PART-BASED 3D FACE RECOGNITION UNDER POSE AND EXPRESSION VARIATIONS

by

Hamdi Dibeklioğlu

B.S, in Computer Engineering, Yeditepe University, 2006

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

Graduate Program in Computer Engineering

Boğaziçi University

2008

ACKNOWLEDGEMENTS

With my deepest gratitude; I would like to thank to my thesis supervisor Prof. Lale Akarun for her guidance, great patience and endless support all through this thesis. I am grateful to Prof. Fikret Gürgen and especially Prof. Bülent Sankur for participating in my thesis jury, their criticism and helpful comments that have greatly improved this thesis.

I would like to thank the members of Media Laboratory, Koray Balcı, Cem Keskin, Umut Konur, Pınar Santemiz, Yunus Emre Kara, and especially Neşe Alyüz for their support throughout the preparation of this work. Additionally, I would like to thank the members of Perceptual Intelligence Laboratory, especially İsmail Arı for his endless support in my most demoralized moments, Mehmet Gönen and Oya Aran for their help in my thesis. I also thank Prof. Arsev Eraslan, Hüseyin Hışıl, Dr. Albert Ali Salah, Erinç Dikici and Oya Çeliktutan.

I would especially like to present my gratitude to Dr. Berk Gökberk for being an encouraging co-advisor and for his boundless support in my thesis.

I am also grateful to Hilal Özbey, Okay Aslan and Murat Kasapsaraçoğlu for their endurance and tenderness. I am specially indebted to my family, Sevim Dibeklioğlu, Mehmet Dibeklioğlu and Dr. Sami Dibeklioğlu for their unconditional love and endless support throughout my life.

Lastly I would like to express my gratitude to TÜBİTAK for supporting me with National Scholarship Programme for M.Sc. Students - 2210.

ABSTRACT

PART-BASED 3D FACE RECOGNITION UNDER POSE AND EXPRESSION VARIATIONS

The advances in sensor technologies and the several years of research in recognition of biometric modalities increased the expectations from 3D face recognition systems. An important reason of scientific interest on 3D face recognition is the ability of acquisition of the facial data nonintrusively. This makes 3D face recognition applicable to real life tasks in terms of security and human computer interaction.

In this study, a fully automatic part-based 3D face recognition system has been proposed. The proposed system is based on pose-correction and curvature-based facial segmentation for recognition tasks. Utilization of facial parts in the recognition step provides robustness to the system even in facial expression variations. Since the nose is anatomically the most stable part of the face, it is largely invariant under expressions. For this reason, we have concentrated on locating the nose tip and segmenting the nose. Furthermore, the nose tip and other nose landmarks enable pose correction. Pose correction feature of the proposed recognition system, allows the identification of people under significant amount of pose variations. For the face recognition task, we try both one-to-all and Average Nose Model (ANM) based methodologies.

Our results show that the utilization of anatomically-cropped nose region in 3D face recognition increases the rank-one recognition success rates up to 94.1 per cent for frontal facial expressions and 79.41 per cent for pose variations in the Bosphorus database.

ÖZET

POZ VE İFADE DEĞİŞİMLERİNDE PARÇA TABANLI ÜÇ BOYUTLU YÜZ TANIMA

Sensör teknolojilerindeki gelişmeler ve biyometrik özellikleri tanıma konusunda son yıllarda yapılan araştırmalar, üç boyutlu yüz tanıma sistemlerinden beklentileri yükseltmektedir. Üç boyutlu yüz tanıma alanına olan bilimsel ilginin önemli bir nedeni yüz taramalarının kişileri rahatsız etmeden yapılabilmesidir. Bu durum, üç boyutlu yüz tanımayı, güvenlik ve insan-bilgisayar etkileşimi konularında kullanılabilir hale getirmektedir.

Bu çalışmada, tam otomatik, parça tabanlı bir üç boyutlu yüz tanıma sistemi önerilmiştir. Önerilen sistem, tanıma konusunda poz düzeltme ve eğrilik tabanlı yüz bölütlemesine dayanmaktadır. Sistemin tanıma basamağında, yüz parçalarının kullanılması, yüz ifadelerindeki değişimlerde bile, sisteme gürbüzlük sağlamaktadır. Burun anatomik olarak yüzün en hareketsiz bölgesi olduğu için ifade değişimlerinden çoğunlukla etkilenmez. Bu sebeple, burun ucu bulma ve burun bölütü çıkarma konuları üzerinde yoğunlaşılmıştır. Ayrıca, burun ucu ve diğer nirengi noktaları poz düzeltmeye imkan vermektedir. Önerilen tanıma sisteminin poz düzeltme özelliği, sistemin poz değişimlerinde kullanılabilirliğini sağlamaktadır.

Sonuçlarımız, doğal sınırlarından bölütlenmis burun bölgesinin, üç boyutlu yüz tanımada tek başına kullanımı ile, Boğaziçi Veri Tabanında, tanıma başarısı oranlarını düz yüz ifadeleri için yüzde 94.1'e kadar ve poz değişimleri için yüzde 79.41'e kadar arttırdığını göstermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS/ABBREVIATIONS

$\mathcal{C}$	Combination map
d	Euclidean distance
$\mathcal{D}$	Difference map
$\mathcal{E}$	Eigen map
E	Error function
E_{PSD}	Total Point Set Difference
G	Gaussian curvature
G_h	Height of a given map
G_w	Width of a given map
H	Mean curvature
$\mathcal{H}$	Correlation matrix
$\boldsymbol{I}$	Facial depth image
L_y	Horizontal line which passes through y-coordinate of nose tip
m	Min-Max Normalization function
$\mathcal{N}$	Depth map of registered training sample to ANM
$\mathcal{N}'$	Depth map of registered gallery sample to ANM
N_G	Number of samples in gallery
N_p	Unit normal vector
N_P	Number of samples in probe
O	Computational complexity operator
p	Three dimensional point
P	Point Set
P_c	A close point to nose tip
P_L	Left side starting point of minimum path estimation
P_R	Right side starting point of minimum path estimation
q_i	Adjacent vertex
r	Radius
R_y	Rotation matrix about y-axis
S	Smooth surface

S'	Gaussian smoothed shape index
SI	Shape index
tr	Trace operator
x_o	Offset on x-axis
x_{offset}	Middle point of facial surface on x-axis
y_o	Offset on y-axis
y_{offset}	Middle point of facial surface on y-axis
z_{offset}	Farthermost z-coordinate to the camera
W	Weingarten curvature matrix
κ_i	Principal curvature
κ_u	Normal curvature on u direction
λ_i	Eigenvalue
Λ	Change map
μ	Center of mass
∇	Gradient operator
Φ	Protrusion map
Θ	Rotation angle
2D	Two Dimensional
3D	Three Dimensional
3D-RMA	A 3D face database
AFM	Average Face Model
ANM	Average Nose Model
ARM	Average Region Model
AU	Action Unit
BOS	Bosphorus database
BOSv1	Bosphorus database version 1
BOSv2	Bosphorus database version 2
BUMM	Boğaziçi University Multimedia Group
FACS	Facial Action Coding System

GavabDB	A 3D dace database
ICP	Iterative Closest Point algorithm
M2VTS	A 3D face database
MC1	Proposed curvature-based nose tip estimator
MC2	Proposed voting-based nose tip estimator
MD	A depth-based nose tip estimator from literature
MIT-CBCL	A 3D face database
RF1	A 3D face recognition method from literature
RF2	A 3D face recognition method from literature
RN1	Proposed 3D face recognition method
RN2	Proposed 3D face recognition method
SIFT	Scale Invariant Feature Transform
SVD	Singular Value Decomposition
SVM	Support Vector Machine
PSD	Point Set Difference
TPS	Thin-Plate Spline
v.1	Version 1
v.2	Version 2

1. INTRODUCTION

Recognition of human face has received a great deal of attention and emerged as an active research area especially over the last 20 years. The major purpose of face recognition is to identify the humans from data acquired from their faces, as humans do. A good recognition system has to be fully automatic and robust enough for real life conditions such as illumination, rotations, expressions and occlusions.

Improvements in sensor technology, and the difficulty of implementing robust 2D face recognition systems have made 3D face recognition an attractive alternative, especially in biometrics applications. Works on 3D face recognition in the last years provide nearly 99 per cent recognition success rates but most of these results are constrained to certain assumptions and the methods are not applicable under realistic conditions. For example, most of the 3D face databases contain only frontal and neutral scans. Therefore, the recognition results obtained on these databases are usually optimistic. With the help of challenging databases such as the Bosphorus database, studies on robust systems under expression, rotation and occlusion variations, have become possible.

The main application areas of face recognition are law enforcement and commercial identity management systems. Governments, academic, and commercial institutions have been working on automated identification systems which recognize human beings from their intrinsic or behavioral characteristics. Biometric-based robust identification is very important in terms of governmental and commercial applications. Commonly used biometrics include fingerprint, face, iris, hand geometry, voice, palmprint, handwritten signatures and gait. The acquisition of faces is a non-intrusive process requiring no contact with the subject. Furthermore, face is a widely acceptable biometric, since it is the natural mode of identification. That is why face recognition is more applicable for the daily life, despite the high accuracy rates of the iris and fingerprint recognition systems.

1.1. Motivation

Human face is a 3D surface and 3D facial geometry can give more discriminative information about human faces than 2D texture images. In 3D, facial expressions can be examined by segmentation into parts, rotations can be rectified and illumination differences can be eliminated. 3D surface geometry enables face recognition under rotation, illumination, and expression variations. Additionally, more accurate registration and recognition of faces can be provided by all these features of 3D facial data. As a result, we are motivated to work on 3D data.

This work focuses mainly on segmentation of facial parts and part-based registration and recognition of 3D faces. Stable parts of the human face can increase recognition success rates under facial expression changes. Furthermore, rotation estimation and pose correction can lead to robust segmentation and recognition systems under realistic pose conditions.

1.2. 3D Face Recognition Literature Review

In this literature review, we mainly focus on part-based 3D face recognition, partial segmentation and feature extraction studies. Recent surveys on 3D face recognition literature are given in [1] by Hamouz *et al.*, in [2] by Scheenstra *et al.*, in [3] by Bowyer *et al.* and in [4] by Zhao *et al.*

Research on 3D face recognition gained impetus with the availability of newer and larger 3D face datasets. There are several 3D face databases such as MIT-CBCL face recognition database [5], 3DRMA [6], GavabDB [7], extended M2VTS database [8], FRGC and the recently collected Bosphorus database [9].

Certainly, the most popular 3D face database is the FRGC dataset which has been collected at University of Notre Dame. FRGC version 1.0a contains 943 near-frontal scans from 277 subjects [10]. For each acquisition, there is a range image and the corresponding registered 2D texture, taken under controlled indoor lighting

conditions. Version 2.0 subsumes version 1.0a, and contains 4,007 frontal scans from 466 subjects recorded at 22 sessions with minor pose, but difficult illumination and expression variations.

The newest 3D face database is the Bosphorus database [9, 11] which is recently collected at Boğaziçi University and will be open for academic usage soon. The Bosphorus database is composed of 3D facial surface point clouds and correlated 2D texture images of the subjects. It consists of 3,396 facial scans of 81 subjects. The database is the first one which has large amounts of pose, expression, and occlusion variations with manually localized 24 fiducial points.

The earliest studies in 3D face recognition area started in the early of 90's [12, 13, 14, 15]. In the recent years, 3D face recognition studies have rapidly increased in number. Most of the earlier studies reported performances that reached up to almost 100 per cent recognition rates, because of the usage of small data sets with a few number of neutral/frontal subjects. Only a small number of studies cover the variations in pose and expression. Pose and expression independent 3D face recognition has started to arouse interest by the introduction of new databases.

1.2.1. Part-Based 3D Face Recognition

There are several 3D face recognition techniques in the literature but component or part-based approaches are very new and this type of approach is an open research topic. Part-based face recognition is useful to alleviate the pose, facial expression and partial occlusion effects on the recognition performance. To deal with these effects, entire 3D model of the face is split into several regions and these regions are processed separately. Recognition scores of components are fused at the final identification phase.

Alyüz *et al.* [16] split the face into 15 patches as shown in Figure 1.1, and carry out an exhaustive search of all possible combinations of these surface to find the best subset of all the patches around the whole facial surface. A region comprised of eye, nose, and central forehead parts equipped with the maximal/minimal curvature direction

features provides 97.28 per cent of rank-one correct classification rate in the Bosphorus database v.2.

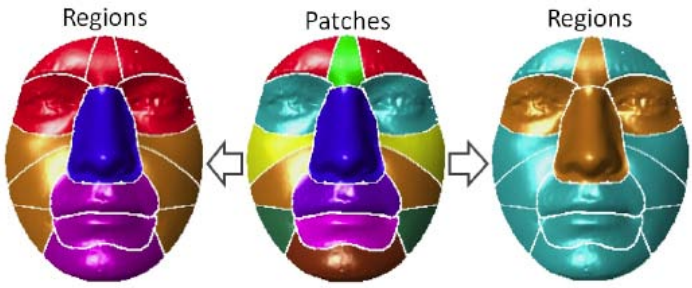


Figure 1.1. Division of a facial surface into two-level components in the method of Alyüz *et al.* [16]: Central image displays first level components, patches, and the left/right images depicts two different alternatives for higher-level components, regions, formation.

In [15], Lee and Milios segment the range images into convex regions by using the sign of the mean and Gaussian curvature at each point. They observed that distinct facial features correspond to convex regions of the range image of the face. Regional matching is obtained by Extended Gaussian Images correspondence. This approach is reported to be able to cope with occlusions and expression variations.

Moreno *et al.* [17] segment the 3D facial surface using signs of mean, Gaussian curvatures and several three dimensional descriptors. 68 descriptors have been obtained from the segmented regions. With the feature discrimination power analysis, the first 35 features of the ordered list of features according to the Fisher coefficients are selected to represent faces in recognition experiments. In frontal views, these 35 features provide 78 per cent rank-one success rate. Rank-five recognition provides 92 per cent of recognition success rate.

Cook *et al.* [18] use Log-Gabor Templates on range images to deal with occlusions, distortions and facial expressions. Range images are divided into multiple regions both in spatial and frequency domains and these observations are each classified individually and then combined at the score level. After application of the Log-Gabor filters, the

face is broken into 49 square windows arranged in a 7x7 grid and each grid is further decomposed by 3 scales to generate 147 subregions. Then, Principal Component Analysis (PCA) is then applied to the Log-Gabor filter responses in each of all subregions. Mahalanobis Cosine distance metric is used for classification, and the classifiers are fused by the sum rule. This approach provides 94.63 per cent recognition performance on neutral expressions in the FRGC v.2.

In [19], Chang *et al.* introduce a new approach called Adaptive Rigid Multi-region Selection. This method is fully automated and uses no manually selected landmark points. It is based on multiple regions selected around the nose area which has the least shape variation due to facial expressions and then the method creates a fused result. Facial surfaces are registered by Iterative Closest Point (ICP) algorithm [20]. FRGC v.2 database is used, where the gallery consists of one neutral image per subject. A rank-one recognition result of 92 per cent was reported using the product rule.

