THREE DIMENSIONAL FACE RECOGNITION UNDER OCCLUSION VARIANCE

by

Neşe Alyüz

B.S., Computer Engineering, Istanbul Technical University, 2005

M.S., Computer Engineering, Boğaziçi University, 2008

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

Graduate Program in Computer Engineering

Boğaziçi University

2013

THREE DIMENSIONAL FACE RECOGNITION
UNDER OCCLUSION VARIANCE

APPROVED BY:

Prof. Lale Akarun (Thesis Supervisor)	. . . . .
Assoc. Prof. Ali Taylan Cemgil	. . . . .
Assist. Prof. Albert Ali Salah	. . . . .
Prof. Bulent Sankur	. . . . .
Prof. Raymond Veldhuis	. . . . .

DATE OF APPROVAL: 11.10.2013

ACKNOWLEDGEMENTS

With gratitude to my thesis supervisor Lale Akarun, for her invaluable guidance, support, and patience throughout all these years; to my thesis core jury members Bülent Sankur and Ali Taylan Cemgil, for their prescient comments and feedbacks; to Raymond Veldhuis and Luuk Spreeuwers, for their guidance and the long hours they have granted me, when I was visiting SAS group.

With gratitude to my co-supervisor Berk Gökberk, for his guidance throughout this thesis, for his patience with my long list of questions, and for encouraging me to keep working on; to my very first mentor Albert Ali Salah, for believing in me and my potential; to Cem Ersoy, for even transforming the long hours of grading into moments of fun.

With gratitude to my friends and colleagues: Yunus Emre Kara, Furkan Kıraç, Heysem Kaya, Umut Konur, Alp Kındıroğlu, Salim Eryiğit, Ali Akkaya, Akın Günay, Pınar Santemiz, Hamdi Dibeklioğlu, Koray Balcı, Onur Dikmen, for their support, and especially to İsmail Arı, for his thoughtful comments and ideas about my work.

With gratitude to the special ones that brightened my days with all those coffee breaks: Şükrü Kuran, Pınar Sağlam, Oya Çeliktutan, Erinç Dikici.

With gratitude to my best friends, Gaye Soykök, Demet Nar, Pınar Karagülle, for their endurance and support; to Bülent Kaplan, for believing in me for our visionary projects.

With gratitude to my parents, Handan and Salim Alyüz, for their endless love and support, and for everything they have done for me; to all members of the Çivitci Family, for always making me feel at home with their warmth.

With my deepest gratitude to Fehmi Çivitci for not letting me get lost in the gloom, with his love, support, and even academic guidance.

ABSTRACT

THREE DIMENSIONAL FACE RECOGNITION UNDER OCCLUSION VARIANCE

With advances in sensor technology, three dimensional (3D) face has become an emerging biometric modality, preferred especially in high security applications. However, dealing with occlusions covering the facial surface is a great challenge. In this thesis, we propose a fully automatic 3D face recognition system, attacking three sequential problems: (i) Registration of occluded surfaces, (ii) detection of occluded regions, and (iii) classification of occlusion-removed faces. For the alignment problem, we propose an *adaptively selected model* based registration scheme, where a model is selected for an occluded face such that only the valid non-occluded patches are utilized in correspondence establishment. After registration, occlusions are detected, where we propose two different occlusion detection approaches. In the first detector, fitness to a pixelwise statistical model of the facial surface is used. In the second approach, in addition to the facial model, neighborhood information is incorporated. For occlusion handling, two different strategies are evaluated: (i) Removal of occlusions, and (ii) restoration of missing parts. In the classification stage, a masking strategy, which we call *masked projection*, is proposed to enable the use of subspace analysis techniques with incomplete data. Experimental results on two databases with realistic facial occlusions, namely, the Bosphorus and the UMB-DB, confirm that: (i) The proposed registration technique based on the adaptively selected model is a good alternative to obtain occlusion robustness; (ii) in occlusion detection, use of a statistical facial model is beneficial to make a pixelwise decision, which can further be improved by incorporating neighborhood relations to model coherency of surfaces; (iii) restoration provides only an approximation of the surface and is not suitable for classification purposes, (iv) masked projection serves as a viable approach to apply subspace techniques on incomplete data.

ÖZET

ÖRTME DURUMUNDA ÜÇ BOYUTLU YÜZ TANIMA

Sensor teknolojisindeki gelişmeler sayesinde, üç boyutlu (3B) yüz tanıma sıklıkla kullanılan bir biyometrik kip haline gelmiştir ve özellikle güvenlik uygulamalarında tercih edilmektedir. Ama yüz yüzeyini kapatan örtme durumları, çözülmesi gereken zor bir sorun olmaktadır. Bu tezde, üç farklı problemi ele alarak tamamen otomatik bir 3B yüz tanıma sistemi önermekteyiz: (i) Örtmeli yüzeylerin kayıtlanması, (ii) örtmeli bölgelerin belirlenmesi, ve (iii) örtmelerin çıkarıldığı boşluklu yüzeylerden öznitelik çıkarılması ve tanıma işleminin gerçekleştirilmesi. Kayıtlama için, adaptif olarak model seçimine dayalı bir yöntem önermekteyiz. Bu yöntemde, örtmesiz yüzeye uygun şekilde model seçilerek, nokta eşleştirmede yalnızca örtmesiz yüz parçalarının kullanılması sağlanmaktadır. Kayıtlama sonrası, örtmeli yüzeyleri bulmak için iki farklı örtme kestirim yöntemi önermekteyiz: İlk yöntemde, piksel bazlı istatistiksel yöntemler kullanılmakta ve herbir pikselin karşılık gelen modele uyumu test edilmektedir. İkinci yöntemde ise, komşuluk ilişkileri de kestirim aşamasına dahil edilmektedir. Örtme durumlarının üstesinden gelmek için iki farklı yaklaşım değerlendirilmektedir: (i) Örtmelerin yüzeyden çıkarılması, (ii) Eksik bölgelerin geri çatma ile doldurulması. Öznitelik çıkarımı ve sınıflandırma aşamasında ise, bir maskeleme stratejisi önermekteyiz. Maskeli projeksiyon adını verdiğimiz bu yöntem sayesinde, alt-uzay yöntemleri boşluklu veri ile kullanılabilir hale gelmektedir. Gerçekçi örtmeli kayıtlar içeren iki farklı 3B yüz veri kütüphanesi (Bosphorus ve UMB-DB) ile elde edilen deneysel sonuçlar sayesinde şu çıkarımlar yapılabilir: (i) Önerilen kayıtlama tekniği örtmeli durumlar için iyi bir alternatif olmaktadır; (ii) Örtme kestiriminde, istatistiksel yüz modelleme piksel bazlı karar vermeyi sağlarken, yüzey devamlılığını ifade eden komşuluk bilgisini dahil etmek sonuçları iyileştirmektedir; (iii) Geri çatma sadece yüzey yaklaştırımı sağladığı için tanımada yarar sağlamamaktadır; (iv) Maskeli projeksiyon alt-uzay tekniklerini boşluklu veriye uygulamaya olanak sağlamaktadır.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS . . . iii
ABSTRACT . . . iv
ÖZET . . . v
LIST OF FIGURES . . . x
LIST OF TABLES . . . xiii
LIST OF SYMBOLS . . . xv
LIST OF ACRONYMS/ABBREVIATIONS . . . xvii
1. INTRODUCTION . . . 1
1.1. Research Overview and Contributions . . . 2
1.2. Outline of the Thesis . . . 5
2. LITERATURE OVERVIEW . . . 7
2.1. Face Recognition Stages and The Occlusion Challenge . . . 8
2.2. 3D Face Databases used to Evaluate Occlusion Robustness . . . 10
2.2.1. UMB-DB Database . . . 10
2.2.2. Bosphorus Database . . . 12
2.2.3. FRGC v.2 Database . . . 12
2.3. Occlusion Handling in the 2D Face Recognition Literature . . . 14
2.4. Occlusion Handling in the 3D Face Recognition Literature . . . 16
2.4.1. Handling of Partial Hair Occlusions: Evaluations on the FRGC v.2 . . . 17
2.4.2. Handling of Complex Occlusions: Evaluations on the Bosphorus and UMB-DB . . . 18
2.4.3. Handling of Incomplete Data at the Classification Stage . . . 20
2.5. Part-based Systems in the 3D Face Recognition Literature . . . 21
3. TECHNICAL BACKGROUND . . . 29
3.1. Curvature Information . . . 29
3.1.1. Principal, Mean, and Gaussian Curvatures . . . 29
3.1.2. Shape Index and Curvedness Maps . . . 30
3.2. Basic Registration Techniques . . . 31
3.2.1. Procrustes Analysis . . . 32

3.2.2. Iterative Closest Point Algorithm 36
3.2.3. Model-based Registration 40
3.3. Subspace Analysis Techniques 43
3.3.1. Principal Component Analysis and the Eigenfaces Method 43
3.3.2. Gappy Principal Component Analysis 46
3.3.3. Linear Discriminant Analysis and the Fisherfaces Method 47
4. MOTIVATIONAL WORK: PART-BASED 3D FACE RECOGNITION 49
4.1. Part-based Face Recognition System 50
4.1.1. Automatic Landmark Localization 51
4.1.2. 3D Face Registration 54
4.1.3. 3D Features 56
4.1.4. Classification: Fusion Techniques 57
4.2. Experimental Results 58
4.2.1. Automatic Landmark Localization Performance 59
4.2.2. Identification Results 60
4.2.3. Fusion of Regional Classifiers 63
4.2.4. Results of Statistical Features 64
4.3. Conclusion 66
5. SURFACE REGISTRATION UNDER OCCLUSION 68
5.1. Model-based Registration 69
5.1.1. Alignment to the Average Face Model 69
5.1.2. Alignment to the Average Nose Model 70
5.2. Proposed System: Adaptive Model-based Registration 70
5.2.1. Nose Detection 71
5.2.2. Patch Selection and Adaptive Registration 73
5.3. Experimental Results 75
5.3.1. Nose Detection Accuracy 76
5.3.2. Patch Validation and Selection Accuracy 77
5.3.3. Registration Accuracy 78
5.3.4. Evaluation of the Initial Alignment Accuracy 81
5.4. Conclusion 83

6. OCCLUSION HANDLING 85
6.1. Occlusion Detection 85
6.1.1. Baseline Occlusion Detector: Difference from the Average Face Model 86
6.1.2. Statistical Facial Modeling via Pixelwise GMMs 87
6.1.3. Occlusion Segmentation via Graph Cut 88
6.2. Restoration of Occlusion-Removed Surfaces 95
6.3. Experimental Results 96
6.3.1. Databases 96
6.3.2. Occlusion Detection Accuracy 97
6.3.3. Classification Accuracy with Occlusion Removal 101
6.3.4. Removal versus Restoration 102
6.4. Conclusion 103
7. FACE RECOGNITION UNDER OCCLUSION 105
7.1. Global Classification using Masked Projection 106
7.2. Regional Classification using Masked Projection 109
7.3. Experimental Results 110
7.3.1. Evaluation of the Global Classification Performance 111
7.3.2. Evaluation of the Regional Classification Performance 113
7.3.3. Comparison of Masked Projection and Masked Training Performances 114
7.3.4. Effect of Occlusion Percentage on Performance 117
7.3.5. Comparison of Masked Projection with Different Occlusion Detectors 118
7.3.6. Different Fusion Schemes 122
7.3.7. Time Complexity 122
7.3.8. Masked Projection for Other Acquisition Scenarios 124
7.4. Conclusion 126
8. CONCLUSION 127
8.1. Contributions and Discussion 127
8.2. Future Directions 130
APPENDIX A: FACTORIAL DESIGN EXPERIMENTS 133
A.1. Analysis of Variance Table for Factorial Design 134
A.2. Response Surface Fitting 134

A.3. Experimental Results 137
A.3.1. Database 137
A.3.2. Factorial Design Experiments and Occlusion Detector Evaluation . 137
REFERENCES 143

LIST OF FIGURES

Figure 1.1. General scheme of the proposed occlusion robust 3D face recognizer. . 3

Figure 2.1. Two types of occlusion that can appear on the facial surface. 8

Figure 2.2. Overall diagram of a traditional face recognizer. 9

Figure 2.3. Samples and occlusion percentage histogram for the UMB-DB. 11

Figure 2.4. Examples to four occlusion types of the Bosphorus database. 13

Figure 3.1. Shape index represents a transition between concave and convex shapes. 31

Figure 3.2. Curvedness defines the rate of curvature of a surface. 31

Figure 3.3. A simple example for Procrustes alignment. 33

Figure 3.4. A visual example for ICP alignment. 37

Figure 3.5. Average face model together with nine landmarks. 41

Figure 3.6. Registration based on average face model. 42

Figure 4.1. Illustrative diagram of the proposed 3D face recognition approach. . . 51

Figure 4.2. Illustration of the automatic landmarking algorithm. 53

Figure 4.3. The AvFM and its landmarks, and facial regions. 55

Figure 4.4. Sample 3D scans for the Bosphorus and FRGC v.2 databases. 59

Figure 4.5. Manual and automatic landmarks shown on a sample set. 61

Figure 5.1. Diagram of the proposed registration method. 71

Figure 5.2. Curvature maps utilized for nose detection are illustrated on an example. 73

Figure 5.3. A diagram summarizing the patch validation and model selection. . . 74

Figure 5.4. Facial patches and adaptive models utilized in registration are given. . 75

Figure 5.5. Correct and incorrect nose detection examples for the UMB-DB. . . . 78

Figure 5.6. An occluded face registered with different models. 78

Figure 5.7. Estimation examples of landmarks with missing points. 82

Figure 6.1. The graph construction and segmentation method for occlusion detection. 90

Figure 6.2. Examples to occlusion masks obtained with different methods. 100

Figure 6.3. An example to restoration obtained with Gappy PCA is given. 103

Figure 7.1. The regional division scheme: (a) patches, (b) regions. 110

Figure 7.2. Illustrative diagram of the proposed 3D face recognition approach. . . 111

Figure 7.3. CMC plots for the Bosphorus and UMB-DB databases. 115

Figure 7.4. Regional recognition rates for the Bosphorus and the UMB-DB databases. 116

Figure 7.5. Regional recognition rates for masked training and masked projection. 117

Figure 7.6. Occluded area histogram for the Bosphorus and UMB-DB databases. . 119

Figure 7.7. Highly occluded samples correctly classified by the proposed method. 120

Figure A.1. Normal probability plot of residuals. 139

Figure A.2. The residual plots. 140

Figure A.3. The normal probability plot of the residuals. 141

Figure A.4. The response surfaces plotted together with observed responses. . . . 141

LIST OF TABLES

Table 2.1. 3D face databases that contain occlusions. 14

Table 2.2. Summary of 3D face recognition methods considering occlusions. . . . 27

Table 2.3. Rank-1 classification rates reported on FRGC v.2. 28

Table 4.1. Average Euclidean distances between manual and automatic landmarks. 60

Table 4.2. Identification results of the AvFM-based approach. 62

Table 4.3. Identification results of the individual regions. 63

Table 4.4. Fusion results of regional classifiers using point cloud features. 64

Table 4.5. Rank-1 classification results of the Fisherfaces based AvRM approach. 65

Table 5.1. Nose detection performances on the Bosphorus and UMB-DB databases. 77

Table 5.2. Identification performances on Bosphorus and UMB-DB databases. . . 80

Table 5.3. Partial Gappy PCA-based landmark estimation performance. 83

Table 5.4. Identification performances with manual and automatic initialization. . 83

Table 6.1. Precision, recall, and F_1 measures on the Bosphorus-70 subset. 99

Table 6.2. Depth-based classification results with different occlusion masks. . . . 102

Table 6.3. Depth-based classification results on occlusion-removed, restored data. 103

Table 7.1. Global identification accuracies with the standard and masked Fisherfaces. 112

Table 7.2. Regional identification accuracies with standard and proposed methods. 113

Table 7.3. Regional identification accuracies with the proposed method. 121

Table 7.4. Regional identification accuracies with different fusion schemes. . . . 123

Table 7.5. Identification accuracies of masked projection on different challenges. 126

Table A.1. The ANOVA table for the three-factor fixed effects model. 134

Table A.2. The factors and sets of levels considered in the factorial experiments. . 137

Table A.3. The ANOVA table for the three-factor fixed effects model. 138

LIST OF SYMBOLS

$C(i)$	Curvedness value at surface point i
$\mathbf{D}$	Absolute difference map
$\mathcal{E}$	Set of edges
F_β	F-measure
$\mathcal{G}$	Graph representation
$H(i)$	Mean curvature at surface point i
$K(i)$	Gaussian curvature at surface point i
$\mathbf{L}$	Geometrical shape matrix
$\mathcal{L}$	Binary vector of assigned labels
$\mathbf{L}_C$	Consensus shape matrix
$\mathbf{m}$	Occlusion mask vector
$\mathbf{P}$	Point set matrix
$\mathbf{p}_i$	Point vector
s	Source node
$\mathbf{S}_B$	Between-class scatter matrix
$\mathbf{S}_W$	Within-class scatter matrix
$SI(i)$	Shape index value at point i
$SI_{cx}(i)$	Shape index value at point i after convexity thresholding
SS	Sum of squares error term
t	Sink node
$\mathcal{V}$	Set of vertices
$\mathbf{W}$	Projection matrix
$\mathbf{w}_1$	First principal component vector
w_e	Weight of edge e
$\mathbf{W}_m$	Masked projection matrix
$\mathbf{W}_\perp$	Orthogonalized masked projection matrix
$w_p^{(s)}$	Weight of edge to the source terminal
$w_p^{(t)}$	Weight of edge to the sink terminal

$w_{(p,q)}^{(n)}$	Weight of edge between neighboring nodes p and q
$WSI(i)$	Curvedness-weighted convex shape index value at point i
$\mathbf{x}$	Observation vector
$\hat{\mathbf{x}}$	Incomplete observation vector
$\tilde{\mathbf{x}}$	Completed version of observation vector
$\mathbf{X}$	Observation matrix
$\mathbf{x}^{(P)}$	Probe face image vector
$\mathbf{x}^{G_k}$	Face image vector for k^{th} gallery subject
$\tilde{\mathbf{y}}$	Coefficient vector corresponding to completed image
$\mathbf{z}$	Projected vector at the subspace
α	Scaling constant
$\boldsymbol{\alpha}$	Eigenvalue vector
$\boldsymbol{\beta}$	Eigenvector coefficients of the completed input vector
$\boldsymbol{\gamma}$	Translation vector
$\boldsymbol{\Gamma}$	Rotation matrix
$\kappa_{min}(i)$	Minimum principal curvature at surface point i
$\kappa_{max}(i)$	Maximum principal curvature at surface point i
λ	Eigenvalue
$\boldsymbol{\Lambda}_m$	Diagonal matrix of occlusion mask
$\boldsymbol{\mu}$	mean vector
π_k	Mixture component coefficient
$\boldsymbol{\sigma}$	standard deviation vector
$\boldsymbol{\Sigma}$	Covariance matrix
$\boldsymbol{\tau}^{(H)}$	Upper threshold
$\boldsymbol{\tau}^{(L)}$	Lower threshold

LIST OF ACRONYMS/ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
AFM	Annotated Face Model
ANOVA	Analysis Of Variance
AvFM	Average Face Model
AvRM	Average Region Model
AU	Action Unit
BL	Baseline
CMC	Cumulative Match Characteristic
CV	Committee Voting
FACS	Facial Action Coding System
FF	Fisherfaces
FN	False Negative
FP	False Positive
FRGC	Face Recognition Grand Challenge
GC	Graph Cut
GMM	Gaussian Mixture Model
ICA	Independent Component Analysis
ICP	Iterative Closest Point
k-NN	k Nearest Neighbor
LBP	Local Binary Patterns
LDA	Linear Discriminant Analysis
LGT	Log-Gabor Template
LOO	Leave-One-Out
MOD-CV	Modified Committee Voting
MOD-PROD	Modified Product Rule
MRF	Markov Random Fields
MS	Mean Square

PCA	Principal Component Analysis
PROD	Product Rule
PSD	Point Set Distance
ROI	Region Of Interest
SDM	Spherical Depth Map
SIFT	Scale-Invariant Feature Transform
TP	True Positive
TPS	Thin Plate Spline
UMB-DB	University of Milano Bicocca 3D Face Database
v.1	version one
v.2	version two

1. INTRODUCTION

In identity management systems, the task of determining the correct identity of a person is critical. Identity representation systems utilizing a password associated with 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 sysftem studies have moved towards the use of biometrics.

The term *biometrics* refers to automated systems where physiological or behavioral characteristics of an individual are used for identification purposes. Face, fingerprint, iris, retinal image, vein, or voice can be listed among the physiological features used in biometric systems. Among others, face is the most familiar-to-human modality, since our cognitive system often utilizes facial data to recognize people. Moreover, face modality is highly preferred for automated systems, since the biometric data can be acquired in a contactless manner and it can be employed for non-cooperative scenarios. Due to these advantages, face recognition has a wide application domain, including surveillance, access control and human-computer interaction practices. Hence, it has been a popular research topic for the last three decades. Further research in the last decade has shown that, face recognition in constrained acquisition scenarios can reach the performance levels of high security modalities such as fingerprint and iris [1].

Initially, face recognition studies focused on identifying people from their two dimensional (2D) facial images [2]. However, when non-cooperative and uncontrolled scenarios are considered, recognizing individuals from their 2D face scans remains as a challenging task. The main challenges, including illumination differences, pose variations, and presence of facial expressions; triggered the shift of face representation from 2D modality to 3D: In the 3D domain, illumination differences, pose and expression variations can be better handled since the true geometric information residing in the 3D data is utilized. This shift was supported by the emerging sensor technology allowing acquisition of the 3D facial geometry. With the advances in sensing technology, large evaluation 3D face datasets became publicly available: In 2006, the Face Recognition Grand Challenge (FRGC) [3] was presented as the

first large evaluation set.

In the three dimensional domain, challenges caused by illumination, pose, and expression variations can be better handled. However, extreme occlusion variations still complicate the task of identification. Handling of occlusions for face recognition is extremely important, when non-cooperative security applications are considered. In this thesis, we propose a complete 3D face recognition system, that is robust under occlusions.

In biometrics, recognition is a general term encapsulating identification and verification scenarios. However, in most of the papers in the literature, recognition often refers to identification, whereas the term authentication is used for verification. In identification, the aim is to find the identity of a person from a gallery set. The gallery set is a previously acquired database, where the biometric data for subjects to be checked are stored. The identification scenario can either be closed-set or open-set. In closed-set identification, it is assumed that all users are included in the gallery and the probe is identified as one of the gallery subjects. In the open-set scenario, however, the probe can be an unknown subject and the identification process should be able to indicate that the probe is not among the gallery set. In contrast to identification, in verification, the probe both provides the biometric data and claims an identity: The verifier checks if the claim is valid. In this thesis, we focused on the identification problem, where the 3D facial surface information is used as the biometric data. In our *closed-set* identification scenario, the identity of the probe is sought among the subjects in the gallery, where the 3D face of the probe is compared with each of the gallery faces to find the closest match. Throughout this thesis, the term *recognition* and *identification* are used interchangeably to denote a closed-set identification scenario.

1.1. Research Overview and Contributions

Presence of occlusions is a new challenge being considered in face recognition scenarios, and there are only a few studies in the 3D face recognition considering occlusion handling. However, it is a vital topic especially for high security applications: In current security systems, the identification process can easily be misled by partially covering the facial surface naturally by hair or hand, or by using exterior objects, such as eyeglasses, hat, and

scarf. In this thesis, our aim is to construct a fully automatic 3D face recognizer, which is robust to realistic occlusion variations.

To be able to compare two faces and find the closest match between the probe and the gallery in the presence of occlusions, first, these surfaces should be brought into a common coordinate frame, and then, the occluded parts should be accurately located to be disregarded in the classification process. Therefore, in this thesis, we treat the occlusion handling problem as a combination of three separate problems: (i) Alignment of occluded surfaces, (ii) detection and handling of occluded parts in the registered faces, and (iii) classification of faces free of occlusions. In Figure 1.1, a general scheme of the developed system is given. As this figure illustrates: (i) The registration process encapsulates face localization (nose detection) and surface alignment; (ii) occlusion handling can be obtained by either simple occlusion removal, or restoration can be applied on the occlusion-removed surface to obtain a completed version of the face; and (iii) the occlusion free faces can be used for classification, where the occlusion mask should be incorporated if incomplete faces after removal are to be used.

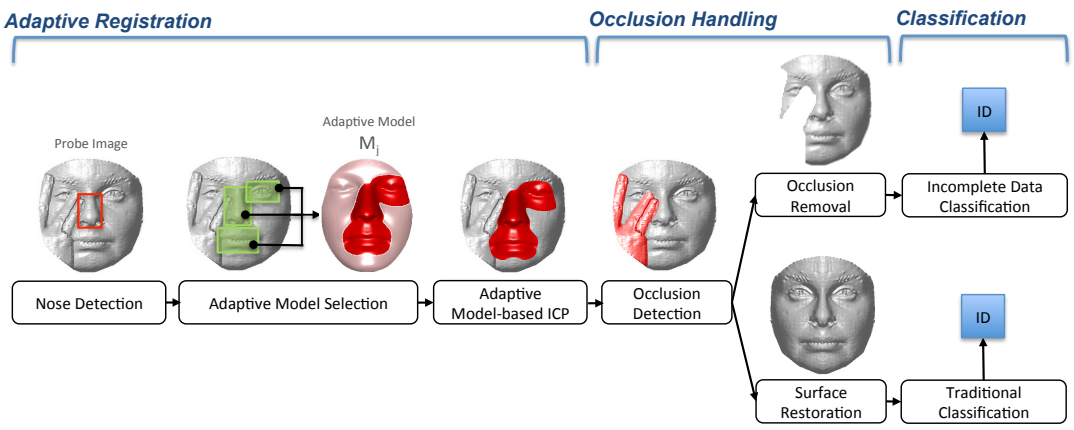


Figure 1.1. General scheme of the proposed occlusion robust 3D face recognizer.

The contributions of this thesis can be listed as follows, sorted according to the process sequence of the implemented system:

- *Face Localization*: For alignment initialization, it is necessary to locate the facial sur-

faces. In the face literature, a number of landmark points are often detected for localization and initialization purposes. However, the process of localizing distinct fiducial points gets complicated in the presence of occlusions. To overcome this problem, instead of fiducial points, we propose to detect the *nose region* for initialization purposes. As presented in [4], the nose area is detected by employing curvature information (shape index and curvedness) together with template matching. Even when the nasal area is partially visible, the initialization obtained is sufficient for the fine alignment to converge.

- *Registration*: Motivated by the registration approach based on aligning to an average model [5], we proposed an adaptive model based registration technique [6]. Based on the idea of nose detection of the previous sub-process, we propose to detect other important regions, such as the eyes and the mouth area. The detected regions are checked for validity as non-occluded parts. Using the validity flags, the registration model is selected adaptively. Hence, disregarding occluded parts in the registration process is achieved without any occlusion detection prior to registration.
- *Occlusion Detection*: After the facial surfaces are aligned, the occluded regions should be accurately located to exclude them from the comparison process. In this thesis, we propose two different occlusion detectors. The first detector uses a statistical approach to model the facial surface, where *Gaussian Mixture Models* (GMMs) are utilized to express the pixelwise structure of a face. The idea of pixelwise GMMs was proposed in [7] for background-foreground separation in video sequences, where the background is modeled. Here, we used their segmentation approach for occlusion detection, where the facial surface is modeled as background. In the second detector, we proposed to incorporate the neighboring pixel-pair relations into a simpler mean-variance modeling of the facial surface to improve the detection at the boundaries. These regional and boundary cues are used to construct a graph representation of the face, where *graph cut techniques* are employed to solve the binary image segmentation problem [8]. These two occlusion detectors are compared in [9].
- *Occlusion Handling*: After the occluded regions are accurately located, they should be handled prior to classification. The occluded regions can be removed to obtain occlusion free surfaces. However, due to missing components, traditional 3D face classifi-

cation methods are not directly applicable. To handle missing components, *restoration* can be applied. We inspected a restoration strategy based on Gappy Principal Component Analysis (Gappy PCA) [10], where the facial surface is approximated using information residing in the non-occluded parts [11]. We compared two occlusion handling alternatives, namely removal vs. restoration, as presented in [4].

- *Classification*: The comparative results obtained in the previous sub-process of occlusion handling, showed that restoration is not capable of reconstructing any discriminative information necessary for classification. Therefore, it was necessary to modify classifiers to work with incomplete data. We introduced a technique called *masked projection* [12], that enables to incorporate occlusion masks into subspace techniques (such as Eigenfaces [13] or Fisherfaces [14]). By masked projection, it is possible to project the non-occluded facial information to the subspace that is specific to the occlusion mask considered. Furthermore, the specific subspaces are obtained using masks, without the need for any extra training.
- *Evaluation*: Individual stages of the proposed system are evaluated on two different 3D face databases including occlusion variations. The databases, namely the Bosphorus [15] and UMB-DB [16] databases, are currently the largest publicly available databases with occlusions.

1.2. Outline of the Thesis

The thesis is organized as follows: In Chapter 2, a literature survey is given: First, the basic processes to be considered for a face recognizer are given. Then, 3D face databases including realistic occlusions are summarized, followed by the overview of the studies both in the 2D and in the 3D face recognition literature considering occlusion variations. In Chapter 3, the technical background on techniques used in this thesis are included: First, curvature information used in landmark or facial region localization are given. Then, alignment methods used to build up the proposed registration technique are summarized, which is followed by the subspace analysis techniques to be utilized in the development of the proposed classifier. In Chapter 4, a motivational work is explained in detail, where a part-based registration and recognition scheme is proposed for expression handling. Motivated from the results obtained

in this work, we have decided to modify the regional model-based registration method and the regional subspace classification technique for adaptation to the occlusion problem.

In Chapter 5, the registration method proposed to obtain occlusion robustness is explained. Experiments to analyze the performance of registration are given. In Chapter 6, the occlusion detection techniques proposed to label the occluded pixels are summarized. Detection performances are compared through experiments. In addition, simple removal of occlusions and restoration of missing components are constrasted for the purpose of classification. In Chapter 7, the proposed classification technique that enables the applicability of subspace technique to incomplete data is summarized. Experimental results to show applicability to missing data is given. Finally, in Chapter 8, conclusions are summarized and possible future directions are pointed out.

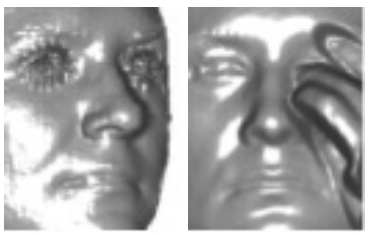
2. LITERATURE OVERVIEW

Interest in 3D face recognition systems caused an enormous growth in research studies focusing on the 3D modality. A thorough survey of previously proposed 3D face recognizers can be found in [17–19] and details of some fundamental concepts can be overviewed in [20–22]. Besides the problem of expression handling, which has been extensively studied in recent years [23–28], occlusion variations remains as a challenging task: Although occlusions appear as a practical problem for realistic scenarios, they are not investigated well in the literature.

Due to the sensor technology, the acquired 3D scans can include small holes. These holes can either be caused by self occlusions appearing as a result of a single viewpoint, such as the nose borders, or by reflectance properties of surface patches, such as surfaces with facial hair. However, usually these holes are quite small and can be filled by interpolation as a sufficient preprocessing step. In this thesis, we focus on the handling of larger holes, that cannot be handled by simple hole filling procedures.

There are two types of occlusions: The first is caused by self-occlusions during acquisition, where a part of the facial surface hinders acquisition of another region shadowed with respect to the sensor. These occlusions appear as missing data in the facial surface. The other type of occlusions can be caused by external objects such as hand, hair, scarf, eyeglasses and other objects. The second class of occlusions is more complex to handle, since the occluding objects alter the 3D facial geometry. Visual examples of two types of occlusions are given in Figure 2.1. In this thesis, we mainly focus on the second class of occlusions, where exterior objects partially cover the facial surface: Hereafter, the term “occlusion” will refer to occlusions caused by exterior objects. In this chapter, we first briefly mention basic processes of a face recognizer and how they can be affected by the occlusion presence. Then, we give an overview of the databases used to evaluate the occlusion challenge. Next, we summarize the studies that consider occlusion handling both in the 2D and the 3D face recognition literature. Moreover, we briefly mention the classification approaches used for the first type of occlusions, where self-occlusions cause incomplete surfaces, since these classification ap-

proaches can be useful after the occlusions are detected and removed. In addition, the studies considering facial surfaces as a combination of parts are reviewed, since we are motivated from part-based systems used for registration or recognition. These part-based face recognition systems focus on expression variations, hence their performances on an expression subset are reported for comparative purposes.



(a) (b)

Figure 2.1. Two types of occlusions can appear on the facial surface: In (a) an example of self-occlusions caused by pose variations, and in (b) an example of occlusions caused by exterior objects, are given.

2.1. Face Recognition Stages and The Occlusion Challenge

In the presence of occlusions over the facial surface, alteration of the geometry complicates the identification process, affecting different stages of face recognition systems. The main steps of a face recognizer can be listed as: face detection, landmark localization, coarse and fine registration, feature extraction, and classification. An overall diagram for a traditional face recognizer is given in Figure 2.2. Face detection is the process of localizing the facial surface and determining its extent. After the facial surface is detected, it is often necessary to detect some fiducial points, referred to as landmarks, such as nose tip, eye or mouth corners. In contrast, some landmark points can be detected beforehand, serving as a guide for the face detector. The landmark points can be beneficial in the initial alignment of surfaces. The process of registration is the process of aligning two surfaces, so that they can be compared for classification purposes. Registration can be divided into two stages, where an initial registration can be used to coarsely align two surfaces, and it can then be followed by a fine alignment to obtain a dense correspondence. After the facial surfaces are registered, facial features can be extracted to represent the discriminative information inher-

ent in the surfaces. In some studies, the registration step is discarded and feature extraction is obtained using transformation independent descriptors. Some of these methods include systems based on keypoints. Keypoints are points with some specific geometrical properties, but are not necessarily at meaningful locations as landmarks. Keypoints are then used to extract transformation invariant features from a region of interest around these locations and the extracted information can be directly utilized to recognize faces. The extracted features, either directly obtained without any alignment or right after the registration process, are incorporated into the classification approach to reveal the most probable identity of the scan in question.

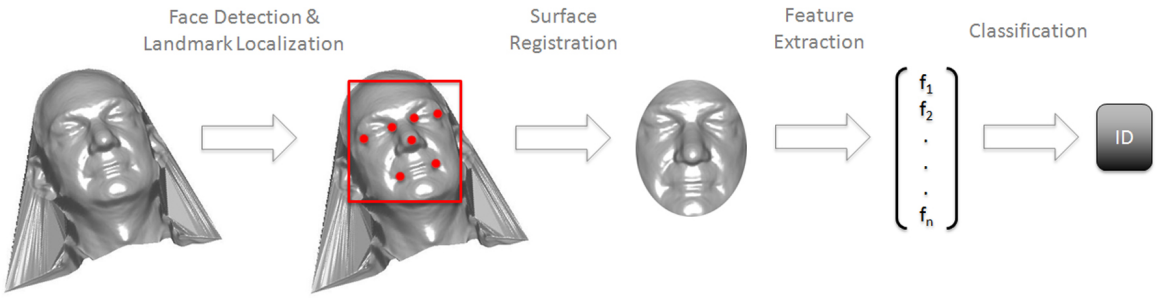


Figure 2.2. Overall diagram of a traditional face recognizer.

When the facial surface is occluded, all of these stages will be affected to some extent. Therefore, when a standard face recognizer is employed on occluded faces, probable errors occurring at each step will accumulate and result in an enormous degradation in the identification performance. A robust face recognizer should handle the occlusion problem at different stages: (i) the registration stage, which can cover face detection, landmark or keypoint localization, determination of region of interest (ROI), and coarse or fine alignment of surfaces; (ii) detection and handling of occlusions, where removal or restoration can be employed; (iii) classification, where either restored or partial faces are used. In this thesis, we focused on the problem of occlusion handling, considering these three stages: registration, occlusion handling, and classification. In this chapter, we summarize solutions in the 3D face recognition literature, proposed to handle occlusion variations.

2.2. 3D Face Databases used to Evaluate Occlusion Robustness

Before giving the literature review of face recognition systems considering occluded facial surfaces, we briefly introduce the publicly available 3D face databases including occlusion variations. The databases included here are the mostly referred databases in the literature to evaluate occlusion presence. Sorted according to the level of challenge, the mostly referred databases can be listed as follows: (i) University of Milano Bicocca 3D Face Database (UMB-DB); (ii) Bosphorus Database; (iii) Face Recognition Grand Challenge Version 2 (FRGC v.2). Although the FRGC v.2 database includes only small occlusions (caused by hair over the forehead region or caused by facial hair such as mustache or beard), some systems report results on it. Furthermore, in the experiments carried out in this thesis, this database is employed for training purposes. Therefore, it is included in the database overview.

2.2.1. UMB-DB Database

The UMB-DB database [16] is collected to evaluate 3D face recognition systems, mainly focusing on the occlusion scenario. As the acquisition device, Minolta Vivid 900 series sensor is used, which is a laser scanner. The database is acquired from a total of 142 subjects, and there are a total of 1473 scans. The non-occluded scans (a total of 883 scans) include neutral and expressive scans. The other 590 scans include occlusions caused by scarves, hats, hands, eyeglasses, and other realistic exterior objects. In the literature, recognition results are reported using the available experimental protocol of [16]: The gallery set contains the first neutral scan of each subject, and the probe set consists of the occlusion subset. The gallery and probe sets contain 142 neutral and 590 occluded scans, respectively. In the classification experiments carried out in this thesis, this experimental protocol is followed to allow a fair comparison. The occlusions in this database are highly challenging, where the location and amount of occlusion vary greatly. Some occlusion examples from the UMB-DB database are given in Figure 2.3, illustrating how challenging the occlusions can be. Furthermore, occlusion percentage histogram is given, clearly showing that this set includes challenging occlusions.

