, Vol. 32, pp. 129–149, 1999.

40. Bookstein, F. L., "The measurement of biological shape and shape change", *Lecture Notes Biomathematics*, Vol. 24, 1978.

41. Gower, J. C., "Generalized Procrustes Analysis", *Psychometrika*, Vol. 40, No. 1, pp. 33–51, 1975.

42. Goodall, C., "Procrustes methods in the statistical analysis of shape", *Journal of the Royal Statistical Society B*, Vol. 53, No. 2, pp. 285–339, 1991.

43. Rohlf, F. J. and D. Slice, "Extensions of the Procrustes Method for the Optimal Superimposition of Landmarks", *Systematic Zoology*, Vol. 39, No. 1, pp. 40–59, 1990.

44. Eckart, C. and G. Young, "The approximation of one matrix by another of lower rank", *Psychometrika*, Vol. 1, pp. 211–218, 1936.

45. Besl, P. and N. McKay, "A Method for Registration of 3-D Shapes", *IEEE Trans.*

Pattern Analysis and Machine Intelligence, Vol. 14, No. 2, pp. 239–256, 1992.

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

47. Bookstein, F. L., "Principal warps: thin-plate splines and the decomposition of deformations", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 11, pp. 567–585, 1989.

48. Salah, A. A., N. Alyüz and L. Akarun, "3D Face Registration with Average Face Models", *Journal of Electronic Imaging*, in press.

49. Salah, A. A., H. Çınar Akakın, L. Akarun and B. Sankur, "Robust facial landmarking for registration", *Annals of Telecommunications*, Vol. 62, No. 1-2, pp. 1608–1633, 2007.

50. Salah, A. A. and L. Akarun, "3D facial feature localization for registration", *International Workshop on Multimedia Content Representation, Classification and Security*, pp. 338–345, 2006.

51. Gauthier, I. and M. J. Tarr, "Becoming a "Greeble expert": Exploring the face recognition mechanism", *Vision Research*, Vol. 37, pp. 1673–1682, 1997.

52. Tong, M. H., C. A. Joyce and G. W. Cottrell, "Are Greebles special? Or, why the Fusiform Fish Area would be recruited for sword expertise (if we had one)", *Procedings of* 27^{th} *Annual Cognitive Science Conference*.

53. Furl, N., P. J. Phillips and A. J. O'Toole, "Face recognition algorithms and the other-race effect: computational mechanisms for a developmental contact hypothesis", *Cognitive Science*, Vol. 26, pp. 797–815, 2002.

54. Goldstone, R. L., "Do we all look alike to computers?", *Trends in Cognitive Sciences*, Vol. 7, No. 2, pp. 55–57, 2003.

55. Valentine, T., "A unified account of the effects of distinctiveness, inversion and race in face recognition", *Quarterly Journal of Experimental Psychology*, Vol. 43A, pp. 161–204, 1991.

56. Phillips, P. J., P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. J. Worek, "Overview of the face recognition grand challenge", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 947–954, 2005.

57. Gökberk, B., H. Dutağacı, L. Akarun and B. Sankur, "Representation Plurality and Decision Level Fusion for 3D Face Recognition", *IEEE Transactions on Systems, Man, and Cybernetics*, in press.

58. Savran, A., O. Çeliktutan, A. Akyol, J. Trojanova, H. Dibeklioğlu, S. Esenlik, N. Bozkurt, C. Demirkır, E. Akagündüz, K. Çalışkan, N. Alyüz, B. Sankur, I. Ulusoy, L. Akarun and T. M. Sezgin, "3D face recognition performance under adversorial conditions", *eNTERFACE'07 Workshop on Multimodal Interfaces*, Istanbul, Turkey, 2007.

59. Ekman, P. and W. V. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement", *Consulting Psychologists Press*, 1978.

60. Solomon, D., *Computer Graphics and Geometric Modeling*, Springer-Verlag, 1999.