In [21] Faltemier *et al.* extend the use of multiple regions. Seven overlapping regions around the nose are extracted as ellipses with different parameters and ICP is used to align each segment separately. Threshold values for regions are determined empirically that allows related regions to vote without significantly affecting the overall false acceptance rate. Then, threshold values are combined with committee voting. On the FRGC v.2 database, a rank-one accuracy of 94.9 per cent is reported.

Faltemier *et al.* [22] use multi-regions to segments extracted from the whole facial surface and it is reported that 28 small regions on the face allow for the highest accuracy. To find the best committee of local regions for optimum results, 38 regions on the face, some of which overlap (see Figure 1.2), are used in experiments that run on the FRGC v.2 database. By selecting multiple small regions on the face, any errors caused by a single region can be compensated for when combining the matching scores from the other regions, making the system more robust to image artifacts, wrinkles, facial hair, or expression variations. The individual regions are aligned with ICP. The best performance is provided by the fusion of 28 regions with a modified borda count method, and the recognition accuracy is reported as 97.2 per cent.

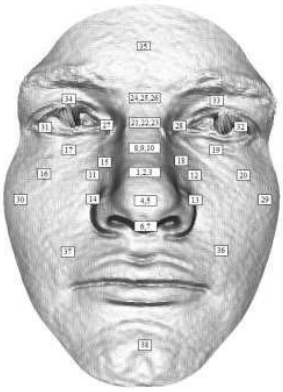


Figure 1.2. Image of probe sphere centroids used by Faltemier *et al.* [22] labeled by region number: Multiple region numbers at a centroid indicate that more than one radius were used for cropping, yielding multiple region probes with the same centroid.

Mian *et al.* [23] develop a multi-modal (2D and 3D) algorithm and perform hybrid (feature based and holistic) matching in order to achieve efficiency and robustness to facial expressions. In their work, six points are manually identified on a gallery face to segment its corresponding 3D face into three disjoint regions; namely, the forehead, the nose, and the cheeks (Figure 1.3). A skin detection algorithm is used to detect the skin pixels in the 2D colored image of the forehead region to remove the possible artifacts caused by the eyes and hair. Then, segmented regions are matched separately using a modified ICP algorithm. All matching results are fused at the metric level to achieve higher accuracy. Experiments are done on the FRGC v.2 database and 98.82

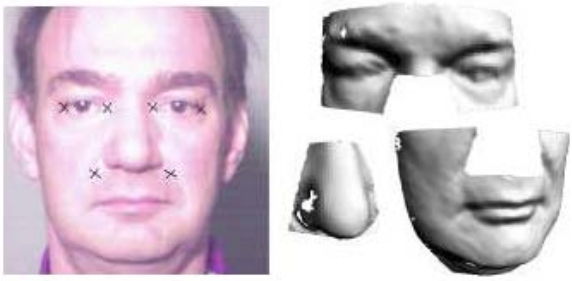


Figure 1.3. Manually identified landmarks on a gallery face to segment its corresponding 3D face into eyes-forehead, nose and cheeks regions in Mian's method [23].

per cent and 92.36 per cent recognition rates for neutral and non-neutral probe sets are achieved by using 3D data only, respectively. Recognition accuracies of 98.20 per cent and 93.74 per cent are achieved for neutral and non-neutral probes by applying the developed multi-modal approach, respectively.

1.2.2. 3D Feature Extraction

3D facial feature extraction has aroused interest with the increasing development of 3D modeling and digitizing techniques. Wang *et al.* [24] used the point signature and the stacked Gabor filter response to identify 3D and 2D features. Boehnen *et al.* [25] explored 2D color information to extract skin tone regions and identify eyes and mouth. Lu *et al.* [26] proposed an algorithm using a shape index to identify the inside of an eye as a saddle location. However, their results were not precise enough. In their following work, they designed a more accurate system [27, 28]. Unfortunately, its time complexity was high due to the exhaustive searching process. Gordon [12] proposed a method using face shape based on curvature to identify facial features. The method worked well, but it was only tested on a small database of 24 range scans. Feature extraction based on Active Appearance Models (AAM) algorithm [29, 30] seems to meet the needs; however, lots of manual work is needed during its training procedure and the efficiency is unsatisfactory.

Senior [31] locates a few high-level features, namely, eyes, nose, mouth and then 26 low-level features such as the parts of the eyes, nose, mouth, eyebrows are located relative to the high-level feature locations, instead of searching for all the facial features directly in the face image. The approximate locations of the high-level features are known from statistics of mean and variance (relative to the nose position) gathered on a training database. The discriminant templates are used to score each potential matching image patch for a given feature. In that work, a constellation of local patches has been used as the representation. They chose the local template approach, in contrast to global identity templates such as those used in eigenface systems. A simple Gabor jet model has been used to describe particular patches of the face corresponding to the 29 facial features found. Each patch is represented by a feature vector consisting

of 40 complex elements each, representing the filter responses of Gabor filters with five different scales and eight different orientations, centered at the estimated feature location.

In [32], Akagündüz *et. al.* suggest using *scale and transform invariant features* (SIFT) to represent 3D facial data. The nose is located first, and candidates for the remaining landmark points (two eye pits and nose saddle) are encoded according to their position with respect to the nose. Mean and Gaussian curvatures of the facial surface are used to detect peaks, pits and saddle regions. Even though relying on a single landmark is a potential threat to the robustness of the system, nose localization can produce very good results in practice, since it is the easiest landmark to locate and has the greatest effect in registration [33].

1.2.3. Face Segmentation Methods

Approaches of face segmentation are commonly based on the anthropometric and anatomic features. Muscle system of a face is shown in Figure 1.4. In [34], Ye manually aligned eye position and segmented a face image into two parts according to the geometric characteristics of the human face. Samaria coarsely subdivided the frontal face image into five horizontal regions: forehead, eyes, nose, mouth and chin [35].

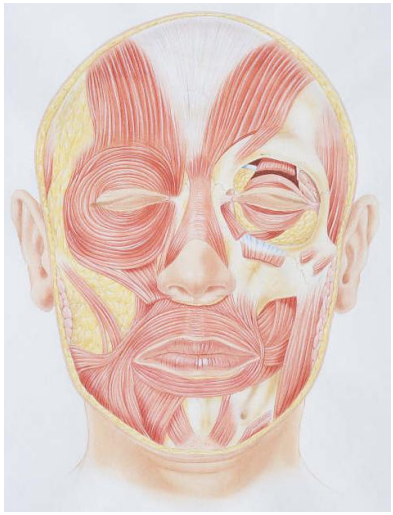


Figure 1.4. Muscle system of the human face.

Several PCA-based face recognition algorithms also require the face to be segmented. Su utilized face patches to calculate their eigenface, eigeneyebrow, eigeneye, eigennose and eigenmouth components [36]. Some algorithms even segmented the face into patches without taking feature distribution into consideration, such as [37].

All the above mentioned methods are based on the 2D face image, which may provide some useful hints for segmenting 3D faces. Segmentation approaches on 3D face images and 2D are not completely identical due to the differences inherent in the data structures. Lengagne *et al.* [38] proposed a global scheme which can produce a segmentation of the human face into 3D patches from a pair of stereo images. A similar work had been done by Gu *et al.* [39]. These works located the features in an interactive way and the segmented patches were irregular.

Gong *et. al.* [40] introduced an automatic 3D face image segmentation algorithm based on feature extraction. Given a 3D face, the 2D texture image is obtained through converting the 3D coordinates to 2D and then, feature regions are extracted. Both 2D color intensity and 3D geometric information are utilized simultaneously to locate the facial features. On a database of 500 3D faces, the system achieved an accuracy of more than 95 per cent in locating the major features (see Figure 1.5).

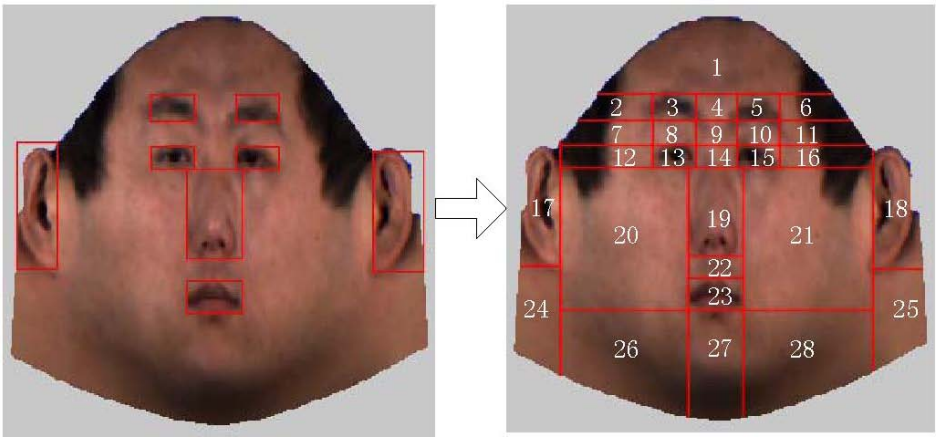


Figure 1.5. 28 patches that are segmented by 17 lines and supporting vectors on 3D model in the method of Gong *et al.* [40].

Eraslan [41, 42, 43] represents 3D nose surface by a sparse set of 14 3D fiducial points and the associated sparse set of 86 interconnecting fiducial point distance lines; 3D upper lip surface by a sparse set of 14 3D fiducial points and the associated sparse set of 86 interconnecting fiducial point distance lines; 3D lower lip surface by a sparse set of 14 3D fiducial points and the associated sparse set of 86 interconnecting fiducial point distance lines; 3D chin surface by a sparse set of nine 3D fiducial points and the associated sparse set set of 36 interconnecting fiducial point distance lines; 3D right and left eye socket surfaces by a sparse set of 11 3D fiducial points and the associated sparse set of 52 interconnecting fiducial point distance lines (see Figure 1.6).

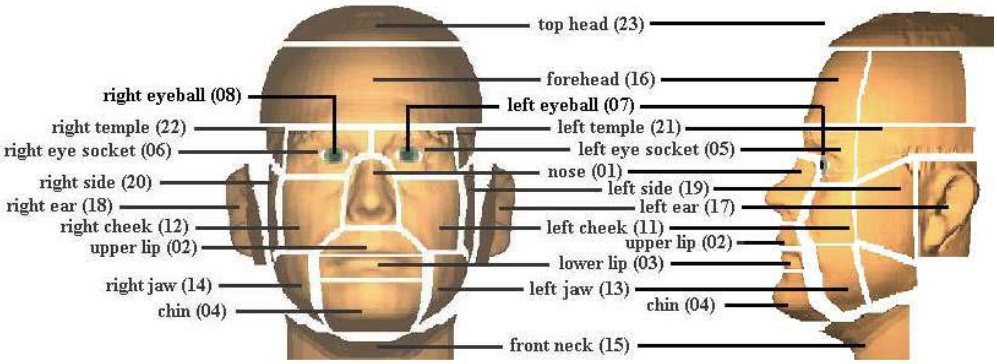


Figure 1.6. Facial parts represented by Eraslan [41, 42, 43].

Weyrauch *et al.* [44] use a two level component-based face detection system. The first level consists of 14 independent component classifiers (linear SVMs). Each component classifier is trained on a set of extracted facial components and on a set of randomly selected non-face patterns. The components can be automatically extracted from synthetic images since the full 3D correspondences between the face models are known. On the second level, the maximum continuous outputs of the component classifiers within rectangular search regions around the expected positions of the components are used as inputs to a geometrical classifier (linear SVM), which performed the final detection of the face. From the original 14 components extracted by the face detector only nine components are used for face recognition. Five components are eliminated because they strongly overlap with other components or contain few gray value structure (e.g. cheeks). In addition, a global component is added to improve recognition. The location of this component is computed by taking the circumscribing square around

the bounding box of the other nine components. After extraction, the squared image patch is normalized to 40×40 pixels. The component-based face detector is applied to each synthetic face image in the training set to extract the ten components (see Figure 1.7).

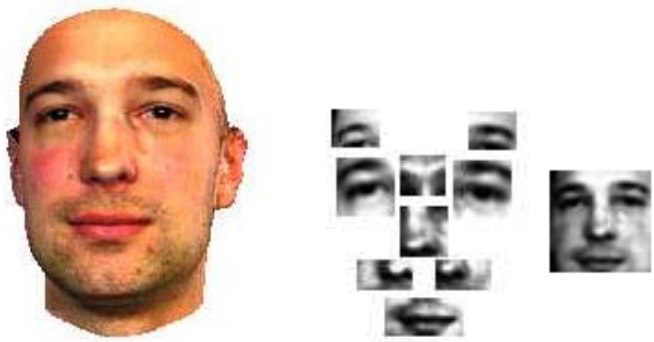


Figure 1.7. The ten components used for face recognition by Weyrauch *et al.* [44].

1.3. Approach and Contributions

Accuracy of a part-based face recognition system depends on the selected parts on the facial surface. Although many partitioning alternatives have been proposed, the nose is the most distinguishing part [22]. It does not change much under expressions and it is the center of the face. For segmentation, we needed to find the tip of the nose, and we proposed two different curvature based heuristic nose tip estimators. Then, we cropped nose regions of the faces with the natural boundaries by the proposed nose segmentation algorithm which is explained in Section 2.1. In the stages of registration and recognition, we observed that an extra portion of facial data in the nose region obviously degraded the registration accuracy; so, the recognition performance also decreases. Therefore, we show that the importance of the anatomical-feature-based segmentation of facial parts in recognition systems.

Additionally, we proposed and tested a coarse rotation estimator for 3D facial models by using the relative nose tip position on the facial data grid. Pose estimations and corrections were limited as 0 to ± 30 degrees of yaw rotations to avoid possible risk of wrong estimations.

Both one-to-all and one-to-one registration approaches are used for comparisons. The Average Nose Model (ANM) is proposed and its usage in one-to-one ICP approach is suggested. Recognition is handled with Point Set Difference (PSD) calculations between gallery and test noses. The registered noses are aligned to a re-sampled grid, prior to PSD calculation. After re-sampling, the registered noses all have the same x and y values, and only z values are used for calculating the PSD. As a result PSD computation is simplified down from three to one dimensional space.

1.4. Outline of the Thesis

The organization of the rest of the thesis is as follows: In Chapter 2, theoretical foundations and mathematical formulations are given. Additionally, all the methods used in landmark localization, nose segmentation, face registration, ANM generation, and face recognition phases are described in detail. Chapter 3 starts with describing the newly introduced Bosphorus 3D face database which is used in this work. Then, the results of landmark localization tests, nose segmentation and part-based face recognition experiments for frontal expressions and rotation poses are given in Chapter 3. Chapter 4 concludes the thesis by summarising the contributions and giving a brief outline of the obtained results, and finally, by giving possible future directions.

2. PART-BASED 3D FACE RECOGNITION

In this chapter we describe details of all the phases of the proposed face recognition system, namely, landmark localization, nose segmentation, registration and part-based classification. Design of the overall 3D face recognition system with ANM-based registration is shown in Figure 2.1.