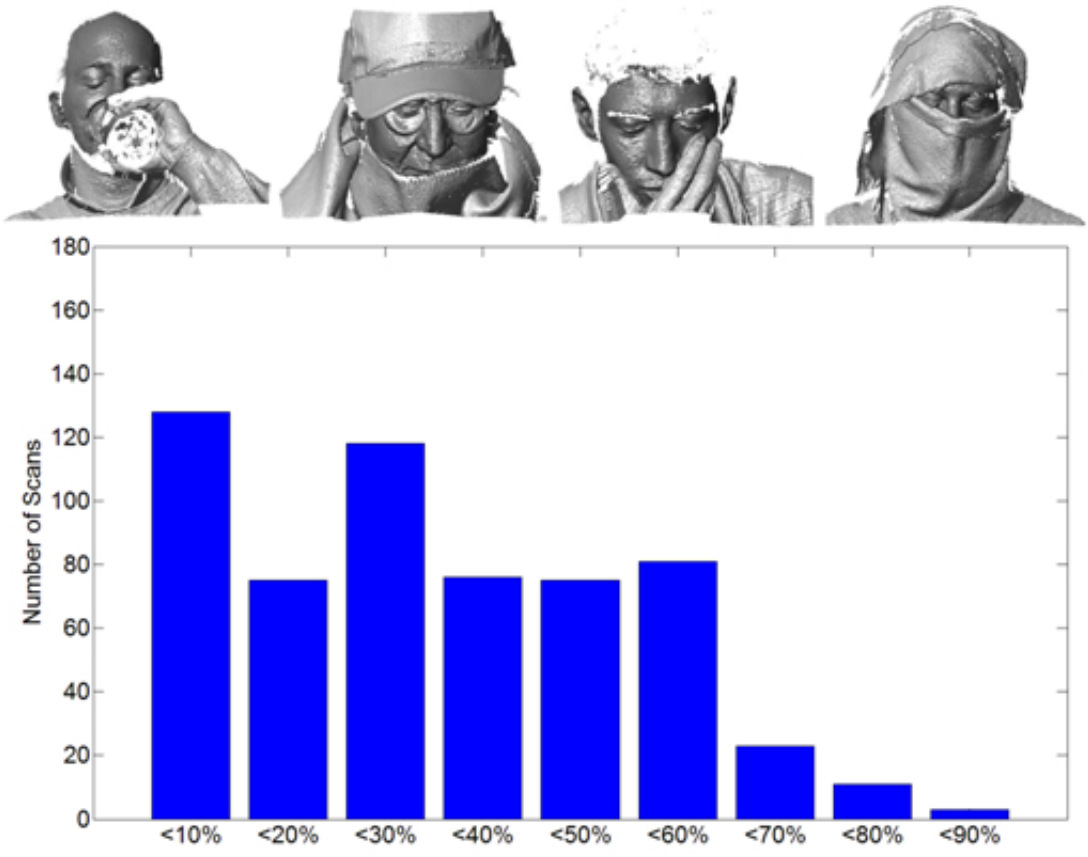


Figure 2.3. Sample faces and occlusion percentage histogram given for the UMB-DB database illustrate how challenging the database is.

2.2.2. Bosphorus Database

The Bosphorus database [15] is acquired using Inspeck Mega Capturor II 3D, which is a digitizer device based on the structured light technology and it has a resolution of about $0.3mm$ in each of three dimensions. The database is collected to enable evaluation of three main challenging scenarios of a realistic 3D face recognizer: The database includes scans of (i) pose variations, including both realistic and extreme poses; (ii) expression variations, including an extensive set of action units in addition to a set of universally accepted expressions; (iii) typical oclusions, that are probable to occur in real life. For a total of 105 subjects, there are 4666 scans. The total number of neutral and occluded scans are 299 and 381, respectively. For the Bosphorus database, a similar experimental protocol for classification to that of the UMB-DB is used: First neutral scan of each subject is used to construct the gallery set, whereas the occluded scans are included in the probe set. There are four different types of occlusions as shown in Figure 2.4, top row: (i) Occlusion of the eye area by eyeglasses; (ii) occlusion of the eye area by a hand, (iii) occlusion of the mouth area by a hand, (iv) occlusion caused by hair. In Figure 2.4, bottow row, the occlusion percentage histogram is included. It is clear that this is a less challenging database, when compared to UMB-DB, as most of the occlusions cover 30% or less of the facial surface.

2.2.3. FRGC v.2 Database

The FRGC v.2 [3] is a database widely used in the literature of 3D face recognition, since it contains a large number of scans collected from a large number of subjects. It is acquired using the same sensor as the UMB-DB, the Minolta Vivid 900 series laser scanner. In total, there are 4007 frontal images of 466 subjects. The neutral subset of 2365 images contains non-occluded images with neutral expression. The remaining faces include expression variations such as happiness, sadness, surprise, anger, disgust, and cheek puffing. Although no occlusion-specific acquisition scenario is considered for occlusion variations, there are several scans which can be considered to include occlusions: Some scans have hair occlusion in small portions of the forehead region, and some others include facial hair. In [29], it is stated that more than 40% of the images include hair occlusions. For most of the studies reporting results on FRGC v.2, the experimental protocol given in [3] is used: The gallery set

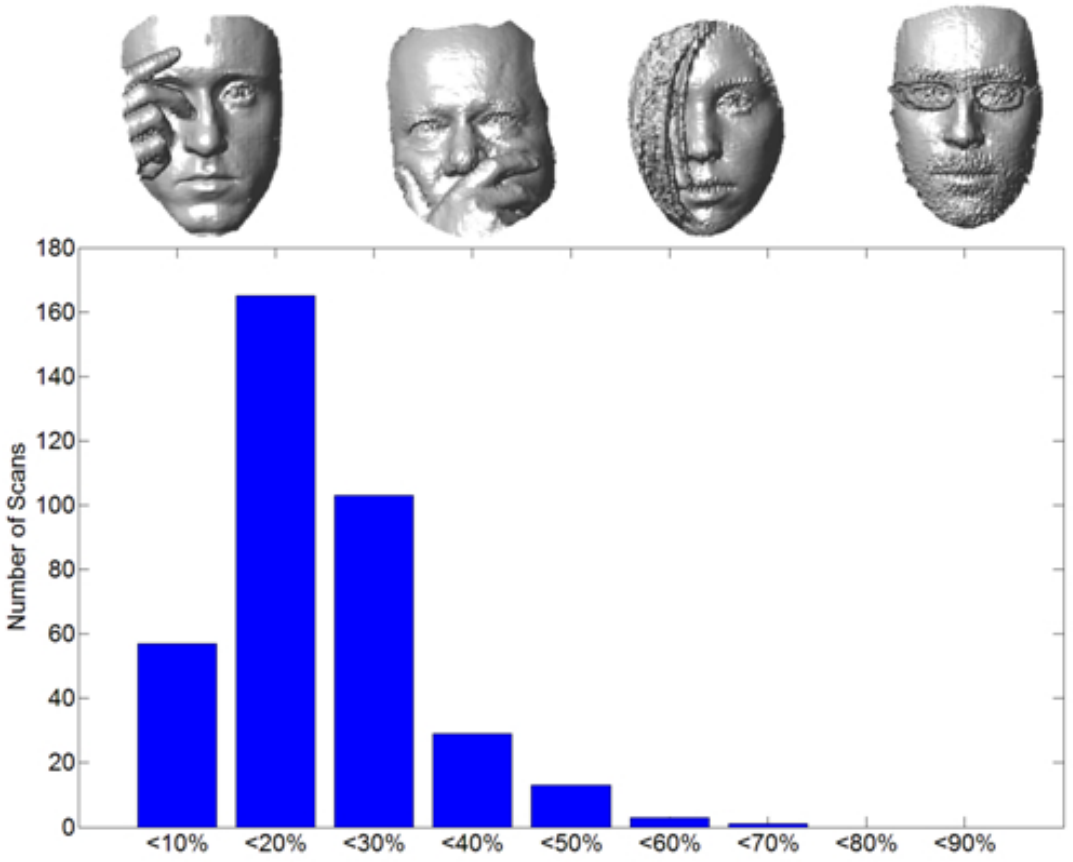


Figure 2.4. Examples to four occlusion types in the Bosphorus database are given: occlusion of the eye area by eyeglasses or by hand, occlusion of the mouth area by hand, and occlusion caused by hair. Additionally, the occlusion percentage histogram is included.

contains the first scan of each subject (a total of 466 scans), and the probe set contains all the remaining 3541 images. Details about three of the databases are summarized in Table 2.1.

Table 2.1. 3D face databases that contain occlusions.

Database	Subjects	Occluded images	Occlusion types	Occlusion difficulty	Pose/Expression variations in occluded images	Presence of non-facial parts (shoulder, torso)	Average 3D points (face only)
UMB-DB	142	590	Hand, hair, scarf, objects	Extreme	Both	Yes	35-40K
Bosphorus	105	381	Hand, hair, eyeglasses	Moderate	Slight pose variations	No	35K
FRGC v.2	466	1400	Hair in forehead region	Low	Only expressions	Yes	35-40K

It should be noted here, that in each of these databases, the 3D facial surfaces include some small holes due to self-occlusions and some outlier points caused by reflectance properties. Therefore, the surfaces are preprocessed to remove spikes using median filtering, and hole filling of small gaps is obtained by interpolation.

2.3. Occlusion Handling in the 2D Face Recognition Literature

Although the aim of this work is to handle occlusions in the 3D modality, in this section, we have reviewed the studies in the 2D face recognition literature, which consider the occlusion variations. In the 2D face recognition studies, there has been a few approaches considering occlusion variations. In most of these studies, the aim is occlusion handling for recognition and the registration problem is not considered: Experimental results are usually reported on databases where the faces are assumed to be accurately registered prior to recognition.

Some studies are based on subspace analysis methods, where the aim is either occlusion robust projection or missing data compensation. In [30], Park *et al.* consider occlusions caused only by eyeglasses and propose a method to compensate for the missing data. Initially, the glasses region is extracted using color and edge information. The offline-generated

Eigenfaces from a set of non-occluded images are then used together with the extracted glasses region for missing data compensation. In [31], occlusion variations are handled by eliminating facial parts where occlusions frequently occur. Several subsets of images are created through masking facial regions both in training and test faces. Using masked training images, different face projection spaces are created through PCA and majority voting is applied to fuse multiple classifiers. In [32], an approach for combining discriminative and reconstructive methods is proposed for better handling of images with outlier pixels. The general discriminative model is rewritten by incorporating the feature vectors corresponding to the reconstructive model. In addition, the truncated projection matrix is extended to retain the complete discrimination power.

Other holistic approaches can be considered as model-based methods. In [33], De Smet *et al.* proposed an iterative approach for the parameter estimation of 3D morphable model fitting procedure. Concurrently, a visibility map defining the occlusions is modeled by Markov Random Fields (MRF), which accounts for spatial coherence of occlusions. The visibility map is used to exclude occluded regions from further computations. Similar to the morphable model formulation, Park *et al.* [34] proposed to encode all the geometric quantities and the structural information residing in a facial surface as an Attributed Relational Graph. Identification is achieved by partial matching of these graphs. In [35], Lin and Tang proposed a method which encapsulates the occlusion detection and recovery problems through a generative process. A Bayesian formulation is proposed, where the quality assessment model is constructed by learning a priori information from a set of images.

Another approach for occlusion handling considers the facial surface as a combination of partitions. When local patches are considered separately, the areas where occlusions occur can be compensated for, in the classifier fusion phase. In [36], the facial surface is divided into local regions. Each region is modeled individually by a mixture of Gaussian distributions, and fusion is achieved by probabilistic evaluation of regional matches. In [37], Kim *et al.* propose a part-based local representation approach based on Independent Component Analysis (ICA). ICA representations are constructed for local regions corresponding to salient parts such as eye, nose, and lip areas. Conservation of discriminative features is achieved by re-ordering of basis images. In [38], a face image is represented by applying

multi-scale and multi-orientation Gabor filters and obtaining the Local Binary Pattern (LBP) map. Recognition is achieved by matching regional histograms.

Recently, there has been increasing interest in the area of sparse representation techniques. For robust face recognition against occlusions and corruptions, Wright *et al.* [39] proposed an identification technique, where the occlusion robustness is obtained by sparsely representing corrupted pixels. Additionally, identification performance for occluded facial images is improved by block partitioning. In [40], a sparse representation technique based on correntropy is proposed for occlusion handling. Nonnegativity constraint is introduced to obtain a more sparse and efficient solution. In [41], Zhou *et al.* proposed to improve sparse representation methods for handling of contiguous occlusions by including prior knowledge about the pixelwise error distribution. The spatial continuity of both corrupted and uncorrupted pixels are modeled by Markov Random Fields. In these approaches, although sparse representation appears beneficial for occluded surfaces, best results are obtained when occlusions are manually removed or compensated for via block partitioning.

2.4. Occlusion Handling in the 3D Face Recognition Literature

Handling expression and pose variations in 3D faces, has attracted wide interest in the literature. Occlusion variations, on the other hand, have only recently been studied by a few groups. In this section we summarize the literature related to occlusion handling in 3D face recognition. We group the studies into three partitions: First, papers considering only partial occlusions over the forehead region (caused by hair) are summarized. Then, details about the studies experimenting on the occlusion datasets (Bosphorus and UMB-DB databases) are given. Lastly, we mention some important studies considering missing data handling, mostly due to self-occlusions. Although self occlusions are outside the scope of this thesis we include these studies here, since if it is possible to detect and remove occluded areas, classification techniques applicable to incomplete data will be beneficial. The mentioned approaches are summarized in Table 2.2 to give a quick overview.

2.4.1. Handling of Partial Hair Occlusions: Evaluations on the FRGC v.2

Some studies in the 3D face recognition literature focus on performance improvement obtained by partial hair occlusions over the forehead region. In [42], the facial surfaces are smoothed to remove spikes using Gaussian filtering, and small holes are filled using interpolation. For face detection and ROI extraction, the nose tip is detected and it is used to center and crop the facial area. For nose tip detection, shape index map is utilized to find nose tip candidates. Nose tip template is fitted to nose candidates, and the best fit is labeled as the nose tip. A predefined radius value is employed to crop the facial surface. For initial alignment, Principal Component Analysis (PCA) is used to normalize facial pose: Here, Y and Z axes appear as the largest and smallest eigenvalued vectors, respectively. Fine alignment is carried out by the Iterative Closest Point (ICP) algorithm. After normalization, the most dissimilar faces in the gallery are rejected using the central profile curve. After narrowing the search space, six facial regions are segmented and curves extracted in these regions are used to map deformations. Once again, ICP is used for partial curve matching. The deformations will result in smaller similarity scores. Hence, they will probably be rejected in the classification process, where curves with high similarity scores are fused to obtain a final identification. The results reported on the FRGC v.2 database (97.5%) make this method a probable solution for small occlusions.

In [29], large pose and expression variations are considered, where results for hair occlusions of FRGC v.2 are reported. First, the facial area is localized using the 2D texture and range images. For localization, the Active Shape Model is utilized over the texture image, and the profile image is extracted. Normalization is handled using the symmetry plane. Next, the nose tip is detected and the facial area is extracted by cropping with a predefined radius value around the nose tip. Fine alignment of the facial surface is achieved using the axis-angle representation and then transforming the point cloud to align it with the reference model. Afterwards, bounding sphere representation is utilized to represent surfaces, where robustness to large expression and pose variations is achieved. After the representation, robust group sparse regression model based on sparse representation [43] is proposed for feature extraction, where the effect of occlusions and corruptions is minimized. The classification is handled by a spectral analysis of graph embedding.

2.4.2. Handling of Complex Occlusions: Evaluations on the Bosphorus and UMB-DB

A few studies attack the occlusion challenge and evaluate their system on the occlusion datasets. In [44], the facial surfaces are represented by Spherical Depth Map (SDM), where a sphere is fitted to the facial point cloud, allowing pose normalization and alignment. For pose normalization, convexities of the facial surface are extracted using an algorithm called Emerging from Sphere. The convexity maxima serve as nose tip candidates. From the candidates, the most probable nose tip and its orientation is extracted using Histogram of Gradient features and Support Vector Machine classifier. The registration is handled by using the nose tip and its orientation, and rotating the face around the center of the fitted sphere. The SDM representation is further utilized for down-sampling and cropping purposes. Fine surface registration is handled by the ICP algorithm, where a rejection strategy is embedded into the original ICP: At each iteration, a predefined percentage of the most distant point pairs are discarded and the remaining point sets are utilized for transformation calculations. This rejection strategy enables the elimination of occluded surface points from the registration process. The overall system is evaluated on the Bosphorus database, where a recognition rate of 97.9% is achieved at a rejection rate of 40%. The performances reported for different occlusion types can be found in Table 2.2. This is the best performance reported on the Bosphorus occlusion subset. However, it assumes visibility of nose tip, and only limited ratio of occluded areas.

In [45–47], the main challenge considered is the handling of expression variations or incomplete facial data. However, they have additionally reported results on the Bosphorus occlusion subset. All of these systems are based on keypoint extraction for obtaining a pair of corresponding salient features to be considered later in the classification process. In [45], meshSIFT is employed to obtain local shape description. In [46], meshDoG is used as the local shape descriptor to describe the local neighborhood around the extracted keypoints. Similarly in [47], various local descriptors are employed around the keypoints: Histogram of gradients, histogram of shape index, and histogram of gradient of shape index are the utilized local descriptors to describe the surface locally. Face similarity is then measured by comparing inlier pairs of matching keypoints. The main assumption is that most stable keypoints will be repeatedly extracted for the scans of an individual. Moreover, the RANSAC

algorithm is shown to be beneficial when eliminating outlier matches, especially caused by occluded facial regions. The recognition result reported on the Bosphorus occlusion subset is 93.2% for [46]. In [47], fusing different types of descriptors, a recognition rate of 99.21% is achieved on the Bosphorus occlusion subset. As the results of [46,47] set forth, employment of keypoints can be considered as a possible solution for occlusion handling.

In [48], a facial surface representation employing radial curves propagating from the nose tip is proposed to handle different types of challenges. Unfortunately, details about nose tip extraction are not sufficient; and in the experiments including occluded surfaces, manually located nose tip locations are used. Prior to extraction of the facial curves, the occlusion detection and removal is handled in corporation with the registration process, namely the recursive ICP algorithm: At each iteration, the surface points that are more distant to the model than a predefined threshold are removed. Therefore, after registration, an occlusion-free facial surface is obtained. Afterwards, using the nose tip, a reference curve vertically passing through the symmetry plane is extracted. Then, several radial curves slicing the facial surface by planes passing through the nose tip are obtained. Using a total of 40 curves, quality filtering is applied to remove curves containing insufficient information and elastic shape analysis is applied for classification purpose. For the Bosphorus occlusion subset, a recognition rate of 78.63% is obtained over the occlusion removed surfaces. They have also reported a recognition rate of 87.06%, where incomplete facial curves are restored using a statistical modeling of radial curves.

In [4], we propose to detect nose area based on curvature information. The detected nose area center is then used for initial registration. In [6], an occlusion-robust registration approach is proposed based on the area localization idea of the previously proposed nose detector. Several regions, such as nose, eyes, and mouth, are detected and checked for validity. Based on the validity of regions, an adaptive model is selected for the fine registration step. Using an adaptive model for registration enables to discard occluded parts and to employ only the non-occluded facial regions. In [12], masked projection is proposed to further improve the occlusion robustness of the face recognizer. In masked projection, the occlusion masks are incorporated into the subspace analysis techniques to extract features only from the available surface information. As the results obtained on the Bosphorus and UMB-DB

databases, nose area can be detected with sufficient accuracy, whereas the adaptive-model based registration improves the registration results. Moreover, masked projection, enabling the use of incomplete data, yields high classification performance.

2.4.3. Handling of Incomplete Data at the Classification Stage

In this section, we summarize studies considering incomplete data from the view of classification: If the occluded areas are localized accurately, removal of detected occlusions will result in incomplete facial surfaces. Therefore, classification approaches proposed for incomplete data handling can be applied to surfaces after occlusion detection and removal.

In [49], an extended version of the Annotated Face Model (AFM) [25], is utilized for fitting the model to the incomplete surface in a non-rigid manner. Here, the problem of missing data is handled by incorporating the facial symmetry property. Therefore, the missing surfaces are filled prior to classification. When high pose variations are present, the faces can be left- or right-half scans. Therefore, in addition to the whole face representation, they have obtained left and right-half representations, resulting in multiple representations for each facial image. On these representations, wavelet analysis is carried out to obtain the classification features. If the facial surfaces are known to have specific types of occlusions (such as occlusions covering left/right or top/bottom halves), and the occlusions are detected and removed accurately, this idea can be applied to extract features prior to classification.

In [46], keypoints are extracted as a first step. Then, the curves connecting pairs of keypoints are used to define the relative change in the corresponding surface regions. This way, the spatial relations between the keypoints are introduced. SIFT features are used to find inlier keypoint pairs between two different surfaces, whereas the facial curves within each surface are used for classification purposes. Handling of incomplete surfaces is automatically handled, since no keypoints will be extracted from these regions, and keypoint pairs from regions missing in at least one of the surfaces will not be chosen. As stated in the previous section, utilizing keypoints for incomplete or occluded surfaces can be beneficial for face recognition purposes. These methods should be evaluated on datasets with challenging occlusions, such as the UMB-DB dataset.

In [11], a face detection and registration method is proposed which is robust to occlusion variations. The facial surface is extracted by detecting nose tip and inner eye corners, assuming that at least two landmarks are available. The detection of landmark candidates is based on curvature analysis. From the candidates, possible encapsulated regions are selected and used together with ICP for alignment. The correct alignment is chosen by the Gappy PCA method [10]. Then a final registration by ICP is performed, which discards any surface point not representing the facial surface well at each iteration. In [50], using this registration strategy, they have focused on an occlusion detection and restoration strategy, so that any standard classification approach can be utilized afterwards. The registered surfaces are projected onto a shape space, which is constructed using a training set of non-occluded and pre-aligned faces. After the back projection to the original face space, the distance between reconstructed and occluded surface is used to find a preliminary occlusion mask. This initial mask is further refined by excluding the detected surface points from the computation of the reconstruction error. After the refinement of the occlusion mask, back projection to the original face space is obtained by the Gappy PCA algorithm, giving a restored version of the originally occluded face. On the Bosphorus occlusion subset, they have obtained a recognition performance of 91.18%, 74.75%, 94.23%, and 90.47% respectively for the eye, mouth, eyeglasses, and hair occlusion types.

2.5. Part-based Systems in the 3D Face Recognition Literature

In this thesis, we are motivated from part-based face recognizers to handle occlusion variations. Therefore, in this section, we have reviewed recent studies that consider the 3D face recognition problem in a part-based manner. The aim of the studies summarized here, is to develop a face recognizer with high performance, even when expression variations are present. They have reported results on the FRGC v.2, since this database is widely used and includes a large number of expression scans. Table 2.3 gives a list of these approaches for comparative purposes, together with rank-1 identification accuracies, that are obtained on the FRGC v.2 database.

Expression insensitive 3D face recognition systems naturally focus on rigid parts of faces. The use of nasal region is a prominent example of such approaches. In [58], three

overlapping nose regions are extracted and matching scores from these different classifiers are combined at the score level. For automatic landmark localization, some fiducial points (namely the nose tip, eye pits, and nose bridge) are located using curvatures. These landmark points are utilized to segment the face into circular regions. For classifier fusion, product and sum rules yield the best performance: On the FRGC v.2 *SuperSet* database with a gallery of 449 subjects with one neutral scan per person, the reported results are 97.1% and 87.1%, respectively for neutral probe and non-neutral probe sets. The results of this study are not included in Table 2.3, since the database is different. In a similar study, Faltemier *et al.* [55] used seven overlapping regions around the nose. The nose tip is located automatically by combining three different algorithms. For regional alignment, they have utilized the ICP algorithm. For each region, classification votes are obtained using registration distances and threshold values. The regional classifiers are fused via committee voting. For the experiments, they have used the FRGC v.2 database, and constructed the gallery with 410 subjects, each with a neutral scan. They provide a rank-1 recognition result of 94.9%. In their later work, Faltemier *et al.* [23] divided the face into a total of 38 regions, distributed over the whole facial area. The nose tip is automatically located and the regions are constructed using x and y offset values from the nose tip location and radius values to define the size of regions. The regional classification results are fused using a modified version of the Borda Count method. They have reported two different recognition results on the FRGC v.2 with different gallery and probe sets. For the first set, the gallery contains 410 neutral scans from different subjects, and for the second set, the gallery consists of the first scan of each subject (which can either be neutral or non-neutral) making a total of 466 scans. The recognition results are 98.1% and 97.2%, respectively for the first and second experimental sets. In all of these studies, the ICP based core matcher should perform alignment for every gallery face, a time consuming task when the gallery set is large.

Passalis *et al.* [57], utilize an annotated deformable face model, that is divided into different facial regions. The facial scans are rigidly registered to the model using ICP to obtain pose-invariance, where the model is elastically deformed to fit the registered scan afterwards. From the deformed model, the deformation image is obtained via UV parametrization and Haar wavelet filtering is applied for compression. For experimental setup, the FRGC v.2 database is divided into gallery and probe sets with 466 (first scan of each subject) and 3541

(remaining scans), respectively. Recognition scores for eye and nose areas, which are relatively resistant to expression variations, are reported as 85.8% and 81.5% respectively. When the regional scores are fused, the recognition rate increases to 89.5%. In [54], this work is improved by using simulated annealing method after the ICP for rigid registration of the scan to the annotated model. They report the recognition result as 96.5% on the FRGC v.2 database. In [25], they further improve their work by adding the construction of a surface normal map. Both the geometry image and the normal map is analyzed using the Haar and Pyramid wavelet transforms, to obtain two sets of coefficients as distance metrics. The classifiers are then fused using the weighted sum approach. On the FRGC v.2 database, they report results for neutral, non-neutral, and full probe sets as 99.0%, 95.6%, and 97.3%, respectively.

In [53], automatic nose tip localization is utilized and a region cropped around the nose tip is obtained. The cropped region is then triangulated and multiple local and global rank-0 tensors are computed. 2D histograms of these tensors are obtained and dimensionality reduction is done with PCA to form a single feature vector for each scan. The FRGC v.2 database is used in experiments, where a gallery of 466 scans (first scan of each subject), and a neutral probe set of 1944 scans are constructed. A recognition rate of 93.78% is obtained for this neutral probe set. Mian *et al.* [24] 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 is handled with ICP and similarity measures are fused at the metric level. The FRGC v.2 database is used for the experiments, with 466 and 3541 scans for gallery and probe sets, respectively. The use of 3D information alone gives 98.82% and 92.36% recognition rates for neutral and non-neutral probe sets, respectively. In [52], they improve their multi-modal method. In the 3D space, they automatically detect key-points at locations with high shape variations. At each key-point, pose-invariant 3D feature extraction is handled via surface fitting and regular re-sampling. PCA is applied on the extracted features and matching is obtained by fusing results at score and feature levels. On the FRGC v.2 database, the gallery and probe sizes are 466 and 3541, respectively. When only the 3D information is used, results on neutral, non-neutral, and full probe sets are 99.0%, 86.7%, and 93.5%.

Mahoor *et al.* [51] propose a method for 3D face recognition from frontal range images. Their approach utilizes ridge images, consisting of points with maximum principal curvatures (points from eyes, nose, and mouth area). For registration, two different methods are applied: Hausdorff distance and the ICP method. They obtain recognition results both on the FRGC v.2 and the GavabDB databases. For the FRGC v.2, they constructed the gallery and probe each with 370 neutral scans. The recognition accuracies are 58.92% and 91.8%, respectively, when Hausdorff distance and ICP methods are utilized for registration. Furthermore, if the whole surface is used for registration via ICP, a recognition rate of 93.7% is obtained.

In [27], the facial surface is divided into four regions: a circular and an elliptical region around the nose, an upper head region containing nose, eyes, and forehead areas, and a region consisting of the entire face. The regional registration is handled via simulated annealing using the most deformation-resistant areas to obtain expression invariance. The regional classifiers are fused via the sum rule. The FRGC v.2 database is used for experiments, where the gallery contains the first scan of each subject and the remaining images constitute the probe set. A recognition result of 98.4% is obtained.

Cook *et al.* [56] used 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 the results are fused at the score level. The facial image is divided into 147 regions and the size of the LGT response features are reduced by the PCA method. 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 scans, yield a recognition performance of 94.63%.

In [59], 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, LDA is applied to 2D textures. 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%, 77% and 90% are obtained, respectively for ICP, LDA, and ICP-LDA combination. In [60], Lu and Jain propose a method to model expression deformations to deal with expression variations. A control group consisting of a small number of subjects, 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 the ICP algorithm is used to find the best fit. Experimental results are reported on a subset of FRGC v.2, with a total of 150 scans from 50 subjects (each with one neutral, one smiling, and one surprise expression). Recognition rates of 97% and 87.6% are achieved, respectively, with and without the deformable models.

In [61], Li and Zhang use multiple intrinsic geometric descriptors such as angles, geodesic distances, and curvatures as features for an expression-invariant 3D face recognition. For each individual feature, a set of weights are trained. To combine the attributes, a different set of weights are also trained. They have experimented with the GavabDB and a subset of the FRGC v.2 containing a total of 180 scans from 30 subjects. For the GavabDB, recognition rates of 97.00% and 94.17% are obtained respectively for the leave-one-out (LOO) approach and for the normal reference (NR) approach. In the LOO approach, one scan for each subject is used as a probe face, and all the other faces constitute the reference system. In the NR method, all the neutral scans constitute the reference set, and the scans with expression variations form the probe set. On the subset of FRGC v.2, they have obtained 96.67% and 98.89% recognition performances for the NR and LOO approaches, respectively. As a cross-database validation experiment, training was performed on the GavabDB to determine weights, and FRGC v.2 subset was used as the probe set. Recognition rates of 85.34% and 95.56% were obtained with the NR and LOO methods respectively.

In [26], we have proposed a part-based registration scheme, followed by a regional subspace approach for classification. The facial surfaces are first aligned to an average face model as an initialization step. Then, for fine alignment, separate regional registrations to individual average region models are obtained. After the facial surfaces are aligned to the regional models using the proposed two-pass alignment approach, statistical features are

extracted by employing the Fisherfaces technique in a regional manner. A final classification is obtained by fusing the regional classifiers. Experimental results on the FRGC v.2 database are reported, where the gallery is constructed using the first scans, and the remaining images constitute the probe set. As the results show, regional registration and regional subspace techniques yield an expression insensitive face recognizer: On the neutral, non-neutral, and the whole probe set, rank-1 identification rates of 98.39%, 96.40%, and 97.51% are obtained, respectively.

Table 2.2. Summary of 3D face recognition methods dealing with the occlusion problem.

Reference	Database	Detection and Alignment Methods	Occlusion Handling	Holistic (H) vs. Regional (R)	Matching	Identification Accuracy
Li & Da, 2012 [42]	FRGC v.2	Nose detection, PCA, ICP	Region selection	R	ICP based curve matching	97.50%
Ming & Ruan, 2012 [29]	FRGC v.2	ASM, nose tip detection, axis-angle representation	Group sparse representation	H	Spectral analysis of graphical embedding	93.59%
Liu *et al.*, 2012 [44]	Bosphorus	Coarse Reg.: SDM for nose tip, Fine Reg.: ICP	Rejection strategy in ICP	H	ICP distance	97.90%
Berretti *et al.*, 2013 [46]	Bosphorus	Keypoint detection, meshDoG	Keypoint matching	H	Curve matching	93.20%
Li *et al.*, 2011 [47]	Bosphorus	Keypoint detection, HoG, HoS, HoGS	Keypoint matching	H	Matched salient point count	99.21%
Drira *et al.*, 2013 [48]	Bosphorus	Nose tip, radial curves	Recursive ICP for occ. detection	H	Elastic shape analysis	78.63% (occ removed), 87.06% (restored curves)
Colombo *et al.*, 2011 [50]	UMB-DB	Nose tip & inner eye coordinates ICP	Detection and restoration by Gappy PCA	H	Fisherfaces	91.18% (eye), 74.75% (mouth), 94.23% (eyeglasses), 90.47% (hair)
Alyuz *et al.*, 2013 [12]	Bosphorus, UMB-DB	Nose detection, adaptive model selection, ICP	Adaptive model based ICP, detection by distance to average face	R	Fisherfaces with Masked Projection	Bosphorus: 93.70% UMB-DB: 74.75%

Table 2.3. Rank-1 classification rates reported on FRGC v.2. N/A stands for not available cases. N and Non-N stand for neutral and non-neutral sets, respectively. The labels fs, ns, and fns denote first scans, neutral scans, and first neutral scans, respectively.

	Identification Results				
Author,Year	Gallery Size	Probe Size	N vs. All	N vs. N	N vs. Non-N
Queirolo *et al., 2010 [27]*	466 (fs)	3541	98.4%	N/A	N/A
Mahoor *et al.*, 2009 [51]	370 (ns)	370 (ns)	N/A	93.7%	N/A
Faltemier *et al.*, 2008 [23]	410 (fns)	N/A	98.1%	N/A	N/A
Faltemier *et al.*, 2008 [23]	466 (fs)	3541	97.2%	N/A	N/A
Mian *et al.*, 2008 [52]	466 (fs)	3541	93.5%	99.0%	86.7%
Mian *et al.*, 2008 [24]	466 (fs)	3541	N/A	98.82%	92.36%
Osaimi *et al.*, 2007 [53]	466 (fs)	1944 (ns)	N/A	93.78%	N/A
Kakadiaris *et al.*, 2007 [25]	466 (fs)	3541	97.3%	99.0%	95.6%
Passalis *et al.*, 2007 [54]	466 (fs)	3541	96.5%	N/A	N/A
Faltemier *et al.*, 2006 [55]	410 (fns)	3451	94.9%	N/A	N/A
Cook *et al.*, 2006 [56]	410 (fns)	N/A	94.63%	98.25%	N/A
Passalis *et al.*, 2005 [57]	466 (fs)	3541	89.5%	N/A	N/A
Alyuz *et al.*, 2010 [26]	466 (fs)	3541	97.51	98.39	96.40

3. TECHNICAL BACKGROUND

In this thesis, we use a variety of techniques from the literature in the occlusion context. In this chapter, we describe those methods which we refer to. First, details about surface curvature information are given. Principal, mean, and Gaussian curvatures, shape index and curvedness maps are introduced. Then, basic registration techniques are included: In this thesis, we have frequently used the following registration techniques: (i) Procrustes analysis, which uses a limited number of fiducial points to find a rigid transformation; (ii) Iterative Closest Point algorithm, which finds a rigid transformation to align two surfaces and a point-to-point correspondence in between; (iii) model-based registration, where surfaces are aligned to an average model to obtain a full correspondence between all surface pairs. Next, subspace techniques, that are used in this thesis, are summarized: (i) Principal Component Analysis is an unsupervised dimensionality reduction technique; (ii) Gappy Principal Component Analysis is a variant of Principal Component Analysis which can cope with incomplete data; (iii) Linear Discriminant Analysis is a supervised dimensionality reduction method used to find a low-dimensional subspace useful for classification.

3.1. Curvature Information

Curvature of a 3D surface measures the amount of local bending. Surface descriptors based on curvature information are frequently used in the 3D domain, since they are advantageous due to their rotation and translation invariance. There are different forms of curvature-based facial representations such as principal curvatures, mean and Gaussian curvatures, shape index, and curvedness maps. These representations can be beneficial, especially in localizing fiducial points or facial areas.

3.1.1. Principal, Mean, and Gaussian Curvatures

Given a point on a surface, we can define two extremal curves passing through that point and lying on that surface; the curves with minimum and maximum curvatures. Therefore, we can represent a surface point with its minimum (κ_{min}) and maximum (κ_{max}) curva-

ture values, whose corresponding directions are orthogonal [62].

Mean and Gaussian curvatures are commonly used descriptors to represent 3D surfaces [63]. They can computed using the minimum (κ_{min}) and the maximum (κ_{max}) curvature values. The mean curvature (H) for a surface point i is defined as:

$$H(i) = \frac{1}{2}(\kappa_{min}(i) + \kappa_{max}(i)), \tag{3.1}$$

whereas the Gaussian curvature (K) is computed as:

$$K(i) = \kappa_{min}(i)\kappa_{max}(i). \tag{3.2}$$

3.1.2. Shape Index and Curvedness Maps

The shape index and curvedness are curvature-based measures of the local surface. They were introduced in [64], and they can be computed using the maximum (κ_{max}) and the minimum (κ_{min}) curvatures. The transformation separates components that are dependent or independent of scale [65]. Scale-independent components, such as shape index, provide the distinction between spherical and cylindrical surfaces. On the other hand, the scale-dependent components, such as curvedness, give the magnitude of the curvature.

The shape index value $SI(i)$ at surface point i can be computed from κ_{max} and κ_{min}:

$$SI(i) = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{\kappa_{max}(i) + \kappa_{min}(i)}{\kappa_{max}(i) - \kappa_{min}(i)} \tag{3.3}$$

The shape index map $\boldsymbol{SI}$ takes values in $[0, 1]$ and provides a smooth transition between concave ($0 < SI(i) < 0.5$) and convex ($0.5 < SI(i) < 1$) shapes, as given in Figure 3.1.

As the scale-dependent counterpart of shape index, curvedness measures the rate of

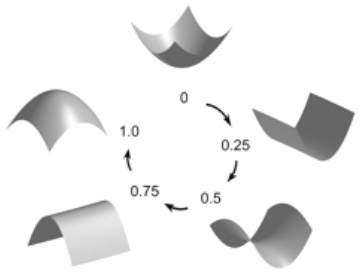


Figure 3.1. Shape index values represent a transition between concave and convex shapes.



Figure 3.2. Curvedness defines the rate of curvature of a surface.

curvature at each point:

$$C(i) = \sqrt{\frac{\kappa_{min}(i)^2 + \kappa_{max}(i)^2}{2}}. \tag{3.4}$$

A planar surface will have a curvedness of zero, whereas a non-planar surface will have a curvedness value proportional to its rate of curvature, as illustrated in Figure 3.2.

3.2. Basic Registration Techniques

The mathematical fundamentals of the registration methods that are mentioned frequently throughout this thesis are provided in this section. Procrustes Analysis, uses only landmark points located on surfaces, and finds the affine transformation between the set of landmarks for alignment. Iterative Closest Point (ICP) algorithm finds a rigid transformation between the surfaces and densely aligns two point clouds, determining a point-to-point correspondence in between. Model-based registration approach uses ICP to register each surface to a model and obtains a complete alignment to the gallery surfaces. The details and the algebraic formulations for these methods are given next.

3.2.1. Procrustes Analysis

Prior to the fine alignment of surfaces, usually a coarse alignment step is necessary. This procedure of *initial alignment* is usually handled by the Procrustes analysis proposed by Gower [66]. Procrustes analysis is a statistical shape analysis technique [67], where the distribution of a set of geometrical shapes are investigated. Procrustes alignment includes the transformations of translation, rotation, and scaling to optimally align the surfaces. Therefore, in Procrustes analysis, both the location in space and the size of the object are adjusted, where the aim is to bring the surfaces into a similar location and size.

As stated before, Procustes analysis analyzes geometrical shapes. A *geometrical shape* refers to the characteristics defining a surface, which remains geometrically unaltered even though a translation, rotation, or scaling is applied to it. A facial surface can be represented as a geometrical shape, using landmark positions: If a facial surface in $\mathbb{R}^n$ is labeled by a set of l landmark points, the corresponding geometrical shape can be represented with a $n \times l$ matrix $\mathbf{L} = \mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_l$, each column presenting a significant landmark point in n dimensions. If two figures are $\mathbf{L}_1 : n \times l$ and $\mathbf{L}_2 : n \times l$, they have the same shape if there exists a similarity transformation that relates them. This special similarity transform can be stated as follows:

$$\mathbf{L}_2 = \alpha \mathbf{\Gamma} \mathbf{L}_1 + \gamma \mathbf{1}_l^T, \tag{3.5}$$

where the parameters of the transformation are defined as $\mathbf{\Gamma} : n \times n, |\mathbf{\Gamma}| = 1$ 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 of size l. By the triple of these parameters $(\gamma, \mathbf{\Gamma}, \alpha)$, the similarity transformation consisting of translation, rotation and scaling that maps the shape $\mathbf{L}_1$ to $\mathbf{L}_2$ is defined. A simple alignment example is given in Figure 3.3, where the shapes are represented by sets of four landmark points in 2D.

The classical Procrustes analysis aligns two objects, whereas generalized Procrustes analysis generalizes the idea of pair-alignment and permits the analysis of multiple shapes [66]. By using generalized Procrustes analysis, a *consensus shape* can be derived from the

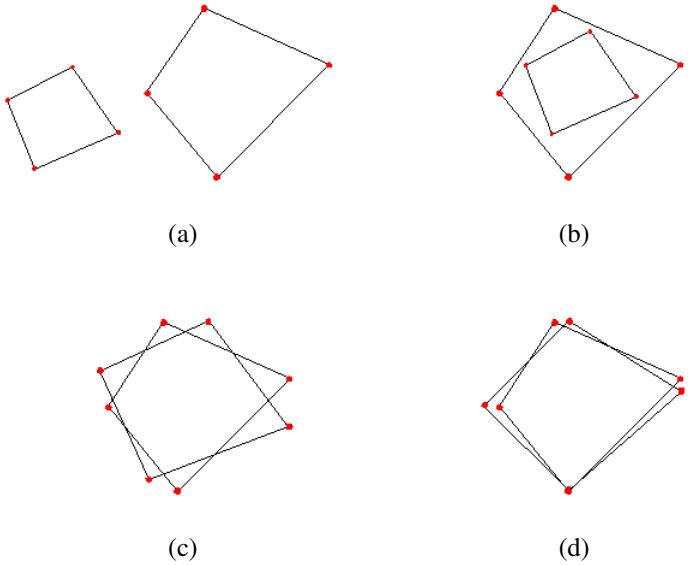


Figure 3.3. A simple alignment example is given. In (a) the raw landmarks of two shapes are shown. (b), (c) and (d) are the transformed landmarks after the translation, scaling and rotation is applied, respectively.

whole set of shapes, by minimizing the sum-of-squares between each shape and the consensus shape through translating, rotating, 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 [66]:

Translation:

(i) Find the centroid of all shapes:

$$\mathbf{C} = \frac{1}{s}\sum_{i=1}^{s}\mathbf{L}_i \tag{3.6}$$

where s stands for the number of shapes to be analyzed.

(ii) Center all shapes $\mathbf{L}_i$ using the centroid $\mathbf{C}$:

$$\mathbf{L}_i = \mathbf{L}_i - \mathbf{C} \tag{3.7}$$

Scaling:

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

(i) Set the *mean* of the squared landmark distances of each shape to unit value [66], where $\mathbf{p}_{i,k}$ stands for the kth landmark of $\mathbf{L}_i$:

$$\mathbf{L}_i = \frac{N * \mathbf{L}_i}{\sum_{k=1}^{l} ||\mathbf{p}_{i,k}||^2} \tag{3.8}$$

(ii) Set the *median* of the squared landmark distances of each shape to unit value [68]:

$$\mathbf{L}_i = \frac{\mathbf{L}_i}{median(\mathbf{D}_i)} \tag{3.9}$$

where $\mathbf{D}_i = d_{j,k} = ||\mathbf{p}_{i,j} - \mathbf{p}_{i,k}||^2 \quad j,k = 1,\ldots,l$ and $median(\cdot)$ is the median operator.

Rotation:

(i) Initialize the consensus shape $\mathbf{L}_C$:

$$\mathbf{L}_C = \mathbf{L}_1. \tag{3.10}$$

(ii) For $i = 2, 3, \ldots, s$, rotate $\mathbf{L}_i$ to fit $\mathbf{L}_C$.

(a) in Gower's method as explained in [66] $\mathbf{L}_C$ is re-evaluated after each update of

$\mathbf{L}_i$ as

$$\mathbf{L}_C = \frac{1}{i}\sum_{j=1}^{i}\mathbf{L}_j \tag{3.11}$$

(b) In Rohlf and Slice's method given in [68], $\mathbf{L}_C$ is updated only once, after the rotation of each $\mathbf{L}_i$.

The rotation matrix $\mathbf{H}$ in two dimensional space can be expressed as:

$$\mathbf{H} = \begin{bmatrix} cos\theta & -sin\theta \\ sin\theta & cos\theta \end{bmatrix} \tag{3.12}$$

To find the best rotation, singular value decomposition [69] can be used:

$$\mathbf{H} = \mathbf{V}\mathbf{S}\mathbf{U}^T \tag{3.13}$$

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

$$\mathbf{L}_i^T\mathbf{L}_C = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{3.14}$$

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

(iii) Update the $\mathbf{L}_i$ and $\mathbf{L}_C$, while monitoring the residual sum-of-squares:

$$SS_r = s(1 - tr(\mathbf{L}_C^{(t)}\mathbf{L}_C^{(t)T} - \mathbf{L}_C^{(t-1)}\mathbf{L}_C^{(t-1)T})) \tag{3.15}$$

where $\mathbf{L}_C^{(t)}$ stands for the consensus shape at iteration t, and $\mathbf{L}_C^{(t-1)}$ is the consensus shape at iteration $t-1$. When SS_r drops below a threshold value, iterations are stopped, and the consensus shape is hence found.

In this thesis, Procrustes Analysis will be used to coarsely align facial surfaces. Since the acquisition of 3D data does not include depth variances, the size of facial surfaces can be treated as a useful hint. Therefore we have eliminated the scaling step of Procrustes in our experiments.

3.2.2. Iterative Closest Point Algorithm

After the initial alignment between two surfaces is obtained, it is necessary to perform a fine registration: In fine registration, point clouds are utilized to align the surfaces, instead of considering a limited number of landmark locations. ICP is an algorithm employed to minimize the difference between two point clouds and achieve a point-to-point correspondence in between [70]. The transformation (translation and rotation) is updated iteratively to minimize the distance between the points of two raw scans. Although the algorithm has a high computational cost, ease of implementation and applicability to several geometrical representations such as point sets, line segments, parametric curves and surfaces makes ICP a frequently referred method for the registration of 3D surfaces.

The idea of the algorithm can be briefly summarized as follows:

- For each point in one scan, find the closest point in the other scan.
- Estimate transformation parameters that would align the associated point pairs.
- Transform all points in one point cloud using the estimated parameters.
- Iterate until the stopping criteria is met.

As this summary points out, the aim of the algorithm is to find both a transformation that best aligns the two point clouds and a point-to-point correspondence in between the two sets. A basic visual example for ICP-based alignment of two 2D surfaces is given in Figure 3.4.

Before going on, some mathematical preliminaries about computing the closest point on a model to a given point of a second surface and finding the correspondence between the two surfaces by the quaternion-based least-squares registration should be reminded. For ease of comprehension, the details are given assuming the surfaces are in 3D.

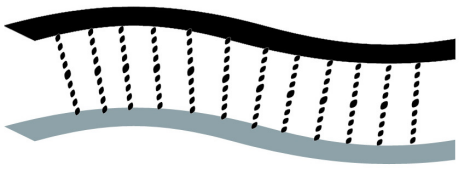


Figure 3.4. A visual example of matching shapes by ICP. A point-to-point correspondence from one surface to the other is found.

Let $\mathbf{p}_1$ and $\mathbf{p}_2$ be two points in such that $\mathbf{p}_1 = (x_1, y_1, z_1)$ and $\mathbf{p}_2 = (x_2, y_2, z_2)$. The Euclidean distance between these two points is formulated as follows:

$$d(\mathbf{p}_1, \mathbf{p}_2) = ||\mathbf{p}_1 - \mathbf{p}_2|| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2} \tag{3.16}$$

If $\mathbf{P}$ is a point set of N_p points such that $\mathbf{P} = \{\mathbf{p}_i\}$ where $i = 1, 2, ..., N_p$, the distance between a given point $\mathbf{p}_g$ and the point set can be defined as:

$$d(\mathbf{p}_g, \mathbf{P}) = min_{i\epsilon 1,...,N_p} \quad d(\mathbf{p}_g, \mathbf{p}_i) \tag{3.17}$$

A closest point $\mathbf{p}_j$ in the point set $\mathbf{P}$ satisfies the definition below:

$$d(\mathbf{p}_g, \mathbf{p}_j) = d(\mathbf{p}_g, \mathbf{P}) \tag{3.18}$$

The closest point computation explained above is in a general form and is applicable to n dimensions. A possible 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.

The *unit quaternion* can be defined as a four-sized vector $\mathbf{q}_R = [q_0 q_1 q_2 q_3]^T$, 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 $\mathbf{R}$ generated by the unit

quaternion is given below:

$$\mathbf{R}(\mathbf{q}_R) = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1q_2 - q_0q_3) & 2(q_1q_3 + q_0q_2) \\ 2(q_1q_2 + q_0q_3) & q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_2q_3 - q_0q_1) \\ 2(q_1q_3 - q_0q_2) & 2(q_2q_3 + q_0q_1) & q_0^2 + q_1^2 - q_2^2 - q_3^2 \end{bmatrix} \tag{3.19}$$

If the translation vector is defined as $\mathbf{q}_T = [q_4 q_5 q_6]^T$ the complete registration state vector can be given as $\mathbf{q} = [\mathbf{q}_R \mathbf{q}_T]^T$. Let $\mathbf{P} = \{\mathbf{p}_i\}$ be a point set to be aligned to the model point set $\mathbf{M} = \{\mathbf{m}_i\}$, where both of the point sets have the same number of points such that $N_p = N_m = N$ and that each point $\mathbf{p}_i$ is in correspondence with point $\mathbf{m}_i$. The mean-square objective function to be minimized by the ICP procedure is:

$$f(\mathbf{q}) = \frac{1}{N} \sum_{i=1}^{N} ||\mathbf{m}_i - \mathbf{R}(\mathbf{q}_R)\mathbf{p}_i - \mathbf{q}_T||^2. \tag{3.20}$$

When minimizing the objective function $f(\mathbf{q})$, first the rotation matrix is computed, which is followed by the estimation of the translation matrix. For the rotation parameter calculations, the following notation should be included: The cross-covariance matrix $\mathbf{\Sigma}_{pm}$ of the point set and the model can be formulated as:

$$\mathbf{\Sigma}_{pm} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{p}_i - \boldsymbol{\mu}_p)(\mathbf{m}_i - \boldsymbol{\mu}_m)^T = \frac{1}{N} \sum_{i=1}^{N} \mathbf{p}_i \mathbf{m}_i^T - \boldsymbol{\mu}_p \boldsymbol{\mu}_m^T. \tag{3.21}$$

where the center of mass of the point set and the model are given by:

$$\boldsymbol{\mu}_p = \frac{1}{N} \sum_{i=1}^{N} \mathbf{p}_i \quad , \quad \boldsymbol{\mu}_m = \frac{1}{N} \sum_{i=1}^{N} \mathbf{m}_i \tag{3.22}$$

The optimal rotation $\mathbf{q}_R$ corresponds to the maximum eigenvalue of the matrix $\mathbf{Q}(\mathbf{\Sigma}_{pm}$, which is a symmetric 4×4 matrix and is formed as:

$$\mathbf{Q}(\mathbf{\Sigma}_{pm}) = \begin{bmatrix} tr(\mathbf{\Sigma}_{pm}) & \boldsymbol{\Delta}^T \\ \boldsymbol{\Delta} & \mathbf{\Sigma}_{pm} + \mathbf{\Sigma}_{pm}^T - tr(\mathbf{\Sigma}_{pm})\mathbf{I}. \end{bmatrix} \tag{3.23}$$

Here, $\mathbf{I}$ is a 3×3 identity matrix and $\mathbf{\Delta}$ is the column matrix, $\mathbf{\Delta} = [\mathbf{A}_{23}\mathbf{A}_{31}\mathbf{A}_{12}]^T$, where $\mathbf{A}_{ij} = (\mathbf{\Sigma}_{pm} - \mathbf{\Sigma}_{pm}^T)_{ij}$.

Lastly, the optimum translation vector $\mathbf{q}_T$ can be computed as follows:

$$\mathbf{q}_T = \boldsymbol{\mu}_m - \mathbf{R}(\mathbf{q}_R)\boldsymbol{\mu}_p. \tag{3.24}$$

The least-squares quaternion operation can be written as:

$$(\mathbf{q}, e) = \Phi(\mathbf{P}, \mathbf{M}) \tag{3.25}$$

where $\mathbf{q}$ denotes the quaternion operation and e is the mean square error. The point set $\mathbf{P}$ can be denoted by $\mathbf{q}(\mathbf{P})$ after the transform represented by $\mathbf{q}$ is applied.

Now that the mathematical preliminaries are given, the ICP algorithm can be summarized. Given a point set $\mathbf{P}$ with N_p points and a model point set $\mathbf{M}$ with N_m points, the following steps should be followed to find the rigid transformation between the point sets:

(i) Initialize: $\mathbf{P}_0 = \mathbf{P}$, $\mathbf{q}_0 = [1, 0, 0, 0, 0, 0, 0]^T$, k=0, where k is the iteration number.

(ii) Repeat until $e_k - e_{k+1} < \theta$, where θ is the predefined convergence threshold, and e_j is the registration error of the jth iteration.

- Compute the closest points: $\mathbf{M}_k = C(\mathbf{P}_k, \mathbf{M})$, where $C(\cdot, \cdot)$ is the closest point operator.
- Compute the registration: $(\mathbf{q}_k, e_k) = \Phi(\mathbf{P}_0, \mathbf{M}_k)$
- Apply the registration: $\mathbf{P}_{k+1} = \mathbf{q}_k(\mathbf{P}_0)$

(iii) Return: $(\mathbf{q}, e)$

In summary, ICP achieves a dense point-to-point correspondence between the point sets of the two surfaces. This correspondence is obtained in an iterative manner, where the closest point on the test surface for each of the points of the model is located and the test surface is rigidly transformed to minimize the total point-to-point distance of the currently estimated correspondence. When convergence is achieved, the total point-to-point distance

gives the point set difference (PSD). The PSD value can be used as the dissimilarity measure when the faces are to be compared.

3.2.3. Model-based Registration

Dense registration techniques such as ICP provide a dense point-to-point correspondence between two surfaces and the finally calculated PSD value serves as the dissimilarity measure. Since for a recognition scenario, a probe surface should be compared against all of the surfaces in the gallery set, it is necessary to align the probe to each of the gallery scans separately. In the literature, this is referred to as the *one-to-all* registration. However, this approach is computationally costly, since the number of registrations to be performed equals the number of gallery scans. In the 3D face recognition literature, a simpler registration technique is often used: In [5], it was proposed to use an average face model, which defines a common coordinate frame for the registration. In this approach, all of the gallery images are aligned to an average face model beforehand. A probe face is registered to the average face model for only once. Therefore, a single registration is sufficient to obtain the point-to-point correspondences between the probe face and all of the gallery faces via the correspondences of these faces to the average model. Further details about model generation and model-based registration are given in the following subsections.

Average Face Model Generation To lower the computational cost of dense registration, it is necessary to construct an average face model. In this thesis, we have used the model generation method proposed in [71], which is based on Thin Plate Spline (TPS) warping algorithm of [72]. This model will be constructed from a set of training faces in an offline manner. For model construction, several accurately labeled landmark points on the training images are needed. Given a set of training images and the landmark locations for each of these images, the average face model can be computed as follows:

(i) Apply Procrustes Analysis on the manually labeled landmark locations to find a *mean* distribution of landmarks.

(ii) Transform the mean landmarks so that they represent a fully frontal face. This is achieved by transforming the landmarks such that the eye plane is approximately par-

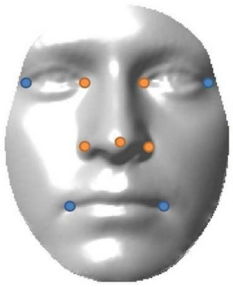


Figure 3.5. Average face model constructed from neutral scans of the FRGC v.2 database. The utilized set of nine landmarks are located on the average model.

allel to the x-axis and the plane vertical to the eye plane is parallel to the z-axis.

(iii) Warp each training face to the mean landmark distribution using the TPS algorithm, where the training landmarks are exactly superpositioned over the mean landmarks, and all other surface points of the training face are interpolated.

(iv) Resample depth values using a regular $x - y$ grid. Regularly resampling ensures that all the training images have points with overlapping x and y values.

(v) Define a cropping mask enclosing the facial area. The mask is determined by first computing the distances from the nose tip to all other landmarks and a threshold value is set which permits a 10% tolerance over the maximum distance to the nose tip. The points that are more distant to the nose tip than the threshold are trimmed off. The remaining point locations constitute the cropping mask, and this cropping mask is used to set the valid parts of the training images.

(vi) Crop all training images according to the cropping mask and average the depth values of corresponding point locations.

Following the above steps, an average face model is constructed. An example average face model is given in Figure 3.5 together with the employed landmark locations, where neutral images of the FRGC v.2 database are utilized. For TPS warping, a set of nine landmarks are used: inner and outer eye corners, mouth corners, nose boarders, and the nose tip.

Alignment to the Average Face Model Once the average face model is constructed, we can proceed to the alignment procedure. The registration of a facial surface to the average

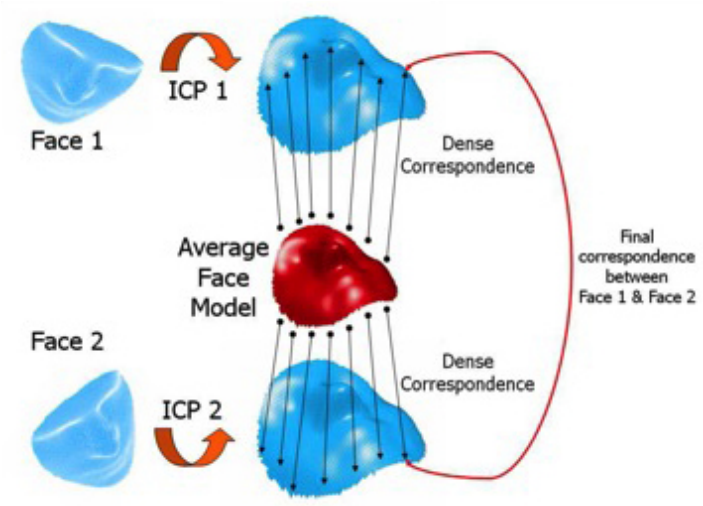


Figure 3.6. Correspondence between any two surfaces can be achieved by registering surfaces separately to the average face model.

face model consists of two phases, namely the coarse and the fine registration steps. The aim of coarse alignment is to achieve a correct convergence in the fine alignment step. If a set of landmark locations are present for the facial surface to be registered, then Procrustes analysis can be used to align the two surfaces coarsely. After the facial surfaces are aligned using only the landmark locations, a more detailed alignment step is necessary. In the fine alignment phase, all of the surface points are taken into account, seeking for a better alignment. In fine alignment, we performed ICP: At each iteration, ICP improves the point-to-point correspondence between the input face and the average face model. Once the ICP algorithm converges, the final point-to-point correspondence gives the valid point set of the input face that best resembles the average model. Thus, it can be stated that the average face model acts as an index, where for each model point, the corresponding point on the input surface is located. Therefore, the registered input point set will have as many points as the model. Given any two faces, if they are both registered to the average face model, their correspondence can be readily found via their correspondences to the model. The idea of registering to the average face model is given in Figure 3.6.

3.3. Subspace Analysis Techniques

In classification applications, observations often are high-dimensional data, including redundancy. To reduce the both time and space complexity of learning and inference algorithms, it is necessary to remove the redundancy, for which dimensionality reduction techniques are employed. Dimensionality reduction is the process of reducing the dimensionality of the data, while preserving any important information needed for decision making. Dimensionality reduction techniques can be divided into two as *feature selection* and *feature extraction* methods [73]. In this thesis, feature extraction methods are utilized, where feature extraction transforms the data in the original high-dimensional space to a subspace with fewer dimesions. Therefore, we refer to these feature extractors as *subspace* techniques. The most widely used subspace techniques are the Principal Component Analysis and the Linear Discriminant Analysis, both of which provide linear transformations. Principal Component Analysis is an unsupervised technique, where only the input data is used to find the transformation. On the other hand, Linear Discriminant Analysis is a supervised method, which additionally benefits from the output information to define the subspace. In this section, we consider three subspace techniques, that will be referred in later chapters: (i) Principal Component Analysis (PCA), where a lower dimensional representation is learned in an unsupervised manner; (ii) Gappy PCA, which modifies PCA to transform incomplete data; and (iii) Linear Discriminant Analysis (LDA), where the lower dimensional space is learned for classification purposes in a supervised manner.

3.3.1. Principal Component Analysis and the Eigenfaces Method

Principal Component Analysis (PCA) [74] is an unsupervised feature extraction technique, since only the input data is used to find the linear transformation. PCA find the orthogonal projection, that maps the input data onto a lower dimensional subspace, while maximizing the variance of the projected data.

Suppose that we have a set of observations $\mathbf{x}$ in the euclidean space with d dimensions with covariance $\mathbf{\Sigma}$. The goal of PCA is to find the projection $\mathbf{W}$, which gives a mapping from the d dimensional space to r dimensional space ($r < d$), while maximizing the variance

between the observations in the new r dimensional space. The matrix $\mathbf{W}$ is a collection of the *principal component* vectors. Let's assume that, $\mathbf{w}_1$ is the first principal component ($||\mathbf{w}_1|| = 1$), suct that projecting onto this direction makes the data points as distant apart as possible. The projection of $\mathbf{x}$ onto the direction defined by a projection vector $\mathbf{w}_1$ is given as:

$$z_1 = \mathbf{w}_1^T \mathbf{x} \tag{3.26}$$

Then, the variance of projected data is:

$$var(z_1) = \mathbf{w}_1^T \mathbf{\Sigma} \mathbf{w}_1 \tag{3.27}$$

In PCA, the aim is to find $\mathbf{w}_1$ such that $var(z_1)$ is maximized subject to $\mathbf{w}_1^T \mathbf{w}_1 = 1$. This can be stated as a Lagrange problem:

$$max_{\mathbf{w}_1}(\mathbf{w}_1^T \mathbf{\Sigma} \mathbf{w}_1 - \lambda(\mathbf{w}_1^T \mathbf{w}_1 - 1)) \tag{3.28}$$

If we take the derivative with respective to $\mathbf{w}_1$ and set it equal to zero, we will have:

$$\mathbf{\Sigma} \mathbf{w}_1 = \lambda \mathbf{w}_1 \tag{3.29}$$

Therefore, $\mathbf{w}_1$ as an eigenvector of $\mathbf{\Sigma}$ with an eigenvalue of λ. The eigenvector with the largest eigenvalue will provide the variance to be maximum. Therefore, the principal vector $\mathbf{w}_1$ is computed as the eigenvector of the covariance matrix with the largest eigenvalue. Similarly, it can be shown that the other principal components are the eigenvectors of $\mathbf{\Sigma}$ with eigenvalues in descending order. Using the set of observations with sample mean $\mathbf{m}$ and covariance $\mathbf{S}$, the projection matrix $\mathbf{W}$ can be constructed, where the columns of the matrix are the d leading eigenvectors of $\mathbf{S}$. Then, the projection onto the subspace can be computed as:

$$\mathbf{z} = \mathbf{W}^T(\mathbf{x} - \mathbf{m}) \tag{3.30}$$

This projection transforms the data onto a subspace whose dimensions are the eigenvectors [73, 75].

When we want to apply PCA on face space, unfortunately the covariance matrix of facial images becomes computationally infeasible due to high dimensionality. However, we know that the rank of the covariance matrix is limited by the number of observation samples: If there are N observations, then there will be at most $N - 1$ eigenvectors with non-zero eigenvalues. Therefore, if the number of training images is smaller than the dimensionality of the face space (which is often the case), there is a feasible way to compute the eigenvectors. This technique is called the *Eigenfaces* approach [13], and the details are given below.

Let's assume that we have an observation matrix $\mathbf{X}$, whose mean is zero. The sample covariance matrix is $\mathbf{S} = \mathbf{X}\mathbf{X}^T$, and the eigenvector decomposition is given by:

$$\mathbf{S}\mathbf{w}_i = \mathbf{X}\mathbf{X}^T\mathbf{w}_i = \lambda_i\mathbf{w}_i \tag{3.31}$$

However, since observations are high-dimensional, the covariance matrix is too large to work with. Instead, we can find the eigenvector decomposition of $\mathbf{X}^T\mathbf{X}$ as:

$$\mathbf{X}^T\mathbf{X}\mathbf{v}_i = e_i\mathbf{v}_i \tag{3.32}$$

If both sides in this equation is multiplied by $\mathbf{X}$, then we get:

$$\mathbf{X}\mathbf{X}^T\mathbf{X}\mathbf{v}_i = e_i\mathbf{X}\mathbf{v}_i \tag{3.33}$$

Therefore, we can conclude that the eigenvectors of $\mathbf{S}$ can be computed by multiplying the eigenvectors of $\mathbf{X}^T\mathbf{X}$ by $\mathbf{X}$. In our experiments, instead of using traditional PCA, we have employed the eigenfaces approach, when necessary. In this thesis, when we refer to PCA, actually we are refering to the Eigenfaces approach for the computation of the subspace.

3.3.2. Gappy Principal Component Analysis

Gappy PCA [10] was proposed as a Principal Component Analysis (PCA) variant to handle data with missing components. With Gappy PCA, it is possible to reconstruct original signal up to a certain degree when the signal contains missing values. In order to estimate the unknown facial data by the Gappy PCA method, locations of the missing components are required. Prior to estimation, Gappy PCA method utilizes PCA to construct the lower-dimensional subspace using a training set of complete observations. The basis vectors are determined using a training set of N observations, $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset \mathbf{R}^n$. A sample $\mathbf{x}$ can then be estimated using a subset ($r < d$) of these basis vectors:

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{W}\boldsymbol{\alpha} \tag{3.34}$$

where the vector $\boldsymbol{\mu}$ defines the mean, and $\mathbf{W}$ is the matrix of eigenvectors whose eigenvalues are given in $\boldsymbol{\alpha}$. The eigenvector coefficients are obtained by the inner product of the input vector and the corresponding eigenvector. Suppose there is an incomplete version of $\mathbf{x}$, namely $\mathbf{y}$, whose missing components are encoded in the occlusion mask. In Gappy PCA, the aim is to find a similar expression that approximates the incomplete data as in (3.34):

$$\mathbf{y} \simeq \tilde{\mathbf{y}} = \boldsymbol{\mu} + \mathbf{W}\boldsymbol{\beta} = \boldsymbol{\mu} + \sum_{i}^{r} \beta_i \mathbf{w}_i \tag{3.35}$$

However the β coefficients cannot be computed by the simple inner product method. Instead, the coefficients minimizing the squared reconstruction error should be sought. A basic definition of the squared reconstruction error would be given as $E = ||\mathbf{y} - \tilde{\mathbf{y}}||^2$.

To improve the error term, only the available information should be involved in the calculations. To discard the missing components, the gappy norm [10] must be used, where the information about the missing components is encoded in the mask $\mathbf{m}$. The gappy norm for a vector $\mathbf{u}$ with the mask $\mathbf{m}$ can be defined as $||\mathbf{u}|| = \sqrt{(\mathbf{u}, \mathbf{u})_m}$ where

$$(\mathbf{u}, \mathbf{u})_m = \sum_{i=1}^{n} u_i u_i m_i. \tag{3.36}$$

Using the gappy norm, the reconstruction error term can be redefined as: $E_m = ||\mathbf{y} - \tilde{\mathbf{y}}||_m^2$. If we rewrite the error term by opening the squared terms and differentiating with respect to each β_i coefficient, we obtain a linear system of M equations:

$$\frac{\partial E}{\partial \beta_i} = -\mathbf{z}_i + \sum_{j=1}^{r} \beta_j A_{ij} = 0. \tag{3.37}$$

where $\mathbf{z}_i = (\mathbf{y}, \mathbf{w}_i)_m$ and $A_{ij} = (\mathbf{w}_i, \mathbf{w}_j)_m$. The linear system can be rewritten as $\mathbf{A}\boldsymbol{\beta} = \mathbf{z}$ and the coefficients can be computed as follows: $\boldsymbol{\beta} = \mathbf{A}^{-1}\mathbf{z}$.

After the coefficients are computed, the incomplete image can be reconstructed by Equation 3.35.

In face recognition, we can find the projection matrix using Eigenfaces approach. Then the Gappy PCA method can be used together with a mask (for this thesis, the mask will be the occlusion mask), to find a projection onto the subspace. This technique was used in the 3D face literature by Colombo *et al.* in [11, 76]. In [4], we propose to use the reconstructed data only for the missing components and the original data for the non-occluded facial regions. We refer to this method as partial Gappy PCA (pGPCA).

3.3.3. Linear Discriminant Analysis and the Fisherfaces Method

Linear discriminant analysis (LDA) is supervised dimensionality reduction method, used for classification purposes. Develop by Fisher [77], it is also referred to as the *Fisher's Linear Discriminant*. Let's assume that we have samples from K classes. The aim of LDA is find a matrix $\mathbf{W}$, such that the projection onto the subspace defined by this matrix separates the samples from different classes, whereas the samples of the same class are grouped together:

$$\mathbf{z} = \mathbf{W}\mathbf{x} \tag{3.38}$$

where $\mathbf{z}$ is a k-dimensional vector and the projection matrix $\mathbf{W}$ is $d \times k$. The $\mathbf{W}$ matrix is selected such that the between-class scatter is maximum, whereas the within-class scatter is maximum. The between-class scatter matrix can be computed as:

$$\mathbf{S}_B = \sum_{i=1}^{k} N_i(\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T \tag{3.39}$$

where N_i is the number of samples in class C_i, $\boldsymbol{\mu}_i$ is the class mean, and $\boldsymbol{\mu}$ is the total mean of all the samples. The within-class scatter can be defined as:

$$\mathbf{S}_W = \sum_{i=1}^{k} \sum_{\mathbf{x}_t \in \mathbf{X}_i} (\mathbf{x}_t - \boldsymbol{\mu}_i)(\mathbf{x}_t - \boldsymbol{\mu}_i)^T \tag{3.40}$$

where $\mathbf{X}_i$ is the observation matrix belonging to the class C_i. The between-class and the within-class scatter matrices of the projected samples will then be $\mathbf{W}^T\mathbf{S}_B\mathbf{W}$ and $\mathbf{W}^T\mathbf{S}_W\mathbf{W}$, respectively. To maximize the between-class scatter and to minimize the within-class scatter, the projection $\mathbf{W}$ should be sought such that:

$$\mathbf{W}^* = \arg\max_{\mathbf{W}} \frac{|\mathbf{W}^T\mathbf{S}_B\mathbf{W}|}{|\mathbf{W}^T\mathbf{S}_W\mathbf{W}|} \tag{3.41}$$

where the k largest eigenvectors of $\mathbf{S}_W^{-1}\mathbf{S}_B$ constitute the solution. Note that, there are at most $K-1$ non-zero eigenvalues. Therefore, the value of k has an upper limit of $K-1$.

In the face recognition problem, the number of images in in the training set is much smaller than the number of pixels in each image (the feature vector), which results in singularity of $\mathbf{S}_W$. In order to overcome this problem, in [14], a method called *Fisherfaces* is proposed: The samples are first projected to a lower dimensional space via PCA (Eigenfaces) to avoid singular $\mathbf{S}_W$. Then, LDA is applied to the lower dimensional data. In this thesis, the Fisherfaces method will be referred frequently in the classification step.

4. MOTIVATIONAL WORK: PART-BASED 3D FACE RECOGNITION

In a face recognition scenario, if the facial surface is occluded by an exterior object, only the available surface information should be utilized to identify the subject. Therefore, it is necessary to perform the classification procedure using only the *partial* non-occluded surface data. A possible alternative to handle occlusions is to consider faces as a combination of surface patches. In the 3D face recognition literature, there has been a great number of part-based systems, as reviewed in Section 2.5. These studies focus on performance improvement, even when facial surfaces are deformed by expression variations. In this chapter, we introduce a part-based 3D face recognition system [26] as a motivational work, where the aim is to obtain expression-robustness for 3D face recognition. Motivated from the performance improvement obtained with this method, in this thesis, we focused on adapting and improving the ideas of part-based registration and recognition to the problem of occlusion handling.

In the work represented in this chapter, there are two separate parts that yield performance improvement: (i) An efficient part-based facial surface registration approach, and (ii) A part-based classification method, where regional discriminative features are extracted by applying part-based subspace techniques. The first phase of any 3D face recognition system, namely alignment/registration of facial surfaces, is the most crucial part and the final accuracy of the system heavily depends on the quality of the alignment module. In this paper, we propose a simple, fast, and effective region-based rigid registration approach. The probe is registered in a two-pass algorithm: First, rigid registration to an average model, followed by registration to individual average region models. The algorithm is preceded by a novel automatic landmark localization module, which provides initialization. After regional registration is performed, we study the benefits of using statistical feature extraction and the application of Fisherfaces method to 3D point cloud features to obtain regional discriminative features.

The experiments included in this chapter evaluate the system performance both on neutral and expression scans. In the experiments, we have used two 3D face databases containing expressions: FRGC v.2 and Bosphorus. FRGC v.2 is the most commonly used database for 3D face recognition and we have obtained comparable performance to the best accuracy reported in the literature: 97.51%. On the Bosphorus database, which contains an extensive range of expressions, a recognition rate of 98.19% was obtained.

In this chapter, first of all, the details of the system proposed in [26] are given. Then, experimental results are reported to show the performance improvement obtained by the part-based registration and recognition stages. Finally, the conclusions are summarized, emphasizing the ideas that can be adapted for the problem of occlusion handling.

4.1. Part-based Face Recognition System

The proposed system consists of four parts: (i) a novel automatic facial landmark detection algorithm, (ii) a robust component based registration that can deal with the large surface deformations caused by expressions, (iii) discriminative 3D feature extraction, and (iv) a classifier fusion module. Automatic landmark detection algorithm locates five points around the nose region and these points are then used at the first phase of the coarse alignment step. Our region based registration method is inspired by the Average Face Model (AvFM) based registration approach [5, 78, 79] and is extended to incorporate independent local regions as in [25], which will be referred to as the Average Region Model (AvRM). The AvRM based alignment offers several advantages such as using the generic facial parts as an index file, reducing the computational cost of registration due to the elimination of pairwise ICP registrations for every gallery image, and finally providing one-to-one correspondence of all surface points. After registration, we study the importance of using *statistical features* obtained from point coordinates. At the last phase, we utilize several classifier fusion techniques, at *abstract* and *score level*, to deduce the identity of the given probe image. An illustration of the general outline of the proposed approach is shown in Figure 4.1.

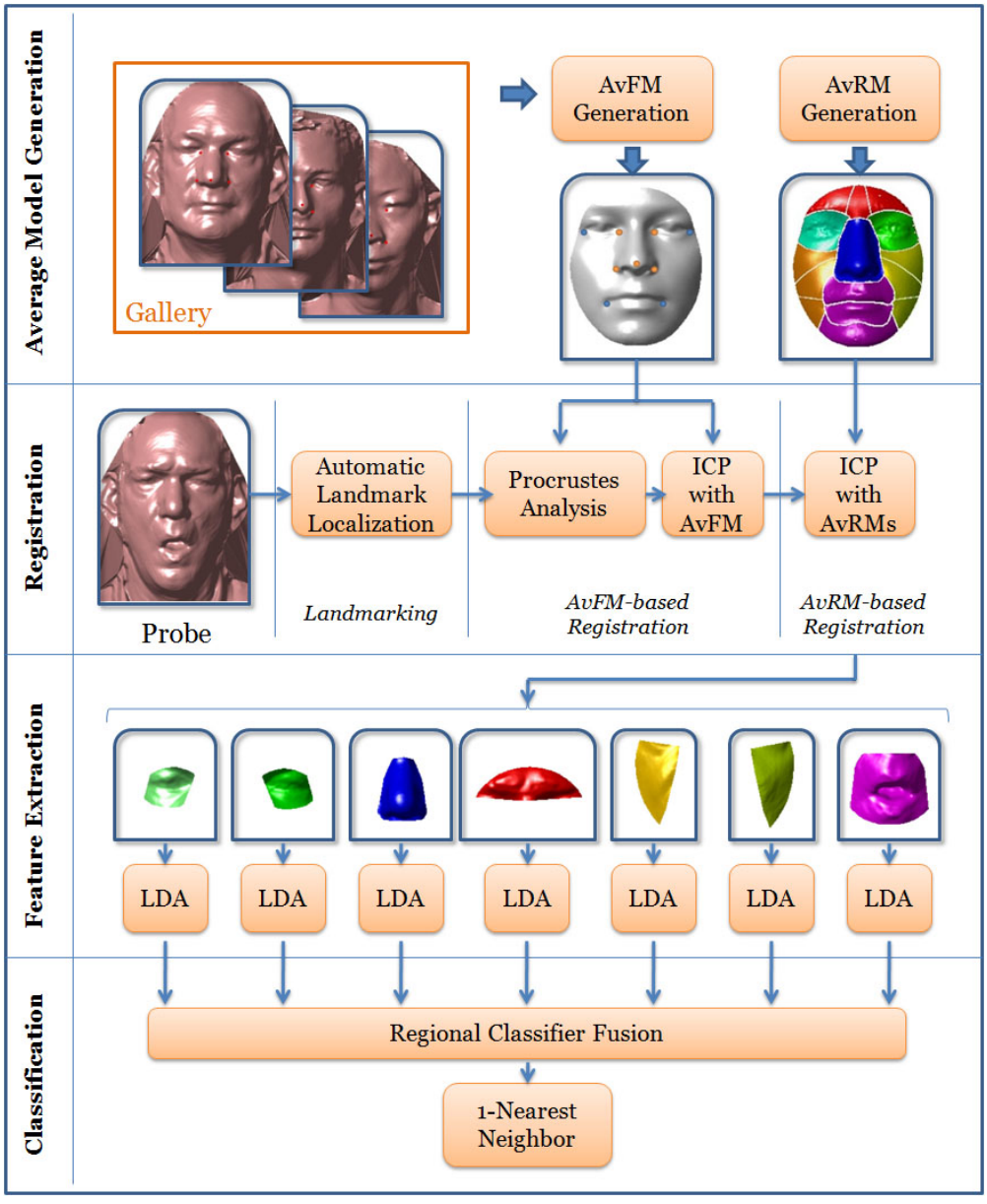


Figure 4.1. Illustrative diagram of the proposed 3D face recognition approach.

4.1.1. Automatic Landmark Localization

The quality of the facial surface alignment methods, especially iterative approaches like the ICP method, relies on initial conditions, such as the starting positions of the facial surface pairs. In order to improve the convergence of the iterative registration methods, pre-alignment is often necessary. Most of the 3D face recognition systems use facial landmarks during the pre-alignment, or *coarse alignment*, phase. Generally, the most distinctive facial features such as the nose tip, eye corners, and mouth corners are located for coarse registration. In this work, we use five fiducial points around the nose region that are mostly stable

even under facial expression variations. These are left/right inner eye pits, nose tip and leftmost/rightmost points of the lower nose border region. Except for the situations where large in-plane rotations are present, all of these points can be localized efficiently and are sufficient for pre-alignment of facial surfaces.