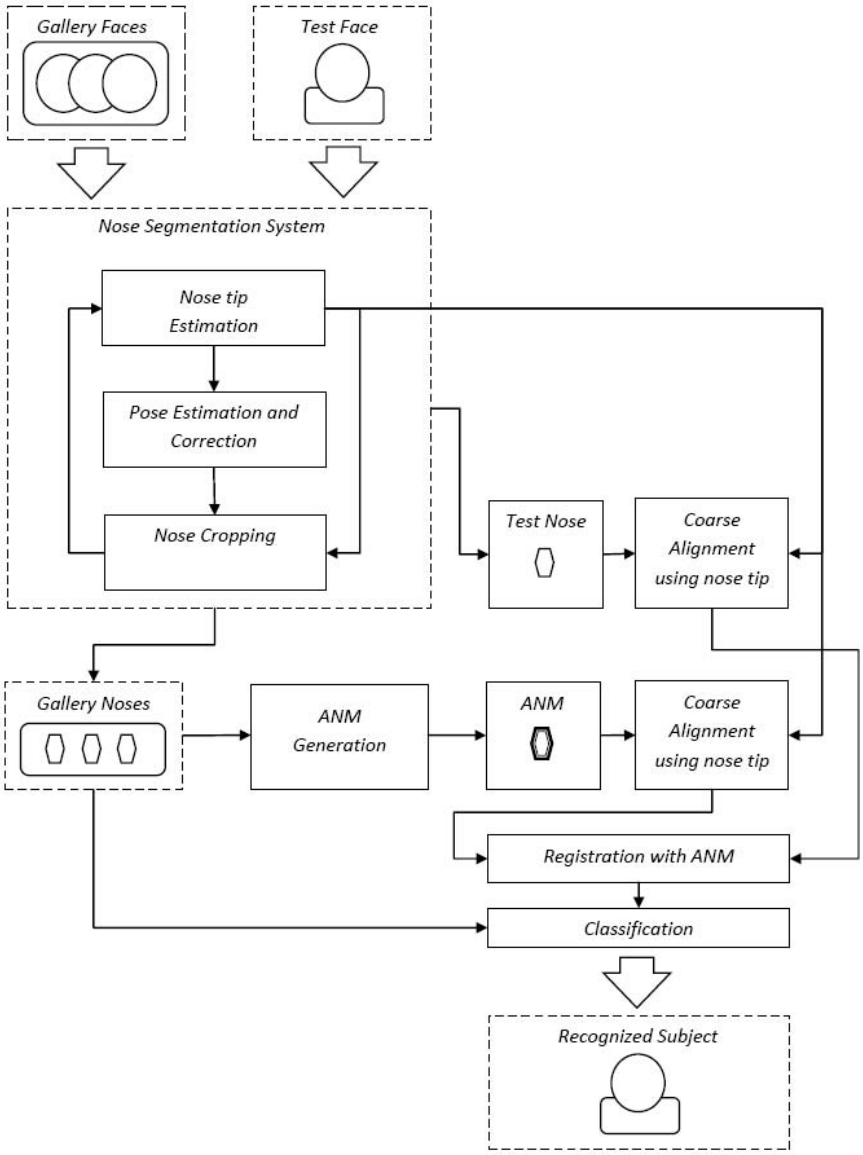


Figure 2.1. Full automatic ANM-based 3D face recognition system.

2.1. Landmark Localization and Nose Segmentation

In this section we describe a heuristic method to localize the nose tip and segment the nose region under pose and expression variations. The proposed algorithm estimates the regions with high protrusion and selects the most probable location based on curvature values. Principle normal curvatures are used since they are invariant to rotation and translation.

2.1.1. Curvature Estimation

In the first step, principle curvatures of the facial surface are needed to be estimated for the proposed algorithms. Assume that p is a point on a smooth surface S, N_p is the unit normal to S at p, and $X(u, v)$ is a local parameterization of S in a neighborhood of p. Principal curvatures of S can be computed by using $X_u(p)$, $X_v(p)$ and N_p as a local coordinate system. The surface normal, N_p is computed by calculating the eigenvalues of the covariance matrix of 3D points' coordinates in a neighbourhood of p. The eigenvector with the minimum eigenvalue gives N_p. The remaining eigenvectors correspond to X_u and X_v. The Weingarten curvature matrix W is defined as:

$$W = \begin{pmatrix} \frac{eG-fF}{EG-F^2} & \frac{fE-eF}{EG-F^2} \\ \frac{fG-gF}{EG-F^2} & \frac{gE-fF}{EG-F^2} \end{pmatrix} \tag{2.1}$$

where,

$$\begin{array}{ccc} e = N_p \cdot X_{uu}(p) & f = N_p \cdot X_{uv}(p) & g = N_p \cdot X_{vv}(p) \\ E = X_u(p) \cdot X_u(p) & F = X_u(p) \cdot X_v(p) & G = X_v(p) \cdot X_v(p) \end{array} \tag{2.2}$$

Here X_{uu}, X_{uv}, and X_{vv} are 3×1 vectors. In the special case that X_u and X_v are orthogonal unit vectors, W becomes a symmetric matrix

$$W = \begin{pmatrix} e & f \\ f & g \end{pmatrix} \tag{2.3}$$

If u is a unit vector in the tangent plane to S at p, then the normal curvature of the surface in the direction of u is:

$$\kappa_u = u^T W u \tag{2.4}$$

The eigenvalues of the Weingarten curvature matrix, λ_1 and λ_2 are the maximum and the minimum normal curvatures of the surface at p, respectively.

The Quadratic Surface Approximation Method is used to solve W [45]. This method is based on fitting a quadratic surface to the adjacent vertices. Initially, each adjacent vertex q_i is transformed to local coordinates (x_i, y_i, z_i). In this coordinate system, p becomes the origin $(0, 0, 0)$ and N'_p is the unit z-axis. If we assume that the quadratic surface is in the form of

$$z = f(x, y) = \frac{A}{2}x^2 + Bxy + \frac{C}{2}y^2, \tag{2.5}$$

then the Weingarten matrix for this surface can be defined as

$$W = \begin{pmatrix} A & B \\ B & C \end{pmatrix} \tag{2.6}$$

If $\boldsymbol{x}$ is defined as

$$\boldsymbol{x} = \begin{pmatrix} A & B & C \end{pmatrix}^T \tag{2.7}$$

then,

$$\begin{pmatrix} \frac{1}{2}x_i^2 & x_i y_i & \frac{1}{2}y_i^2 \end{pmatrix} \boldsymbol{x} = z_i \quad i = 1, ..., n \tag{2.8}$$

where, n is the number of points in the neighbourhood of p. This equation system is

solved by least-squares surface fit method as follows:

$$\boldsymbol{x} = (U^T U)^{-1} U^T z_i \tag{2.9}$$

where, U is the $3 \times n$ matrix defined as:

$$U = \begin{pmatrix} \frac{1}{2}x_i^2 & x_i y_i & \frac{1}{2}y_i^2 \end{pmatrix}. \tag{2.10}$$

2.1.2. Initial Estimation of a Close Point to the Nose Tip

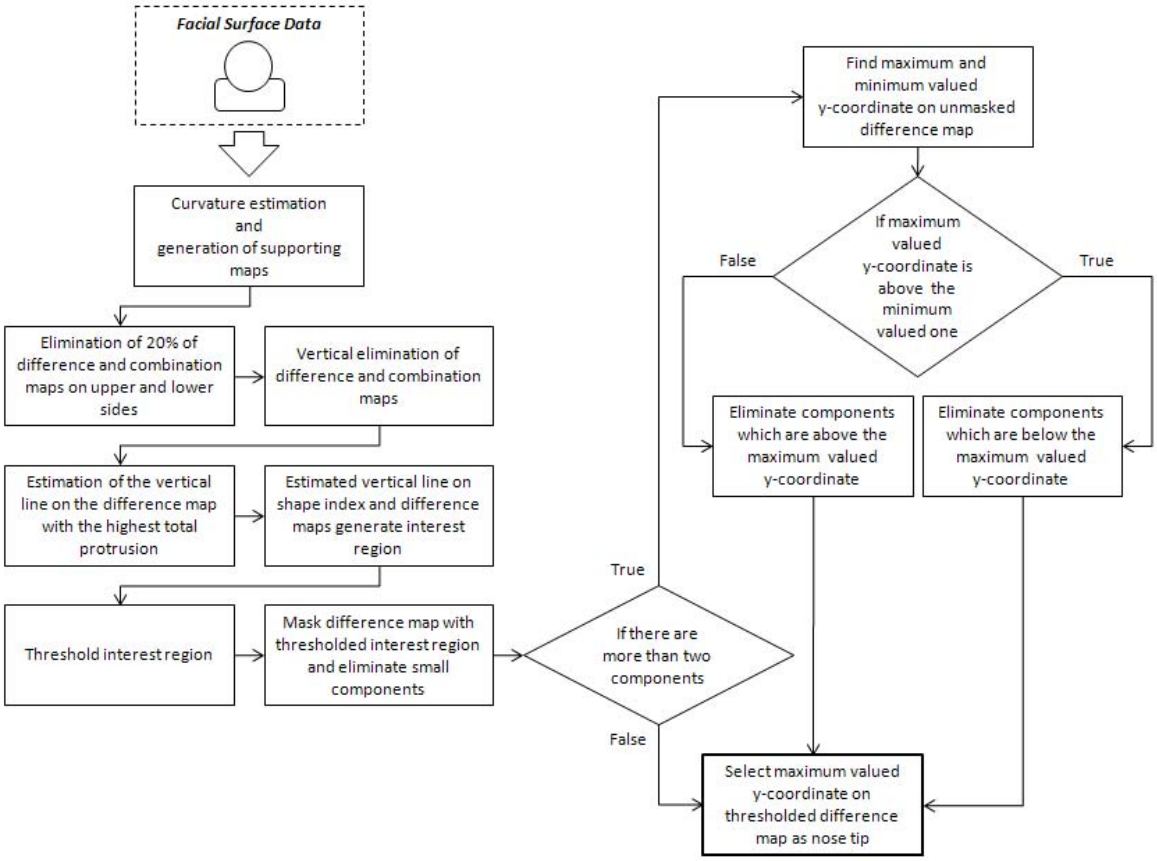


Figure 2.2. Flowchart of the initial nose tip estimation process.

A relative close point to the nose tip is required for segmentation of the nose region, so the initial nose tip localization is the first step. The overall diagram of the initial nose tip estimation procedure is shown in Figure 2.2.

In the first stage of the landmark localization algorithm, Gaussian (K) and mean curvature (H) values of the 3D facial surface are estimated, by computing the principal curvatures of the range image as follows:

$$K = \kappa_1(p) \times \kappa_2(p) \tag{2.11}$$

$$H = \frac{\kappa_1(p) + \kappa_2(p)}{2} \tag{2.12}$$

where, κ_1 and κ_2 denote the minimum and maximum principal curvatures, respectively. The principal curvatures are approximated by sliding a 57×57 pixels window on the range image. After this process, Gaussian and Mean curvature values are used to determine the fundamental elements on the facial surface. Fundamental elements and the related H and K values are shown in Table 2.1. In theory, a HK map can be comprehended as a label map. Although this kind of a labeling would be useful to restrict the search area for statistical methods, they entail a certain loss of information. Our approach is based on the more informative soft values of H and K maps. We compose a difference map $\mathcal{D}$, such that

$$\mathcal{D}_{i,j} = K_{i,j} - H_{i,j} \tag{2.13}$$

Table 2.1. Curvature analysis of facial surface based on H and K.

	$K < 0$	$K = 0$	$K > 0$
$H < 0$	Saddle Ridge	Ridge	Peak
$H = 0$	Minimal	Flat	Impossible
$H > 0$	Saddle Valley	Valley	Pit

Proposed $\mathcal{D}$ map reveals nose tip because of its special shape feature, like the peaked cap. An example K, H and the related $\mathcal{D}$ map can be seen in Figure 2.3. Then the difference map is smoothed by a Gaussian filter to cope with curvature sensitivity

and surface noise. The filter size is selected as 15×15 pixels, with a standard deviation of three pixels.

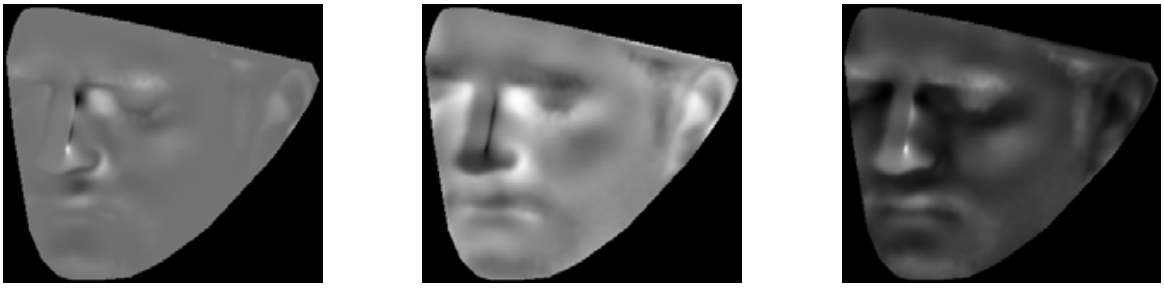


Figure 2.3. The mean, Gaussian, and the difference maps, respectively.

Although $\mathcal{D}$ can be defined as a protrusion map on the data, it is not a suitable method for finding the nose region on the facial surface with big rotations as shown in Figure 2.4. It shows only peaks on the sides of the nose under rotation variations. Therefore, we need a rotation invariant feature which gets higher values around the nose tip. To handle the rotation conditions we estimate an Eigen Map $\mathcal{E}$. A sliding 3×3 pixels block is used for $\mathcal{E}$ value estimation of each pixel on the depth image. X,Y and Z coordinates of nine 3D points on each block are taken as inputs and their covariance matrix is calculated. Then the eigenvalues and eigenvectors are estimated for the calculated covariance matrix. The biggest eigenvalue is selected and the ratio of the selected eigenvalue to sum of all eigenvalues is calculated. This is the value of $\mathcal{E}$-map for the center point of the related block. Estimated $\mathcal{E}$-map gives relatively higher values for points with quick depth changes such as points on nose slopes (see Figure 2.5). Additionally we know that $\mathcal{D}$ has higher values for ridges and peaks. Eventually, all the nose region can be highlighted by the combination of these two indicators. A combination map $\mathcal{C}$ is formed such that:

$$\mathcal{C}_{i,j} = \frac{\mathcal{D}'_{i,j} + \mathcal{E}'_{i,j}}{2} \tag{2.14}$$

where, $\mathcal{D}' = m(\mathcal{D})$, $\mathcal{E}' = m(\mathcal{E})$, and m denotes the min-max normalization function:

$$m(x) = \frac{x - min(x)}{max(x) - min(x)} \tag{2.15}$$

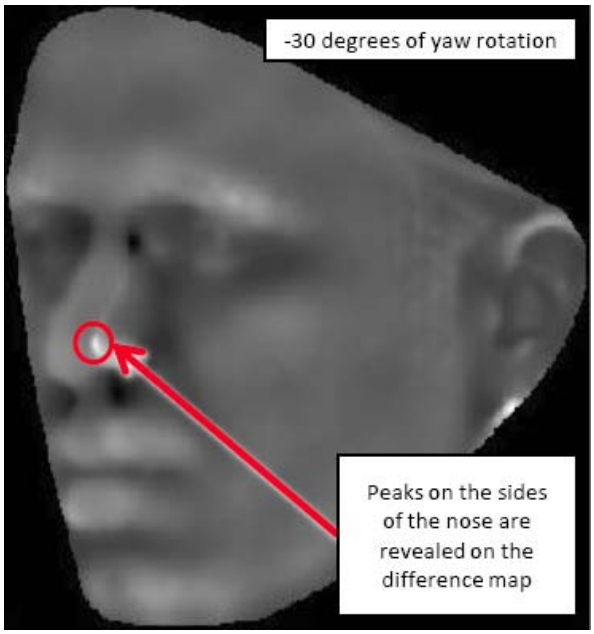


Figure 2.4. Peaks under -30 degree of yaw rotation on difference map.