Our landmark localization algorithm uses 3D shape data only. We first detect the central profile contour, *facial symmetry axis*, and then search for the nose tip on the profile contour. In order to extract the vertical symmetry axis, we employ a symmetry operator that uses shape index values computed from surface curvatures. The use of curvature-based symmetry axis detection is advantageous since it is invariant to rotations and translations. The facial profile curve detection algorithm works as follows: First, for every point on the 3D facial surface, we compute the shape index values, as given in Section 3.1. Then, we use a local sliding window-based symmetry operator which computes a *symmetry map*, I_S, using shape index map $\boldsymbol{SI}$. The symmetry value of a pixel at the $(i,j)^{th}$ location is computed by a local window W of size $2N \times 2M$ centered at pixel (i,j):

$$I_S(i,j) = \sum_{m=-M}^{M} \sum_{n=0}^{N} |SI(i+m, j-n) - SI(i+m, j+n)| \tag{4.1}$$

In I_S, pixels having smaller values denote regions of high symmetry. In our system, we set N and M to 15 pixels. A frontal 3D face image without rotation variations is expected to have high symmetry map values along the vertical facial profile. With this assumption, we locate the vertical position of the symmetry line in the symmetry map by selecting the vertical line which gives the minimum column-wise symmetry value sum.

In order to account for in-plane rotations of faces, we carry out the same procedure for different projection axes, i.e., by not only summing up symmetry values along the vertical lines but also using rotated lines. The projection axis producing the minimum symmetry sum gives the rotation angle of the face together with the position of the central profile line (See Figure 4.2).

After finding the facial vertical profile contour, the nose tip location is found. For that

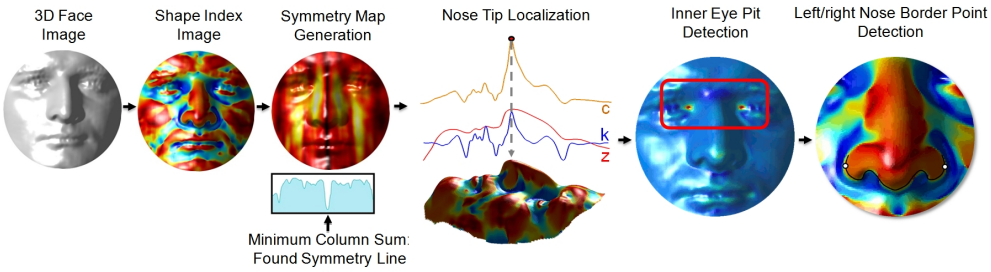


Figure 4.2. Illustration of the automatic landmarking algorithm.

purpose, we use both the depth measurements and the Gaussian curvature (Section 3.1) values along the profile axis: Using a simple heuristic such as selecting the point having the biggest depth value as the nose tip position is not sufficient since in some cases, forehead or mouth may be closer to the camera due to expression or rotation variations. Therefore, we propose to combine Gaussian curvature values with depth measurements. Dome-like shape structures such as the nose tip region produce large Gaussian curvature values, thereby in combination with the depth information, the localization of the nose becomes more reliable. Let $z = (z_1, z_2, \ldots, z_n)$ and $k = (k_1, k_2, \ldots, k_n)$ be the normalized, depth and Gaussian curvature value vectors along the profile line, respectively. We define a function of a combination of z and k as

$$c_i = z_i^2 k_i, i = 1 \ldots n \tag{4.2}$$

and select $\arg\max_i c_i$.

The third step in automatic landmark localization is to find the inner eye pit locations. We observe that these points have cone-like shape structures around the upper nose area. Given the central facial profile and the nose tip position, it is easy to estimate a local search region for eye pits. In Figure 4.2, the search window on a sample face can be seen. Since faces may have in-depth rotations, we use Gaussian curvature values to estimate the locations of eye pits, instead of using depth measurements. In the Gaussian curvature surface, cone-like structures produce values close to zero. Therefore, we search for the local minimum inside the search window and output these locations as the positions of eye pits.

Left and right outermost nose borders can be detected with the use of shape index de-

scriptors efficiently as well. Saddle rut structures such as the nose border regions produce shape index values around 0.375. Therefore, we extract the nose border outline by a contour following approach where the pixels have saddle rut like shapes. Using this approach it is easy to extract the lower nose border contour, as shown in Figure 4.2. Given the nose border contour, we select the rightmost and leftmost pixels along this curve as the rightmost and left-most nose border points, respectively. More formally, let $C = \{(x_1, y_1), (x_2, y_2), \ldots (x_l, y_l)\}$ denote the contour points where $|I_S(x_i, y_i) - 0.375| \leq \delta$, where δ is a small constant. It follows that the (x, y) locations of the left-most and right-most nose border points can be found by $x_{left} = \arg\min_x C$ and $x_{right} = \arg\max_x C$, respectively. y_{left} and y_{right} coordinates are the corresponding indices.

4.1.2. 3D Face Registration

3D registration establishes a one-to-one correspondence between the surface points of two given 3D faces. Human face is a non-rigid surface which deforms in the presence of expressions initiated by muscle movements. The accuracy of rigid registration methods decrease when test scans with expressions are introduced. Region-based approaches try to overcome this difficulty by using smaller regions of faces [24, 25, 55, 58]. In region-based face recognition, a face is represented by a single robust region or it is considered as a composition of facial components. Rigid methods use ICP and rely on positions of landmarks on the face for initialization [24, 55, 58], while non-rigid methods elastically deform the surface to overcome the effect of expressions [25]. Our region-based registration approach provides a simple, fast and robust two-pass alternative: We first employ ICP to register the facial surfaces to a common model, called the Average Face Model (AvFM). This approach was previously used in [5, 78, 79]. The use of AvFM ensures that all gallery faces are in one-to-one correspondence. A second registration phase uses ICP registration to register individual regions to their respective Average Region Models (AvRM), starting from the initialization provided by the first phase. Further details on AvFM-based registration can be found in Section 3.2.3. An example AvFMs generated for the FRGC database is shown in Figure 4.3 together with the landmarks.

For the AvRM based registration, the construction of regional models is necessary. For

this purpose, first of all an AvFM is generated as described in Section 3.2.3. Then, regional masks are created to divide the facial surface into patches that constitute the basic building blocks. The facial patches are constructed by manually labeling corresponding areas on the AvFM. Patch construction on the AvFM is performed only once. The patches are collected into higher level components, namely the AvRMs. These regional models act as index files for the regional registration approach. In this work, we divided the face into a total of 15 patches, and from these patches we constructed seven meaningful regions: nose, left/right eye, forehead, left/right cheek, and mouth-chin. We also constructed a regional model for the area which is considered to be the region least affected by facial expression variations. This region is referred to as the upperface region and covers patches belonging to eye, nose, and forehead areas. The division of the facial surface into patches and the construction of regions from these patches are illustrated in Figure 4.3.

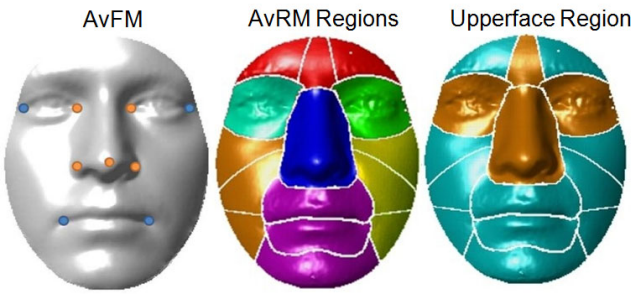


Figure 4.3. The AvFM and its landmark points computed from the FRGC database (leftmost image). Center and rightmost images show seven facial regions and upperface region for the AvRM, respectively.

The dense correspondence obtained between the face and the whole facial model acts as a coarse alignment for the regional approach. The aligned probe face, $P_{reg} = \{\mathbf{p}_1, \ldots, \mathbf{p}_t\}$, has the same number of points with the AvFM in exactly the same order. In AvRM-based registration, a second ICP is performed between the regional model, $M^{(k)} = \{\mathbf{m}_1^{(k)}, \ldots, \mathbf{m}_{t_k}^{(k)}\}$ and the face previously registered to the AvFM, P_{reg}, to construct a local one-to-one correspondence of individual regions. The regions to be registered are considered independently of each other and for each component, different transformation parameters are calculated. The steps of our registration technique are summarized in the upper part of Fig-

ure 4.1.

4.1.3. 3D Features

Point Cloud Features

After the alignment phase, 3D facial surfaces can be compared since they lie on the same coordinate system. A simple method is to use the coordinate differences between two corresponding surfaces. If the registered facial surfaces are resampled at the same (x,y) coordinates, then it suffices to use only the depth (z-coordinates) measurements in the computation of the PC value. More formally, let Φ be the whole facial surface composed of N local regions, ϕ_i, then $\Phi = \cup_{i=1...N}\phi_i$. With the assumption of regular resampling in the point cloud method, each region ϕ_i is represented by a vector of z-depth measurements: $\phi_i = [z_1, z_2, \ldots, z_{M_i}]$ where each region ϕ_i contains M_i z-depth values. The dissimilarity between any two corresponding facial region then can be computed for person A and B as

$$D(\phi_i^A, \phi_i^B) = \frac{|\phi_i^A, \phi_i^B|}{M_i} \tag{4.3}$$

where $|.|$ denotes L_1-norm.

Statistical Point Cloud Features

A useful property of the generic AvRM based registration is that 3D facial features, particularly ϕ_i, are ordered vectors. In order to have a more compact and discriminative feature space, we propose to utilize Fisherfaces (details given in Section 3.3.3) for the point coordinate features. Basically, we form a separate Fisherfaces space for every facial region. Construction of the subspace, i.e., the computation of the transformation matrix, is carried out by using an independent training set. Let Λ_i be the transformation matrix found by region i. Then the regional Fisherface features, γ_i can be found by the projection of ϕ_i: $\gamma_i = \Lambda_i\phi_i$. The dissimilarity between any two facial regions can be computed by the angular cosine

distance measure in the respective subspace as:

$$D(\gamma_i^A, \gamma_i^B) = 1 - \frac{\gamma_i^A \cdot \gamma_i^B}{|\gamma_i^A||\gamma_i^B|} \tag{4.4}$$

4.1.4. Classification: Fusion Techniques

In region based techniques, each region acts as an independent classifier, and recognition results can be fused to obtain an improved overall performance. The fusion techniques can be grouped into three basic categories, namely score level, rank-level, and abstract-level approaches [79]. In this work, we use score and abstract level fusion. In score-level fusion, the similarity measures obtained from different classifiers are combined using basic arithmetic rules. Two score-level methods are considered: sum rule (SUM) and product rule (PROD). Both of these rules operate on normalized distances. For distance normalization, we utilized the min-max normalization method.

In abstract-level fusion, each individual classifier produces a class label. The individual class labels are combined to provide a single label. In this category, committee voting (CV) and modified committee voting (MOD-CV) methods are considered. In CV, 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. In MOD-CV, the approach of committee voting is improved, where for each classifier, a confidence value is estimated together with the class label. When there are ties, the decision is based on the confidence values. The confidence value is based on normalized scores. If $d = [d_1, d_2, \ldots, d_N]$ denotes the sorted dissimilarity values to N gallery samples in ascending order, a second score normalization is performed by

$$d_i' = \frac{(d_i - d_1)}{median(d) - d_1}, i = 2, \ldots, N \tag{4.5}$$

After this 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 about its decision on the rank-1 class.

For further details on confidence estimation, please refer to [79].

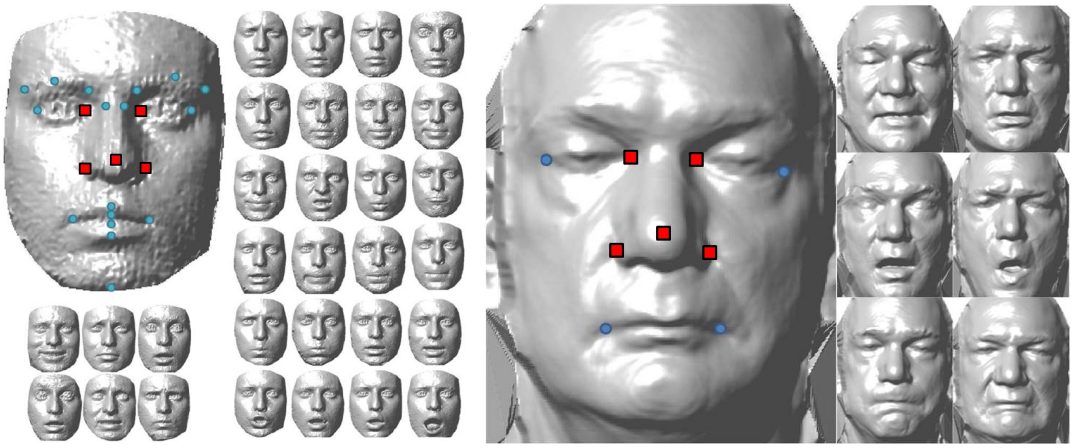
4.2. Experimental Results

The main purpose of this work [26], is to develop a 3D face recognition system that is resistant to expression variations. For this purpose, two 3D face databases containing scans with facial expressions are employed: (i) The Bosphorus database, which has a large variety of expressions; (ii) the FRGC v.2 database, which is the most widely used database in the literature.

The expression subset of the Bosphorus database has a total of 2919 scans, with roughly 34 different expression scans per subject. 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 [80]. Among the 28 AUs, 20 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 is related to common emotions: happiness, surprise, fear, sadness, anger, and disgust. In Figure 4.4a, the manual landmark points present for each scan and the expression variability are illustrated. As the experimental setup, we constructed a gallery set containing one neutral scan for each subject. The remaining scans constitute the probe set. Hence, gallery and probe set sizes are 105 and 2814, respectively.

The expression variations in the FRGC v.2 database are frontal containing a number of facial expression variations such as happiness, sadness, surprise, anger, disgust, cheek puffing. For this database, we manually labeled a set of nine landmark points on each facial surface. In Figure 4.4b, an example subject is shown with manual landmarks available and the expression variability is illustrated.

For the identification scenario, we designed an experimental setup with one image per subject in the gallery set and all the other images in the probe set. The gallery set constitutes



(a) Bosphorus (b) FRGC v.2

Figure 4.4. Sample 3D scans for the (a) Bosphorus and (b) FRGC v.2. Manual landmarks are shown on a neutral face. The red-colored landmarks are used in coarse alignment. For the Bosphorus, emotional expression variations and action units are shown on the bottom left, and on the right, respectively. For the FRGC v.2, expression examples are given.

a total of 466 scans, one scan per each subject. The images contained in the gallery are not restricted to be neutral, they are the first appearing scan of each subject. This experimental protocol is also used in [23, 25, 52] and we have chosen the same setup to allow a direct comparison with the techniques proposed in those studies (summarized in Section 2.5).

4.2.1. Automatic Landmark Localization Performance

Good landmarks are needed for convergence of the registration algorithm. The performance of the automatic landmark localization is thus, crucial. The average Euclidean distances of the automatically labeled landmarks to the corresponding manual landmarks are given in Table 4.1 for the Bosphorus and FRGC v.2 databases. The average Euclidean distance between the eyes is 64mm for the Bosphorus database. Therefore it is seen that automatic landmark localization algorithm has an average error rate of 4% to 6% of the inter eye distance. This accuracy is sufficient for a coarse registration, as will be shown by the identification accuracies in later sections. To evaluate the performance of the proposed automatic landmarking method better, we designed an experiment to observe the variability of manual landmarking subject to the precision of the annotators. The five-point landmark set (the inner eye corners, nose tip, and the nose corners) is labeled by ten different annota-

tors on a subset of the Bosphorus database, consisting of 20 scans. The average Euclidean distances given in the second row of Table 4.1 correspond to the *manual labeling variability* (MLV), which is the average distance of the manually labeled landmarks by the ten different subjects to the original manual landmarks. The results show that the variability of the automatic landmark locations and the variability that can be caused by the annotators are not significantly different. It is also evident that the outer nose corners are located more precisely both automatically and manually.

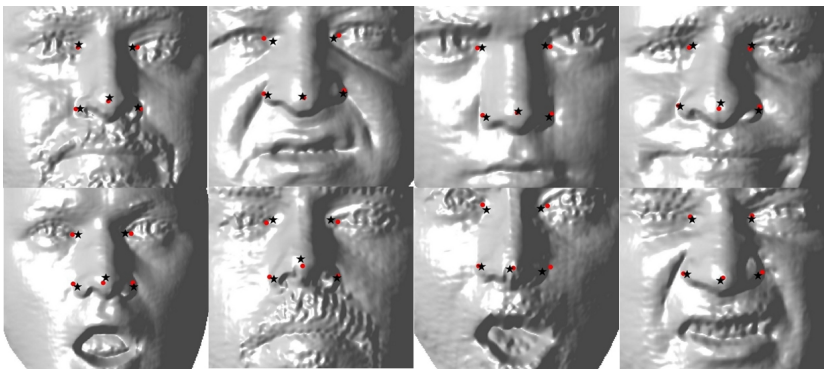
Table 4.1. The average Euclidean distances (mm) between manual and automatically found landmarks.

	Left Inner Eye Corner	Right Inner Eye Corner	Nose Tip	Left Nose Corner	Right Nose Corner
Bosphorus	3.96	3.43	3.05	3.19	3.00
MLV (Bosphorus)	2.70	2.32	2.96	1.68	1.82
FRGC v.2	4.90	5.05	3.26	4.68	4.51

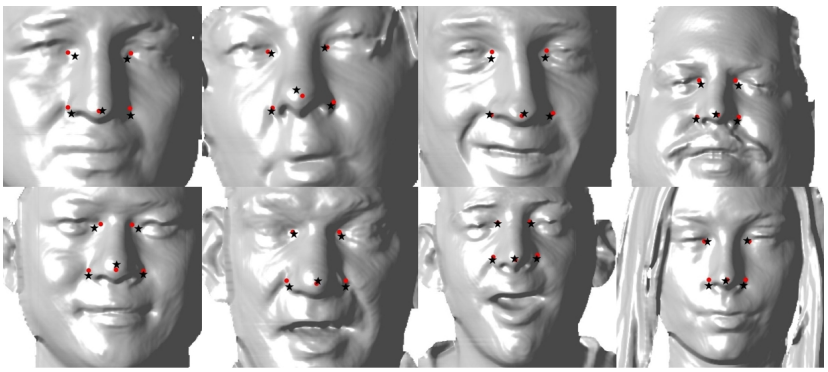
The automatic landmark localization results are illustrated in Figure 4.5a and Figure 4.5b on a sample set of scans with facial expression variations for the Bosphorus and FRGC v.2 databases, respectively. The original manually labeled landmarks are also shown, to permit visual interpretation of the results. It can clearly be seen that the eye and nose corner points can be located efficiently in the presence of expressions, enabling adequate results for the coarse registration phase.

4.2.2. Identification Results

Expression variations give rise to deformations on the facial surface. These deformations cause performance degradations of the registration approaches that treat the faces as rigid and *global* surfaces. To substantiate our assertion, we examined the AvFM-based rigid registration method on both the Bosphorus and the FRGC v.2 databases. For coarse alignment of faces, Procrustes analysis utilizing the five-point landmark set is performed. The coarse alignment is followed by a fine registration step via the ICP algorithm. Subsequent



(a) Bosphorus



(b) FRGC v.2

Figure 4.5. Manually (red dots) and automatically (black stars) located landmarks shown on a sample set of scans for the Bosphorus and the FRGC v.2 databases.

to registration of the faces, the surfaces are considered as point clouds and the Euclidean distances between a probe face and each of the gallery faces are computed. As a classification approach, the nearest neighbor algorithm is utilized to obtain identification results. In Table 4.2, the rank-1 recognition rates obtained via AvFM-based registration are reported on both databases, using manual and automatic landmarks in the coarse alignment phase. The first, second, and third rows are the identification performances for neutral, non-neutral, and for the full probe set respectively. These results support our claim that the rigid registration accuracy decreases in the presence of facial expression variations. For the FRGC v.2 database, which contains a large probe set of neutral and non-neutral scans, the performance degradation due to expression is about 40%. The performance decrease is also quite significant (30%) for the Bosphorus face database. Regarding the effect of manual and automatic landmarking on the identification performance, we see that performance decrease is quite small if landmarks are found automatically. By looking at the whole probe set (neutral + non-neutral), it is observed that rank-1 accuracies decrease by 0.14% and 0.36% for the

FRGC v.2 and Bosphorus databases, respectively when automatic landmarks are used.

Table 4.2. Identification results of the AvFM-based approach for manual and automatic landmarks.

	FRGC v.2 (Gallery Size: 466)			Bosphorus (Gallery Size:105)		
	Probe Size	Manual	Automatic	Probe Size	Manual	Automatic
Neutral Probes	1984	84.07	83.92	193	99.48	100.00
Non-Neutral Probes	1557	48.62	48.49	2621	69.71	69.29
All Probes	3541	68.48	68.34	2814	71.75	71.39

After showing that global ICP-based registration is not sufficient for non-neutral faces, we can now proceed to analyze local AvRM-based registration performances. As explained before, in the AvRM-based alignment, first a global ICP alignment is performed and then average region models are independently registered to a given probe facial surface. In Table 4.3, the independent regional identification results using a single region are given. For comparative reasons, the identification rates obtained using the AvFM-based registration are given in the first row. As these results exhibit, some regions are less affected by facial expressions, such as the nose, eye, and forehead regions. When the combination of these regions is used as a single AvRM, namely the upper-face AvRM, the best regional recognition rates are obtained. The cheek regions and the region containing the mouth and chin are the worst performing areas. This is basically due to the fact that facial expressions deform the mouth greatly and subsequently the cheek regions are affected by the mouth movement. Although their regional deformations are less than the mouth and chin area, cheek regions perform even worse, implicating their low discriminative ability.

An important observation from the results in Table 4.3 is that using only the nose region, it is possible to significantly improve the identification rates, compared to using the whole face with the standard ICP approach (the AvFM method). This finding is also compliant with the other studies that focus on the nasal region. However, we see that incorporating the forehead and eye regions with the nose, by forming a bigger upperface region, it is possi-

Table 4.3. Identification results of the individual regions.

	FRGC v.2		Bosphorus	
	Manual	Automatic	Manual	Automatic
AvFM	68.48	68.34	71.75	71.40
Nose AvRM	85.12	84.98	86.96	86.70
Left Eye AvRM	65.15	65.04	60.23	60.27
Right Eye AvRM	64.90	64.78	62.30	62.26
Forehead AvRM	62.92	62.84	77.36	77.19
Left Cheek AvRM	34.11	33.95	32.76	32.66
Right Cheek AvRM	31.69	31.74	36.28	36.17
Chin-Mouth AvRM	41.51	41.46	36.64	36.57
Upperface AvRM	86.78	86.59	91.22	91.05

ble to improve the accuracy obtained by the nasal region alone. In terms of the landmarking method used in the coarse registration phase, we see that automatically located landmarks only slightly reduce the rank-1 identification accuracy for all the regions.

4.2.3. Fusion of Regional Classifiers

Although some regions are deformed less in the presence of facial expression variations, use of a single region is not sufficient for identification purposes. To improve the recognition results obtained by independent regional classifiers further, we propose to fuse the classification results. We have eight regional classifiers: nose, left/right eye, forehead, left/right cheek, chin-mouth, and upperface classifiers. In Table 4.4, we present the fusion results using sum, product, committee voting and modified committee voting fusion schemes. In addition to the reported fusion schemes in Table 4.4, we have also tried several other fusion mechanisms such as highest confidence and Borda count method. However, they have performed worse than the reported results in Table 4.4.

For the FRGC v.2 database, the best identification accuracies are obtained by fusing individual classifiers with the modified committee voting scheme. If automatically found landmarks are used, MOD-CV achieves 91.16% rank-1 identification rate. This is significantly better than the best individual classifier, namely, the upperface classifier (86.59%, Table 4.3). If we compare the fusion methods, we see that voting schemes perform better

than the arithmetic rules such as sum/product rules for the FRGC v.2 database. However, for the Bosphorus database, this performance improvement is not visible: If voting based fusion mechanisms are used, having a large number of base classifiers leads to a performance improvement [26]. However, here we have a limited number of regional classifiers, and arithmetic fusing schemes works sufficiently well for the Bosphorus database.

Table 4.4. Fusion results of regional classifiers using point cloud features.

	FRGC v.2		Bosphorus	
	Manual	Automatic	Manual	Automatic
SUM	81.11	61.25	86.11	76.15
PROD	88.39	88.11	95.91	95.56
CV	90.39	90.14	93.75	93.57
MOD-CV	91.39	91.16	94.92	94.63

4.2.4. Results of Statistical Features

In this section, we provide the classification results of using statistical point set based features using the Fisherfaces technique. As explained before, ordered z coordinates of the independently registered facial surfaces are used to construct Fisherface subspaces per region. In order to determine the regional transformation matrix, we use separate training sets. For the FRGC v.2 experiments, we use the FRGC v.1 set which includes a total of 943 3D scans. For the Bosphorus face database, we divide the whole database into two parts: 643 scans of 20 subjects are used to construct Fisherface subspaces and the 2265 scans of 85 subjects are used to form an evaluation set (gallery and probe sets) for identification tests. The 20 subjects that are used for the training are different from the ones in the evaluation set. In the Bosphorus evaluation set, there are 85 gallery images (single neutral image per person) and 2180 probe images. The rank-1 identification rates obtained by the product fusion of individual regional Fisherface classifiers are given in Table 4.5. The results are provided in terms of Neutral vs. Neutral and Neutral vs. Non neutral comparisons in order to analyze the behavior of the proposed scheme under expression variations. If we look at the FRGC v.2 results with automatic landmarking, we see that 97.51% rank-1 rate is achieved. Compared to the best performance of fusing point cloud features with the MOD-CV method, we improve

the accuracy from 91.16% (See Table 4.4) to 97.51%. This rank-1 classification rate obtained on the FRGC v.2 database is one of the best reported accuracy in the literature. For the Bosphorus face database, fusion of regional Fisherface classifiers also provides very high identification rates: on the independent evaluation set 99.31% of the probe set is correctly classified. It should be noted that this performance value cannot be directly compared to the results provided in Table 4.4 since the evaluation set is a subset of the whole database used in Table 4.4. A very important observation about our Fisherface based regional approach is that non neutral probes are identified quite accurately compared to neutral probes. This proves that our proposed scheme, with the help of i) regional registration and ii) the statistical subspace analysis is very beneficial and is even insensitive to expression variations. A very practical advantage of the regional Fisherface approach is the compactness of the feature vectors. The results shown in Table 4.5 are obtained by Fisherfaces feature dimensionality of 90 per region. In real-world biometric applications, where the template size and matching speed are important, the use of such compact features is very crucial.

Lastly, in order to further analyze the generalization ability of the Fisherfaces approach, we perform *cross database* training for the FRGC v.2 set. Basically, we train the Fisherfaces subspace with the Bosphorus training set and form the feature vectors for the FRGC v.2 set by using the Fisherfaces space trained with the Bosphorus database. With cross database training, the rank-1 identification rate is 94.55% for the FRGC v.2 database. This result implicates that even with such a challenging scenario of training with a completely different database with different sensor and different composition, it is possible to achieve quite acceptable recognition accuracy.

Table 4.5. Rank-1 classification results of the Fisherfaces based AvRM approach.

	FRGC v.2		Bosphorus (Evaluation Set)	
Gallery vs Probe	Manual	Automatic	Manual	Automatic
Neutral vs Neutral	98.59	98.39	99.47	100.00
Neutral vs Non neutral	97.11	96.40	99.60	99.25
Neutral vs All	97.94	97.51	99.59	99.31

4.3. Conclusion

In this chapter, we present a fully automatic 3D face recognition system which exploits facial surface characteristics to infer the identity of a person in a regional manner. Here, our focus was to design a part-based face recognition system with a special emphasis on expression insensitivity. In order to achieve an accurate identification system under severe expression variations, it is essential to employ an efficient facial surface registration scheme. The main contribution of this work is the *utilization of component based regional registration methodology with the help of a generic face model and generic region models* which has advantages for (i) better registration under local facial surface deformations, (ii) fast search in identification mode, and (iii) the applicability of statistical feature extraction methods for unordered 3D point data. While the regional registration can cope with facial expression variations effectively, registering to an average model brings the ability to use dimensionality reductions techniques such as Fisherfaces. By registering each facial region to a common regional model, we perform *regional* Fisherfaces in a smaller space where the main mode of variation is based on identity. Hence, the Fisherfaces in the regional spaces is able to capture identity variations better.

As the experimental results show, with respect to the registration method utilized, AvRM based regional registration significantly improves the classification rates when compared to AvFM based global registration. By merging the power of regional registration through generic facial region models with the statistical feature extraction methods, the discriminative ability of 3D features can be highly improved. The application of Fisherfaces, as a statistical feature extractor, improves the classification rates of the point set features from 88.11% to 97.51% for the FRGC v.2 database, and from 96.24% to 99.31% for the Bosphorus database (see automatic landmarking results).

Some important conclusions drawn from the work represented in this chapter, gave direction to our work on occlusion handling: Since occlusions will cause the facial surface to change partially, a part-based registration and recognition scheme can be beneficial. However, the herein given region-based registration approach is not directly applicable to occlusion variations: When facial surfaces are occluded, automatic landmarking scheme will fail,

since the symmetry plane cannot be computed. Moreover, the coarse alignment is based on registration to a whole face model, which is not a viable alternative when the probe surface is altered due to occlusions.

Besides the registration problems, there are two other issues to be condisered for occlusion handling: First, the occluded regions should be accurately detected and removed before any comparisons can be made. Second, in order to use the regional statistical features for classification, the incomplete surfaces cannot be directly used. To handle incomplete surfaces, the missing parts can be filled. However, since our aim is to recognize subjects from facial surface information, approximation to fill in surfaces can cause incorrect information to be used as discriminative data. Therefore, if restoration of facial surfaces are found to have negative influence on classification scenarios, then the statistical feature extraction method should be adapted to work on incomplete data.

5. SURFACE REGISTRATION UNDER OCCLUSION

Humans recognize familiar faces even under different poses, with different illumination conditions, with varying facial expressions, and even under occlusion presence. However, automatic face recognition by machines is not a straight forward task: 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 location or pose variations of the acquired subject. In either case, the 3D shapes need to be located in the acquired scene and should be brought into a common coordinate frame before they can be compared to determine the identity of the subject. Therefore, preprocessing steps of face detection and registration are necessary: Face detection is the process of localizing the facial surface in the acquired scene, whereas registration is the alignment procedure of two similar shapes.

Since face is a 3D surface, locating and registering surfaces in the 3D domain is advantageous. Recent studies have shown that in the 3D domain; challenges such as illumination and pose can be better handled. However, dealing with extreme occlusion variations remains a challenging task: When occlusions are present, it is problematic to detect the partially occluded 3D facial surfaces. Furthermore, even when the faces are detected, 3D face registration algorithms fail to provide accurate facial point correspondences due to occluding surface points. The resulting alignment between facial surfaces is usually incorrect, leading to low recognition rates.

In this thesis, we propose an occlusion invariant registration approach, referred to as the adaptive-model based registration, which includes face detection and alignment procedures. For detection, instead of detecting the whole facial surface, only the nose area is considered. This way, facial surfaces even with partial nose occlusions can easily be located in the acquired scene. Afterwards, the non-occluded regions are estimated to decide on a patch-based registration model. Thus, every facial surface is registered using a model that can eliminate the occluded regions from the alignment process.

In this chapter, we first summarize the model-based registration approach, which serves as the baseline method to construct a computationally feasible technique. Next, we give details about the adaptive-model based registration approach, proposed for occlusion invariant alignment of facial surfaces. Finally we give experimental results and conclude this chapter.

5.1. Model-based Registration

Since for a face recognition scenario, a probe face should be compared against all of the images in the gallery set, it is necessary to align a probe face to each of the gallery faces separately. *One-to-all* registration can be obtained by dense alignment techniques such as ICP (given in Section 3.2.2) and some of the previous face recognition studies employ this technique for similarity computations [81–84]. However, this approach is computationally costly, since the number of registrations to be performed equals the number of gallery images. As proposed in [5], model-based registration (given in Section 3.2.3) can be used for computational feasibility, where all the probe and gallery images are registered to an average model, for only once, to obtain full correspondence between all image pairs. Further details about model-based registration, including average face model generation, can be found in Section 3.2.3.

5.1.1. Alignment to the Average Face Model

In this part of our work, we have experimented with registering to an average *face* model, where the whole facial surface is taken into account for alignment. As given in Section 3.2.3, the registration of a facial surface to the average face model consists of two phases, namely the coarse and the fine registration steps. If a set of landmark locations are present for the facial surface to be registered, then Procrustes analysis can be used to align the two surfaces coarsely. If only a single point location, such as the nose tip, is available, then a simple translation to coincide the nose tip to the nose tip of the model can be sufficient. In the fine alignment phase, all of the surface points are taken into account, seeking for a better alignment. In fine alignment, we performed ICP.

5.1.2. Alignment to the Average Nose Model

Since occlusions present over the facial surface alter the surface geometry, aligning to an average face model will not be sufficiently accurate: The surface points corresponding to occluding object will cause an incorrect alignment in between the surfaces. Moreover, a transformation computed to align the incorrectly paired surface points will cause the actual facial parts to become distant from each other. If we can assume that the nose area is visible for the occluded input face, then using an average *nose* model for alignment will be a good alternative. The nose-based registration was proposed in [4]. The average nose model is constructed from the average face model by manually cropping the nose area. Since registration based on average nose model will be handled by the ICP algorithm, coarse alignment is once again necessary. Two different coarse alignment approaches can be followed due to the available landmark locations. If a single landmark location such as the nose tip is present and the faces are known not to have great pose variations, then translation according to the single landmark can be provide a sufficient coarse alignment. If there are a multiple of landmark points, then Procrustes Analysis can be employed to coarsely align the input face to the average nose model using these fiducial points. In Figure 5.7 (first image), a set of nose landmarks are colored with green, that can be used to align a surface to a nose model. After the input face is coarsely superimposed over the model using either of these methods, ICP is utilized to finely register the input face to the nose model.

5.2. Proposed System: Adaptive Model-based Registration

Although registration via an average nose model appears as a viable alternative, partial occlusions over the nose area can disrupt the overall alignment process. Furthermore, the small size of the nose model can cause inaccurate registration. Therefore, it would be beneficial to include other facial parts into the alignment phase: Motivated from the model-based registration idea, in this thesis, we propose an occlusion invariant 3D facial registration method. Instead of using a single holistic facial model or a small-sized average nose model to register any input face, we propose to coarsely detect the non-occluded parts of the input surface and choose an alignment model accordingly. In this way, the occluded parts will be eliminated from the registration process, allowing a better surface alignment and correspon-

dence establishment.

This registration approach was presented as a conference paper [6]: We handle registration by an adaptive model-based approach which assumes partial visibility of the nose. Prior to registration, nose detection is employed and is used to locate eye and mouth patches. Detected patches are then evaluated for their validity. The corresponding valid (occlusion-free) patches of the average face model are selected to construct an adaptive face model. ICP alignment with the adaptive model is able to discard the occluded surface points for point matching.

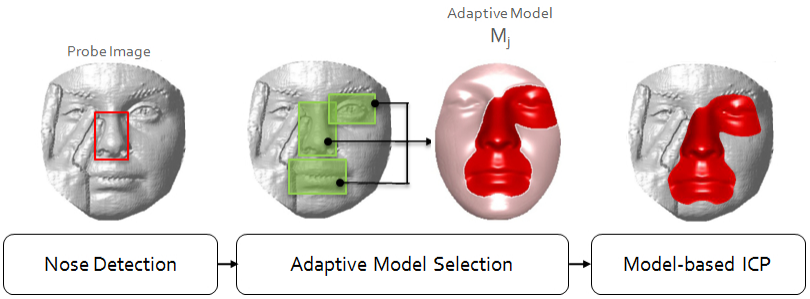


Figure 5.1. Diagram of the proposed registration method.

The proposed face registration system has three phases: (i) nose detection via curvature maps, providing an initialization for fine registration; (ii) facial patch localization and validation to form an adaptive face model; (iii) model based fine registration via ICP. The overall diagram of the system is given in Figure 5.1. Details about each phase are given in the following subsections.

5.2.1. Nose Detection

As stated before, like many of the other iterative approaches, performance of ICP relies greatly on the initial conditions. Therefore, an initial alignment should be provided, which will be improved in further iterations. For the surface initialization, most of the 3D face recognition systems depend on accurate localization of facial landmark points [76, 85, 86]. However, when occlusions are present over the facial surface, localization of fiducial points fails. Since facial occlusions may occur over the nose area, our nose detector assumes partial visibility of the complete nose structure with the help of local nasal surface sub-patches (See

Section 5.2.2 for further details).

The nose detection algorithm [4] utilizes surface curvature information, which provides an advantage due to its rotation and translation invariance. Two curvature maps are computed for a given surface, namely the shape index map and the curvedness map (given in Section 3.1). These measures of the local surface, separates components that are dependent or independent of scale. Scale-independent components, such as shape index, provide the distinction between spherical and cylindrical surfaces. On the other hand, the scale-dependent components, such as curvedness, give the magnitude of the curvature. The shape index map $\boldsymbol{SI}$ takes values in $[0, 1]$ and provides a smooth transition between concave ($0 < SI(i) < 0.5$) and convex ($0.5 < SI(i) < 1$) shapes. As the scale-dependent counterpart of shape index, curvedness measures the rate of curvature at each point. The nose detector first constructs shape index and curvedness maps. Since nose is a convex structure, the $\boldsymbol{SI}$ map is thresholded (by 0.5) to eliminate concave regions. The convex $\boldsymbol{SI}$ map, denoted as $\boldsymbol{SI}_{cx}$, is defined as

$$SI_{cx}(i) = \begin{cases} 0 & \text{if } SI(i) < 0.5 \\ SI(i) & \text{otherwise.} \end{cases} \tag{5.1}$$

After concave regions are eliminated, $\boldsymbol{SI}_{cx}$ is weighted with curvedness [87] to integrate scale-dependent and scale-independent components:

$$WSI(i) = SI_{cx}(i) * C(i) \tag{5.2}$$

Here, $\boldsymbol{WSI}$ denotes the curvedness-weighted convex shape index. In Figure 5.2, the maps constructed at each step are given for an example facial image. The maps illustrated are: $\boldsymbol{SI}$, $\boldsymbol{SI}_{cx}$, $\boldsymbol{C}$, and $\boldsymbol{WSI}$.

As illustrated in Figure 5.2, the nose region appears as a distinct fork-shaped structure in the $\boldsymbol{WSI}$ map. To locate the nose area, template matching is employed. For the construction of the nose template, the average nose model is obtained by manually cropping the face model. Then, the $\boldsymbol{WSI}$ map for the nose model is constructed to serve as the nose template.