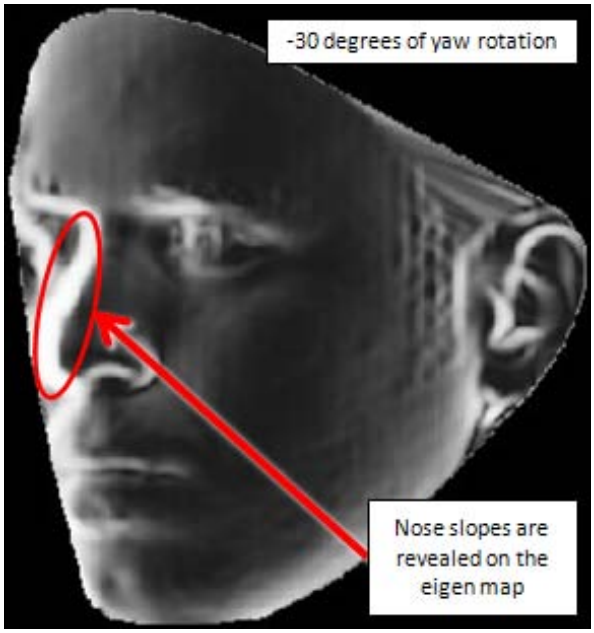


Figure 2.5. An example eigen map under -30 degree of yaw rotation.

An example of the combination map is shown in Figure 2.6. In the next step 20 per cent of difference and combination maps on upper and lower sides are horizontally eliminated to get rid of the outliers on the facial boundaries, by using the center point of the range image (see Figure 2.7).

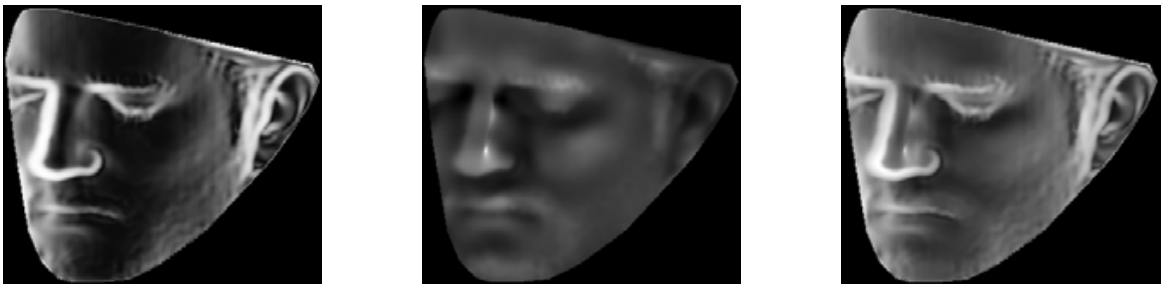
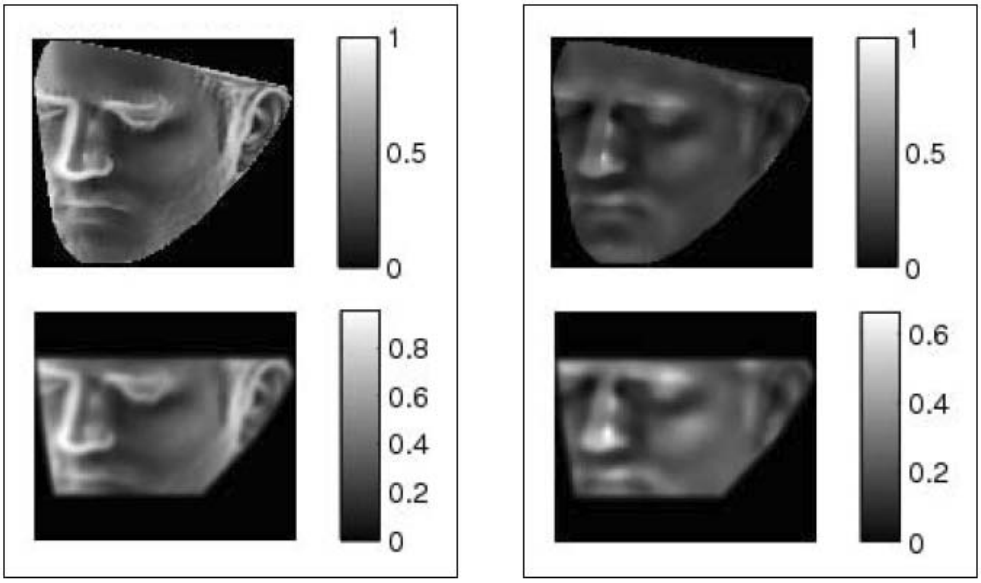
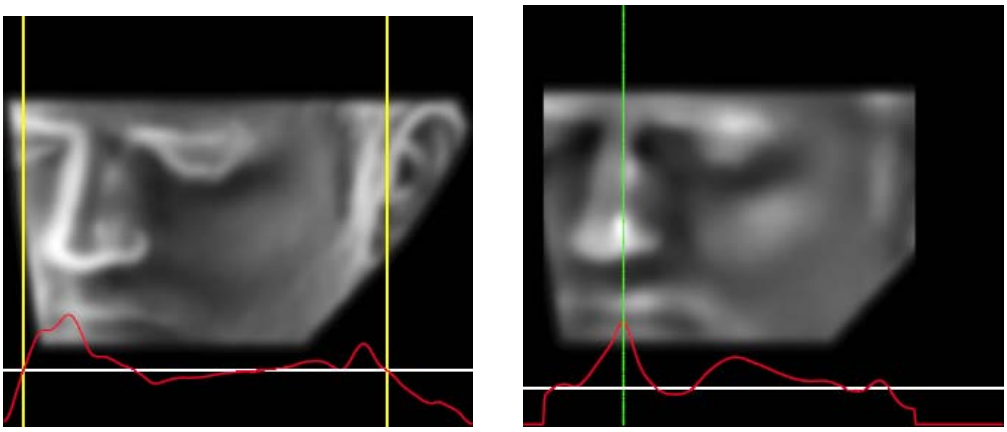


Figure 2.6. The eigen, difference and the combination maps, respectively.



(a) (b)

Figure 2.7. The horizontal elimination of (a) $\mathcal{C}$ and (b) $\mathcal{D}$ maps.



(a) (b)

Figure 2.8. (a) Vertical elimination and (b) maximum value selection, respectively. Red curves show the sum of the vertical values and the white line shows the mean values of the curves. Yellow lines denote the vertical elimination limits.

After the horizontal elimination process, vertical sum of the each column of the $\mathcal{C}$ map values are calculated. Estimated values for each column are used to determine the total protrusion of the related column of the difference map. The nose region displays more protrusion so the sum of the difference map values of the related column is expected to have the maximum value. Left and right sides of the $\mathcal{C}$ map are eliminated at the first and last vertical sum values which are smaller than the average of vertical sums. Although this assumption is valid in general, the nose tip may be outside of the face region on $\mathcal{C}$ map. Because of that, the vertical elimination step is skipped if the absolute maximum of the vertical projection curve is closer than five per cent of the width of the map to the border of the image. Lastly, the elimination is applied to the $\mathcal{D}$ map. Then, the sum of the squares of each point on the map is calculated vertically like previously. The x-coordinate of the maximum value is selected as the vertical line which passes through the nose region. The process can be seen in Figure 2.8.

After vertical and horizontal elimination steps, y coordinate of the point, close to the nose tip is required to be estimated. For this purpose, shape index of the related surface data is estimated (see Figure 2.9). The shape index (SI) of the face image is

the quantitative measure of the shape of a surface at a point [46], and it is defined as:

$$SI(p) = \frac{1}{2} - \frac{1}{\pi} tan^{-1} \frac{\kappa_1(p) + \kappa_2(p)}{\kappa_1(p) - \kappa_2(p)} \tag{2.16}$$

where $\kappa_1(p)$ and $\kappa_2(p)$ are the principal curvatures.

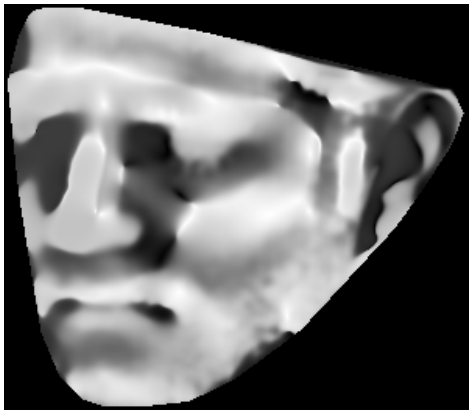


Figure 2.9. Shape index map.

Shape index is more discriminative on saddles and ridges on the face surface and the nose has a big saddle ridge. Therefore SI is used to reveal candidate regions. Previously estimated vertical line and two pixels neighbourhood of it on the shape index map generates interest region. Then the interest region is thresholded by the mean value of the whole shape index map (see Figure 2.10).

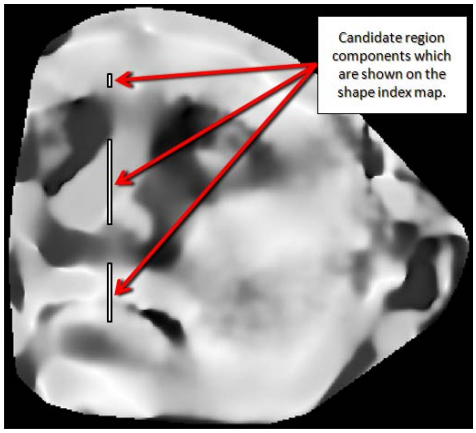


Figure 2.10. Nose tip candidate region components on the shape index map.

Narrow regions are formed by these operations. If there are more than two regions, regions which are smaller than half of the biggest region are eliminated. Generally the eliminated regions are on the lips and the forehead. If the number of regions is still more than one after elimination, maximum and minimum $\mathcal{D}$ values on the interest region are computed. Interest region is a narrow strip that is the superset of all components as shown in Figure 2.11. Points with the minimum $\mathcal{D}$ value are generally on the upper

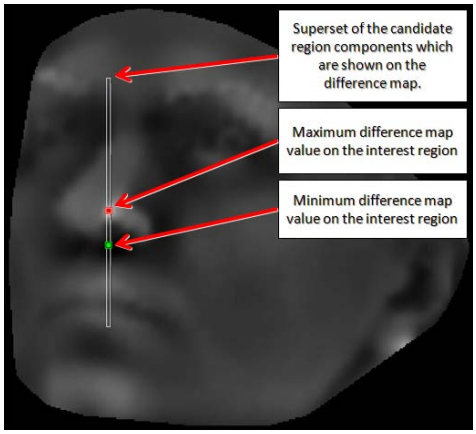
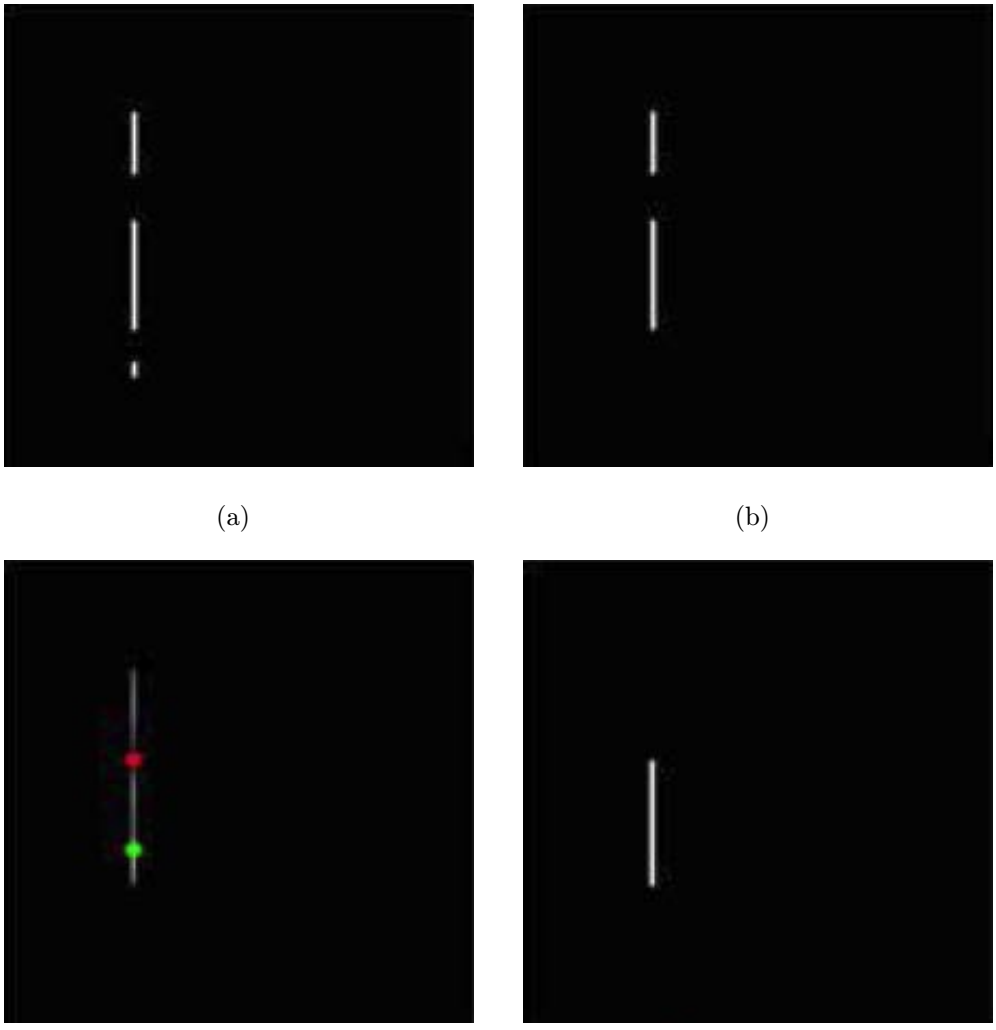


Figure 2.11. Superset of candidate region components on the difference map.

nose saddle or between the upper lip and the lower part of the nose. Maximum $\mathcal{D}$ value indicates the nose tip. If the minimum point is under the maximum, minimum point is assumed as located under the nose and its lower side is eliminated. If it is above the maximum, it is assumed as nose saddle and its upper side is eliminated.

Lastly, horizontal projection of the $\mathcal{D}$ values are calculated for each row of remaining points and the point with the maximum value is selected as a close point to the nose tip. In Figure 2.12(a), three candidate regions are shown. Small candidate regions elimination and maximum/minimum $\mathcal{D}$ values estimation on the interest region are illustrated in Figure 2.12(b) and Figure 2.12(c), respectively. The red point is the minimum and the green one is the maximum-valued point. Minimum-valued point is above the maximum in the given example, so the it is assumed as nose saddle. Finally upper side of the nose saddle elimination is shown Figure 2.12(d). Initial estimation of the nose tip is shown in Figure 2.13.



(a) (b)

(c) (d)

Figure 2.12. Elimination of candidate regions: (a) Initial candidate regions, (b) after elimination of small regions, (c) superset of the candidate regions. Red and green points denote minimum and maximum values, respectively. (d) Final region.

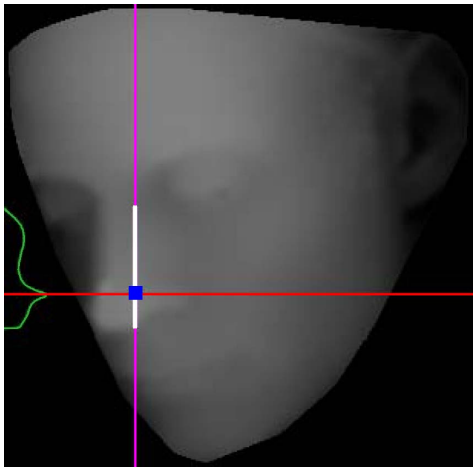


Figure 2.13. Initial estimation of the nose tip: Vertical curve on the left hand side, shows the horizontal projection of the $\mathcal{D}$ values for the interest region.

2.1.3. Coarse Nose Cropping

After the estimation of a close point to the nose tip, points on the initial $\mathcal{D}$ map are eliminated if they are located away from the estimated nose tip point. Elimination distance threshold is 17.5 per cent of the map diagonal (see Figure 2.14). The horizontal line which passes through the y-coordinate of the nose tip is determined as L_y. The points, with minimum $\mathcal{D}$ value, on the L_y, are estimated. Then, the left minimum point P_L and right minimum point P_R are used as starting points for nose cropping by greedy search. L_y, P_L, and P_R are shown in Figure 2.14. Coarse cropping is based on estimation of the minimum energy path on the $\mathcal{D}$ map. Left and right-hand side minimal nose paths are determined and the region between the minimal paths are cropped. Minimum energy path estimation starts at the point, P_i where P_i is initialized as P_L for left side and P_R for right side of the nose. Path estimation goes through data upwards by checking three neighborhood points on the upper horizontal line. The neighbor point with the minimum $\mathcal{D}$ value is selected as new P_i, and the procedure iterates till P_i arrives the upper elimination limit. Minimal path estimation procedure is illustrated at the right side of Figure 2.15. Same operations are done from initial P_i to the lower elimination limit. Coarse cropping can be seen in Figure 2.15.

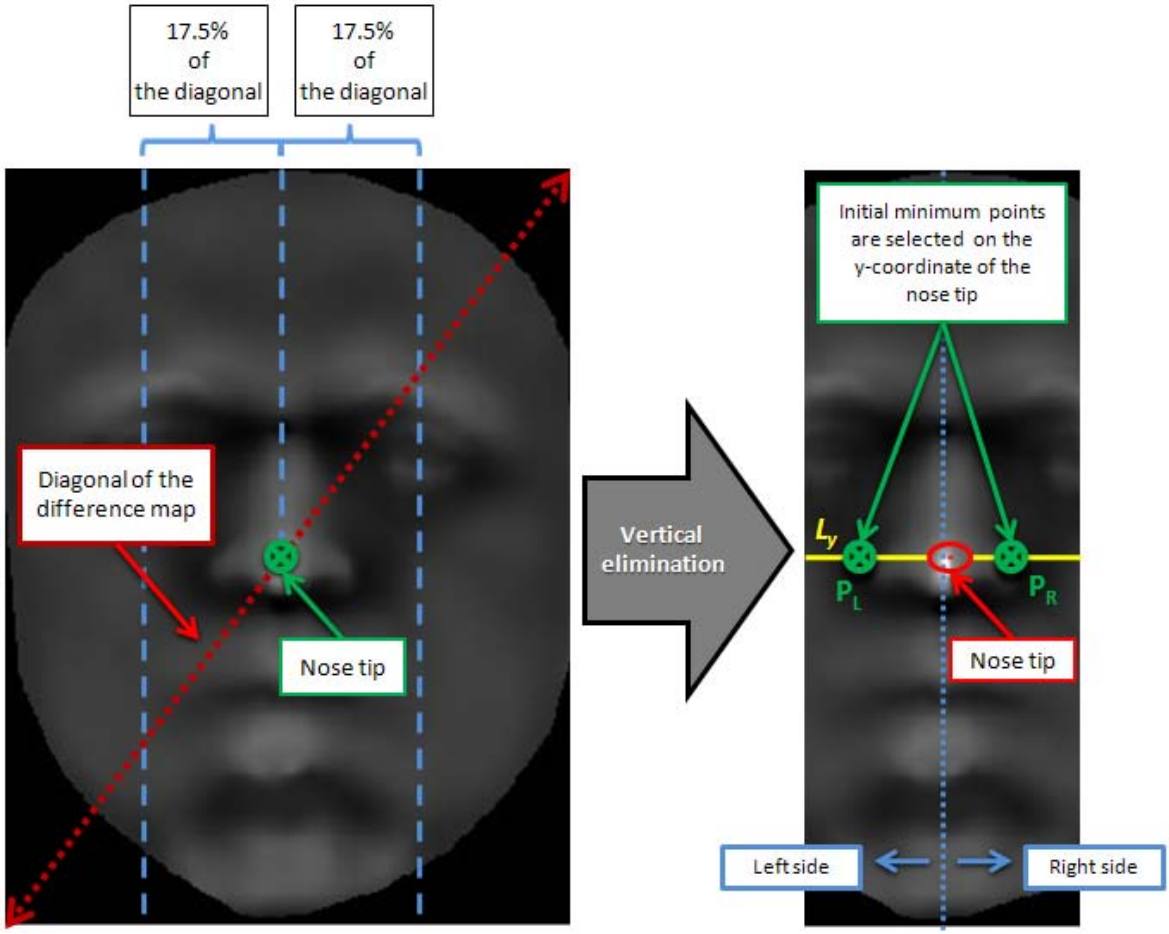


Figure 2.14. Starting points of the minimum energy path estimation.

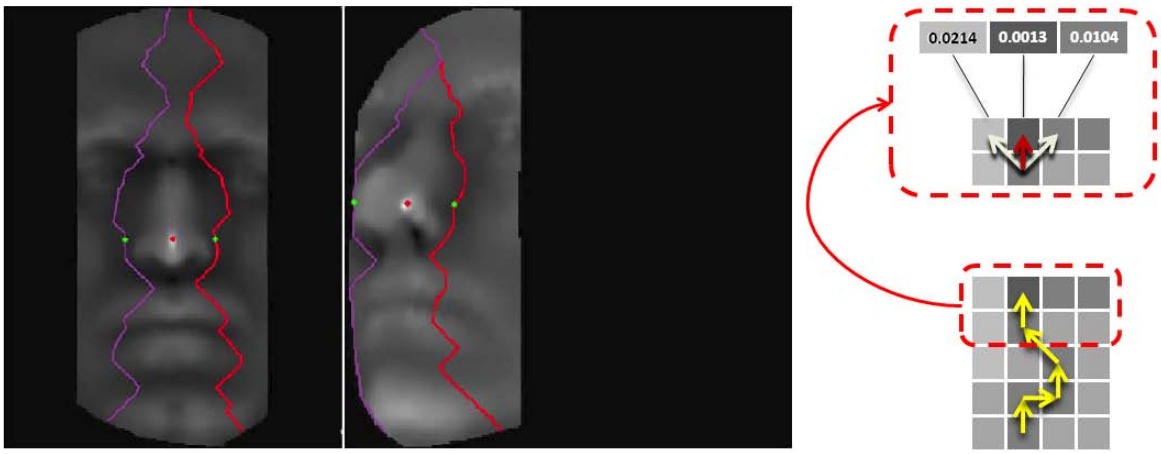


Figure 2.15. Minimum energy path estimation: Green points show the initial minimum points. Red points denote the estimated nose tip.

2.1.4. Yaw Rotation Estimation

Segmentation of the nose in the rotated facial surface data is a hard problem and robust segmentation based on curvatures is practically impossible in big yaw rotations. Rotated nose peak in big yaw rotations can not be estimated accurately by curvatures. Therefore, yaw rotations have to be estimated to rotate the facial 3D surface data.

It can be assumed that the relative horizontal position of the nose tip on the surface data gives the yaw rotation angle of the face. Hence, normalized nose tip positions for subjects are determined relative to the left most valid point of the facial surface. In other words, minimum x coordinate of the facial surface is subtracted from the x coordinate of the nose tip, and the ratio of the calculated value to the width of the facial surface gives the normalized nose tip position. Since the nose tip position of the frontal poses is approximately at the middle, its value is 0.5. 90 degrees to the left and 90 degrees to the right rotated surfaces' nose tip positions are 0 and 1, respectively, as illustrated in Figure 2.16.

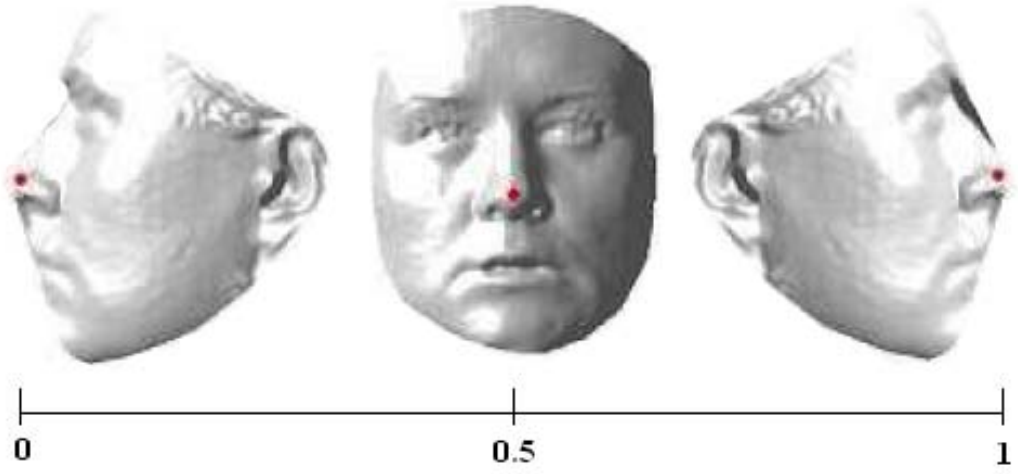


Figure 2.16. Relative nose tip positions on the facial surfaces.

Consequently, a rough scale for nose tip position to rotation angle is introduced. Rotations higher than 30 degrees are dismissed to avoid the possible risk of wrong high rotation estimations. Introduced scale is given in Table 2.2.

Table 2.2. Scale of yaw face rotations for different nose tip positions.

Yaw Rotation Angle(Degrees)	Relative Horizontal Nose Position
+30	0.75 to 1
+20	0.65 to 0.75
+10	0.55 to 0.65
0	0.45 to 0.55
-10	0.35 to 0.45
-20	0.25 to 0.35
-30	0 to 0.25

Introduced scale assumption has been tested in Bosphorus database v.1 and v.2. In Figure 2.17, y-axis shows the relative nose tip positions of the subjects in BOSv1 which are indicated at the x-axis. The mean value of frontal poses in BOSv1 is approximately 0.5. The relative nose tip positions are between 0.1 and 0.4 for 10 to 45 degrees of yaw rotations. Similarly, Figure 2.18 shows the results for BOSv2. Obtained results verify that the relative nose tip position can be used as a coarse rotation estimation metric for facial surfaces.

2.1.5. Yaw Pose Correction

In this step the estimated yaw angle Θ is used to correct the rotation and to bring the facial points to a frontal pose. The rotation matrix Ry is defined as:

$$R_y = \begin{bmatrix} \cos(-\Theta) & 0 & \sin(-\Theta) \\ 0 & 1 & 0 \\ -\sin(-\Theta) & 0 & \cos(-\Theta) \end{bmatrix} \tag{2.17}$$

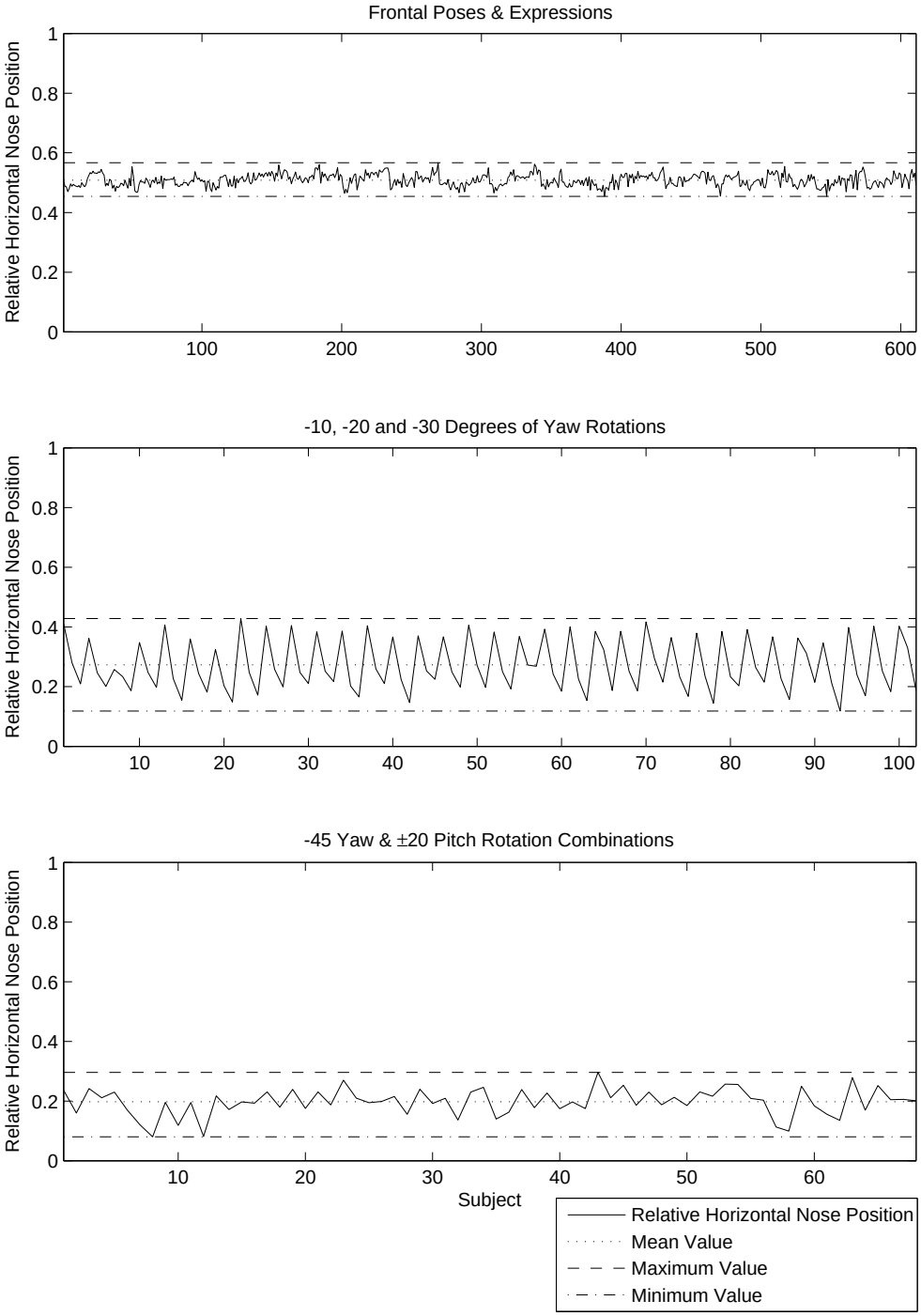


Figure 2.17. Relative nose tip positions in BOSv1.