(a) (b) (c) (d) (e)

Figure 5.2. Curvature maps utilized for nose detection are illustrated on an example image: (a) depth image, (b) shape index, (c) convex shape index, (d) curvedness, and (e) weighted convex shape index.

Given a test image, template matching is performed by normalized cross-correlation, and the region which mostly resembles the nose structure is located.

5.2.2. Patch Selection and Adaptive Registration

In [4], only local nose regions were considered for occlusion invariant registration. After nose detection, the probe surface was registered using an average nose-region model. However, this approach has shortcomings. Relying solely on the nasal region for the overall face alignment might be suboptimal; especially if the borders of the nose region are affected by occlusions. Additionally, any problems on the nose surface structure, either due to acquisition errors or uncommon nose shapes, may lead to inaccurate facial surface registration. Here, we propose to utilize an adaptive face model. The idea is to adaptively detect and include other non-occluded facial regions such as eyes and mouth automatically to form an adaptive face model for registration. For instance, if the left side of a face is occluded by a hand (See Figure 5.1), our adaptive face model will automatically be constructed using the non-occluded regions such as right eye, mouth and nose. Then, combined regional models are used for alignment estimation instead of using only the nasal region.

In Figure 5.3, an overall diagram for patch validation and model selection procedure is visualized. Using the detected location of the nose area as a start point, we find other patch locations. In Figure 5.4, the patch division scheme is shown on the first image. However, not all of the facial patches are beneficial for registration. Therefore, we use a subset of these patches. The patches we use are: nose, left/right eye, and mouth. We also have sub-patches such as left/right nose halves, upper/lower nose halves. Hierarchical division of

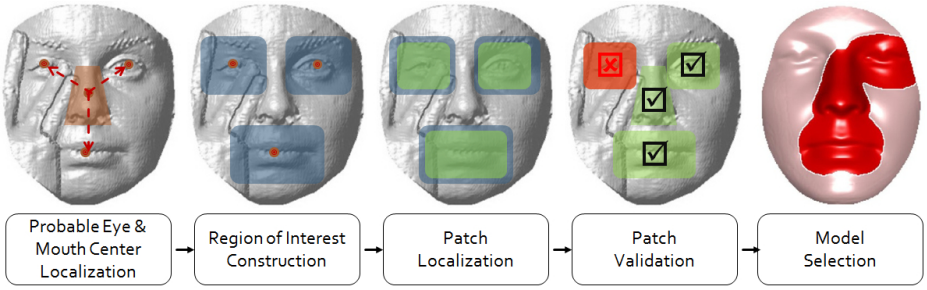


Figure 5.3. A diagram summarizing the patch validation and model selection is given: Probable eye and mouth centers are localized and validated using detected nose area and the predefined region of interest. An adaptive model is selected according to the valid parts.

patches into sub-patches enables us to discard regions where occlusion artifacts are present. To construct average patch models, for each patch, an average patch model is constructed by cropping the average face model. From each model, the $\boldsymbol{WSI}$ map is computed to define the patch template. Using these templates, corresponding patch regions on a given face are detected via template matching based on normalized cross correlation. To limit the search space for the localization of each patch, we compute the probable patch center of a probe face using the relative displacements vectors between patch centers of the average face model. Additionally, a predefined bounding box around each patch center is utilized. Due to occlusions over the face, some patches will not be visible and cannot be located correctly. Therefore, in order to determine the validity of each patch, thresholding is applied on template matching scores. The thresholds used for patch validity are calculated from patches of a separate non-occluded neutral database, namely the neutral subset of the FRGC v.2 [3]. The probe patches that have dissimilarity scores below the threshold define the valid parts. Here, the patch localization and validation steps are not used to detect patches of the probe face to be used in registration. The validity information of patches are only used for the model selection: The respective valid patches are selected from the average face model to constitute the adaptive patch-based model for the respective probe face. In Figure 5.4, the 17 adaptive models utilized in the registration process are shown (the first image was included to show patch division scheme). After adaptive model construction, the whole probe surface is aligned to the adaptive model via ICP, where ICP estimates the alignment parameters using only the non-occluded regions. Hence, the overall registration approach becomes insensitive to occlusions.



Figure 5.4. Facial patches and the adaptive models utilized in registration are given. The first image shows the division scheme utilized for patch construction. To construct the adaptive models, combination of nose, eye, and mouth patches are considered.

5.3. Experimental Results

In our experiments, we have used three face databases: (i) The FRGC v.2 [3], including non-occluded acquisitions; (ii) the Bosphorus databse, including realistic occlusions; and (iii) the UMB-DB 3D database [16], including challenging occlusion variations. In the experiments summarized in this chapter, the FRGC v.2 is used for the construction of the average face and patch models, and for the determination of threshold values used for validity check over template matching scores. We have used the neutral subset consisting of 2365 scans. To evaluate the performance of the registration methods, we have utilized the other two databases including occlusion scans. In our experiments, we have employed neutral gallery sets, occlusion probe sets, and neutral probe sets. The Bosphorus database includes 105 scans, hence there are 105 neutral scans in the gallery. The neutral and occluded probe sets include 194 and 381 scans, respectively. The UMB-DB includes 142 neutrals scans, one for each subject, in the gallery set. The neutral and occluded probe sets are formed from 299 and 590 images, respectively.

5.3.1. Nose Detection Accuracy

The automatic nose detection results are inspected on all three databases. First, the performance of the nose detector is evaluated on the whole FRGC v.2 database. When inspected visually, the nose detector shows 100% accuracy both on the neutral subset (2365 scans) and the non-neutral subset (1642 scans with expression variations). For quantitative evaluation of the detection performance on the other databases, the ground truth nose landmarks are employed to estimate the nose area centers. Then, the distances between the estimated and automatically located nose area centers are thresholded. The detections that are distant from the ground truth centers within the predefined threshold value are counted as correct. The threshold value is set empirically on FRGC v.2: Using the neutral subset of the FRGC v.2 database, the distances between the estimated and automatically located nose area centers are computed. The maximum distance (after trimming the outlier distances) is set as the threshold value ($11.5mm$). The automatic nose detection results indicate that the nose areas for non-occluded scans can be successfully detected: For the Bosphorus neutrals (299 scans) and for the UMB-DB neutrals (441 scans), 100% nose detection accuracies are obtained. The nose detection performance on the occlusion subset of the Bosphorus database is 98.69%. These results are verified by visually inspecting the detected noses on the respective subsets. For the UMB-DB database, the occlusions often cover the nose area partially. Therefore, manually labeled landmarks are incomplete, preventing a similar quantitative evaluation on this data set. Although the nose area is not fully visible for the UMB-DB occlusion subset, the nose localization performance, obtained by visual inspection, is still quite high: 93.90%. The performance of the nose detector is similar to the face detection performance (93.7%) reported on the UMB-DB database in [16]. When the erroneous detections are inspected, it is seen that the nose area is highly occluded for those scans. The nose detection results for the Bosphorus and the UMB-DB databases are summarized in Table 5.1, where the detection performance even for the non-occluded non-neutral scans are included. Furthermore, results for occluded scans are analyzed according to different types of occlusions, indicating the level of challenge for the two databases: As the detection results show, UMB-DB database include more challenging occlusion, where the nose area is partially occluded especially for the scarf, hair, and hand occlusions. In Figure 5.5, some correct and incorrect nose detection examples from the UMB-DB are given for challenging occlusions.

Table 5.1. Nose detection performances on the Bosphorus and UMB-DB databases.

	Bosphorus		UMB-DB	
Acquisition Type	Sample Count	Detected Noses (Detection Rates)	Sample Count	Detected Noses (Detection Rates)
Neutral (Gallery)	105	105 (100.00%)	142	142 (100.00%)
Neutral (Probe)	194	194 (100.00%)	299	299 (100.00%)
Non-neutral	2620	2620 (100.00%)	442	442 (100.00%)
Occlusion	381	376 (98.69%)	590	554 (93.90%)
Occlusion Type				
Scarf	N/A	N/A	151	140 (92.72%)
Glasses	104	104 (100.00%)	75	74 (98.67%)
Hair	67	64 (95.52%)	33	27 (81.82%)
Hand	210	208 (99.05%)	165	152 (92.12%)
Hat	N/A	N/A	183	181 (98.91%)
Other	N/A	N/A	38	33 (86.84%)

5.3.2. Patch Validation and Selection Accuracy

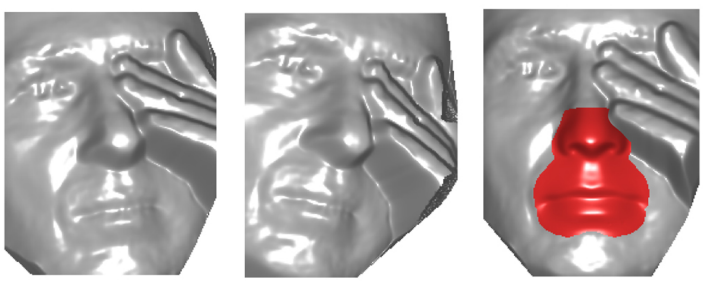
After the nose detection phase, the patches of a probe face are estimated and checked for validity and corresponding models are constructed adaptively. The thresholds used for patch validation are determined from the template matching scores of the FRGC v.2 neutral subset: For a specific patch, the scores are sorted and the smallest 10% of them are discarded and the smallest score of the remaining 90% is set as the threshold for that patch. The thresholds are used to set patch validity flags of the Bosphorus and the UMB-DB scans. When the model selection results are analyzed, it is seen that for the Bosphorus occlusions, 54 out of 381, and for the UMB-DB occlusions, 77 out of 590 scans, the model selection is erroneous: Some patches appear as invalid due to their template matching scores, even though they are non-occluded and their patch localization is correctly handled. Therefore, erroneous model selection is often caused by choosing a smaller model including less patches than available. Most of the errors for UMB-DB (40 out of 77) are caused by the prior nose detection failures. Note that, even when patch selection is erroneous, faces may be registered well enough to be recognized.



Figure 5.5. Correct and incorrect nose detections for the UMB-DB database are given in the first and second rows, respectively.

5.3.3. Registration Accuracy

To visualize the effect of using an adaptive model for alignment when occlusions are present, an example face is given in Figure5.6. Here, the facial surface registered with face, nose, and adaptive model are given in (a), (b), and (c), respectively. As this example illustrates, when the facial surface is partially occluded, registering to a holistic face model is problematic. Employing a nose model is expected to be beneficial if the nose region is not occluded. However, even when a very small portion of the region is occluded, the registration process will be affected greatly (as shown in Figure 5.6b). In such a case, including other non-occluded parts into the model is beneficial, where a greater part of the available surface is utilized.



(a) (b) (c)

Figure 5.6. An occluded face registered with (a) the face model, (b) the nose model, and (c) the adaptive model. In (c), automatically selected adaptive model is shown in red.

To evaluate the registration performance quantitatively, a baseline recognition exper-

iment is performed using depth information: Using the ground truth occlusion masks, the occluding parts on the registered images are discarded. The occlusion mask is applied both to the probe and to the gallery images. It should be noted that the depth-based identification performances reported here, with manually removed occlusions are provided to indicate the relative standing of the registration approaches.

To formally present the depth-based classifier, let the facial surface be represented by a vector of depth measurements: $\boldsymbol{x} = [z_1, z_2, \ldots, z_d]$, where each surface $\boldsymbol{x}$ contains d valid depth values obtained after regular resampling and occlusion masking. The dissimilarity between any two corresponding facial regions can be computed as:

$$D(\boldsymbol{x}^{(P)}, \boldsymbol{x}^{(G_k)}) = \frac{|\boldsymbol{x}^{(P)} - \boldsymbol{x}^{(G_k)}|}{d} \tag{5.3}$$

where P is a probe image and G_k is the kth gallery face, and $|.|$ denotes L_1-norm. For identification, a 1-NN classifier is employed on the masked images. Since in the previous registration phase, a specialized model is selected for each probe face, the adaptive approach should be imposed in the classification stage. Therefore, when the dissimilarities are computed, the probe face is compared against the gallery images registered using the corresponding model.

The depth-based identification experiment is conducted with three different registration approaches: (i) global face model-based ICP, as a baseline approach; (ii) nose model-based ICP, which was previously used in [4]; and (iii) the model-based ICP, where the model is selected adaptively, as initially proposed in [6]. In Table 5.2, recognition rates for the Bosphorus and the UMB-DB databases are given.

When the identification results in Table 5.2 are compared, it is clear that using a bigger model is beneficial for the non-occluded scans: For the neutral subsets, best performances are obtained when the whole face model is utilized. Nevertheless, the adaptive model-based registration has comparable results with the facial model, even though the adaptively selected model has at least 47.7% fewer surface points. This shows that the considered patch regions (eyes, nose, and mouth) provide sufficient information for registration. When the results on the occluded subsets are compared, it is clear that the face model-based registration is not

Table 5.2. Identification performances on the Bosphorus and the UMB-DB database to indicate the relative standing of the registration approaches.

Acquisition	Bosphorus			UMB-DB		
Type	*Face Model*	*Nose Model*	*Adaptive Model*	*Face Model*	*Nose Model*	*Adaptive Model*
Neutral (Probe)	100.00	97.14	100.00	98.66	85.28	97.32
Occlusion	60.63	79.00	83.99	47.29	46.27	65.25
Occlusion Type						
Scarf	N/A	N/A	N/A	19.21	27.15	43.05
Glasses	97.12	83.65	87.50	88.00	60.00	84.00
Hair	76.12	77.61	82.09	57.58	36.36	63.64
Hand	37.62	77.14	82.86	22.42	29.70	56.97
Hat	N/A	N/A	N/A	73.22	72.13	84.15
Other	N/A	N/A	N/A	31.58	36.84	55.26

applicable to occluded faces, and the advantage of the adaptive model over the nose model is clearly visible: For the Bosphorus database, the improvement is from 79.00% to 83.99%; and for the UMB-DB, the results are significantly improved from 46.27% to 65.25% with the baseline depth-based classifier. The nose model-based registration fails on the UMB-DB, since in most of the occlusions in this data set, the nose area is partially covered. However, for the adaptive approach, the valid patches are used instead of using a single nose patch, and the identification rate is improved. On the other hand, in the Bosphorus database, occlusions over the nose area are small, yielding acceptable results even with the nose-model-based registration. It should be noted that this registration method assumes partial visibility of the nose area, since the initial alignment is based on nose detection. Nevertheless, the experimental results show that even the samples with over 50% nasal area occlusions are aligned. This is obtained by incorporating validity together with eye and mouth patches. Furthermore, analysis of performances for different occlusion types are included in Table 5.2. In most of the scarf occlusions, the lower half of the face including the nose area is occluded. Therefore using a face or a nose model cannot provide acceptable registration. However, for the adaptive approach, the valid eye patches are used and the identification rate is improved. In the hair, hand, and hat occlusions, the adaptive model is always better than face and nose model registrations. In comparison, the nose model covers a much smaller area, and is less prone to occlusions. However, even a small portion of an occlusion appearing in the nasal area will affect the final registration significantly. When valid eye and mouth regions are included in

the model, alignment disruptions will be corrected. For the eyeglasses case, the registration scheme depending on a face model is slightly better than the adaptive method since glasses can sometimes invalidate the eye regions.

5.3.4. Evaluation of the Initial Alignment Accuracy

In the previous experiments, the initialization necessary for the convergence of ICP algorithm is obtained by translating the probe face to the average model using only the nose area center locations. Although using multiple of landmarks with Procrustes algorithm is usually preferred for initialization of facial surfaces, presence of occlusions complicates the automatic landmark localization task. Using center locations of the detected nose area provides the necessary coarse alignment information, since the acquired probe faces often has limited pose variations and these rotational transformations can be easily handled in the fine registration step. To evaluate the performance of initialization, additional experiments are constructed on the Bosphorus database, where facial surfaces are initialized using ground truth landmark points: Procrustes Analysis [67] (given in Section 3.2.1) of five manually labeled fiducial points around the nose area is employed for initial alignment of faces. The considered landmark points are the nose tip, inner eye corners, and nose corners, which are colored in orange in Figure 3.5. It should be noted here, that the original ground truth landmark points are incomplete due to occlusions, where at most two of the landmarks are missing. To be able to evaluate the initialization approach, the incomplete landmarks were estimated using the partial Gappy PCA, which was given in Section 3.3.2: Using the manual landmark locations of the FRGC v.2 neutral subset, a PCA subspace was trained. Since the missing landmarks of the Bosphorus landmarks are known, the Gappy PCA algorithm can be applied to project the incomplete landmark sets to the learned subspace. Then, back projecting to the original space, we will obtain a completed version of the landmark set. As done in partial Gappy PCA, we have used the estimated locations of the missing landmarks to complete the originally incomplete landmark sets. In Figure 5.7, the first image illustrates the average face model together with the five landmark points. The last four images visualize the results of missing landmark estimation: the ground truth landmarks are visualized with green labels, whereas the estimated ones are shown in red.

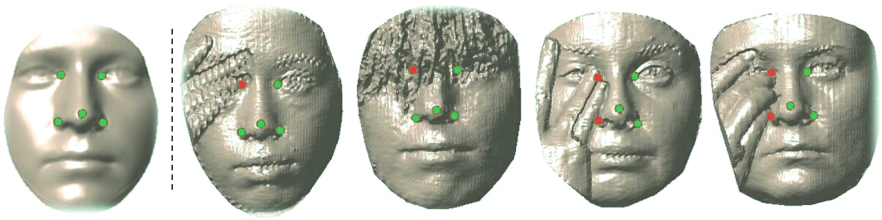


Figure 5.7. Manual landmark points: First image shows the average face model with five landmark points. The last four images are landmark estimation examples (ground truth and estimated points are shown with green and red labels, respectively).

For the performance evaluation of the proposed partial Gappy PCA-based landmark estimator, we have constructed landmark estimation experiments on the Bosphorus gallery set, which consists of 299 neutral scans with complete manual landmarks[1] . The gallery set is divided into two random groups, where 250 scans constitute the training set to train the PCA space and the remaining 49 scans form the test set. For each scan of the test set, we randomly selected a maximum number of two landmark points as missing and estimated them using partial Gappy PCA. Then, the Euclidean error between the estimated and the original landmarks are computed in 3D. For performance evaluation, four experimental setups are considered: Either one or two landmarks can be missing, and the nose tip point can appear in the missing set or not. Each of the four experiments are performed in 10 folds, where for each fold, the gallery is separated into training and test sets randomly. In Table 5.3, the mean Euclidean error values are provided. The ratio of the mean Euclidean distance to the average interocular distance (71.88mm) is also provided in the last column. As these results indicate, the missing landmarks can be estimated with sufficient accuracy. If the nose tip is visible, missing landmarks are estimated more accurately. Furthermore, the estimation performance is higher if fewer landmarks are missing.

In Table 5.4, the depth-based global classification experiments are conducted on the Bosphorus database, where three different models are considered for ICP. Prior to ICP, the initialization is handled by either manual landmark points or automatically detected nose area centers. The results are given both for neutral (second column) and occluded scans (third column). As these results indicate, the performance differences are not significant.

[1] It should be noted that we use Bosphorus database here only for performance evaluation. In our actual system where we perform recognition experiments, FRGC v.2 database is used to learn PCA model for landmark estimation

Table 5.3. Partial Gappy PCA-based landmark estimation performance.

Missing Landmark Count	Nose Tip Missing or Not	Average Euclidean Error (mm)	Error Ratio
1	Visible	5.56	7.73%
1	Missing	6.48	9.01%
2	Visible	6.85	9.52%
2	Missing	7.24	10.07%

Therefore we can conclude that using automatically detected nose area centers provides a sufficient initialization prior to fine registration by ICP. For this experiment, we have only considered the Bosphorus database, since for all of the scans at least three of the considered five landmarks were visible.

Table 5.4. Identification performances with manual and automatic initialization.

Test Set	ICP Model	Recognition Rates (%)	
		Manual Initialization	Automatic Initialization
Neutral	Face	99.48	100.00
	Nose	97.42	97.94
	Adaptive	99.48	100.00
Occluded	Face	61.42	60.63
	Nose	79.53	79.00
	Adaptive	83.99	83.99

It should be stressed that depth-based identification performances with manually removed occlusions are only provided to indicate the relative standing of the registration approaches. A recognition approach based on a more advanced representation method is expected to give better recognition performance.

5.4. Conclusion

In this chapter, we have summarized the proposed 3D face registration approach which is robust to occlusions: For the experiments, we have used the Bosphorus and the more challenging UMB-DB databases. Our experiments show that the adaptive model based registra-

tion is beneficial for occluded faces. The detection part, which is handled by nose detection can detect nose area with 100% accuracy for non-occluded faces, whereas for the occluded scans, the performance of the nose detector is still very high: 98.69% and 93.90% for the Bosphorus and the UMB-DB databases, respectively. We have conducted a simple identification experiment, where the depth-based classifier is employed over the occlusion-removed surfaces. We have shown that, under extreme occlusions, face and nose model-based registrations fail. The proposed scheme, on the other hand, is able to cope with occlusions: The depth-based classifier on occlusion-removed faces shows an improvement: From 83.65%)nose model) to 87.50% (adaptive model) for the simpler Bosphorus database, and from 46.27% to 65.25% for the more challenging UMB-DB.

6. OCCLUSION HANDLING

Occlusions covering the facial surface alter the 3D surface information and degrade the traditional 3D face recognition performance. Even if we assume that face detection and registration approaches detailed in the previous chapter provide promising results and yield accurately aligned faces, occluded regions disrupt the process of comparison. Therefore, it is important to handle the occluded surface regions.

In this thesis, the occlusion handling is done after the surfaces are registered. The proposed registration approach automatically discards the occluded regions from finding the pixel-pair correspondences and computing the necessary transformation. However, occluded pixels are not detected or handled in the registration step. Before the faces can be classified, it is vital to detect the occluded parts accurately. The process of occlusion detection can also be thought of as a binary segmentation problem, where the surface pixels are labeled as either face or occlusion. After the occlusions are detected, they should be handled to perform classification: The occluded parts can either be removed, leaving an incomplete surface, or they can be restored, yielding a completed facial image. In this thesis we have experimented with both occlusion removal and restoration. In this section, we first outline the proposed occlusion detection techniques. Then, we briefly explain the restoration strategy used. In the experimental results section, we summarize and compare the results obtained by occlusion removal and restoration.

6.1. Occlusion Detection

In this thesis, we have implemented three different occlusion detection methods: (i) The basic occlusion detection method that is based on the difference from the generic face model (which was previously used in [4]); (ii) the probabilistic occlusion detection method that is based on training pixelwise Gaussian Mixture Models (GMMs); and (iii) the graph cut method, that incorporates boundary cues into regional cues for better detection performance. In the baseline approach, the difference between a generic face model and the test face is computed and a predefined threshold value is used discard pixels that are distant from the

average. The second occlusion detector is similar to the baseline technique, where each pixel is checked for validity as a facial surface point. However, instead of checking the difference to an average and using a single threshold for each facial pixel, we propose to model the facial surface using Gaussian Mixture Models (GMMs): For each pixel, we learn a separate GMM, which is later used to evaluate the fitness of a test pixel. The third detection technique incorporates neighboring pixel relations, when modeling faces: The face is represented as a graph, where node weights are set using fitness to pixelwise and neighboring pixel-pairwise relations. Then, the occlusion detection problem is solved using graph cut techniques. Further details about these two occlusion detectors are given in Section 6.1.2 and Section 6.1.3.

The proposed occlusion detection methods assume that the facial surfaces are registered and regularly resampled to give depth images with exactly the same dimensions. In our experiments, we have employed the registration approach proposed in [6] to align 3D faces with occlusion variations, which is outlined in Chapter 5. In this section, the occlusion detection strategies are outlined. The experimental results, evaluating the detection techniques, are summarized in Section 6.3.

6.1.1. Baseline Occlusion Detector: Difference from the Average Face Model

For occlusion detection, the most straightforward approach is to analyze the differences between a mean face template and the input face, which has also been implemented by Colombo *et al.* in [11] and Alyuz *et al.* in [12]: If there is an exterior object appearing as a part of the facial surface, the difference for this specific area will be more evident. Therefore occlusion detection can be handled by thresholding the difference map obtained by computing the absolute difference between face template and the input face. If we denote the input image and the average face model used in comparison by $\mathbf{I}$ and $\mathbf{M}_{av}$, respectively, the difference map $\mathbf{D}$ is the absolute difference:

$$\mathbf{D} = |\mathbf{I} - \mathbf{M}_{av}| \tag{6.1}$$

On the difference map, the absolute difference value at each pixel i is compared with a predefined threshold value, T:

$$m_i = \begin{cases} 1 & \text{if } D_i > T \\ 0 & \text{otherwise} \end{cases} \tag{6.2}$$

The occlusion mask, $\mathbf{m}$ is then post-processed by morphological dilation and connected component analysis operations. Throughout this thesis, we will refer to this occlusion detector as the *baseline* approach.

6.1.2. Statistical Facial Modeling via Pixelwise GMMs

In the baseline occlusion detection approach, the difference between an image and the average face model is used to decide whether each pixel is from the facial surface or the occluding object. Furthermore, for each distinct pixel a single threshold value is used. However, the depth variation at each surface point is different and sometimes the variation is too large to represent with a single average value. Therefore, we propose to use pixelwise GMMs to model each surface point. From a set of non-occluded training images, we train pixelwise GMMs, where a probability density function is obtained for each depth image pixel separately. Utilizing 2D pixelwise GMMs were previously proposed for background subtraction in video image sequences [7], where background was modeled with pixelwise GMMs. Here, we employ GMMs to detect occlusions, since we can model our 'background', which is actually the face.

The proposed occlusion detection method involves the decision whether a pixel belongs to the facial surface or to the surface of an occluding object. If z_i denotes a pixel of the depth image, the decision can be made by evaluating the following ratio:

$$R = \frac{p(\mathcal{S}_F|z_i)}{p(\mathcal{S}_O|z_i)} = \frac{p(z_i|\mathcal{S}_F)p(\mathcal{S}_F)}{p(z_i|\mathcal{S}_O)p(\mathcal{S}_O)} \tag{6.3}$$

Here, $\mathcal{S}_F$ and $\mathcal{S}_O$ denote the facial surface and the surface of the occluding object, respectively. Since we do not have any prior information about the occluding object or about

the location and amount of the occlusion, the probabilities of $\mathcal{S}_F$ and $\mathcal{S}_O$ are set equal: $p(\mathcal{S}_F) = p(\mathcal{S}_O)$. Therefore, the decision ratio can be simplified as follows:

$$R = \frac{p(z_i|\mathcal{S}_F)}{p(z_i|\mathcal{S}_O)} \tag{6.4}$$

Furthermore, we do not have any prior information about the surface of the occluding object or the location and amount of the occlusion. Hence, uniform distribution for the occluding surface appearance can be assumed: $p(z_i|\mathcal{S}_O) = c_O$. The decision simplifies to:

$$p(z_i|\mathcal{S}_F) > th \tag{6.5}$$

where $th = R * c_O$ denotes a threshold value. Hence, pixels with a likelihood lower than the threshold are labeled as occlusion.

The facial surface model is obtained separately for each pixel, where a set of registered non-occluded training images are used to estimate pixelwise GMMs. The pixelwise facial surface model can be denoted as:

$$p(z|\mathcal{S}_F) = \sum_{k=1}^{K} \pi_k N(z|\mu_k, \Sigma_k) \tag{6.6}$$

where $N(z|\mu_k, \Sigma_k)$ denotes a mixture component with mean m_k and covariance matrix Σ_k. For simplicity, we assume diagonal covariance matrices: $\Sigma_k = \sigma_k^2 I$. Furthermore, we estimate facial surface model with a predefined number of mixtures ($K = 3$) and use the same number of components for each pixel. Therefore, the models can be estimated by the Expectation Maximization (EM) algorithm [88].

6.1.3. Occlusion Segmentation via Graph Cut

In this section, technical details about the graph-cut technique [89] utilized for occlusion detection are given. We consider the occlusion detection as a binary image segmentation problem, where "face" and "occlusion" pixels form two distinct sets of surface pixels. Start-

ing from an initial labeling of pixels as face or occlusion, we solve an energy minimization problem to converge to the final segmentation. The surface energy is defined using two main cues about the facial surface: (i) Regional cues, where each facial pixel is modeled; and (ii) boundary cues, where the relationship for each neighboring facial pixel pair is modeled. These models will be jointly employed to detect regions not resembling the general facial surface. Algorithmic details about the graph cut segmentation and the modeling of boundary and regional cues are given in the following subsections.

Graph Cut Method:

The pre-registered and regularly resampled depth images can be directly used to construct a graph representing the surface relations. A graph representation includes two terms, the vertices (nodes) and the edges:

$$\mathcal{G} = \langle \mathcal{V}, \mathcal{E} \rangle \tag{6.7}$$

Here, $\mathcal{V}$ and $\mathcal{E}$ represent the nodes and the edges, respectively. When defining the facial surface as a graph, each of the depth pixels corresponds to a node in the graph. Additionally, we have two terminal nodes, namely the source and the sink nodes, represented as s and t nodes, respectively. The face or occlusion pixels will be connected to one of these terminals: Source represents the object (occlusion) terminal, whereas sink represents the background (face) terminal. After convergence, all face pixels will be connected only to the sink, and all the occlusion pixels will be linked only to the source. The node set can be summarized as:

$$\mathcal{V} = \mathcal{P} \bigcup \{s, t\}. \tag{6.8}$$

where the set $\mathcal{P}$ represents the pixel depth values. In the constructed graph, there are two kinds of edges: (i) terminal links, called t-links, which are the edges between any pixel node and a terminal node; and (ii) neighborhood links, called n-links, which are the edges

connecting neighboring pixel pairs[2] . In summary, the edge set can be defined as follows:

$$\mathcal{E} = \mathcal{N} \bigcup_{p \in \mathcal{P}} \{\{p, s\}, \{p, t\}\} \tag{6.9}$$

Here, $\mathcal{N}$ represents the neighborhood system. All edges in the graph are assigned non-negative weights. The edge weights of t-links define the belief about a node being a face or an occlusion pixel: For example, a pixel highly resembling the face pixel at the corresponding location will be strongly connected to the sink, whereas its t-link to the source will be weak. On the other hand, the edge weights of n-links define the neighborhood system: The pairs of neighbors conforming to the relation between respective pixel pairs will be strongly connected, whereas unexpected relations will be represented with weak edges. A representative example to illustrate the graph construction is given in Fig. 6.1. Here, the n-links and t-links for a 3×3 pixel set are shown in the left subfigure, where the thickness of the line segments represent the edge weights.

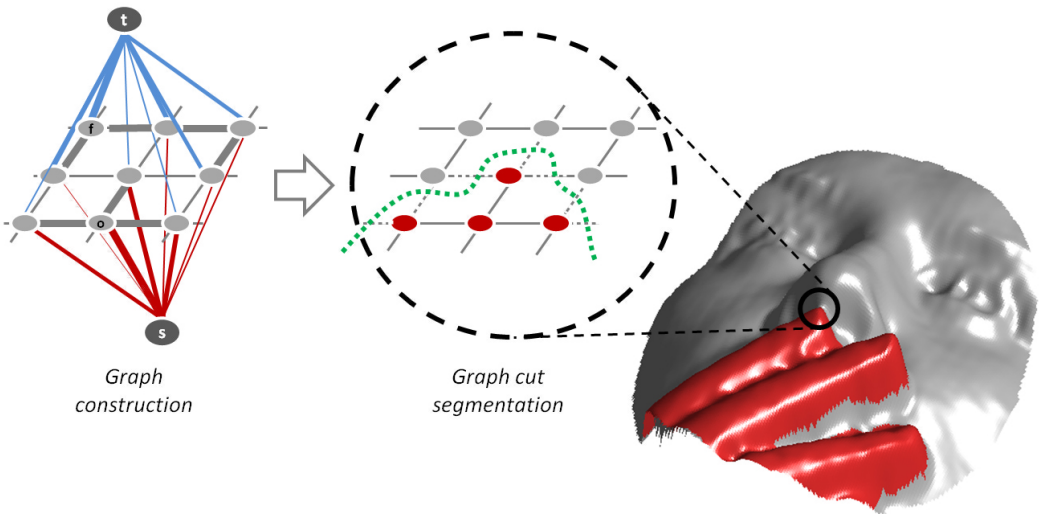


Figure 6.1. The graph construction and segmentation method is illustrated: On the left, the constructed graph for a 3×3 pixel set is shown, 'o' and 'f' representing 'occlusion' and 'face' seeds. On the right, segmentation found is shown by a green dashed line.

After representing the regional and neighborhood relations of pixels as a graph, the binary segmentation of pixels into face and occlusion segments is handled by the graph cuts method [8]: Here, the aim is to find an s-t cut $\mathcal{C}$ on the graph, where the cut is a set of edges

[2]In this work, we consider undirected edges for simplicity.

and the removal of these edges gives two disjoint subsets of nodes, namely $\mathcal{S}$ and $\mathcal{T}$, where they include the source and the sink terminals, respectively. The graph cut can be defined as follows:

$$\mathcal{G}(\mathcal{C}) = \langle \mathcal{V}, \mathcal{E} \backslash \mathcal{C} \rangle \tag{6.10}$$

The cost of a cut is defined by the total of the weights of the edges included in the cut:

$$|\mathcal{C}| = \sum_{e \in \mathcal{C}} w_e \tag{6.11}$$

Here, w_e defines the weight of the edge e included in the cut. The optimal segmentation can be found by solving the *min-cut* problem, where the cut with the minimum cost is searched among all possible cuts in the graph. The energy minimization problem, where the energy to be minimized is the total cost of the s-t cut separating the two terminals, can be solved to give a *global* minimum in polynomial time [90]. Alternatively, the energy to be minimized can be defined as

$$E(\boldsymbol{\mathcal{L}}) = \lambda \mathcal{R}(\boldsymbol{\mathcal{L}}) + \mathcal{B}(\boldsymbol{\mathcal{L}}) \tag{6.12}$$

where $\boldsymbol{\mathcal{L}} = \{\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_p, \ldots, \mathcal{L}_{|\mathcal{P}|}\}$ is a binary vector of assigned labels defining the segmentation. Here, the energy of the segmentation $E(\boldsymbol{\mathcal{L}})$ encapsulates two energies: (i) The regional term $\mathcal{R}(\boldsymbol{\mathcal{L}})$, reflecting how a pixel value fits into the given model of object or background; and (ii) the boundary term $\mathcal{B}(\boldsymbol{\mathcal{L}})$, defining the discontinuities residing in the neighborhood model given. Additionally, we have a non-negative coefficient λ, which gives the relative importance given to the regional term versus the boundary term.

As mentioned above, the total energy to be minimized is defined using both regional cues about each pixel and boundary cues about each neighboring pixel pair. To set these cues for a given 3D surface, predefined surface models can be used. These models are either derived from a set of training facial surfaces, or set heuristically. When defining regional cues, a pixel is linked both to the face (background) and to the occlusion (object) terminal.

The degree of the pixel's resemblance of either model will be expressed in the weights of the t-links. On the other hand, the degree of neighborhood resemblance will be expressed in the n-link weights. Strong connections will be assigned larger weights, whereas weak connections will have smaller weights. All these soft-constraint cues will be drawn from the 3D surface. In addition, we can have some pure regional beliefs about some pixels, on whether they are from the facial surface or from the occluding object. Then, these restrictions will be defined as hard constraints: For example, if we know that a pixel belongs to the face, then its connection to the face terminal will be the strongest, and the link between this pixel and the occlusion terminal will be removed. The hard constraints can again be drawn from the 3D surface and they will be used to restrict the search space of the energy optimization problem.

In this thesis, we have utilized the min-cut/max-flow algorithm proposed in [89]. The contribution of this work, is to apply this image segmentation approach to occlusion detection, and we have proposed to utilize statistical models to set the regional and the neighborhood cues.

Pixelwise Face Modeling for Hard Constraints and Regional Cues:

Surface registration followed by regular resampling gives a set of facial images of same dimensions and each pixel corresponds approximately to the same location on the facial surface. Using this property, the facial surface can be statistically modeled.

A face surface $\boldsymbol{x}$ can be represented by a vector of $|P|$ valid depth values:

$$\boldsymbol{x} = [z_1, \ldots, z_p, \ldots, z_{|P|}] \tag{6.13}$$

Here, the z_p's correspond to pixel depth values. Using a training set of non-occluded neutral faces, mean (μ_p) and standard deviation (σ_p) of each pixel depth is learned to give a basic

pixelwise model of the face. The statistical model is given as follows:

$$\begin{aligned} \boldsymbol{\mu} &= [\mu_1, \ldots, \mu_p, \ldots, \mu_{|P|}] \\ \boldsymbol{\sigma} &= [\sigma_1, \ldots, \sigma_p, \ldots, \sigma_{|P|}] \end{aligned} \tag{6.14}$$

The above pixelwise model can be utilized to define hard constraints: For each pixel, a threshold is set by employing the corresponding mean and standard deviation. The $|P|$-dimensional vector of pixelwise thresholds can be defined as

$$\boldsymbol{\tau}^{(H)} = \kappa^{(H)} \cdot \boldsymbol{\sigma} \tag{6.15}$$

where $\kappa^{(H)}$ is a predefined constant. The $\boldsymbol{\tau}^{(H)}$ threshold defines the upperbound for the fitness of pixels to the facial surface: If the difference between the depth value and the corresponding pixel mean is above the threshold value, then the pixel will be assigned as an initial occlusion seed.

$$\mathcal{L}_p = \begin{cases} 1 & \text{if } |\mu_p - z_p| > \tau_p^{(H)} \\ 0 & \text{otherwise} \end{cases} \tag{6.16}$$

Here, the binary labels of 1 and 0 correspond to occlusion and non-seed pixels respectively. Similarly, initial face seeds can be set, where a constant $\kappa^{(L)}$ will be used to define the lower bound threshold $\boldsymbol{\tau}^{(L)}$:

$$\boldsymbol{\tau}^{(L)} = \kappa^{(L)} \cdot \boldsymbol{\sigma} \tag{6.17}$$

Then, the pixels with distance to the face model smaller than $\boldsymbol{\tau}^{(L)}$ will be labeled as initial face seeds:

$$\mathcal{L}_p = \begin{cases} -1 & \text{if } |\mu_p - z_p| < \tau_p^{(L)} \\ 0 & \text{otherwise} \end{cases} \tag{6.18}$$

Here, the face seeds and non-seeds are labeled as -1 and 0, respectively. However, in our experiments, we have only employed occlusion seeds, discarding face seeds due to their low accuracy of detection.

The regional cues of non-seed pixels, given by the edge weights to each terminal, can be set using the statistically defined $\boldsymbol{\tau}^{(H)}$ and $\boldsymbol{\tau}^{(L)}$ thresholds[3] : The t-links to the terminals will be computed proportionally to the distance between the depth value and the thresholds. For the source (occlusion) terminal, the edge weight is given by:

$$w_p^{(s)} = \frac{|\mu_p - z_p| - \tau_p^{(L)}}{\tau_p^{(H)} - \tau_p^{(L)}} \cdot 255 \tag{6.19}$$

Similarly, for the sink (face) terminal, edge weights can be computed as follows:

$$w_p^{(t)} = \frac{\tau_p^{(H)} - |\mu_p - z_p|}{\tau_p^{(H)} - \tau_p^{(L)}} \cdot 255 \tag{6.20}$$

Hence, the edges connecting pixels to the source or the sink will have values in the range $[0, 255]$. The initial seeds should be connected to their respective terminal with a weight of infinity, so that these links will never be broken in the min-cut computation.