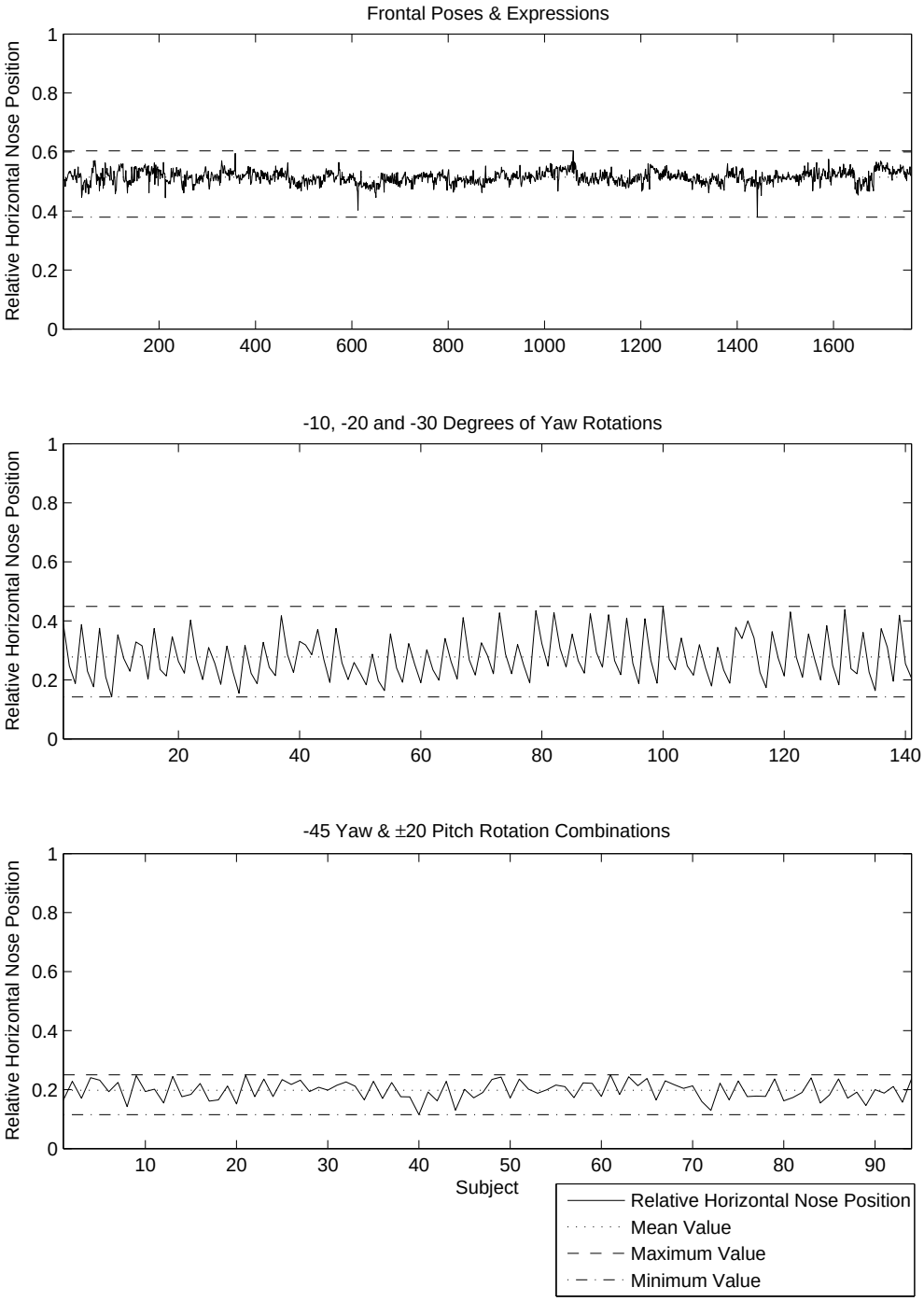


Figure 2.18. Relative nose tip positions in BOSv2.

The key point in this step is the normalization of the point cloud before the rotation procedure. Rotation axis of the facial surface has to be translated to the y-axis to get correct results. Anatomically, yaw rotation axis is the middle of the neck for human head (see Figure 2.19), so the required rotation axis normalization of the point cloud $P = \begin{pmatrix} X & Y & Z \end{pmatrix}$, can be approximated as follows:

$$X_i = X_i - x_{offset} \tag{2.18}$$

$$Y_i = Y_i - y_{offset} \tag{2.19}$$

$$Z_i = Z_i - z_{offset} \tag{2.20}$$

where,

$$x_{offset} = \frac{\max(X) + \min(X)}{2} \tag{2.21}$$

$$y_{offset} = \frac{\max(Y) + \min(Y)}{2} \tag{2.22}$$

and z_{offset} is the farthermost z coordinate of the facial surface to the camera, derived as:

$$z_{offset} = \min(Z) \tag{2.23}$$

After the rotation process, normalization offsets are taken back to conserve real 3D coordinates. Sample rotation from -30 degrees to 0 degree is shown in Figure 2.20.

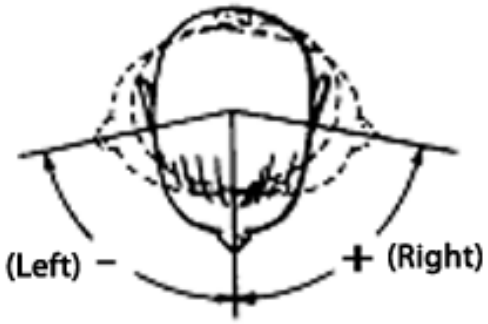
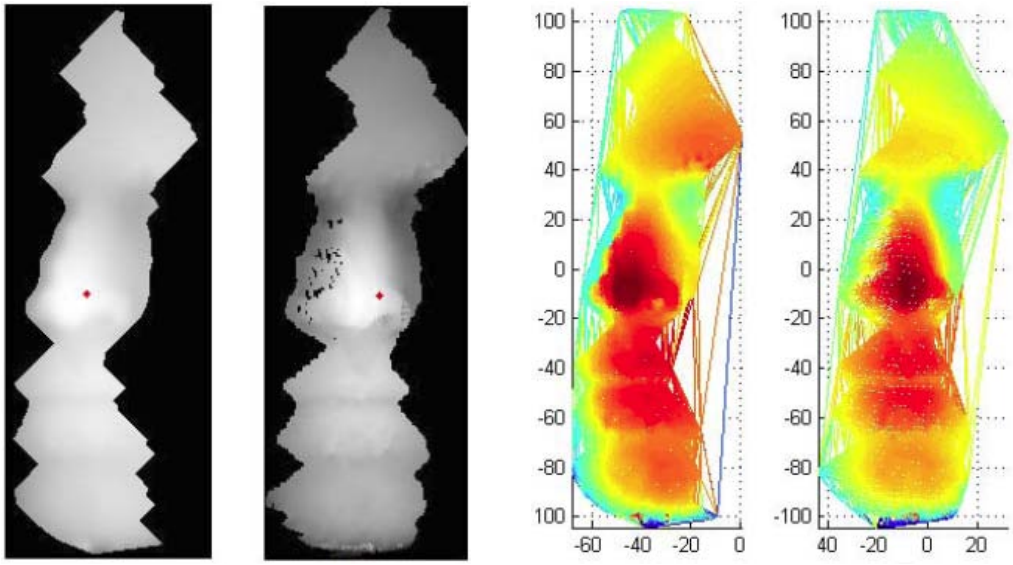


Figure 2.19. Rotation axis and directions of human head.



(a) (b)

Figure 2.20. Sample rotation from -30 degrees to 0 degree: (a) On the range image and (b) on the point cloud vertices.

2.1.6. Nose Tip Correction and Fine Level Cropping

When the data is successfully rotated, curvatures are estimated again for the current situation. $\mathcal{D}$ map is calculated again by the new curvature values. Figure 2.21 displays the new $\mathcal{D}$ and SI maps .

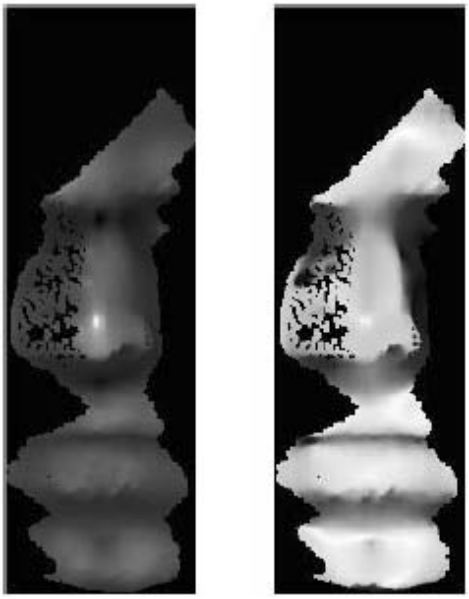


Figure 2.21. Rotated distance and shape index maps, respectively.

Nose tip correction is performed as explained in the initial nose point estimation procedure. Additionally, upper side of the nose is eliminated horizontally where the horizontal average of $\mathcal{D}$ values is minimum, also lower side of the nose is eliminated by the same way. Nose tip correction and elimination are shown in Figure 2.22. After the nose tip point is updated, revealed side of the nose after pose correction is cropped vertically where the holes are significantly increased. Before the elimination of holes, the map is scanned through the revealed side horizontally. When a hole is reached with size more than two pixels, all the remaining pixels are eliminated horizontally. This procedure is shown in Figure 2.23. Lastly, the total number of the invalid points for each column on the range map is computed. The side revealed by rotation is eliminated vertically where the total number of invalid points changes abruptly. These points on the nose are eliminated instead of interpolation because interpolation degrades the quality of the facial surface data.

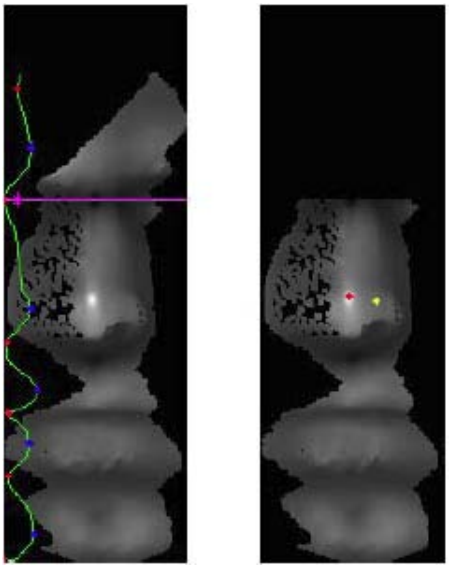
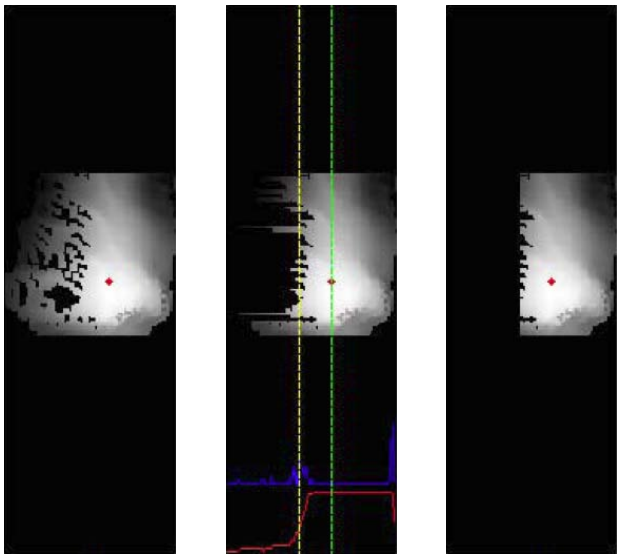


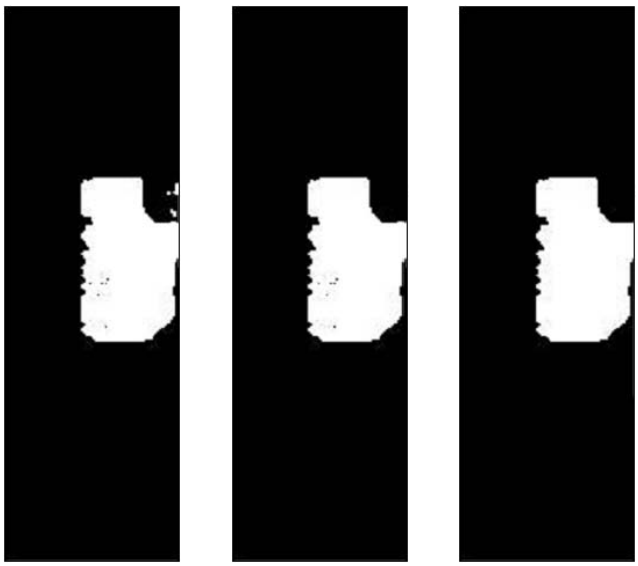
Figure 2.22. Upper/lower side elimination and the nose tip correction, respectively. Green curve denotes the average $\mathcal{D}$ value for each row. Yellow and red points show the initial and the corrected nose tip points, respectively.



(a) (b) (c)

Figure 2.23. (a) The rotated range image, (b) vertical elimination procedure, and (c) the cropped nose. Red and blue curves denote valid pixels for each column and change in valid pixel numbers, respectively. Yellow line shows the vertical elimination limit.

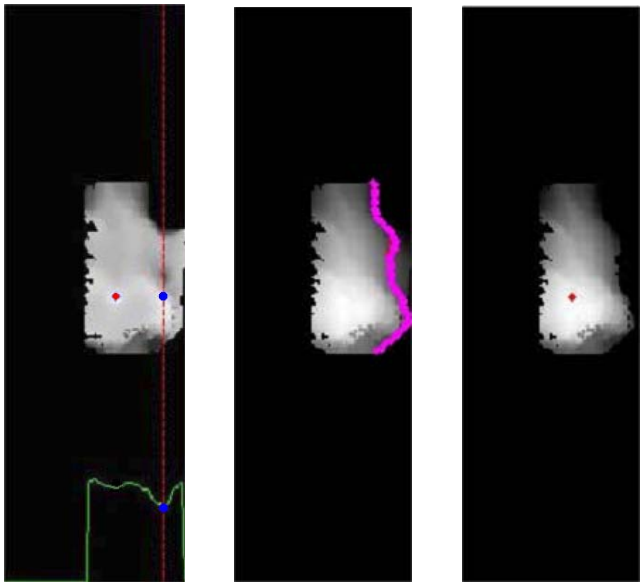
After the elimination of holes and noisy regions for rotated poses, $\mathcal{D}$ map is thresholded. The threshold level is selected as 10 per cent empirically. Additionally binary operations are done to eliminate disconnected smaller regions and fill the holes (see Figure 2.24). In the next step, minimum energy path following algorithm is applied again. It is applied to the nose sides which are not affected by rotation. The minimum path following method differs from the previous usage of it. In the fine level cropping, it is started at two different initial points for both of the related nose sides. First initial point is the same as in the coarse cropping procedure. Hence, the other initial start point is the minimum-valued $\mathcal{D}$ on the uppermost horizontal line of the nose region. Both of the minimum paths are voted and the closer point to the nose tip is selected for each row of the map. Details can be seen in Figure 2.25.



(a) (b) (c)

Figure 2.24. (a) Thresholded binary map, (b) elimination of disconnected regions, and (c) holes filling.

At the last step, vertical limits of the nose are cropped. Initially y-coordinate of the nose tip is corrected by selecting the maximum averaged horizontal line on the eliminated $\mathcal{D}$ map. Afterwards, minimum averaged row above the nose tip is selected as the initial estimation of the upper limit of the nose region. Then, the profile view of the facial surface depth data is generated on the vertical line which passes through



(a) (b) (c)

Figure 2.25. (a) The nose tip and the minimum-valued $\mathcal{D}$ on the right side of the nose are denoted by red and blue points, respectively. Green curve shows the difference map values on the y-coordinate of the nose tip. (b) The minimum energy path and (c) the cropped nose.

the nose tip. Uppermost and lowermost points of the profile are found. 20 pixels of the lower side and the five pixels of the uppermost sides are eliminated to get rid of outliers on the face boundaries. After this process, a line which passes through the uppermost and the lowermost points on the eliminated profile is estimated. The line gives us the approximate axis of the profile (see Figure 2.26). The profile data is transformed to the y-axis by using profile axis. This process is called profile normalization. It is assumed that the maximum-valued profile point gives the nose tip on the normalized profile. Normalized profile is shown at the middle of Figure 2.26. By using this assumption, final nose tip coordinates are estimated. Then the local minima are calculated on the profile curve. The first minimum point above the nose tip is selected as uppermost limit of the nose region. By the same manner, first minimum point below the nose tip is selected as the lowermost limit. However, the lowermost limit cannot be estimated correctly for every subject. Therefore two more methods are introduced to handle these conditions. First one is checking the gradient of the profile curve. The first

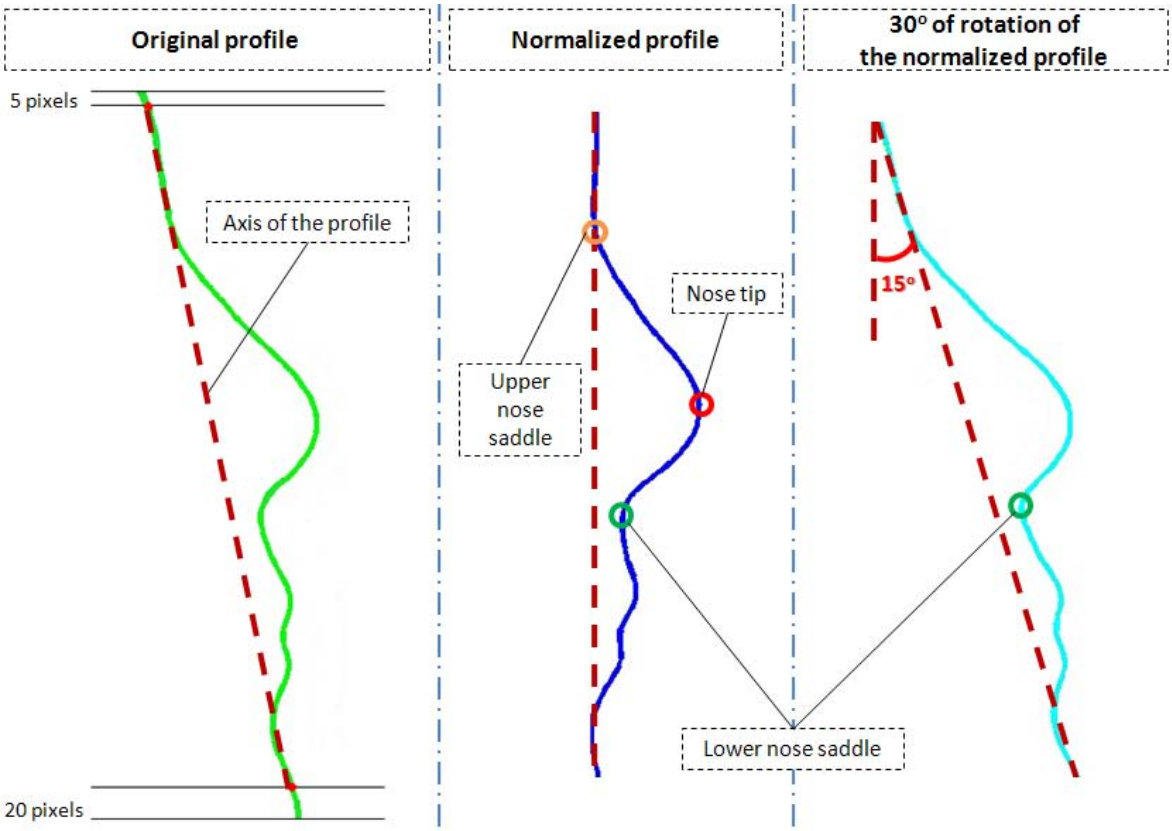


Figure 2.26. Upper and lower limits estimation procedure of the nose region.

local minimum below nose tip on the profile gradient denotes the lowermost limit. The other method is to rotate the normalized profile data 15 degrees about the uppermost point. Rotation is done to reveal the lowermost limit of the nose as a local minimum below the nose tip on the profile curve. Estimated three different lowermost limits are voted for unrotated conditions and the upper one is selected. Estimation on the rotated profile is used for rotated data. This is needed because, rotated data is noisy in spite of Gaussian smoothing on range data. The Gaussian filter size is selected as 7×7 pixels, with one standard deviation corresponding to three pixels. All the steps of this procedure are shown in Figure 2.27. Figure 2.27(a) displays the nose tip

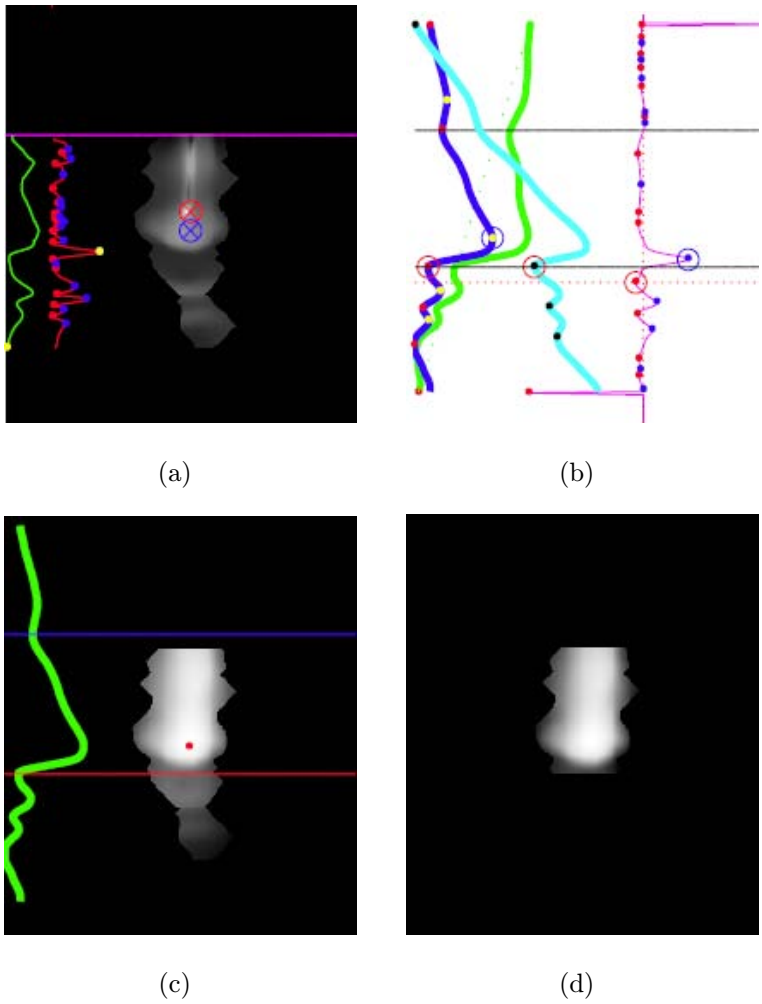


Figure 2.27. (a) Nose tip correction, (b) original and rotated face profiles, (c) upper/lower limits of the nose and normalized profile. (d) Cropped nose.

correction and initial uppermost limit estimation on $\mathcal{D}$ map. Green curve shows the row averages of the $\mathcal{D}$ map and the red curve denotes the change in $\mathcal{D}$. Red and blue markers show the previous and updated nose tip points, respectively. Magenta line denotes the estimated uppermost limit. In Figure 2.27(b) green, blue and cyan profiles are original, normalized and rotated face depth profiles, respectively. Magenta curve shows the gradient of the normalized profile. Red 'o' shape markers denote estimated lowermost limits. In Figure 2.27(c) upper and lower limits of the nose on the range data and normalized profile are shown. Red point shows the final estimation of the nose tip point. Blue and red lines denote upper and lower limits, respectively. Lastly the cropped nose is seen in Figure 2.27(d).

Finally, cropped nose data is ready for the recognition phase. An example of 3D cropped nose is shown in Figure 2.28.

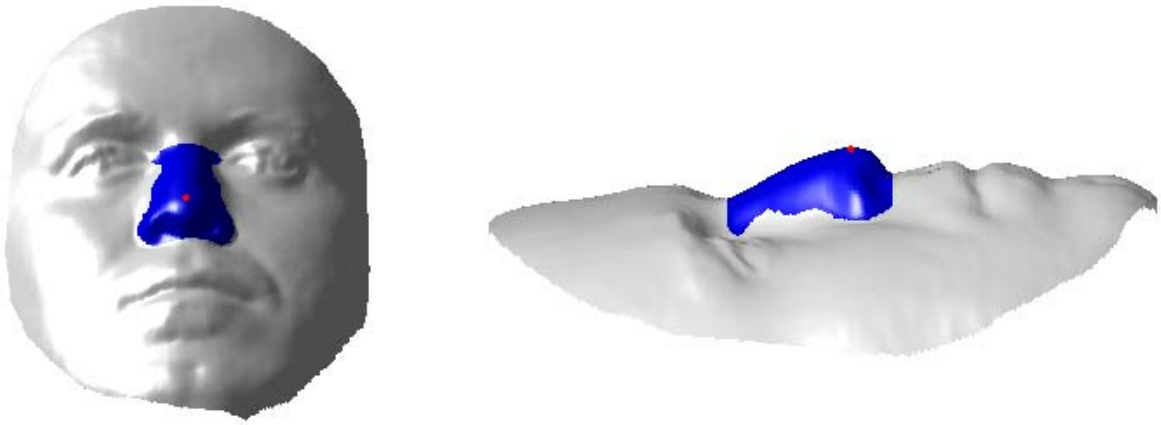


Figure 2.28. Cropped nose region and nose tip are shown in blue and red, respectively.

2.2. Voting-Based Nose Tip Estimator

In this section we describe another curvature based heuristic method to localize the tip of the nose under severe conditions. The proposed algorithm finds labelled candidate regions for the nose, and selects the most plausible location based on curvature values, which makes it rotation invariant. Overall nose tip estimation procedure is shown in Figure 2.29.

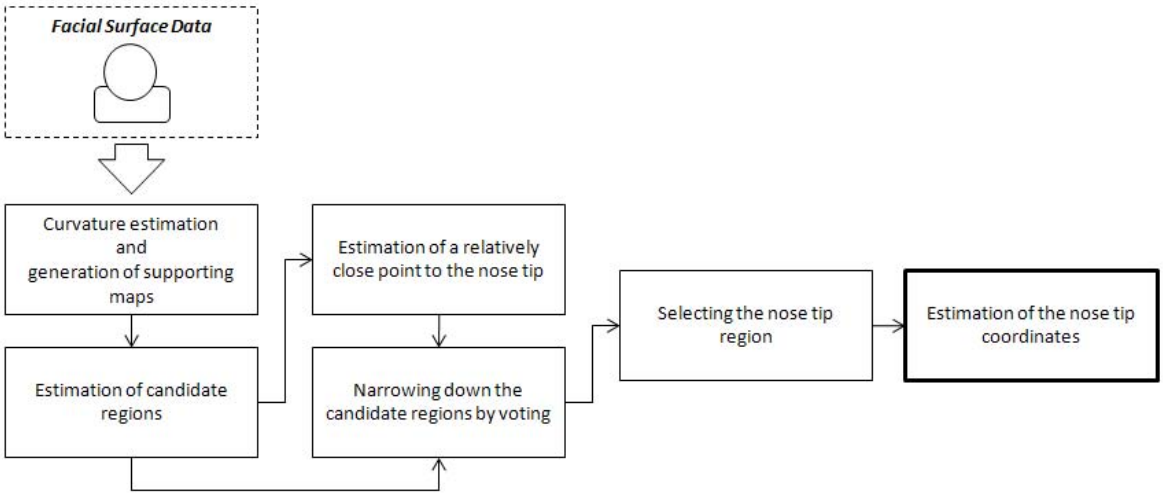


Figure 2.29. Flow of the voting-based nose tip estimation.

2.2.1. Estimation of Candidate Regions

Initially, Gaussian and mean curvature values of the 3D facial surface are calculated, by computing the principal curvatures of the range image as explained in Section 2.1. However, the window size is set as 3×3 in this method. After this process, a difference map $\mathcal{D}$ is produced and smoothed (see Figure 2.30) as explained in Section 2.1. Gaussian filter size is set as 15×15 and standard deviation is selected as three pixels, empirically.

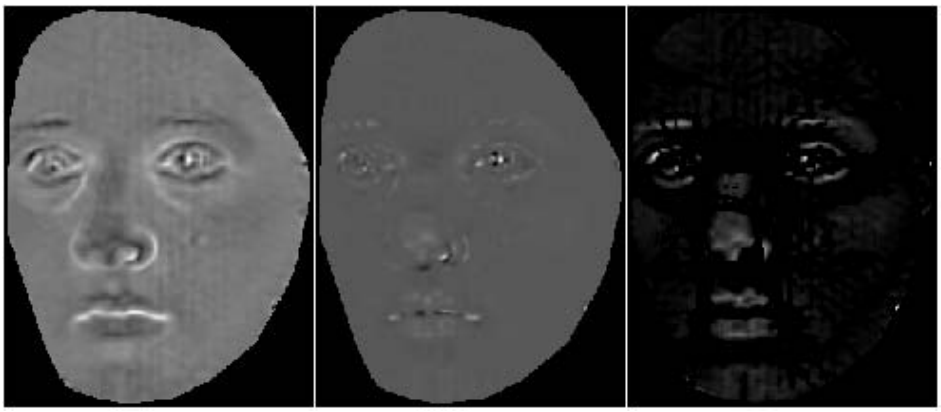


Figure 2.30. The mean, Gaussian, and the difference maps, respectively.