Neighborhood Modeling for Boundary Cues:

In binary image segmentation literature, the computation of n-link weights is based on local intensity differences, Laplacian zero-crossing, gradient direction, or any other basic edge detection method. Unlike other segmentation problems, for occlusion detection prior to face recognition, we are certain that the background is a facial surface. This apriori knowledge enables us to set n-link weights more elaborately: The neighborhood relations of faces can be used to define a background model, and the pixel pairs not fitting the model will denote the occlusion boundaries. Using a background model instead of using basic depth differences, will enable to differentiate between boundary pixels and non-boundary facial pixels with depth differences (such as nose, mouth or eye corners). Below, the background

[3] In our experiments, we do not consider face seeds. Therefore $\boldsymbol{\tau}^{(L)}$ is set to zero.

model and the methodology to set n-link weights are given.

For a 2D grid, we can either consider a first-order (4-neighbors) or a second-order (8-neighbors) neighborhood model. Here, 8-neighborhood model is employed. Using a training set of facial depth images, the background model is constructed: For each neighboring pixel pair (z_p, z_q), mean $\mu_{(p,q)}$ and standard deviation $\sigma_{(p,q)}$ are computed. Using these parameters, the depth difference in between the pixel pair can be used as a proportion to set the n-link weights:

$$w_{(p,q)}^{(n)} = \begin{cases} 0, \text{ if } |d_{(p,q)} - \mu_{(p,q)}| > \kappa^{(B)} \cdot \sigma_{(p,q)} \\ 255 \cdot \left(1 - \frac{|d_{(p,q)} - \mu_{(p,q)}|}{\kappa^{(B)} \cdot \sigma_{(p,q)}}\right), \text{otherwise} \end{cases}$$

Here, $\kappa^{(B)}$ is a preset constant defining the limit of variance from the mean depth difference. The weights are set to be in the same range as the non-seed t-links ($[0, 255]$): The edge weights of more distant pixel pairs are limited to have a maximum value of 255, so that the t-links of the initial seeds persist to have the strongest connection.

6.2. Restoration of Occlusion-Removed Surfaces

Instead of applying classification on incomplete facial surfaces after occlusion removal, a possible alternative to handle missing components is to apply restoration. If the facial surfaces can be accurately restored, then any traditional classification strategy can be applied on completed surfaces. In this thesis, we have inspected restoration of facial surfaces after occlusion detection and removal, using the partial Gappy PCA approach (given in Section 3.3.2): Gappy PCA [10] is a PCA variant capable of handling missing components, and partial Gappy PCA was used in [4] to improve the Gappy PCA approach. With Gappy PCA, it is possible to reconstruct the original facial surface up to a certain degree when the surface contains missing values (due to occlusion). In partial Gappy PCA, reconstructed data is used only to recover the missing parts of the surface. In order to estimate the unknown facial data by the Gappy PCA method, locations of the missing components are required. Prior to estimation, a lower-dimensional subspace is learned using a training set of non-occluded images. Then, the projection of the incomplete data to this subspace is handled using the occlusion

mask together with the gappy norm [10]. Then, the projected version is used to compute the back projection to the original face space. This back projected surface is complete and is an approximation of the original surface. In partial Gappy PCA, the reconstructed version is used only to complete the missing parts in the original facial surface. Further details about the Gappy PCA algorithm are given in Section 3.3.2.

6.3. Experimental Results

In this chapter, we have experimented with four different occlusion masks: (i) manually labeled ground truth masks (GT); (ii) masks obtained by thresholding the difference from an average face model (BL); (iii) masks obtained by facial modeling with pixelwise GMMs (GMM); and (iv) masks obtained by the graph cut technique (GC), where μ-σ modeling is used to set both the regional and the boundary weights. The results with the ground truth masks are included for comparative purposes. The results obtained using the difference from an average model is included as a baseline approach, since this technique is used in the literature. The third approach is expected to yield a better pixelwise facial modeling, where in the fourth approach neighboring relations are taken into account. Next, we summarized the databases used in the experiments. Then, we report occlusion detection accuracy results and perform simple face recognition experiments with occlusion removal to consolidate our conclusions about the occlusion detection performances. Furthermore, we compare occlusion removal and surface restoration approaches as two occlusion handling alternatives.

6.3.1. Databases

In the analysis of occlusion detectors, three databases[4] are employed, namely: (i) FRGC v.2, (ii) Bosphorus, and (iii) UMB-DB. The FRGC v.2 [3] neutral subset, containing a total of 2365 images of 466 subjects, serves as a separate training set for: (i) the construction of the statistical models defining regional and boundary relations of non-occluded facial surfaces, and for (ii) the construction of the average face model used in the baseline approach. The Bosphorus [15] and UMB-DB [16] databases are employed for the evaluation of the

[4]Although detailed information about these databases are given in Chapter 2, some necessary details are included here for the completeness of the experimental results section.

occlusion detection performance. For the simple classification experiment run for occlusion detection evaluations, the gallery and probe sets are constructed as follows: The first neutral scan of each subject is used to construct the gallery set, whereas the images with occlusion variations form the probe set. For the Bosphorus database, 105 and 381 images are included in gallery and probe sets, respectively. For the UMB-DB, there are 142 gallery images and 590 probe images.

For the occlusion databases, manually labeled occlusion masks are available. However, these masks are coarsely labeled and using these masks to evaluate the occlusion detection for the whole databases would be misleading. Therefore, for evaluation of the occlusion detectors, we have selected a subset of 70 facial surfaces from the Bosphorus database: These samples are selected such that the ground truth occlusion masks are accurately labeled, which is important when evaluating the performance of occlusion detectors via the F-measure. Furthermore, we have selected a subset so that some examples are challenging for occlusion detection, whereas some can be easily detected by the baseline occlusion detector. Throughout the experimental results section, this subset will be referred to as the *Bosphorus-70 subset*.

6.3.2. Occlusion Detection Accuracy

Let's assume that we have the manually labeled occlusion masks for the occluded surfaces. Using these ground truth masks, the performance of the automatic occlusion detector can be evaluated: In this paper, we have utilized precision and recall values to compute F-measure [91], which will serve as the evaluation measure of the occlusion detection module. In a classification scenario, precision is the ratio of the number of true positives to the total number of positives, whereas recall is the ratio of the number of true positives to the total number of positives. Therefore, precision gives the fraction of the retrieved examples that are relevant, and recall gives the fraction of the relevant instances that are retrieved. In the context of occlusion detection, precision defines the percentage of the correct ones among all the detected pixels, whereas recall is the percentage of the detected ones among all the

occlusion pixels. The precision and recall measures can be summarized as follows:

$$precision = \frac{TP}{TP + FP} \tag{6.21}$$

$$recall = \frac{TP}{TP + FN} \tag{6.22}$$

Here, TP, FP, and FN refer to true positives, false positives, and false negatives, respectively. In occlusion detection, it is important to detect most of the occluded pixels (high recall). In addition, it is not desirable to label non-occluded pixels as occlusion (high precision). Hence, there is a trade-off in between these two measures, and neither precision nor recall will be enough to evaluate the detection accuracy. Therefore, we have utilized the *F-measure*, which is a measure combining precision and recall. In general, the measure can be computed as,

$$F_\beta = (1 + \beta^2) \cdot \frac{precision \cdot recall}{\beta^2 \cdot precision + recall} \tag{6.23}$$

using the precomputed precision and recall values. Here, β defines the weight given to precision versus recall. In our experiments, F_1 measure is employed: F_1 gives the harmonic mean of precision and recall, and it is referred to as the *balanced* F-score. For face recognition, it is important to exclude almost all of the occluded parts, whereas the discarded facial parts should be minimal. Hence, we have considered a balanced measure.

Before evaluating the accuracy of different occlusion detectors, we first utilize F_1 measures to optimize the parameters used in the mathematical formulations of the graph cut technique: When setting the edge weights in the detector based on graph cut technique, we have a set of three parameters, namely ($\kappa^{(H)}$, $\kappa^{(B)}$, λ). Using a full-factorial experimental design [92] on the Bosphorus-70 subset, we have checked the effect of each parameter and the interactions in between. After finding which elements and interactions are important, we have modeled the relationship between the factors and the response, fitting a regression model. The parameter set maximizing the response (F_1 measure) is selected to be used for further occlusion detection experiments. Further details on the factorial design analysis are given in Appendix A.

Table 6.1. Precision, recall, and F_1 measures on the Bosphorus-70 subset for different occlusion detectors.

Occlusion Detector	Precision	Recall	F_1 measure
BL	0.933	0.706	0.785
GMM	0.910	0.826	0.849
GC	0.848	0.907	0.868

On Bosphorus-70 subset, we have evaluated different automatic occlusion detectors (BL, GMM, GC). In Table 6.3.2, precision, recall, and F_1 measure values are given for all three occlusion detectors. When the results are inspected, it is clear that both of the proposed occlusion detectors perform better than the baseline approach: F_1 measures are significantly better for the newly proposed detectors. Due to the inclusion of neigborhood relations in the graph cut method, GC detector performs better than the GMM detector, where only pixel-specific information is employed. The recall values show that GC can capture a larger ratio of the occluded parts than GMM. However, when we check the precision values, we see that GC performs poorer to exclude non-occluded parts when detecting most of the occlusions. This is mainly due to the fact that neighboring relations can cause inclusion of some neighboring non-occluded pixels. In Figure 6.2, some examples are given, where different occlusion masks are plotted. In the first column, ground truth masks are included for comparative purposes. The second, third, and fourth columns show results of BL, GMM, and GC detectors. For each mask, the correctly detected (true positive), incorrectly detected (false positive), and incorrectly missed (false negative) pixels are colored in green, blue, and red, respectively. As these results illustrate, the baseline approach cannot detect a large number of occluded pixels (labeled in red), whereas the better performing GMM and GC detectors can include some nonoccluded pixels in the occlusion mask (labeled in blue). This explains the high precision and low recall values for the baseline technique, whereas lower precision and higher recall values are obtained for the better performing GMM and GC detectors.

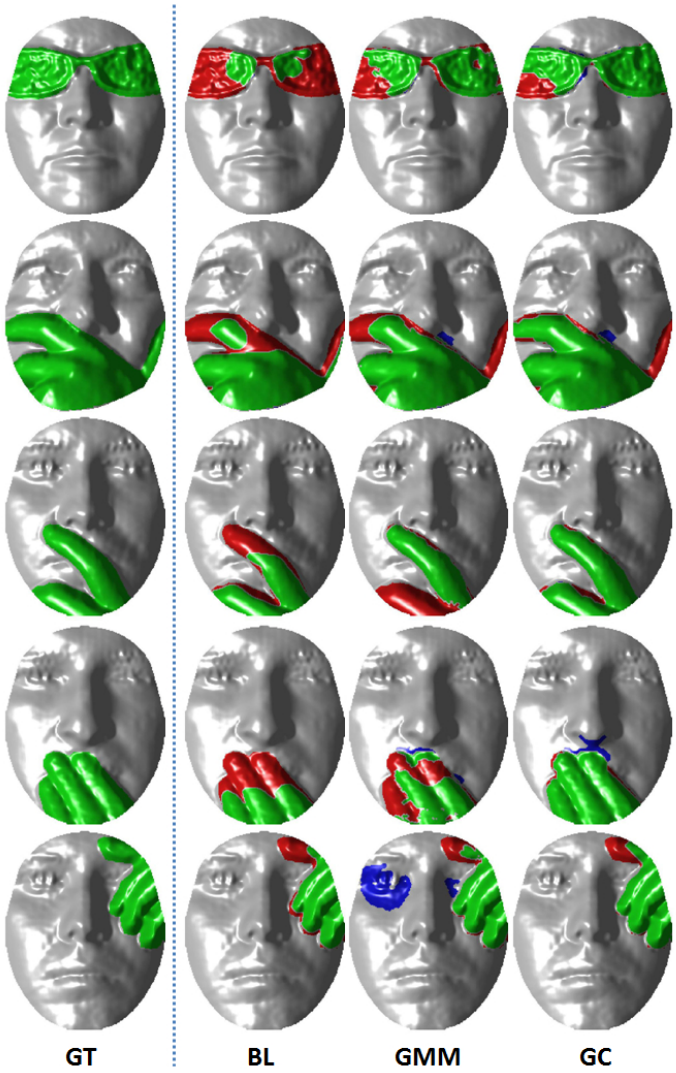


Figure 6.2. Examples are given, where GT (ground truth), BL (baseline), GMM (Gaussian Mixture Models), and GC (graph cut) masks are shown in first, second, third, and fourth columns. The TP, FP, and FN pixels are colored in green, blue, and red, respectively.

6.3.3. Classification Accuracy with Occlusion Removal

Although the main problem addressed in this paper is occlusion detection, our aim is to detect occluded surface regions for robust face recognition. Moreover, in the previous experiments, only a small selection of the Bosphorus database is analyzed due to inapplicability to the whole datasets with incomplete or incorrect manually labeled masks. To evaluate the performance of the occlusion detection approaches better for the two occlusion databases, we have constructed a simple classification experiment, where *depth-based* classifier is utilized: In the depth-based classifier, the depth information is used to calculate the mean Euclidean distance between a probe and a gallery face, serving as the dissimilarity score. Here the occlusion mask is employed to discard the pixels labeled as occlusion from both of the surfaces. Formal definition for this classification approach, referred to as the depth-based classifier, was previously included in Section 5.3.3.

In Table 6.3.3, the classification results for the Bosphorus and the UMB-DB databases are given using different occlusion masks. When the recognition rates are inspected, it is clear that these results are consistent with the occlusion detection accuracy results reported in Table 6.3.2: Both of the proposed occlusion detectors perform better than the baseline approach, whereas the graph cut technique yields higher recognition rates than the one using pixelwise GMMs. The results obtained using the automatically detected occlusion masks validate that the UMB-DB database includes highly challenging scans: For UMB-DB, classification results obtained by using automatically detected masks perform poorly when compared with the results reported by employing manually labeled masks. For the Bosphorus database, all of the performances are quite similar. Furthermore, the GC detector outperforms the classification results obtained using the ground truth masks. When the correctly identified scans are investigated, it is apparent that automatic occlusion detection has an additional benefit: Some minor registration errors can reside in the facial surfaces, yielding a decrease in the fitness criterion for some specific non-occluded surface points. When automatic occlusion detector is employed, surface points not resembling the corresponding facial depth values are located. Hence in addition to occlusions, parts that are not sufficiently similar to training facial surfaces are labeled for removal. In other words, regions that are mostly affected by registration errors are discarded from the classification comparison. As a conse-

Table 6.2. Depth-based classification results with different occlusion masks.

Occlusion Masks	Bosphorus	UMB-DB
GT	83.99	65.25
BL	83.20	56.78
GMM	83.99	57.80
GC	84.51	58.14

quence, the identification accuracies are higher than the results obtained using ground truth masks.

6.3.4. Removal versus Restoration

In this section, we investigate the restoration by partial Gappy PCA as an occlusion handling alternative. In Figure 6.3 a reconstruction example obtained by Gappy PCA is given for a challenging example from the UMB-DB database. As visualized in the second row, the quality of face restoration depends on the subspace dimensionality of the Gappy PCA: As the dimensionality increases, the restored facial surface gets more similar to the original surface.

Here, it should be noted that although the restored face appears as an appropriate resemblence, it is only an approximation and the discriminative information needed for classification can be lost. Partial Gappy PCA can be an alternative to reduce the negative effect of restoration, where the restored surface information is utilized only to fill the missing parts. In Table 6.3.4, we have included global depth-based classification results, where ground truth occlusion masks are utilized: In the second row, the original occluded surfaces are used without any removal or restoration. In the third row, the results with occlusion removal are included for comparative purposes. In the last row, the classification performance is reported, where restoration is handled by partial Gappy PCA. Here, the basis vectors are learned from a separate training set (FRGC v.2). As these results indicate, it is better to handle missing parts. However, restoration does not provide sufficient performance: Since restoration gives only an approximation of the surface, it is not appropriate to restore missing parts for

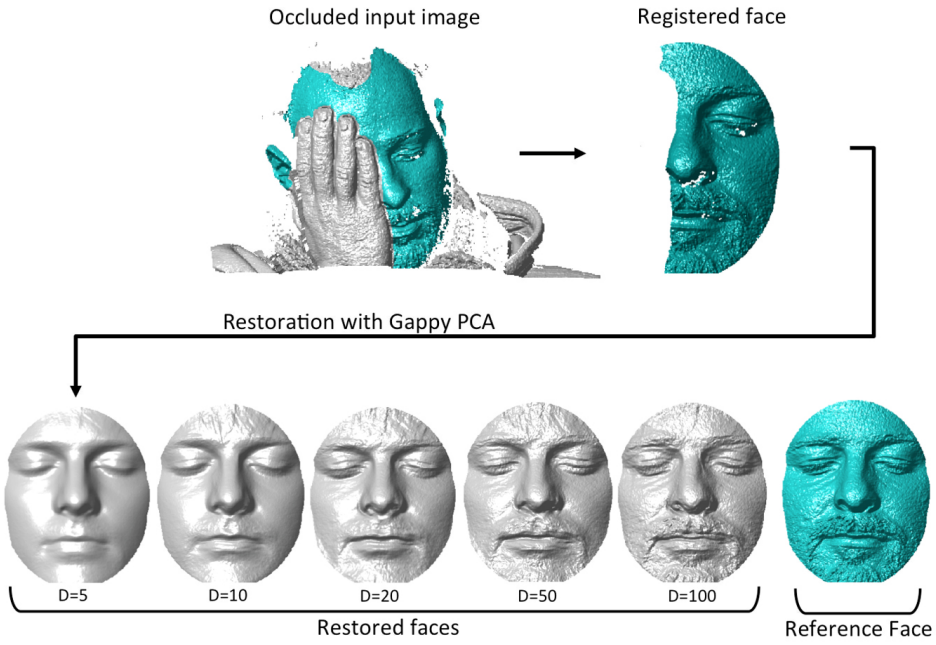


Figure 6.3. An example to restoration obtained with Gappy PCA is given, where different subspace dimensionalities are utilized.

Table 6.3. Depth-based classification results on occlusion-removed, restored data.

Occlusion Handling	Bosphorus	UMB-DB
None	63.52	46.10
Occlusion Removal	83.99	65.25
Restoration (partial Gappy PCA)	76.90	47.80

a classification scenario, and inferior results are obtained with restoration.

6.4. Conclusion

In this study, we focused on the problem of occlusion detection and handling prior to surface classification, where the surfaces are assumed to be accurately aligned. We have proposed two main occlusion detectors: One of the detectors is based on complex modeling of the facial surface by using pixelwise Gaussian Mixture Models, where pixels are checked for their fitness to the corresponding mixture model. The other detector, incorporates the information residing in neighborhood relations into the pixelwise cues. The facial

surface is represented as a graph, where the regional and boundary cues are embedded as edge weights. Occlusions are detected solving the binary segmentation problem on the constructed graph. For comparative purposes, a baseline detector using difference from an average face is utilized. When the performance of occlusion detectors are compared with the results of ground truth occlusion masks and the baseline technique, it is clear that the facial modeling with GMMs yields better results than the baseline approach. Further improvement can be achieved by considering both the regional and neighborhood information: GC results outperform the results of GMM. Furthermore, we have experimented with two different occlusion handling alternatives: removal versus restoration. The experimental results showed that, even though the restored images appear as good approximations of the true surface, they should not be employed for classification: Since restoration is only an approximation of the surface, discriminative information cannot be reconstructed. Therefore, restoration should not be preferred for recognition systems.

7. FACE RECOGNITION UNDER OCCLUSION

In a face recognition system, the aim is to infer the identity of a subject from the acquired image. For an identification scenario, the identity of the subject is searched among the subjects present in the gallery: Features are extracted from the input image, after the preprocessing steps of detection and registration. The extracted features are then used to compare the input image against all of the images of the gallery set. For a closed-set system, the identity of the gallery sample that is closest to the input scan is set as the estimated identity of the probe.

When occlusions are present over the facial surface, however, standard classification approaches are not applicable directly to infer the identity of subjects: The probe face has missing parts, whereas the gallery images are acquired in a cooperative manner and therefore are complete. After the preprocessing steps of detection, registration, occlusion detection and removal are applied on the input image, a probable solution would be to restore the incomplete parts, so that standard classification approaches can be employed. However, as the experiments reported in Chapter 6 point out, restoration should not be preferred to complete large surface holes caused by occluding objects when classification is to be applied afterwards: Although face-like surfaces are obtained via restoration, the estimated surface information does not embody discriminative information, thus is inadequate for comparative purposes. An alternative approach is to alter the classification approaches, so that they can work on incomplete data. Since subspace techniques are often utilized in classification scenarios, we investigated the applicability of subspace classification methods to incomplete probe data.

This chapter introduces a new technique called *masked projection* for subspace analysis with incomplete data. The preliminaries on subspace techniques were given in Section 3.3. Here, we give details about the proposed technique: First of all, the algebraic derivations of the masked projection are given and the idea of incorporating local regions into this new technique are outlined. Then, experimental results on two occlusion databases, namely the Bosphorus and the UMB-DB databababases, are given; where the conclusions drawn from

the experiments are summarized.

7.1. Global Classification using Masked Projection

A useful property of the model-based registration scheme is that the extracted facial features, $\boldsymbol{x}_i$, are ordered vectors of the same size, enabling the use of subspace analysis techniques. However, subspace approaches assume complete facial feature vectors. Therefore standard subspace approaches cannot be applied directly on occlusion-free faces. The first idea to deal with incomplete data, would be to remove the pixels that are not present in the probe image from all of the training and gallery images, as in [31]. Using the masked training images, the subspace representing the partial surfaces can be learned by the Fisherfaces approach [14]. However, this approach is not feasible, since each probe face will have different pixels missing and a separate training phase is required. In this work, we propose a projection masking approach to obtain the adaptive subspace: The general projection matrix is learned using a set of non-occluded complete training images. Then, the adaptive projection matrix is obtained by masking. The masked probe and gallery images are projected onto the subspace, and classification is performed. The algebraic details of the approach are given below.

Let $\boldsymbol{x}$ be the registered facial surface vector, and $\mathbf{W}$ represent a projection matrix[5] . The surface vector can be defined as $\boldsymbol{x} = \boldsymbol{\mu} + \mathbf{W}\boldsymbol{y}$, where $\boldsymbol{\mu}$ is the mean of the training images, and $\boldsymbol{y}$ is the coefficient vector residing in the subspace defined by $\mathbf{W}$. To simplify equations, we assume that $\boldsymbol{\mu}$ is zero. In practice, this is assured by a change of coordinates. The coefficients are computed as

$$\boldsymbol{y} = \mathbf{W}'\boldsymbol{x}, \tag{7.1}$$

where $\mathbf{W}'$ is the transpose of $\mathbf{W}$. Now, suppose there is an incomplete version of $\boldsymbol{x}$, namely $\hat{\boldsymbol{x}}$, whose missing components are encoded in the occlusion mask $\mathbf{m}$. Let's assume that we

[5] In our experiments, we have used the Fisherfaces projection.

have the coefficient vector $\tilde{\boldsymbol{y}}$, where the input image can be approximated as

$$\tilde{\boldsymbol{x}} = \mathbf{W}\tilde{\boldsymbol{y}} \tag{7.2}$$

where $\tilde{\boldsymbol{x}}$ is the approximated complete version of the input image $\hat{\boldsymbol{x}}$. Our objective is to find the coefficient vector $\tilde{\boldsymbol{y}}$, minimizing the error term $E = ||\hat{\boldsymbol{x}} - \tilde{\boldsymbol{x}}||^2$. In this formulation, the missing components in $\hat{\boldsymbol{x}}$ will augment the total error term. To improve the error term, the masked norm [10] is used[6] , where the information about the missing components is encoded in the mask $\mathbf{m}$. The masked norm for a vector $\mathbf{u}$ with the mask $\mathbf{m}$ is defined as $||\mathbf{u}||_m = \sqrt{(\mathbf{u},\mathbf{u})_m}$ where

$$(\mathbf{u},\mathbf{u})_m = \mathbf{u}'_m\mathbf{u}_m. \tag{7.3}$$

Here $\mathbf{u}_m$ is the masked version of $\mathbf{u}$, $\mathbf{u}_m = \mathbf{\Lambda}_m\mathbf{u}$, where $\mathbf{\Lambda}_m$ is a diagonal matrix, whose diagonal elements constitute the mask: $\mathbf{m} = diag(\mathbf{\Lambda}_m)$. When the masked data is inserted into E, we obtain:

$$\begin{aligned} E_m &= ||\hat{\boldsymbol{x}} - \tilde{\boldsymbol{x}}||^2_m \\ &= \hat{\boldsymbol{x}}'_m\hat{\boldsymbol{x}}_m - \hat{\boldsymbol{x}}'_m\mathbf{W}_m\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{y}}'\mathbf{W}'_m\hat{\boldsymbol{x}}_m \\ &\quad +\tilde{\boldsymbol{y}}'\mathbf{W}'_m\mathbf{W}_m\tilde{\boldsymbol{y}} \end{aligned} \tag{7.4}$$

where $\mathbf{W}_m = \mathbf{\Lambda}_m\mathbf{W}$. The error is minimized with respect to $\tilde{\boldsymbol{y}}$:

$$\frac{\partial E_m}{\partial \tilde{\boldsymbol{y}}} = -2\mathbf{W}'_m\hat{\boldsymbol{x}}_m + 2\mathbf{W}'_m\mathbf{W}_m\tilde{\boldsymbol{y}} = 0. \tag{7.5}$$

$$\hat{\boldsymbol{x}}_m = \mathbf{W}_m\tilde{\boldsymbol{y}} \tag{7.6}$$

To calculate the coefficients $\tilde{\boldsymbol{y}}$, the inverse of $\mathbf{W}_m$ is needed. Since $\mathbf{W}_m$ is constructed from $\mathbf{W}$ by setting occluded regions to zero, $\mathbf{W}_m$ is no longer orthogonal. Since inverse of an

[6] In [10], this measure is referred to as the *gappy* norm.

orthogonal matrix is just the transpose it, for ease of calculations, we first orthogonalize the $\mathbf{W}_m$ matrix [32]:

$$\mathbf{W}_{\perp} = \mathbf{W}_m(\mathbf{W}'_m\mathbf{W}_m)^{-1/2} \tag{7.7}$$

Then, the coefficient vector can be computed as:

$$\bar{\boldsymbol{y}} = \mathbf{W}'_{\perp}\hat{\boldsymbol{x}}_m. \tag{7.8}$$

where $\bar{\boldsymbol{y}}$ is the coefficient vector obtained by projection onto the space defined by $\mathbf{W}_{\perp}$. The masked projection matrix can be applied to the masked gallery image matrix $\boldsymbol{X}$, whose columns correspond to observations:

$$\boldsymbol{Y} = \mathbf{W}'_{\perp}\boldsymbol{X} \tag{7.9}$$

Here, it should be noted that gallery vectors should also be projected using the masked projection matrix $\mathbf{W}_{\perp}$, rather than the original projection matrix $\mathbf{W}$, since these two matrices define different subspaces: The subspace of $\mathbf{W}$ is trained using a subset of complete facial surfaces. When $\mathbf{W}_m$ is constructed, parts corresponding to occlusions are eliminated from the original matrix. Therefore, the orthogonal vector sets defining the subspaces are different for $\mathbf{W}$ and $\mathbf{W}_m$ (hence for $\mathbf{W}$ and $\mathbf{W}_{\perp}$). The idea of projecting the gallery images with masked projection, in addition to the probe images, is the main difference between the proposed approach and the Gappy PCA method of [10].

After the projection to the adaptive subspace, the dissimilarity between the probe coefficients $\bar{\boldsymbol{y}}$ and the coefficients of any gallery image; $\boldsymbol{y}_{G_k}$ which is the kth column of $\boldsymbol{Y}$; can be computed by the angular cosine distance measure:

$$D(\bar{\boldsymbol{y}}, \boldsymbol{y}_{G_k}) = 1 - \frac{\bar{\boldsymbol{y}} \cdot \boldsymbol{y}_{G_k}}{||\bar{\boldsymbol{y}}|| \cdot ||\boldsymbol{y}_{G_k}||}. \tag{7.10}$$

To obtain the final identification rates, the regional dissimilarity measures are fused by the

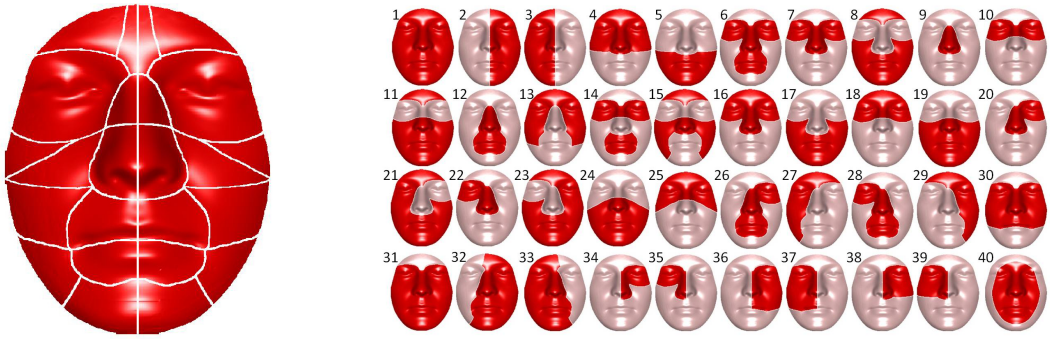
product rule and 1-NN classification is employed.

It should be noted here, that since we are using Fisherfaces subspace technique, we are applying PCA and LDA sequentially. The masked projection approach is applied to the PCA projection matrix to obtain the feature vector that will be input to the LDA. Since this feature vector is complete, the traditional LDA can then be applied directly, without applying the masked idea.

7.2. Regional Classification using Masked Projection

For further improvement in the classification phase, we propose to consider the 3D surface as a combination of several regions. If the facial area is partially occluded by external objects, the incorrect information regarding the covered regions will cause the global classification approaches to fail. Therefore, in the presence of occlusions, it is beneficial to incorporate separate regional classifiers. In regional techniques, each region acts as an independent classifier, and the regional recognition results are fused to obtain an improved overall performance. For the construction of the regions, we have divided the facial surface into several non-overlapping patches. Then, combination of these patches are merged to generate facial regions. The proposed regional division scheme consists of 40 regions as illustrated in Figure 7.1. In Figure 7.1a, the 24 symmetrical patches defined on the average face model are given. The facial surface is partitioned considering both the semantic structure (eyes, mouth, forehead, cheeks) and the facial symmetry. When the patch sizes and locations are set, the extent of the local regions to be constructed are taken into account. For the determination of patch combinations, possible real life occlusion scenarios are considered. In Figure 7.1b, the regions created using different subsets of patches are visualized (except for the last region, which is obtained by eroding the global face model).

To incorporate regional classifiers with the proposed subspace method, a separate regional subspace should be learned [26]. Therefore, for each alignment model and for each region, a separate projection matrix, $\mathbf{W}$, is trained. Each projection matrix defines a separate subspace for the corresponding region, where the training images are registered with the corresponding alignment model. When a probe face is examined, all of the regional subspaces of



(a) (b)

Figure 7.1. The regional division scheme: (a) patches, (b) regions (in red). The regions in (b) are constructed as combinations of patches of (a) (except for region 40, which is obtained by eroding region 1).

the corresponding model are employed: Regional features are computed by regional masked projections, where the occlusion and regional masks are merged to obtain the final masks employed in the projection stage. Then, separate regional subspace features are compared against corresponding feature sets of gallery images, and the regional classification results are fused. Although training of separate subspaces appears as a time consuming process, it is handled in an offline manner and does not affect the duration of the classification phase.

7.3. Experimental Results

For our experiments, we use the system outlined in Figure 7.2: The preprocessing module includes the registration and occlusion removal steps. For alignment, the adaptive registration module given in Chapter 5 is utilized, which registers the occluded surfaces. By adaptively selecting the model, it is possible to discard the effect of occluding surfaces on registration. For evaluation of the proposed system, we have experimented with ground truth occlusion masks. Additionally, we have included a comparison of different occlusion detectors (given in Chapter 6) integrated into the proposed masked projection. The training module works offline to learn the projection matrices from the training set of non-occluded faces for different regions. The classification module uses the occlusion mask of the probe image to compute the masked projection, and projects the probe image to the adaptive subspace. The identification is handled in the subspace by 1-nearest neighbor (1-NN) classifier. The proposed system is evaluated on two main 3D face databases that contain realistic oc-

clusions: (i) The Bosphorus, and (ii) the UMB-DB databases.

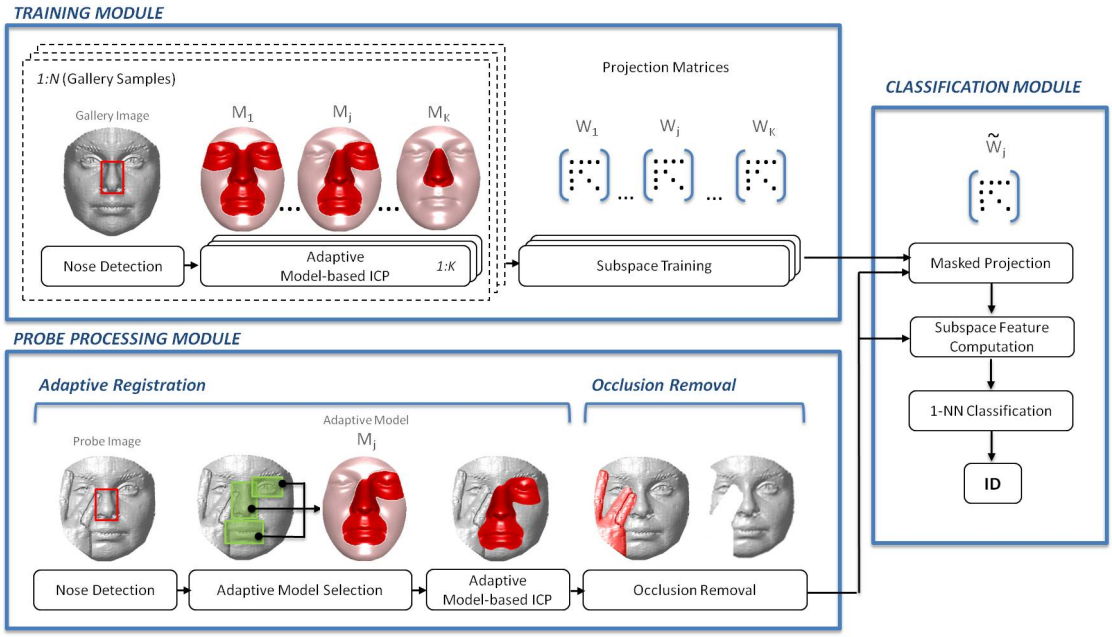


Figure 7.2. Illustrative diagram of the proposed 3D face recognition approach.

7.3.1. Evaluation of the Global Classification Performance

First, we start by considering the whole facial surface in an holistic manner. We compare two approaches to deal with missing data, where the occluded parts are either removed or restored (as given in Chapter 6). These approaches are evaluated in comparison with a baseline classifier, where the surfaces are considered without any preprocessing of the occluded parts. For removal, we have utilized global masked projection. For restoration, we have employed the partial Gappy PCA method of [4], which was summarized in Chapter 6: In partial Gappy PCA, the occluded parts are first removed from the surface. Then the whole facial surface is estimated using eigenvectors computed by PCA, where the estimated parts corresponding to the missing components are used to complete the facial surface.

In summary, we compare four different classification strategies in Table 7.1, using the ground truth occlusion masks: (i) The standard Fisherfaces [14] on original data, where no occlusion removal or restoration is applied (first row); (ii) the standard Fisherfaces on restored data, where the missing parts are restored by partial Gappy PCA of [4] (second row); (iii) the standard Fisherfaces applied on the *masked* probe features obtained by Gappy

PCA [10] (third row); and (iv) the proposed *masked* Fisherfaces, where the globally learned projection matrices are masked to obtain projections of both the gallery and the probe images (last row). As stated in Section 7.1, the gallery images should also be projected using the masked projection approach, since the subspaces defined by $\mathbf{W}$ and $\mathbf{W}_m$ are different. This is the main difference between the proposed approach and the idea of the Gappy PCA, and a quantitative comparison between the two approaches are included in the last two rows of Table 7.1. For the training of the Fisherfaces, the FRGC v.2 neutral subset is employed. As the results in Table 7.1 indicate, restoring occluded parts offers an improvement over original surfaces: For the Bosphorus, the performance is improved by 30%; for the more challenging UMB-DB, the improvement is 17%. However, we see that it is beneficial to remove the occluded parts, instead of restoring them: For the last two rows, the occlusions of the probe images are removed, whereas for the second row, restoration is employed: For the Bosphorus, $2-4\%$ further increase is obtained; whereas for the UMB-DB, a more significant performance improvement (about $6-9\%$) is achieved. Furthermore, the proposed approach (last row) yields better results than of Gappy PCA (third row): For a fair comparison, the parameters used for the compared dimensionality reduction techniques are identically chosen (the dimensions used for PCA and Fisherfaces are 150 and 100, respectively). As these results indicate, instead of restoring occluded areas, it is beneficial to employ the masked projection, which incorporates only the non-occluded surface regions. Moreover, the masked projection should be used to project both the gallery and the probe images.

Table 7.1. Global identification accuracies with the standard and masked Fisherfaces.

Method	Fisherfaces Approach	Gallery Data	Probe Data	Bosphorus	UMB-DB
Fisherfaces [14]	Standard	Original	Original	53.28	43.56
Fisherfaces with Restoration [4]	Standard	Original	Restored	83.46	60.34
Fisherfaces with Gappy PCA [10]	Masked	Original	Masked	85.83	66.10
Proposed (Global)	Masked	Masked	Masked	87.40	69.15

7.3.2. Evaluation of the Regional Classification Performance

Next, we evaluate the performances obtained by fusing the 40 separate regions, where the regional classifiers are fused at the score level by the product rule. Here, we compare three different classification strategies: (i) the standard regionally trained Fisherfaces on the restored data, which is included for comparative purposes; (ii) the regionally trained masked Fisherfaces, where a set of regional projection matrices are learned and then masked by the occlusion mask; and (iii) the globally trained masked Fisherfaces, where only a single projection matrix is learned and then masked by both the region and the occlusion mask. Once again, the FRGC v.2 neutral set is used for training of the Fisherfaces. For the analysis, we have conducted the experiments only with manually labeled occlusion masks. The performances are reported in Table 7.2, where the results obtained by the ground truth occlusion masks are given.

Table 7.2. Regional identification accuracies with both the standard and the newly proposed Fisherfaces approaches (results are reported with manual occlusion masks.)

Method	Fisherfaces Approach	Training Approach	Gallery Data	Probe Data	Bosphorus	UMB-DB
Fisherfaces with Restoration	Standard	Regional	Original	Restored	93.44	71.19
Proposed (Occlusion Masking)	Masked	Regional	Masked	Masked	93.18	73.90
Proposed (Occlusion&Region Masking)	Masked	Global	Masked	Masked	93.18	73.56

As the results in Table 7.2 indicate, better performances are obtained by the proposed masked projection approach. The cumulative match characteristic (CMC) plots are given in Figure 7.3 to verify the behavior of the considered classifiers. As these plots show, the CMC curves for the Bosphorus database are very similar, since the occlusions are relatively small and the regional division scheme can compensate for badly restored regions. For the UMB-DB database, the impact of masking is more visible: The standard Fisherfaces method on restored images performs poorly when compared with the masked approach. Moreover, these results indicate that employing masking to obtain regional projection matrices by combining the occlusion and regional masks is a viable alternative. Furthermore, the performance of

the proposed projection scheme is superior, when compared with the results reported in the literature on the UMB-DB database [16], where a PCA based classifier attains 56.50% identification rate on restored faces.

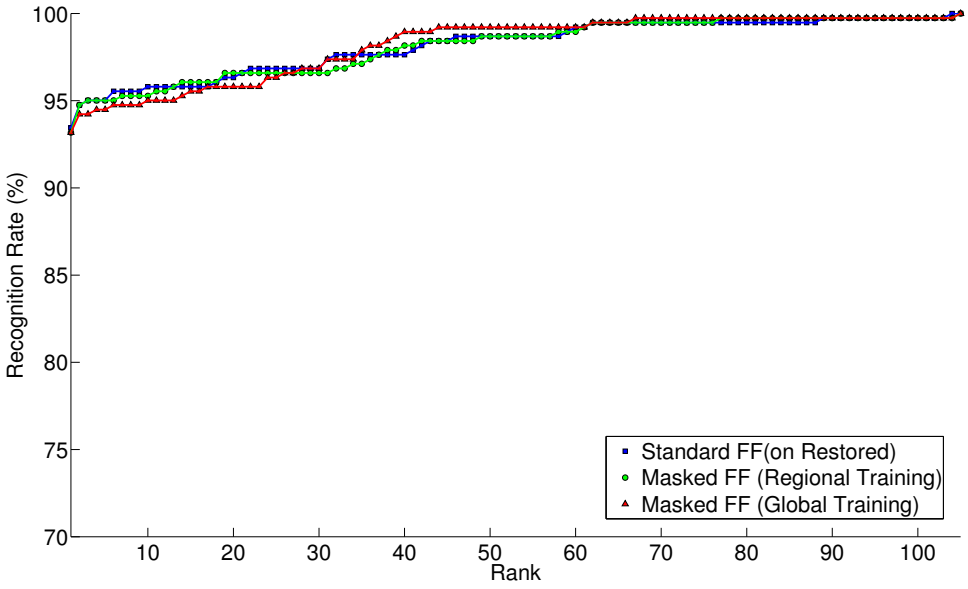
In Table 7.2, the identification rates for the standard Fisherfaces over the restored images appear as comparable. However, further analysis show that this is a result of the successful regional division scheme considered: Since the regions are determined by considering possible facial occlusions, for an occluded (or restored) probe face, a number of regional classifiers consider only the non-occluded parts. Therefore, they rectify the overall classification results. The regional classification results are reported using manual occlusion masks in Figure 7.4a for the Bosphorus database, and in Figure 7.4b for the UMB-DB, where three different classification approaches are compared[7] . The bars represent the performance improvement of the proposed masked Fisherfaces over the standard method, for different regions of the face. We observe that a performance improvement of 2 to 14% is obtained. When the globally trained masked Fisherfaces is compared with the regionally trained masked Fisherfaces, we see that performances are comparable and neither method performs better for all regions. We observe, however, that global training is superior for small regions due to the availability of more data in training. Global training is also preferred, since the learning is performed only once. The differences between the standard and the masked Fisherfaces performances are more prominent for the UMB-DB database, since this database contains more challenging occlusions.

7.3.3. Comparison of Masked Projection and Masked Training Performances

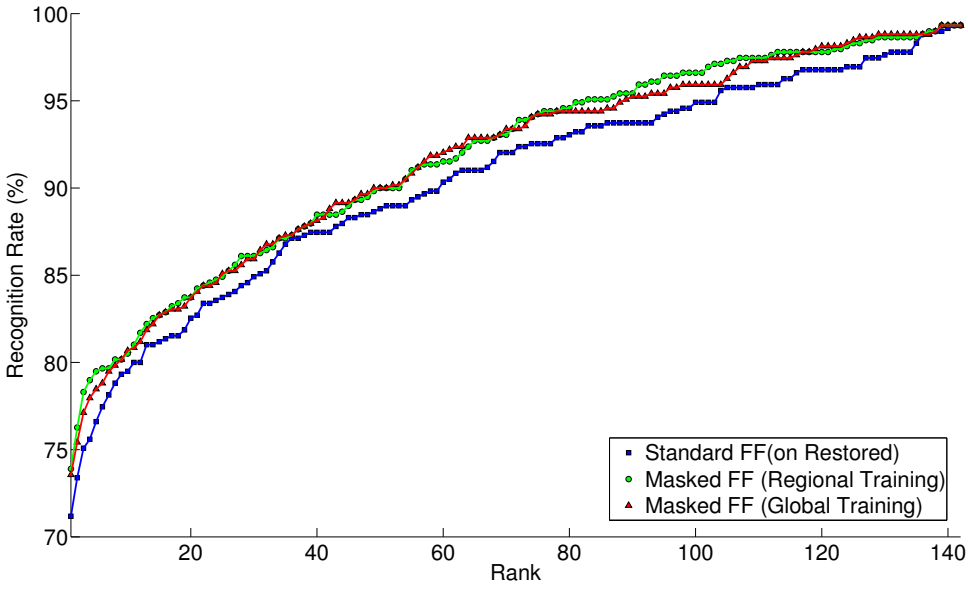
As stated in Section 7.1, a possible approach to deal with missing data in subspace analysis, is to remove the corresponding missing pixels from the training data and to learn the projection matrix from the masked training samples. Although this approach is not practical in occluded faces since each occlusion is unique, for comparative purposes, we have obtained recognition rates on the Bosphorus database using this masked training idea[8] . The

[7]The corresponding region for each region number is given in Figure 7.1b.

[8]The masked training results are included only for comparative purposes. Due to its high computational cost, the results are reported only on the Bosphorus database.

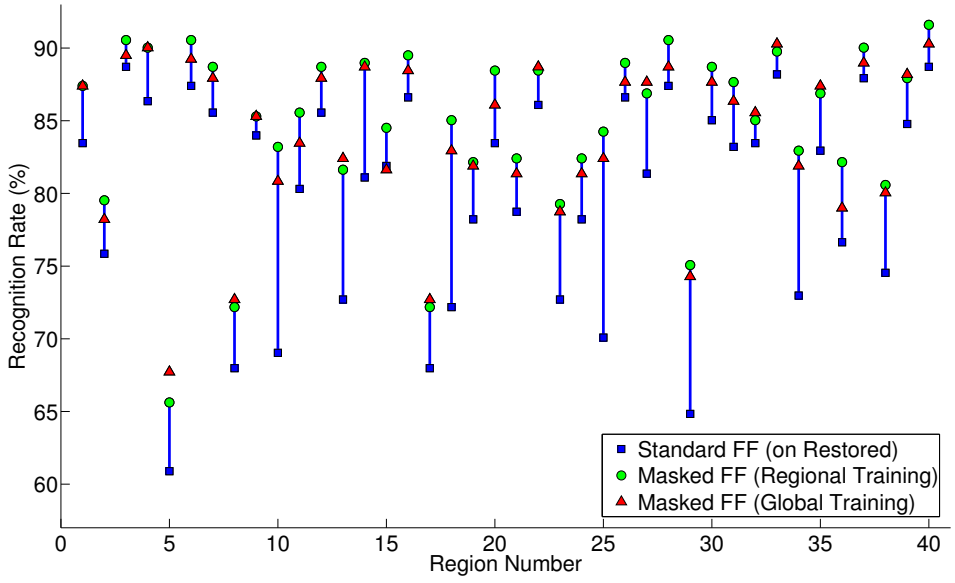


(a)

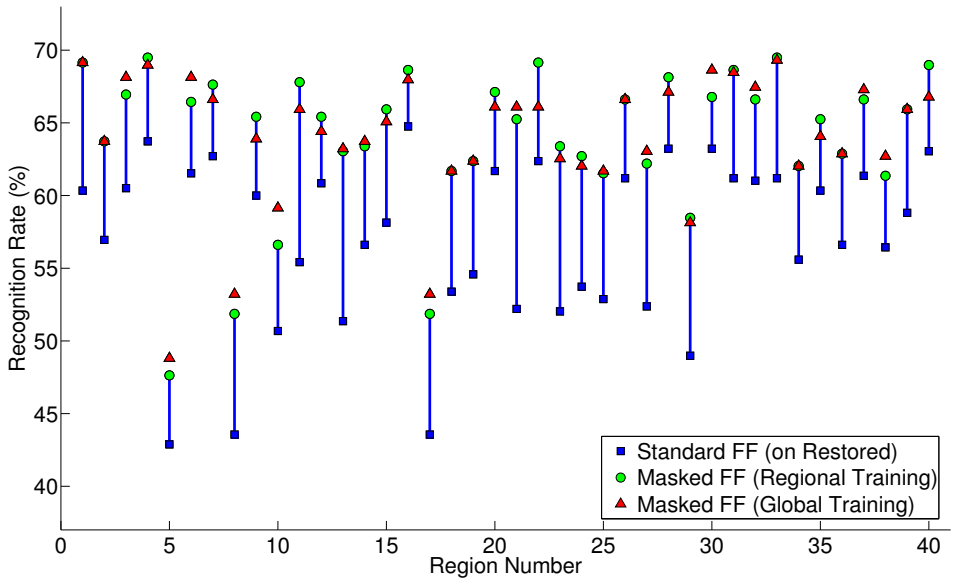


(b)

Figure 7.3. CMC plots for (a) the Bosphorus, and (b) the UMB-DB databases, with different approaches: (i) Standard Fisherfaces (FF) on restored images, (ii) masked FF with regional training, and (iii) masked FF where regional projection matrices are obtained by masking.



(a)



(b)

Figure 7.4. Regional recognition rates for (a) Bosphorus, and (b) UMB-DB. Blue lines indicate performance improvement of regionally trained masked FF over standard FF after restoration. Results with globally trained masked FF are included for comparative purposes.

regional results of masked training and masked projection approaches are compared in Figure 7.5, where manually labeled occlusion masks are utilized. In contrast to our expectations, the newly proposed masked projection strategy gives better recognition results for all of the 40 different regions: Since in the masked projection approach, the regional projection matrix is learned from the complete training regions, the relation between the original face space and the lower-dimensional subspace is represented better. Therefore, instead of training the projection matrices separately for each probe face, it is beneficial to obtain a complete regional projection matrix in an offline manner and then to compute the corresponding projection matrix using the occlusion mask. In addition, the masked projection strategy is a more feasible method: Instead of re-training a projection matrix separately for each probe face, the corresponding masked projection matrix is computed from the complete projection matrix.

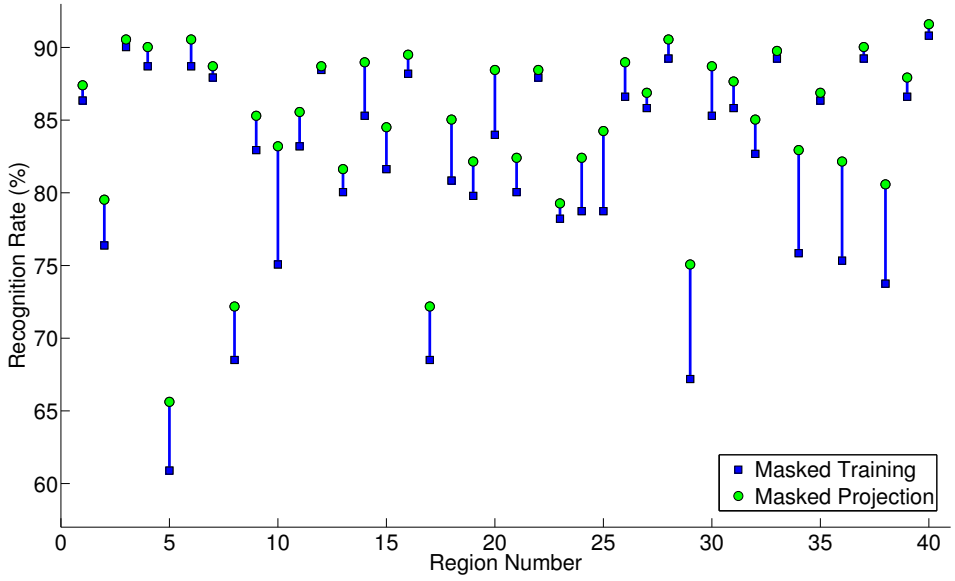


Figure 7.5. Regional recognition rates using manual occlusion masks for masked training and masked projection. The results are obtained for the Bosphorus database using manual occlusion masks, where training is handled at the regional level.

7.3.4. Effect of Occlusion Percentage on Performance

For evaluating the impact of occlusion on the recognition performance, we have analyzed the correctly and incorrectly identified samples for varying sizes of occluded areas.

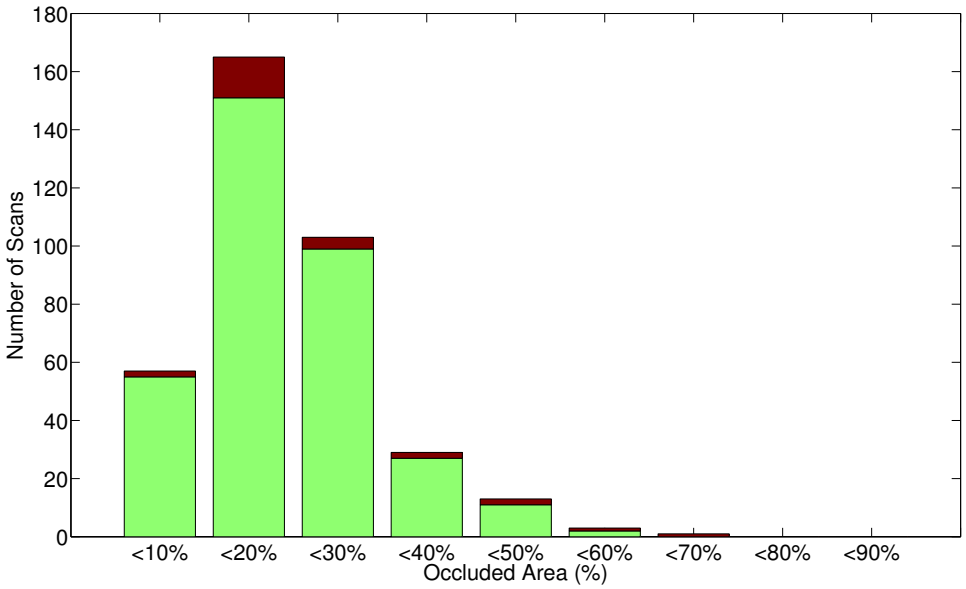
In Figure 7.6, the histograms of occlusion percentages are given for (a) the Bosphorus, and (b) the UMB-DB databases, where correctly and incorrectly classified sample counts are shown respectively in green and red. As these figures indicate, the UMB-DB database has more extensive occlusions when compared with the Bosphorus database. Nevertheless, both of the databases have some occlusions covering more than 50% of the facial area; and as the occluded areas expand, the recognition performances drop. However, when the highly occluded and correctly classified examples are investigated, it is clear that the proposed registration and recognition scheme serves as a viable approach: In Figure 7.7, highly occluded and correctly classified examples are given (with both manually labeled and automatically detected occlusion masks) for the Bosphorus (Figure 7.7a) and the UMB-DB (Figure 7.7b)[9] . Even when the nose area is partially occluded or the facial surface has low visibility, the facial surface can still be classified correctly.

It is interesting to note that an abnormality appears for the Bosphorus database, as can be seen in Figure 7.6a: The incorrectly classified examples for the occlusions up to 20% are more than the ones with larger occlusions. When the erroneous examples that are up to 20% occluded are checked, it is clear that the registration process fails: Some eye or mouth area occlusions cause small interruptions to the nose area. Although small in size, these occluded parts result in incorrect model selection and registration convergence.

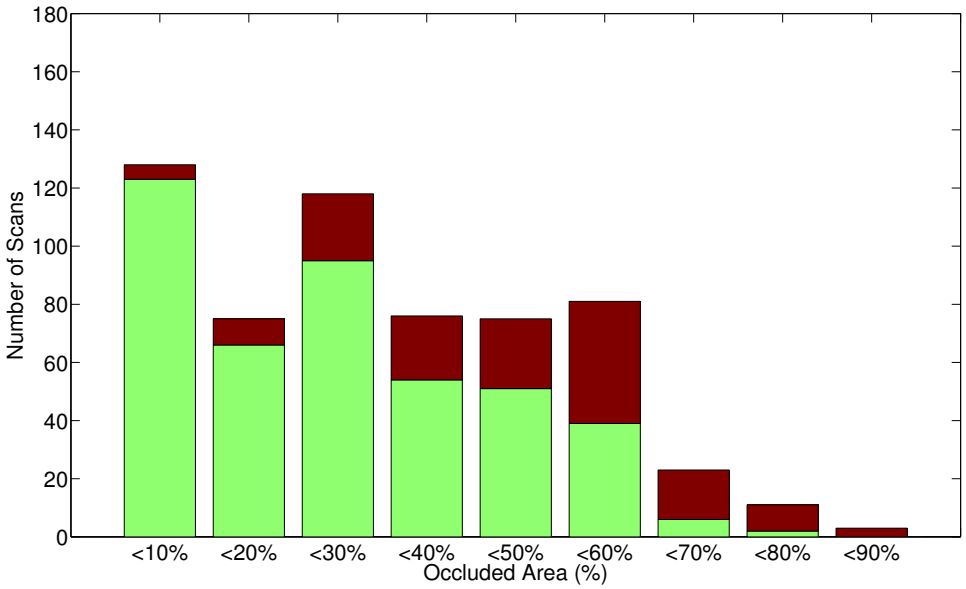
7.3.5. Comparison of Masked Projection with Different Occlusion Detectors

In the experiments reported until here utilized ground truth occlusion masks to give a better evaluation of the masked projection technique. In this section, we compare results obtained using different occlusion detectors to find the occlusion masks. Here, we have utilized three different automatic occlusion masks, that are detailed in Chapter 6: (i) Masks obtained by thresholding the difference from an average face model (BL); (ii) masks obtained by facial modeling with pixelwise GMMs (GMM); and (iii) masks obtained by the graph cut technique, where μ-σ modeling is used to set both the regional and the boundary weights (GC). The results for both the global and the regional masked projection strategies

[9]Note that, for the UMB-DB database, the occlusion masks are shown on the average face model, due to publishing constraints.

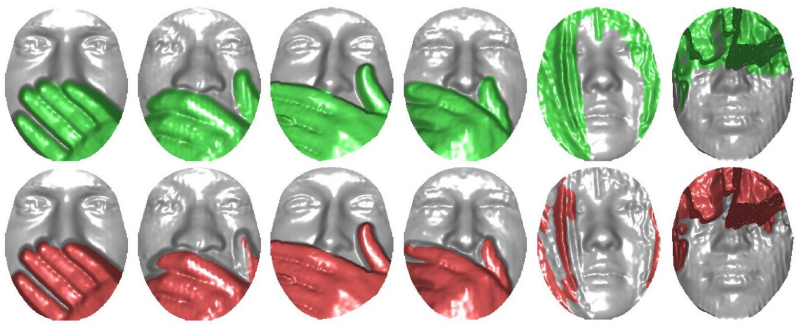


(a)

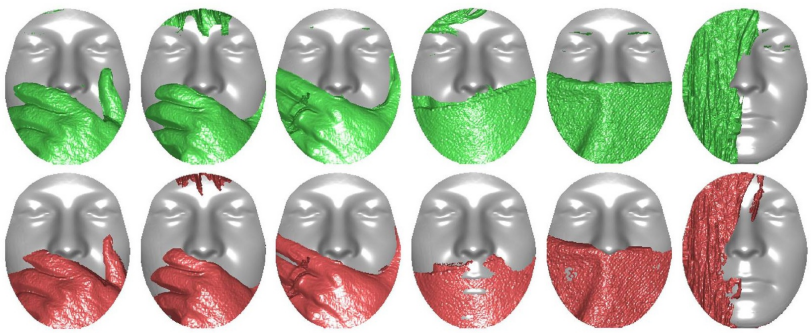


(b)

Figure 7.6. Occluded area histogram for (a) the Bosphorus, and (b) the UMB-DB database. Correctly (green) and incorrectly (red) classified sample counts are shown for different occlusion percentage ranges.



(a)



(b)

Figure 7.7. Highly occluded samples correctly classified by the proposed masked FF: Examples from (a) Bosphorus, and (b) UMB-DB, where top and bottom rows show corresponding manually labeled and automatically detected occlusion masks.

are reported in Table 7.3: (i) Using a single globally trained masked Fisherfaces, where the surfaces are masked with occlusion masks (global masked projection); (ii) using regionally trained masked Fisherfaces, where a set of regional projection matrices are learned and then masked by the occlusion mask (regional masked projection); and (iii) using globally trained masked Fisherfaces, where only a single projection matrix is learned and then masked by both the region and the occlusion mask (regional masked projection). When these results are inspected, we see that both of the proposed occlusion detectors (GMM and GC) perform superior when compared with the baseline (BL) approach. It is clear that incorporating neighborhood information (GC) yields better than using only regional cues (GMM). These conclusions are consistent with the results reported in Chapter 6. Furthermore, it is clear that the proposed classifier using masked projection outperforms the depth-based classifier used in Chapter 6, and further improvement is obtained by considering surfaces as a combination of multiple regions.

Table 7.3. Regional identification accuracies with the newly proposed Fisherfaces approach for different occlusion masks.

Method	Training Approach	Mask	Bosphorus	UMB-DB
Global Masked Projection (Occlusion Masking)	Global	GT	87.40	69.15
Regional Masked Projection (Occlusion Masking)	Regional	GT	93.18	73.90
Regional Masked Projection (Occlusion&Region Masking)	Global	GT + Regional	93.18	73.56
Global Masked Projection (Occlusion Masking)	Global	BL	83.73	65.25
Regional Masked Projection (Occlusion Masking)	Regional	BL	93.18	70.51
Regional Masked Projection (Occlusion&Region Masking)	Global	BL + Regional	92.91	68.47
Global Masked Projection (Occlusion Masking)	Global	GMM	87.66	66.27
Regional Masked Projection (Occlusion Masking)	Regional	GMM	92.91	72.03
Regional Masked Projection (Occlusion&Region Masking)	Global	GMM + Regional	92.13	70.68
Global Masked Projection (Occlusion Masking)	Global	GC	89.24	67.63
Regional Masked Projection (Occlusion Masking)	Regional	GC	92.65	71.86
Regional Masked Projection (Occlusion&Region Masking)	Global	GC + Regional	93.18	71.36

7.3.6. Different Fusion Schemes

Next, we have experimented with the fusion scheme, to check if the overall performances can be improved: Until now the product rule was used to merge the regional dissimilarity measures, and all of the regional results were employed in the fusion. However, since the regional surfaces are occluded, some regions will produce erroneous measures, and they can be discarded in the fusion stage for further improvement. Here, we have employed the confidence estimation technique of [79] (given in Section 4.2.3) to decide on the regional classifiers to be taken into account in fusion: For a probe face, the dissimilarity scores are first normalized and sorted in ascending order. Then a second normalization is performed such that the first score becomes one. After the second normalization, the second dissimilarity value denotes the slope between the normalized scores of the first two top-ranked classes. Therefore, this value defines the confidence of the classifier. Using this approach, separate confidence values are computed for regional classifiers for the considered probe face. In fusion, which we will refer to as the modified product rule, only the classifiers that have confidence values more than the preset threshold value are considered. In Table 7.4, masked projection results obtained with ground truth occlusion masks are given. Here, the fusion results obtained with the basic product rule (PROD), the modified product rule (MOD-PROD) are given. Additionally, we have reported results with the committee voting (CV) and the modified committee voting (MOD-CV) schemes that were introduced in Chapter 4. When the results are inspected, it is clear that the product and the modified product rule performs better than the voting schemes. For the results where confidence thresholding is employed with the product rule (MOD-PROD), a threshold of 0.75 is used. For the Bosphorus database, the results are not affected significantly by using the modified product rule, since the occlusions are small in size. For the UMB-DB database, which has more challenging occlusion variations, we were able to obtain up to 1.69% improvement by employing confidence values in the fusion stage, using the modified product rule.

7.3.7. Time Complexity

It is worth noting the overall time complexity of the proposed registration and classification scheme: In the registration stage, a single model-based ICP is necessary, where the

Table 7.4. Regional identification accuracies with different fusion schemes.

Method	Fusion Method	Bosphorus	UMB-DB
Proposed (Occlusion Masking)	CV	92.65	74.07
	MOD-CV	92.65	74.07
	PROD	93.18	73.90
	MOD-PROD	93.44	75.59
Proposed (Occlusion & Regional Masking)	CV	92.13	72.71
	MOD-CV	92.13	72.71
	PROD	93.18	73.90
	MOD-PROD	93.70	74.75

models include at most half of the whole facial surface. In the classification stage, the probe-specific projection is computed by simply masking the globally trained matrix. The most time-consuming part of the pipeline is the construction of the curvature map. The detailed average timings for processing a single test face, with an unoptimized MATLAB code running on a 64-bit Core i7 2.67GHz PC with 12GB RAM, are as follows: The nose detection stage, including the curvature map generation and template matching steps, takes about 21 seconds. Adaptive model selection and model-based registration takes a total of about 6 seconds. The subsequent occlusion detection stage is negligible (about 2 ms). The final masked projection and classification steps take a total of 3γ ms, where γ is the number of images in the gallery set, e.g., 315 and 423 ms for the Bosphorus and UMB-DB respectively[10] .

Since most of the time is consumed at the nose detection stage, we further examined the timing measures and checked if any improvements can be obtained. The most important factor influencing the computation durations, is the number of model points, which can be reduced by downsampling. Therefore, we have further analyzed the timing measures of our system, where the resampling rate of the depth maps is altered. For the nose detector, the resampling rate can be lowered to obtain significant computational time improvement, while maintaining the exact nose detection performance: If the regular resampling rate is lowered by a factor of 16 (a grid with four times larger step is employed), time consumption of the system can be lowered significantly without any performance degradation: Curvature map generation and template matching durations drop from 21 seconds to four seconds, where

[10]The classification time is given in terms of the size of the gallery set. Nevertheless, its contribution to the overall time can be kept low by parallelizing the distance computation.

the template matching takes about only 0.1 seconds. Although working on a low-resolution model is computationally beneficial without sacrificing the nose detection accuracy, we resort to a high resolution grid for subsequent stages. With a low resolution model, the alignment process cannot converge well and lower recognition rates are obtained due to worse registration, even when the resampling rate is lowered only by a factor of 1.5. Thus, sparser depth maps can be beneficial for the nose detection stage, whereas for the registration process, denser depth maps should be preferred.

7.3.8. Masked Projection for Other Acquisition Scenarios

The adaptive model-based registration and the masked projection approaches proposed in this paper are motivated by large occlusions causing a high proportion of missing points. In previous subsections, we have shown the viability of the proposed approach for occlusion scenarios. A natural question that arises is the applicability to other acquisition scenarios. To answer this question, we have conducted experiments on different subsets of the Bosphorus database: neutral, expression, and pose (up to 30 degrees) variations. As in the previous Bosphorus experiments, the gallery contains the first neutral image of each subject, and has a total of 105 scans. The neutral probe subset consists of 194 scans, whereas the expression variations are a total of 2620 faces. The Bosphorus database includes 13 pose variations for each of the 105 subjects, and six of the variations have extreme poses (45 or 90 degrees of rotation). The other seven pose types include: three yaw rotations (10, 20, and 30 degrees) and four pitch variations (slightly up, slightly down, up, down). For the experiments, the extreme poses labeled as 45 and 90 degrees are discarded, and the pose subset contains the remaining 734 scans.

Table 7.5 summarizes the performance of the two proposed approaches, namely the global and the regional (occlusion and region masking) methods, on neutral, occlusion, expression, and pose subsets. On the neutral subsets, we obtain 100% identification accuracy. We repeat the occluded subset results here for comparative purposes: For global and regional methods, recognition accuracies of 83.73% and 92.91% are obtained respectively, where automatically detected occlusion masks are employed. The performance on the expression subset is higher: 88.24% and 95.04%. The last column shows the performance on

the challenging pose subset: 85.83% and 88.15%. As these results indicate, our proposed system can be directly applied to neutral and expression scans. The results obtained on the expression subset are similar to the results of [26], for which the system is implemented directly for expression handling. The herein proposed approach is advantageous for expression variations, when both the registration and classification methodologies are considered: The idea of adaptive selection of the alignment model is beneficial for expression variations, since some patches can have extreme surface deformations and can mislead the alignment process. By using patch validity values, patches with expressive deformations are discarded from the registration process. Furthermore, occlusion detector automatically finds regions that do not resemble a neutral face. Therefore, expressive deformations are discarded from the classification process. When pose variations are considered, acceptable recognition rates are achieved. When the misclassified examples are examined, we saw that some of the faces exhibit pose variations greater than 30 degrees due to mislabeling during the database acquisition: For around 70% of the incorrect classifications, the face was altered from the frontal pose by more than 20 degrees. When handling *extreme* pose variations, the bottleneck of the system is the registration process: Since patch templates are obtained from the frontal average model, patch localization accuracy will be degraded for images with extreme rotations. Furthermore, the ICP algorithm will not be able to converge. If the facial surfaces are registered correctly to the adaptively selected alignment model, the classification stage can directly be applied: The mask detection procedure will easily and correctly locate the missing parts of the facial surface. With accurate occlusion masks defining the missing parts, the subsequent masked projection will perform well. As the results in Table 7.5 indicate, even though the system proposed is especially for occlusion variation handling, acceptable recognition results are obtained for other acquisition challenges such as expression and pose variations. Furthermore when results for each scenario are examined, it is clear that considering the faces as a combination of multiple regions further improves the recognition performance.

Table 7.5. Identification accuracies of masked projection on Bosphorus neutral, occlusion, expression, and limited pose subsets.

	Training		Classification	Bosphorus Subsets			
Method	Approach	Mask	Approach	Neutral	Occlusion	Expression	Pose
Proposed	Global	Automatic	Global	100.00	83.73	88.24	85.83
Proposed	Global	Automatic + Regional	Regional	100.00	92.91	95.04	88.15

7.4. Conclusion

In this chapter, we introduced the proposed *masked projection* approach, which incorporates a masking scheme into a subspace analysis technique, namely the Fisherfaces, to enable applicability to incomplete data. Subspace training is handled offline; and at the classification stage, the occlusion mask of the probe face is applied to the projection matrix. The masked projection matrix is used to project the gallery set and the probe face to the corresponding subspace, and identification is achieved by 1-nearest neighbor classifier. To further improve the overall identification performance, a regional classification scheme is employed: The facial surfaces are considered as a collection of overlapping regional parts; and each region acts as a separate classifier. In the regional level, the masked projection is applied in two different strategies: (i) each regional subspace is trained; and (ii) the regional subspaces are obtained by applying the region mask to the global projection matrix. As the experimental results indicate, by masked projection an improvement up to 14% can be achieved at the regional level. Furthermore, employing the masking approach to obtain regional subspaces appears as a viable alternative over regional training. Additionally, the proposed system can be directly applied to handle expression and small pose variations.

The proposed system is able to work with good performance under substantial occlusions, expressions, and small pose variations. When we examine the failures, we see that if occlusions are so large that the nose area is totally invisible, the initial alignment becomes impossible. Similarly, if the face is rotated by more than 30 degrees, it becomes difficult to accomplish the initial alignment.

8. CONCLUSION

8.1. Contributions and Discussion

Three dimensional face recognition has become an emerging biometric technique, due to advances in the sensor technology. For applicability to security systems, uncooperative scenarios should be considered, where pose, expression, or occlusion variations can complicate the task of identifying people from their facial data. In particular, when the occlusion challenge is considered, any exterior object can be used easily to mislead an identification system. Nevertheless, there are only a few studies in the 3D face recognition literature considering occlusion variations.

In this thesis, we have developed a fully automatic face recognizer, which is robust under the presence of occlusion variations. We have utilized the 3D modality, since the actual face geometry is in 3D space and occlusions can be better handled using the 3D information. The overall face recognition system is composed of three main parts: (i) Registration, (ii) occlusion detection and handling, and (iii) feature extraction and classification. For each module, we have proposed novel methods to handle the occlusion challenge at different stages of the system.

Registration:

Before any two facial surfaces can be compared for identification, they should be aligned to each other and a dense correspondence should be obtained. Therefore, registration plays a vital role in any face recognition system. Iterative Closest Point algorithm is a widely used method to rigidly align two surfaces and to obtain the point-to-point correspondence. Furthermore, incorporating a model-based registration to the ICP method, computational cost of the registration module can be highly reduced. However, for iterative techniques like ICP, good initialization is necessary. In traditional registration approaches, landmark points are located on the facial surface to guide the coarse alignment of surfaces. However, when occlusions are present over the face, landmark localization methods become inapplicable.

Therefore, instead of localizing fiducial surface points, we propose to detect fiducial *areas* over the surface, such as the nasal region: The nose detection technique based on curvature information serves as an efficient face localization approach even when the nose area is more than 50% occluded. Coarse alignment based on nose works sufficiently well even for scans with pose variations up to 30 degrees of yaw.

Initialization is not the only problem of registering occluded scans. Even though a sufficient coarse alignment is achieved, ICP cannot be directly applied on occluded scans to obtain a fine alignment: Occluded surface points will mislead the distance computations and the algorithm will not be able to converge to a correct point-to-point correspondence. Therefore, it is necessary to discard the occluded parts from the fine registration process. In this thesis, we propose a model-based registration technique, where a patch-based model is selected according to the non-occluded parts of the surface to be registered. The registration module coarsely detects additional fiducial regions (such as eyes and mouth), and checks their validity using curvature information. Therefore, nonoccluded patches will be automatically detected as valid, and the validity measures will be used to select an appropriate alignment model. Since after initialization, the occluded surface and the model are coarsely aligned, aligning the occluded surface to the adaptive model automatically enables to use non-occluded points for correspondence establishment. Therefore, without accurately detected occluded pixels, it is possible to obtain a fine registration.

Although registration performances obtained with different models (face, nose, or adaptive model) point out a significant improvement with the proposed adaptive technique, an erroneous alignment will result in problems in the subsequent steps of occlusion handling and classification. Therefore, the registration module is the bottleneck of the proposed system. The main weakness of this module is its dependence on the nose area. If the nose area is incorrectly detected, initialization and patch validations steps will fail.

Occlusion Detection and Handling:

In traditional face recognition systems, after registration, feature extraction and classification steps can be performed. However, when occlusions are present, occluded surface

points should be accurately labeled. In this thesis, we have proposed two different occlusion detection techniques: The first occlusion detector is based on a statistical model, where each facial surface point is represented using pixelwise Gaussian Mixture Models. Then, each point on the occluded surface is checked for fitness to the corresponding GMM. The surface points that cannot be well represented by the model are labeled as occluded. However, this detector employs only pixelwise relations. In the second detector, we propose to incorporate neighboring pixel-pair relations into pixelwise models. By employing neighborhood relations, the boundaries of the occlusions can be better detected.

It should be noted that the performance of the occlusion detection is significantly dependent on the accuracy of the preceding registration step. If a sufficiently well alignment is not achieved, neither the regional nor the boundary cues will be well represented due to shifts in pixel locations.

After occlusion detection, we have evaluated two occlusion handling approaches: occlusion removal and surface restoration. Although restoration can yield visually good results, the restored surfaces are not appropriate for classification purposes. As the experiments have validated, instead of restoring missing parts, the feature extraction module should handle incomplete data.

Feature Extraction and Classification:

Following the occlusion detection stage, the facial parts that are labeled as occlusions are removed to obtain occlusion-free surfaces. However, these surfaces are incomplete, and feature extraction and classification techniques cannot be directly applied. In this thesis, we have proposed a technique called *masked projection*, which incorporates a masking scheme into a subspace analysis technique, enabling to extract features from incomplete data if the missing parts are labeled. Subspace training is handled offline using complete surfaces. In the classification stage, masked projection enables to use the occlusion mask of the occluded probe face while computing the appropriate subspace from the trained subspace. Projections to the subspace constructed with the occlusion mask, constitute the feature space specific for that mask. Therefore, gallery images are also projected to this feature space to obtain

classification.

The proposed masked projection is a novel technique to obtain subspaces from partial data: Instead of training masked gallery images to learn the projection, the initially learned subspace is masked. Therefore, using occlusion-specific subspaces becomes computationally feasible. Moreover, if the non-occluded part is quite small when compared to the whole surface, accurate training cannot be obtained using masked training images. However, in the masked projection approach, the subspace training is handled accurately using the whole facial surface, and the subsequent masking yields better representation specific for the occlusion mask. In addition, the masking approach can be applied to obtain regional projection matrices. Moreover, our experiments have shown that this approach performs better than the gappy approach. Although the idea in gappy projection is similar, it is based on an incorrect assumption that the masking gives the originally trained subspace.

Although we have utilized masked projection for classification of occlusion-free faces in this thesis, this technique can be used with any subspace technique to obtain dimensionality reduction or feature extraction of incomplete data.

8.2. Future Directions

As stated above, there are some weaknesses of the proposed system and some basic modifications can result in performance improvement.

The main weakness of the system is in the registration module, and the performance of the subsequent steps are highly affected by the accuracy of the alignment procedure. Therefore, an improvement in registration, would boost the overall system performance: The proposed system can be further improved by applying the adaptive model-based registration and occlusion detection/removal stages in an iterative manner. At each iteration, the patch detection and validation steps should be performed to decide on the most appropriate model.

The registration performance of the current system can be improved by considering multiple average face models: Since faces have varying size, different sized face models will

enable better convergence of the alignment procedure. A better alternative for the decision of multiple templates, would be to include cross-cultural models, since different cultural averages can be considered to represent varying sized and shaped prototypes. Furthermore, use of multiple models can be beneficial if pose variations are considered: Rotated versions of an average face be obtained and used to align facial surfaces with pose variances. Further improvement in the registration module can be achieved by considering additional local regions (besides nose) in the alignment initialization step. If the nose area is highly occluded, nose detection can fail, and checking for additional local regions can be beneficial.

Improving the occlusion detection module can result in better classification accuracy. For the GMM-based occlusion detector, the number of mixtures used for modeling is predefined and is held constant for each pixel. However, variations for each pixel is different. Therefore, the occlusion detection performance can be enhanced by learning the mixture size separately for every surface pixel. The other occlusion detector, which is based on the graph cut technique, uses simple mean-standard deviation models to represent both the regional and the boundary cues. For further improvement, GMMs can be embedded instead of the simplistic representation used for pixelwise and neighborhood modeling. Moreover, since neighborhood relations are considered in the graph cut technique, it can be beneficial to develop a multi-resolution system: The neighborhood relations will change with varying resolution, and considering a collection or a hierarchy of graphs representing faces at different resolutions will lead to a better representation of surface coherency.

In addition to the improvements to be considered at the algorithmic level, there are also some future directions to follow for better evaluation of the modules. We have drawn conclusions, experimenting on the available occlusion databases, namely the Bosphorus and the UMB-DB databases. We have evaluated registration performances using several alignment models. However, since we do not have ground truth registration transformations, we cannot evaluate how well the proposed registration module performs. Moreover, the subsequent modules are greatly affected by the performance of the alignment, and we cannot evaluate the performance of specific modules accurately. Therefore, it is necessary to obtain the ground truth registration transformations.

In addition to improving the current system, alternative modifications in the overall setup can be considered to obtain better performances. The most important change to consider is the ordering of the registration and occlusion detection stages: Instead of registering occluded surfaces and then detected occluded parts on the registered faces, a better alternative would be to detect occlusions prior to registration. However, accurate detection occluded pixels prior to surface alignment is not a straightforward task: Regional or neighborhood models proposed in this work, are no longer valid in that situation. Instead, edge detectors can be employed to detect large depth differences between neighboring pixels. On the edge map, the expected edges such as nose borders can be eliminated, such that only the occlusion boundaries are left. This boundary information can be beneficial to detect the occluded regions. If the occluded regions can be detected accurately prior to registration, alignment procedure should be modified to cope with partial surface data.

Another future direction would be to consider using soft labeling of the detected occlusion masks: Instead of labeling pixels as either occlusion or not, a soft weighing scheme can be incorporated to represent the detection confidence. To be able to incorporate soft masks into the classification stage, the masked projection approach should be adapted accordingly: The proposed masked projection approach utilizes a binary mask. Hence, it should be modified to work with a soft-valued mask.

For better applicability to real-life scenarios, the system can be modified to implement an open-set identification scenario: Since the occlusion problem can be an important problem for a watch-list scenario, the probe should not be restricted to be among the gallery subjects. Therefore it would be better to modify the overall face recognizer, making it applicable to such a an open-set application.

APPENDIX A: FACTORIAL DESIGN EXPERIMENTS

In the statistics literature, a factorial design is an experiment where at least two factors (parameters) are involved, and each of the factors has discrete "levels" (values) [92, 93]. The experiment includes a several number of replicates (experimental units), where each unit is tested on all possible level combinations of all factors. A fully crossed experiment designed in such a way, enables us to study both the *main effect* of each of the factors and the *interactions* in between the factors.

In this thesis, factorial design is employed to find the optimum parameter set for the occlusion detector based on graph cut technique, which was explained in detail in Section 6.1.3. In summary, the facial surface is represented as a graph, where pixelwise and neighboring pixel pair-wise models are utilized to define regional and boundary cues. Using these cues, occlusion detection is regarded as a binary image segmentation problem, where graph cuts is utilized to find the optimum s-t cut giving the occlusion and face segments. The regional and boundary relations are expressed in terms of t-link and n-link weights; and when setting these weights, we have utilized constants, namely $\kappa^{(H)}$ and $\kappa^{(B)}$. Furthermore, to give a relative importance either to regional or boundary cues, we have employed another constant, namely λ. Eventually, we have a set of three parameters, where each constant should be set empirically, optimizing the overall occlusion detection performance. The most trivial approach to find the best settings for the constants, would be to keep two of the parameters at some fixed values, while changing the other and finding the optimum value for that parameter. Finding the optimum value for each of the parameters in this way, we would hope to find the best combination for the whole parameter set. However, this approach is based on the assumption that the parameters are *independent* of each other. Instead, for a better analysis of the parameters, we have designed a *full factorial experiment*, where all crossed combination of parameters are considered.

A.1. Analysis of Variance Table for Factorial Design

In our optimization problem, we have a set of three factors: Let's refer to these factors ($\kappa^{(H)}$, $\kappa^{(B)}$, and λ) as A, B, and C, respectively. Initially, we have selected a discrete set of possible levels for each of the factors, where a feasible number of values considered to span the range of appropriate settings. By fixing all factors to some levels, main effects and interactions can be checked using ANOVA (ANalysis Of VAriance) [92, 93]. If A, B, and C factors have a, b, and c levels, respectively, then a single replicate will include a total of abc combinations. If we denote the number of replicates by n, the total number of observations is $abcn$. Test statistics for a main effect or an interaction can be constructed as follows: First, the sum of squares for the corresponding effect or interaction is divided by the respective degrees of freedom to give the corresponding mean square. Then, the test statistic of the main effect or the interaction term is computed by dividing the corresponding mean square by the mean square error. The general ANOVA table is given in Table A.1. The test statistics computed from the experimental measurements are then compared from F distribution tables to evaluate significance of the main effects and the interactions.

Table A.1. The ANOVA table for the three-factor fixed effects model.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Test Statistics
A	SS_A	$a-1$	MS_A	$F_0 = \frac{MS_A}{MS_E}$
B	SS_B	$b-1$	MS_B	$F_0 = \frac{MS_B}{MS_E}$
C	SS_C	$c-1$	MS_C	$F_0 = \frac{MS_C}{MS_E}$
AB	SS_{AB}	$(a-1)(b-1)$	MS_{AB}	$F_0 = \frac{MS_{AB}}{MS_E}$
AC	SS_{AC}	$(a-1)(c-1)$	MS_{AC}	$F_0 = \frac{MS_{AC}}{MS_E}$
BC	SS_{BC}	$(b-1)(c-1)$	MS_{BC}	$F_0 = \frac{MS_{BC}}{MS_E}$
ABC	SS_{ABC}	$(a-1)(b-1)(c-1)$	MS_{ABC}	$F_0 = \frac{MS_{ABC}}{MS_E}$
Error	SS_E	$abc(n-1)$	MS_E	
Total	SS_T	$abcn-1$		

A.2. Response Surface Fitting

In the factorial design experiments, it is possible to determine factors with main effect and any substantial interaction in between the factors. After finding which elements are

important, it is usually of interest to model the relationship between the factors and the response (output). The relationship in between can be expressed by a mathematical model, namely the *regression model*. Using a set of training data, we can fit a linear regression model and compute the best factor combination to be used with the probe data. Suppose that as a result of the factorial design, we concluded that two of the factors (say A and B) have main effects and they interact with each other. The third factor has no effect and it does not interact with the other factors[11] . Therefore, a second-order model with two regression variables (namely x_1 and x_2 for factors A and B, respectively) can be used to describe the relationship between the parameters and the response (denoted as y):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 \tag{A.1}$$

Here, the parameters β_j, $j \in 0, 1, 2, 11, 12, 22$, are the regression coefficients. The main effects will be represented by the β_1 and β_2 parameters, whereas β_{12} is for the interaction term. The second-order terms of β_{11} and β_{22} are included so that the response surface can better fit to the given data. This model can be written in terms of the observations,

$$\begin{aligned} y_i = \quad & \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_{12} x_{1,i} x_{2,i} \\ & + \beta_{11} x_{1,i}^2 + \beta_{22} x_{2,i}^2 + \boldsymbol{\epsilon}_i \end{aligned} \tag{A.2}$$

where $i = 1, 2, \ldots, n$ represents the observation number. The error term $\boldsymbol{\epsilon}_i$ is the difference between the observed response and its actual value. The aim of least squares method is to choose the β parameters such that the sum of the squares of the errors is minimized. To find the solution, the model in terms of the observations can be expressed in matrix notation as follows:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{A.3}$$

[11] For our parameter set, factorial design experiments showed that $\kappa^{(B)}$ factor has no effect and it does not interact with the other factors. The detailed experimental analysis is included in the Experimental Results section (Section A.3.2)

where

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_{12} \\ \beta_{11} \\ \beta_{22} \end{bmatrix}, \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}, \text{and}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & x_{1,1}x_{2,1} & x_{1,1}^2 & x_{2,1}^2 \\ 1 & x_{1,2} & x_{2,2} & x_{1,2}x_{2,2} & x_{1,2}^2 & x_{2,2}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1,n} & x_{2,n} & x_{1,n}x_{2,n} & x_{1,n}^2 & x_{2,n}^2 \end{bmatrix}. \tag{A.4}$$

In general, $\boldsymbol{y}$ is the vector of observed responses, $\boldsymbol{X}$ is constructed using the levels of the factors, $\boldsymbol{\beta}$ is the vector of regression coefficients to be estimated, and $\boldsymbol{\epsilon}$ is the vector of random errors. By least squares estimation, our aim is to find $\hat{\boldsymbol{\beta}}$ minimizing the squared error, which is given as

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{A.5}$$

If the above equation is written in open form, the least squares estimators satisfying the partial derivative equation $\frac{\partial L}{\partial \beta} = \mathbf{0}$, can be given as

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}. \tag{A.6}$$

The least squares estimates can then be used as the optimal parameter values in further occlusion detection experiments.

A.3. Experimental Results

A.3.1. Database

For optimization of parameters employed in occlusion detection (namely $\kappa^{(H)}$, $\kappa^{(B)}$, and λ), we have selected a subset of 70 facial surfaces from the Bosphorus database: These samples are selected such that the ground truth occlusion masks are accurately labeled, which is important when evaluating the performance of occlusion detectors via F-measure. Furthermore, we have selected a subset so that some examples are challenging for occlusion detection, whereas some can be easily detected by basic occlusion detectors (e.g. by the baseline approach given in Section 6.1.1). Throughout this thesis, this subset is referred to as the Bosphorus-70 subset.

A.3.2. Factorial Design Experiments and Occlusion Detector Evaluation

The experiments for the factorial design and response surface modeling are run on the Bosphorus evaluation subset, where each of the facial surfaces stands for an experimental unit. Therefore, we have $n = 70$ replicates. On this subset, a set of initial experiments are run to limit the search space of parameter values and select a set of fixed levels for each of the factors, namely $\kappa^{(H)}$, $\kappa^{(B)}$, and λ. In Table A.2, the set of fixed levels are given for the factors.

Table A.2. The factors and sets of levels considered in the factorial experiments.

Factor	Number of Levels	Levels
$\kappa^{(H)}$	$a = 6$	$5 : 10$
$\kappa^{(B)}$	$b = 5$	$3 : 7$
λ	$c = 12$	$0.1, 0.2, 0.5, 1 : 9$

Using the levels of factors given in Table A.2 and F_1 measures as response outputs, factorial design experiments are held to obtain the ANOVA table given in Table A.3. As the results in the table indicate, the $\kappa^{(B)}$ factor has no main effect, whereas the other two factors,

$\kappa^{(H)}$ and λ, have main effects and they also interact with each other.

Table A.3. The ANOVA table for the three-factor fixed effects model.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	Test Statistics	Prob¿F
$A\,(\kappa^{(H)})$	0.014	5	0.0027	0.45	0.9878
$B\,(\kappa^{(B)})$	12.047	4	3.0118	170.48	0
$C\,(\lambda)$	8.488	11	0.7716	6.72	0
AB	0.096	20	0.0048	0.13	1
AC	0.093	55	0.0017	0.09	1
BC	5.954	44	0.1353	1.99	0.9999
ABC	0.080	220	0.0004	0.01	1
Error	560.855	24840	0.0226		
Total	587.627	25199			

Before adopting conclusions from the ANOVA, we should check the adequacy of the underlying model by residual analysis. Therefore, the residuals are computed as,

$$e_{ijkt} = y_{ijkt} - \bar{y}_{ijk.} \tag{A.7}$$

where i, j, and k indices stand for levels of factors $\kappa^{(H)}$, $\kappa^{(B)}$, and λ, respectively. The t index is for the observation number. Here, $\bar{y}_{ijk.}$ represents the average of the observations in the ijkth cell. The normal probability plot of residuals are given in Figure A.1: This figure shows clearly that the residuals are not from a normal distribution. In Figure A.2a, the residuals are plotted versus fitted values (which is given by the observation mean $\bar{y}_{ijk.}$): It is evident from the figure that the variance decreases as the fitted value increases. In Figure A.2b, A.2c, and A.2d, the residuals are plotted versus each of the factors: As these figures demonstrate, $\kappa^{(B)}$ has no main effect, since the respective residual plot does not reveal anything troublesome. On the other hand, the variance decreases for the middle values of both $\kappa^{(H)}$ and λ.

Through the factorial experiments, we concluded that only two of the parameters, namely $\kappa^{(H)}$ and λ, have significant effect and substantial interaction in between. We can now move on to response surface modeling to decide on parameter levels to be used in further occlusion

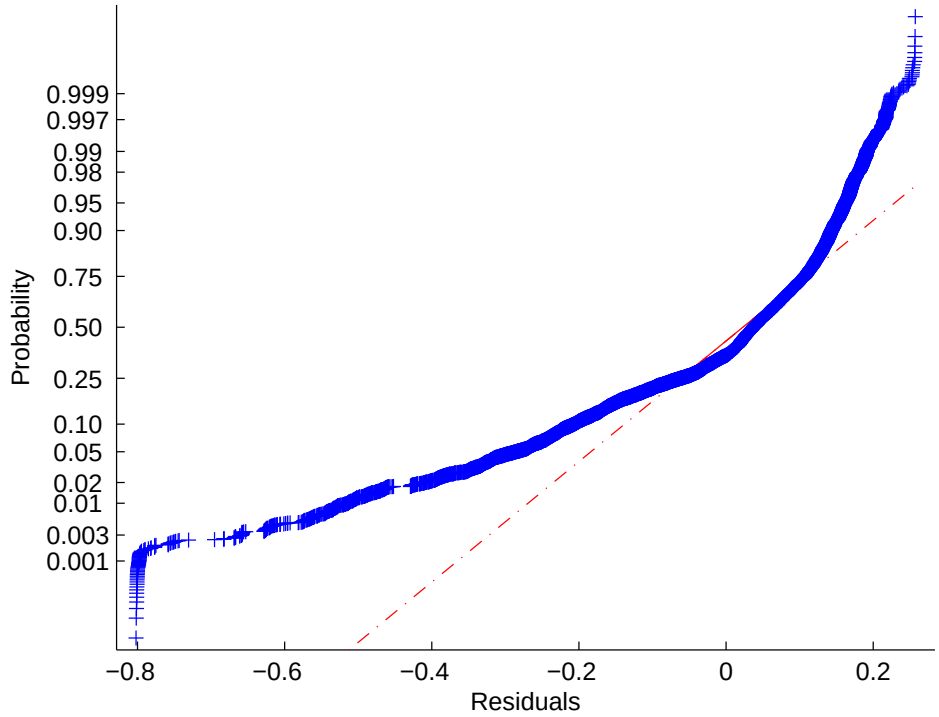
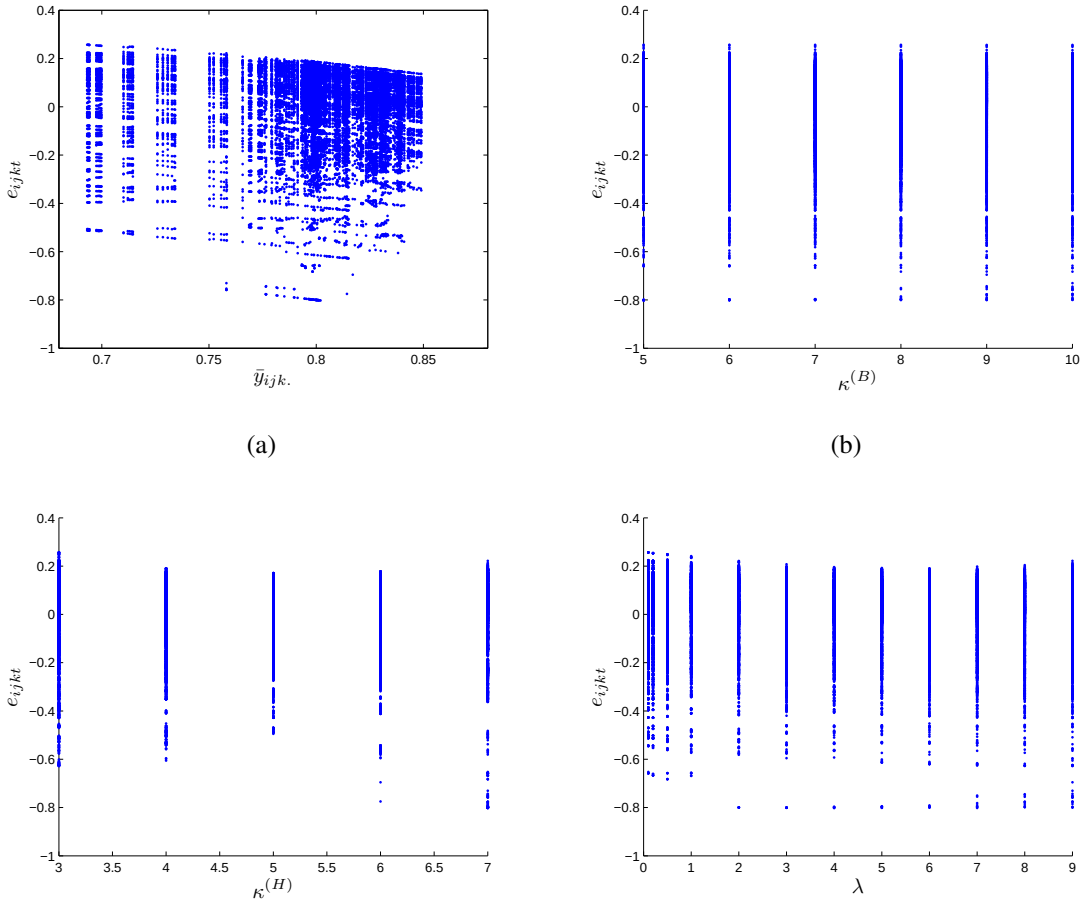


Figure A.1. Normal probability plot of residuals.

detection experiments. For modeling the response surface, the output response is modified for a better representation: Since the training facial surfaces include different occlusions at various different locations with varying sizes, the occlusion detector on different faces can have divergent performance. Therefore, instead of considering each face as a separate replicate, we decided on to use the average F_1 measure over the whole training set as a single replicate. In order to have multiple replicates, the average F-measures at different levels of the $\kappa^{(B)}$ factor are considered as different experimental units. Therefore, for the response surface modeling, the number of replicates is $n = 6$. Since there are two main factors and their interactions, the surface can be modeled by the second-order model given in Eq. A.1. Another model that can be used is,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 \tag{A.8}$$

where only the main effects and interactions are considered. However, the experiments show that when the second-order terms are included, the surface is better modeled. The normal probability plots of the residuals are given for both of the models in Figure A.3: From these

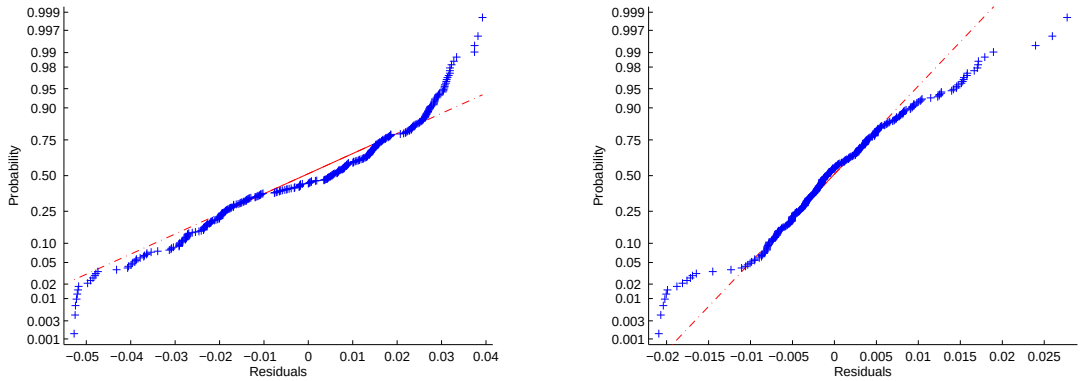


(a) (b)

(c) (d)

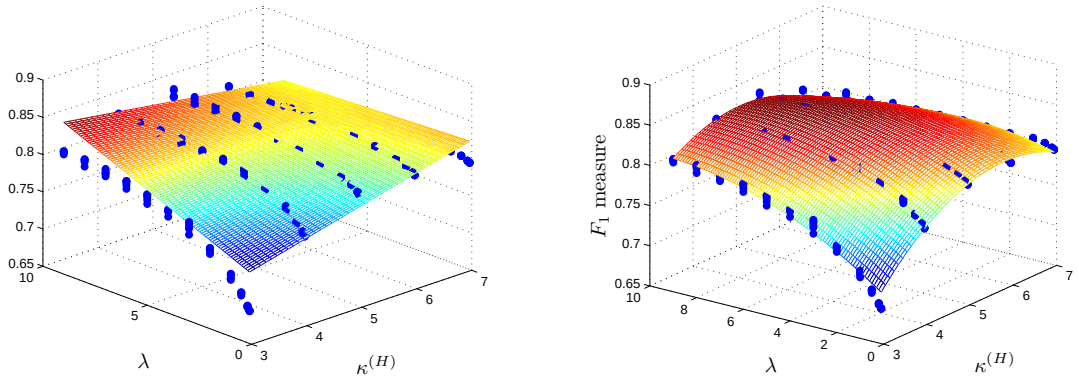
Figure A.2. The residual plots: residuals versus (a)fitted values, (b) $\kappa^{(H)}$ levels, (c) $\kappa_{(B)}$ levels, (d) λ levels.

figures, it is evident that the model with the second order terms gives a better representation of the response. Furthermore, the response surfaces plotted together with the observed responses are given in Figure A.4. Once again, it is clear that the model given in Eq. A.1 should be preferred.



(a) (b)

Figure A.3. The normal probability plot of the residuals are given when the response surface is modeled with: (a) only the main effects and the interaction term, and (b) when the second-order terms are included in addition.



(a) (b)

Figure A.4. The response surfaces plotted together with the observed response values, where the model includes: (a) only the main effects and the interaction term, and (b) the second-order terms are included in addition to the main effects and the interaction.

Using the second-order response surface model visualized in Figure A.4b, the values of

parameters $\kappa^{(H)}$ and λ maximizing the response (F_1-measure) is selected as: $\kappa^{(H)} = 4.8$, $\lambda = 7.0$. We have set $\kappa^{(B)} = 5$, where different levels have no significant effect. These settings are used for occlusion detection experiments via graph cut technique given in Chapter 6.

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", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 5, pp. 831–846, 2007.

2. Zhao, W., R. Chellappa, P. J. Phillips and A. Rosenfeld, "Face Recognition: A Literature Survey", *ACM Computing Surveys*, Vol. 35, No. 4, pp. 399–458, 2003.

3. Phillips, P., P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. Worek, "Overview of the Face Recognition Grand Challenge", *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 947–954, 2005.

4. Alyuz, N., B. Gokberk, L. Spreeuwers, R. Veldhuis and L. Akarun, "Robust 3D Face Recognition in the Presence of Realistic Occlusions", *Proceedings of International Conference on Biometrics*, pp. 111–118, 2012.

5. Gokberk, B., M. O. Irfanoglu and L. Akarun, "3D Shape-based Face Representation and Feature Extraction for Face Recognition", *Image and Vision Computing*, Vol. 24, No. 8, pp. 857–869, 2006.

6. Alyuz, N., B. Gokberk and L. Akarun, "Adaptive Model based 3D Face Registration for Occlusion Invariance", *Proceedings of European Conference on Computer Vision - Workshops*, 2012.

7. Friedman, N. and S. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach", *Proceedings of Conference on Uncertainty in Artificial Intelligence*, pp. 175–181, 1997.

8. Boykov, Y. Y. and M.-P. Jolly, "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in ND Images", *Proceedings of International Conference on*

Computer Vision, Vol. 1, pp. 105–112, 2001.

9. Alyuz, N., B. Gokberk and L. Akarun, "Detection of Realistic Facial Occlusions for Robust 3D Face Recognition", *International Conference on Computer Vision - Workshops*, 2013 (submitted).

10. Everson, R. and L. Sirovich, "Karhunen–Loeve Procedure for Gappy Data", *Journal of the Optical Society of America A*, Vol. 12, No. 8, pp. 1657–1664, 1995.

11. Colombo, A., C. Cusano and R. Schettini, "Gappy PCA Classification for Occlusion Tolerant 3D Face Detection", *Journal of Mathematical Imaging and Vision*, Vol. 35, No. 3, pp. 193–207, 2009.

12. Alyuz, N., B. Gokberk and L. Akarun, "3D Face Recognition under Occlusion Masked Projection", *IEEE Transactions on Information Forensics and Security*, Vol. 8, No. 5, pp. 789 – 802, 2013.

13. Turk, M. and A. Pentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71–86, 1991.

14. Belhumeur, P., J. Hespanha and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using Class Specific Linear Projection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, pp. 711–720, 1997.

15. Savran, A., N. Alyuz, H. Dibeklioglu, O. Celiktutan, B. Gokberk, B. Sankur and L. Akarun, "Bosphorus Database for 3D Face Analysis", *Biometrics and Identity Management*, pp. 47–56, 2008.

16. Colombo, A., C. Cusano and R. Schettini, "UMB-DB: A Database of Partially Occluded 3D Faces", *Proceedings of International Conference on Computer Vision - Workshops*, pp. 2113–2119, 2011.

17. 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, No. 1, pp. 1–15, 2006.

18. Scheenstra, A., A. Ruifrok and R. C. Veltkamp, "A Survey of 3D Face Recognition Methods", *Proceedings of International Conference on Audio-and Video-Based Biometric Person Authentication*, pp. 891–899, 2005.

19. Abate, A. F., M. Nappi, D. Riccio and G. Sabatino, "2D and 3D Face Recognition: A Survey", *Pattern Recognition Letters*, Vol. 28, No. 14, pp. 1885–1906, 2007.

20. Gokberk, B., A. A. Salah, L. Akarun, R. Etheve, D. Riccio and J. L. Dugelay, "3D Face Recognition", D. Petrovska-Delacretaz, G. Chollet and B. Dorizzi (Editors), *Guide to Biometric Reference Systems and Performance Evaluation*, pp. 1–33, Springer, 2008.

21. Abate, A. F., S. Ricciardi and G. Sabatino, "3D Face Recognition in a Ambient Intelligence Environment Scenario", K. Delac and M. Grgic (Editors), *Face Recognition*, I-Tech, 2007.

22. Papatheodorou, T. and D. Rueckert, "3D Face Recognition", K. Delac and M. Grgic (Editors), *Face Recognition*, I-Tech, 2007.

23. Faltemier, T. C., K. W. Bowyer and P. J. Flynn, "A Region Ensemble for 3D Face Recognition", *IEEE Transactions on Information Forensics and Security*, Vol. 3, No. 1, pp. 62–73, 2008.

24. Mian, A., 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, 2008.

25. 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, 2007.

26. Alyuz, N., B. Gokberk and L. Akarun, "Regional Registration for Expression Resistant 3-D Face Recognition", *IEEE Transactions on Information Forensics and Security*, Vol. 5, No. 3, pp. 425–440, 2010.

27. Queirolo, C. C., L. Silva, O. R. P. Bellon and M. P. Segundo, "3D Face Recognition using Simulated Annealing and the Surface Interpenetration Measure", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 2, pp. 206–219, 2010.

28. Spreeuwers, L., "Fast and Accurate 3D Face Recognition", *International Journal of Computer Vision*, pp. 1–26, 2011.

29. Ming, Y. and Q. Ruan, "Robust Sparse Bounding Sphere for 3D Face Recognition", *Image and Vision Computing*, Vol. 30, No. 8, pp. 524–534, 2012.

30. Park, J., Y. Oh, S. Ahn and S. Lee, "Glasses Removal from Facial Image using Recursive Error Compensation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 805–811, 2005.

31. Tarres, F., A. Rama and L. Torres, "A Novel method for Face Recognition Under Partial Occlusion or Facial Expression Variations", *Proceedings of ELMAR International Symposium*, pp. 163–166, 2005.

32. Fidler, S., D. Skocaj and A. Leonardis, "Combining Reconstructive and Discriminative Subspace Methods for Robust Classification and Regression by Subsampling", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 3, pp. 337–350, 2006.

33. De Smet, M., R. Fransens and L. Van Gool, "A Generalized EM Approach for 3D Model Based Face Recognition under Occlusions", *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 1423–1430, 2006.

34. Park, B., K. Lee and S. Lee, "Face Recognition using Face-ARG Matching", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 12, pp. 1982–

1988, 2005.

35. Lin, D. and X. Tang, "Quality-driven Face Occlusion Detection and Recovery", *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1–7, 2007.

36. Martinez, A., "Recognizing Imprecisely Localized, Partially Occluded, and Expression Variant Faces from a Single Sample per Class", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 6, pp. 748–763, 2002.

37. Kim, J., J. Choi, J. Yi and M. Turk, "Effective Representation using ICA for Face Recognition Robust to Local Distortion and Partial Occlusion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1977–1981, 2005.

38. Zhang, W., S. Shan, W. Gao, X. Chen and H. Zhang, "Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A Novel Non-statistical Model for Face Representation and Recognition", *Proceedings of International Conference on Computer Vision*, Vol. 1, pp. 786–791, 2005.

39. Wright, J., A. Yang, A. Ganesh, S. Sastry and Y. Ma, "Robust Face Recognition via Sparse Representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 2, pp. 210–227, 2009.

40. He, R., W. Zheng and B. Hu, "Maximum Correntropy Criterion for Robust Face Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 8, pp. 1561–1576, 2011.

41. Zhou, Z., A. Wagner, H. Mobahi, J. Wright and Y. Ma, "Face Recognition with Contiguous Occlusion using Markov Random Fields", *Proceedings of International Conference on Computer Vision*, pp. 1050–1057, 2009.

42. Li, X. and F. Da, "Efficient 3D Face Recognition Handling Facial Expression and Hair Occlusion", *Image and Vision Computing*, Vol. 30, No. 9, pp. 668—-679, 2012.

43. Peng, Y., A. Ganesh, J. Wright, W. Xu and Y. Ma, "RASL: Robust Alignment by Sparse and Low-Rank Decomposition for Linearly Correlated Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 2233–2246, 2012.

44. Liu, P., Y. Wang, D. Huang and Z. Zhang, "Recognizing Occluded 3D Faces using an Efficient ICP Variant", *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 350–355, 2012.

45. Smeets, D., J. Keustermans, D. Vandermeulen and P. Suetens, "meshSIFT: Local Surface Features for 3D Face Recognition under Expression Variations and Partial Data", *Computer Vision and Image Understanding*, Vol. 117, No. 2, pp. 158—-169, 2012.

46. Berretti, S., A. Del Bimbo and P. Pala, "Sparse Matching of Salient Facial Curves for Recognition of 3D Faces with Missing Parts", *IEEE Transactions on Information Forensics and Security*, Vol. 8, No. 2, pp. 374 – 389, 2013.

47. Li, H., D. Huang, P. Lemaire, J.-M. Morvan and L. Chen, "Expression Robust 3D Face Recognition via Mesh-based Histograms of Multiple Order Surface Differential Quantities", *Proceedings of IEEE International Conference on Image Processing*, pp. 3053–3056, 2011.

48. Drira, H., B. Ben Amor, A. Srivastava, M. Daoudi and R. Slama, "3d Face Recognition under Expressions, Occlusions and Pose Variations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, No. 9, pp. 2270 – 2283, 2013.

49. Passalis, G., P. Perakis, T. Theoharis and I. A. Kakadiaris, "Using Facial Symmetry to Handle Pose Variations in Real-world 3D Face Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 10, pp. 1938–1951, 2011.

50. Colombo, A., C. Cusano and R. Schettini, "Three-Dimensional Occlusion Detection and Restoration of Partially Occluded Faces", *Journal of Mathematical Imaging and Vision*, Vol. 40, No. 1, pp. 105–119, 2011.

51. Mahoor, M. H. and M. Abdel-Mottaleb, "Face Recognition Based on 3D Ridge Images Obtained from Range Data", *Pattern Recognition*, Vol. 42, No. 3, pp. 445–451, 2009.

52. Mian, A. S., M. Bennamoun and R. Owens, "Keypoint Detection and Local Feature Matching for Textured 3D Face Recognition", *International Journal of Computer Vision*, Vol. 79, No. 1, pp. 1–12, 2008.

53. Al-Osaimi, F. R., M. Bennamoun and A. Mian, "Integration of Local and Global Geometrical Cues for 3D Face Recognition", *Pattern Recognition*, Vol. 41, No. 3, pp. 1030–1040, 2007.

54. Passalis, G., I. A. Kakadiaris and T. Theoharis, "Intraclass Retrieval of Nonrigid 3D Objects: Application to Face Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 2, pp. 218–229, 2007.

55. Faltemier, T. C., K. W. Bowyer and P. J. Flynn, "3D Face Recognition with Region Committee Voting", *Proceedings of International Symposium on 3D Data Processing, Visualization, and Transmission*, pp. 318–325, 2006.

56. Cook, J., V. Chandran and C. Fookes, "3D Face Recognition using Loggabor Templates", *Proceedings of British Machine Vision Conference*, pp. 83–92, 2006.

57. Passalis, G., I. A. Kakadiaris, T. Theoharis, G. Toderici and N. Murtuza, "Evaluation of 3D Face Recognition in the Presence of Facial Expressions: an Annotated Deformable Model Approach", *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition - Workshops*, p. 171, 2005.

58. Chang, K. I., K. W. Bowyer and P. J. Flynn, "Multiple Nose Region Matching for 3D Face Recognition under Varying Facial Expression", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 10, pp. 1695–1700, 2006.

59. 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.

60. Lu, X. and A. K. Jain, "Deformation Modeling for Robust 3D Face Matching", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 8, pp. 1346–1357, 2008.

61. Li, X. and H. Zhang, "Adapting Geometric Attributes for Expression-invariant 3D Face Recognition", *Proceedings of IEEE International Conference on Shape Modeling and Applications*, pp. 21–32, 2007.

62. Do Carmo, M. P. and M. P. Do Carmo, *Differential Geometry of Curves and Surfaces*, Vol. 2, Prentice-Hall, Englewood Cliffs, 1976.

63. Gordon, G., "Face Recognition based on Depth and Curvature Features", *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 808–810, 1992.

64. Koenderink, J. and A. van Doorn, "Surface Shape and Curvature Scales", *Image and vision computing*, Vol. 10, No. 8, pp. 557–564, 1992.

65. Tittle, J. and V. Perotti, "The Perception of Shape and Curvedness from Binocular Stereopsis and Structure from Motion", *Attention, Perception, & Psychophysics*, Vol. 59, No. 8, pp. 1167–1179, 1997.

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

67. Goodall, C., "Procrustes Methods in the Statistical Analysis of Shape", *Journal of the Royal Statistical Society, Series B (Methodological)*, pp. 285–339, 1991.

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

69. Eckart, C. and G. Young, "The Approximation of One Matrix by Another of Lower

Rank", *Psychometrika*, Vol. 1, No. 3, pp. 211–218, 1936.

70. Besl, P. J. and H. D. McKay, "A Method for Registration of 3D Shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, pp. 239–256, 1992.

71. Salah, A. A., N. Alyuz and L. Akarun, "Registration of 3D Face Scans with Average Face Models", *Journal of Electronic Imaging*, Vol. 17, No. 1, 2008.

72. Bookstein, F. L., "Principal Warps: Thin-plate Splines and the Decomposition of Deformations", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 6, pp. 567–585, 1989.

73. Alpaydin, E., *Introduction to Machine Learning*, MIT Press, 2004.

74. Jolliffe, I., *Principal Component Analysis*, Spring-Verlag, New York, 1986.

75. Bishop, C. M. and N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, Vol. 1, Springer, New York, 2006.

76. Colombo, A., C. Cusano and R. Schettini, "3D Face Detection using Curvature Analysis", *Pattern Recognition*, Vol. 39, pp. 444–455, March 2006.

77. Fisher, R. A., "The use of Multiple Measurements in Taxonomic Problems", *Annals of Eugenics*, Vol. 7, No. 2, pp. 179–188, 1936.

78. Irfanoglu, M. O., B. Gokberk and L. Akarun, "3D Shape-based Face Recognition using Automatically Registered Facial Surfaces", *Proceedings of International Conference on Pattern Recognition*, Vol. 4, pp. 183–186, 2004.

79. Gokberk, B., H. Dutagaci, A. Ulas, L. Akarun and B. Sankur, "Representation Plurality and Fusion for 3-D Face Recognition", *IEEE Transactions on Systems Man and Cybernetics, Part B*, Vol. 38, No. 1, pp. 155–173, 2008.

80. Ekman, P. and W. V. Friesen, "Facial Action Coding System (FACS): A Technique for

the Measurement of Facial Action", *Palo Alto, CA: Consulting*, 1978.

81. Medioni, G. and R. Waupotitsch, "Face Recognition and Modeling in 3D", *Proceedings of IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pp. 232–233, 2003.

82. Koudelka, M. L., M. W. Koch and T. D. Russ, "A Prescreener for 3D Face Recognition using Radial Symmetry and the Hausdorff Fraction", *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition-Workshops*, pp. 168–168, 2005.

83. Mian, A., M. Bennamoun and R. Owens, "2D and 3D Multimodal Hybrid Face Recognition", *Proceedings of European Conference on Computer Vision*, pp. 344–355, 2006.

84. Wang, Y., G. Pan, Z. Wu and Y. Wang, "Exploring Facial Expression Effects in 3D Face Recognition using Partial ICP", *Proceedings of Asian Conference for Computer Vision*, pp. 581–590, 2006.

85. Lu, X., A. 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.

86. Chang, K., W. Bowyer and P. Flynn, "Multiple Nose Region Matching for 3D Face Recognition under Varying Facial Expression", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 10, pp. 1695–1700, 2006.

87. Lo, T. and J. Siebert, "Sift Keypoint Descriptors for Range Image Analysis", *Annals of the BMVA X*, pp. 1–18, 2009.

88. Ghahramani, Z., G. E. Hinton *et al.*, *The EM Algorithm for Mixtures of Factor Analyzers*, Tech. rep., Technical Report CRG-TR-96-1, University of Toronto, 1996.

89. Boykov, Y. and G. Funka-Lea, "Graph Cuts and Efficient ND Image Segmentation", *International Journal of Computer Vision*, Vol. 70, No. 2, pp. 109–131, 2006.

90. Kolmogorov, V. and R. Zabin, "What Energy Functions can be Minimized via Graph Cuts?", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 2, pp. 147–159, 2004.

91. Powers, D. M. W., "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation", *Journal of Machine Learning Technologies*, Vol. 2, No. 1, pp. 37–63, 2011.

92. Montgomery, D. C., *Design and Analysis of Experiments*, Wiley, New York, 2008.

93. Montgomery, D. C. and G. C. Runger, *Applied Statistics and Probability for Engineers*, Wiley, New York, 2010.