Additionally, the smoothed values are mapped into a logarithmic scale to accentuate details on the facial surface, and compress depth differences as follows:

$$\mathcal{D}'_{i,j} = ln(\mathcal{D}_{i,j} + 1) \tag{2.24}$$

As a result of this transformation, details on the facial surface are revealed. Finally, the map is min-max normalized to the range [0, 1], and thresholded. To annihilate the unrelated regions, thresholding is done. A small elimination threshold is selected to decrease the the probability of missing the nose. A binary image is produced after thresholding. (See Figure 2.31).

(a) (b)

Figure 2.31. Interest Region estimation process: (a) Thresholded $\mathcal{D}'$ and (b) thresholded binary maps.

2.2.2. Narrowing down the Candidate Regions

Two different features are combined to narrow down the candidate regions for the nose. First, we compose a change map Λ of the face image $\boldsymbol{I}$ by:

$$\Lambda = m(ln\nabla(\boldsymbol{I}))^2 \tag{2.25}$$

where ∇ is the gradient operator, and m denotes the min-max normalization function.

(a) (b) (c)

Figure 2.32. Change map Λ formation: (a) Range image $\boldsymbol{I}$, (b) gradient of $\boldsymbol{I}$, and (c) change map (Λ).

Then, the following point P_c is a relatively close point to the nose tip:

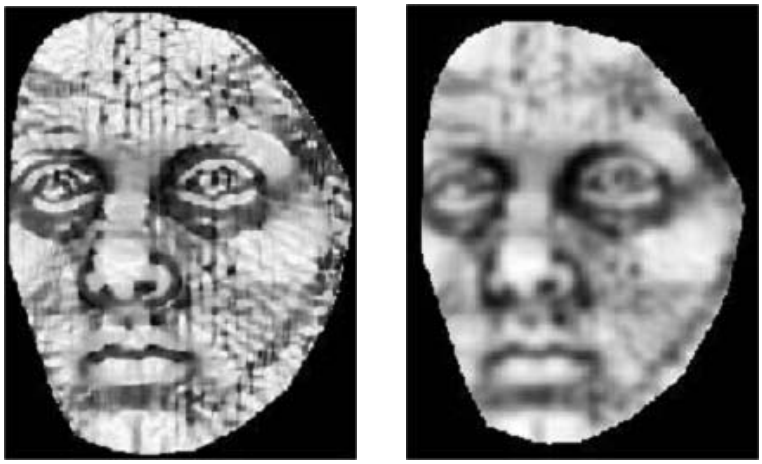
$$P_c = \frac{1}{ij} \sum_i \sum_j \Lambda_{i,j} * S'_{i,j} \tag{2.26}$$

which connotes the weighted centroid of points described by the Gaussian smoothed shape index S' (See Figure 2.33) multiplied by the corresponding change map value at each point.

A circular region is cropped with a dynamic radius r around P_c (See Figure 2.34). The radius is approximated to include roughly the same proportion of data, regardless of scale after thresholding $\mathcal{D}'$. Determination of the radius size depends on the eigenvalues of the three-dimensional covariance matrix of the 3D points. After eigenvalues have been computed, ratio of the biggest eigenvalue to the sum of eigenvalues, E_{rate} is calculated. The radius is defined as:

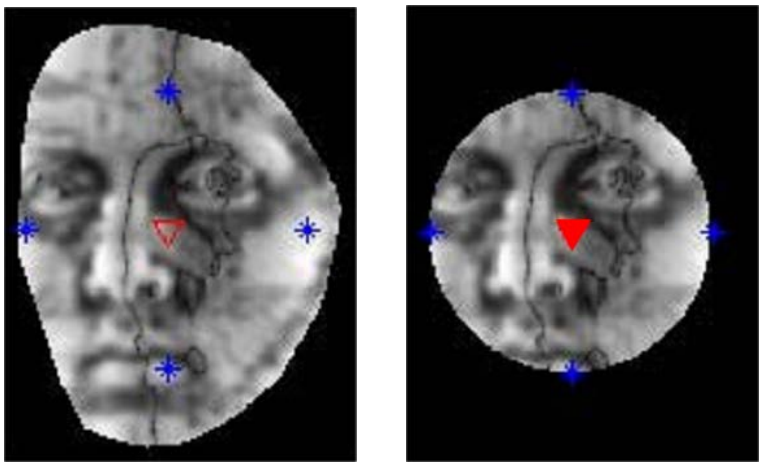
$$r = d_I \times [0.25 + (E_{rate} \times 0.4)] \tag{2.27}$$

where, d_I is the diagonal length of $\boldsymbol{I}$.



(a) (b)

Figure 2.33. (a) Shape index and, (b) smoothed shape index maps. Lighter colors denote higher values.



(a) (b)

Figure 2.34. Cropping a circular region around P_c: (a) Initial face, and (b) cropped face.

After narrowing the interest region thence, P_c is updated for the remaining region by using Eq 2.26. Then the candidate region components which are smaller than one per cent of the largest region component are eliminated. To get rid of the holes and the fractures on the binary image, dilation and erosion operations are done consequently by a two pixels sized disk. Components smaller than the one per cent of the largest region component are eliminated after every binary operation. Lastly the remaining components on the binary map is dilated by five pixels sized disk and all holes are filled inside of it. This processing step is shown in Figure 2.35.

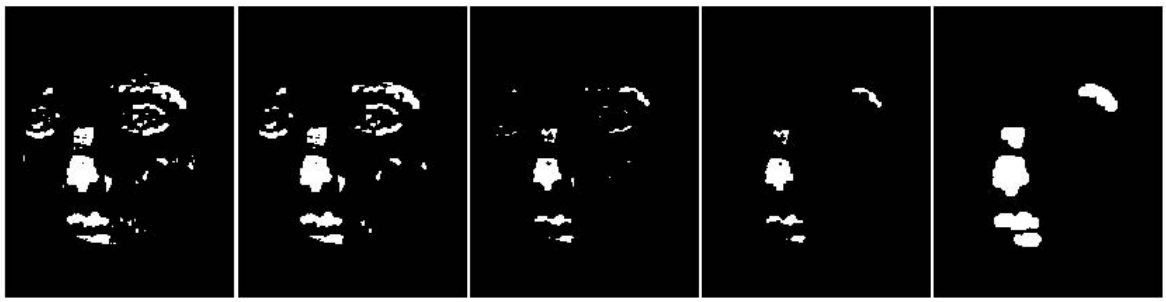


Figure 2.35. Binary operations, respectively.

2.2.3. Selecting the Nose Tip Region

When a small number of candidates that contain the nose region are left, P_c is updated by the weighted centroid of the smallest circular region which contains all components. Each point within this circle is weighted by its square shape index value. Then, a component which contains the nose tip is selected by a voting process. The voting procedure is based on three different descriptors. The first one is the roundness of the components and it is estimated by the components' ratios of width to height of their bounding box, eccentricity and the proportion of the pixels in the bounding box. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. It is assumed that a circle's width to height rate of its bounding box, eccentricity and the proportion of the pixels in the bounding box are 1,0 and $4-\pi$ respectively. The second descriptor is average protrusion value of the components. It is defined as the mean value of $\mathcal{D}$ for every component. The last one is the distance to P_c of component centroids(See Figure 2.36).

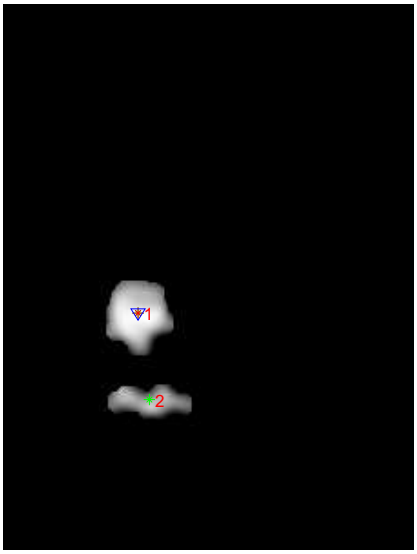


Figure 2.36. Nose tip region selection.

Nose tip is the point with highest protrusion within the nose region, so a protrusion map Φ is composed by using SI and $\mathcal{D}$ maps, also it is defined as:

$$\Phi_{i,j} = SI_{i,j} * \mathcal{D}_{i,j} \tag{2.28}$$

Nose tip coordinates are estimated by the weighted centroid of the selected region's protrusion map values. As an heuristic, if the estimated nose tip point is so close to the face boundaries horizontally, closest boundary point at the same y-coordinate is selected as nose tip. Threshold distance for this rule is used as five per cent of the face width. Figure 2.37 shows the correctly located nose tip.

2.3. 3D Face Registration

Registration is the alignment of the similar point sets before the recognition process. It is a very critical procedure because 3D facial surfaces, initially, have different rotations, translations and scales which depend on data acquisition conditions and pose variations. For the comparison of point sets, they have to be aligned first, because 3D point set comparisons are generally computed due to point-to-point or point-to-surface distances.

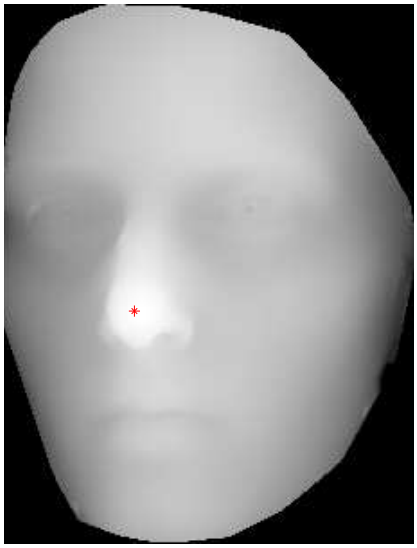


Figure 2.37. Correctly located nose tip on the range image.

3D face registration has two steps in general; coarse registration and fine registration. Coarse registration step is used to initialize 3D facial data for the fine registration step. It increases the convergence accuracy of point sets. Most of the coarse registration procedures depend on fiducial points called landmarks [47] such as nose tip [48, 49, 26, 50] and the eyes [12, 51]. Generally, iterative methods are used for fine registration steps, and the most popular technique is ICP.

Highest accuracy rates, ever, have been achieved in the literature by one-to-all registration methods. In these type of methods, all the face samples in the probe set are registered to each training face sample in the gallery before comparisons. One-to-all registration is not preferable in spite of its high accuracy, because its time complexity is too high. Irfanoğlu et al. proposed to use an Average Face Model (AFM) for 3D face registration to increase computational performance. In the proposed method, test faces are registered to the AFM only, because all the gallery faces are initially aligned to AFM as well.

In this thesis, both the one-to-all and one-to-one registration methods have been used as described above. Automatically cropped nose segments are used as test and training data, instead of using whole facial data. All the cropped nose segments in

the gallery and the probe set are initially translated as the nose tip will be the origin (0,0,0). Then all of them are registered to a Average Nose Model (ANM) by ICP. In the ANM-based registration approach, a test nose need only be registered to the ANM, since the correspondence of the gallery to the ANM is already established.

By contrast, in the one-to-all ICP, each of nose samples in the probe set are compared with every sample in the gallery set. Additionally, all probe samples are registered to each training nose sample in the gallery by ICP and the comparisons are done for every probe and gallery face couple. The mathematical fundamentals of the Iterative Closest Point algorithm, Procrustes analysis and Average Nose Model generation will be explained in the following sections.

2.3.1. Iterative Closest Point Algorithm

Iterative Closest Point algorithm is employed to match two point sets. This matching is generally used to find relative transformation of 3D surfaces from different scans. The algorithm has a high time complexity but its simple implementation and applicability to three dimensional geometric shape representations make it popular for 3D object registration. It iteratively estimates the transformation (translation, rotation) between two point clouds. The inputs are two point clouds, initial estimation of the transformation and the criteria for stopping. As an output of the algorithm, refined transformation between two scans is given.

Let us start with explaining a general form of the closest point computation which is applicable to n dimensions. Assume that P is a point set such that $p_i \in P$ where $i = 1, 2, ..., N_p$ and p_i is a point in three dimensional space such as $p_i = (x_i, y_i, z_i)$. Then the Euclidean distance between p_i and p_j can be calculated as follows:

$$d(p_i, p_j) = ||p_i - p_j|| = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2} \tag{2.29}$$

The distance between a reference point p_r and the point set can be calculated as:

$$d(p_r, P) = min_{i\epsilon 1,\ldots,N_p} \quad d(p_r, p_i) \tag{2.30}$$

A closest point $p_{closest}$ in the point set P has to satisfy the following definition:

$$d(p_r, p_{closest}) = d(p_r, P) \tag{2.31}$$

Let us assume M is the model point set and D is the data point set which corresponds to M. The main steps of the ICP algorithm, proposed by Besl and McKay [20] can be given as: (a) Selection of point correspondences $w_{i,j}$ in {0,1}. (b) Minimization of the error function E for rotation R, and translation t, where E is defined as:

$$E(R,t) = \sum_{i=1}^{N_m} \sum_{j=1}^{N_d} w_{i,j} \|m_i - (Rd_j + t)\|^2 \tag{2.32}$$

(c) Iteration of step (a) and (b) till stopping criteria holds.

Minimization computation of the error function can be calculated in closed from. First, the double sum is canceled and E becomes:

$$E(R,t) = \frac{1}{N} \sum_{i=1}^{N} \|m_i - (Rd_i + t)\|^2 \tag{2.33}$$

Then, center of mass of M, μ_m and D, μ_d are subtracted from points of each other before transformation calculation. The resulting point sets are formulated as:

$$M' = \{m_i - \mu_p\} = \{m_i'\} \tag{2.34}$$

and

$$D' = \{d_i - \mu_d\} = \{d_i'\} \tag{2.35}$$

where,

$$\mu_m = \frac{1}{N_m}\sum_{i=1}^{N_m} m_i \quad , \quad \mu_d = \frac{1}{N_d}\sum_{i=1}^{N_d} d_i \tag{2.36}$$

After these steps the error function can be rewritten as:

$$E(R,t) = \sum_{i=1}^{N} \|{m'}_i - Rd'_i\|^2. \tag{2.37}$$

A singular value decomposition (SVD) based solution is proposed in [52] for minimization of the error function. If a 3×3 correlation matrix $\mathcal{H}$ is given as

$$\mathcal{H} = \sum_{i=1}^{N} {m'_i}^T d'_i = \begin{pmatrix} S_{xx} & S_{xy} & S_{xz} \\ S_{yx} & S_{yy} & S_{yz} \\ S_{zx} & S_{zy} & S_{zz} \end{pmatrix} \tag{2.38}$$

with $S_{xx} = \sum_{i=1}^{N} {m'_{ix}}^T d'_{ix}$, $S_{xy} = \sum_{i=1}^{N} {m'_{ix}}^T d'_{iy}$, ..., then the optimal solution for $E(R,t)$ is $R = VU^T$ from the SVD.

2.3.2. Procrustes Analysis

Procrustes analysis is a form of statistical shape analysis used to analyze the distribution of a set of shapes. The name Procrustes refers to a bandit from Greek mythology who made his victims fit his bed either by stretching their limbs or cutting them off. Procrustes analysis finds the best match between two configurations, necessary rotation matrix and translation vector for the match and distances between configurations.

Suppose there are two configurations (data matrices) $X = (x_1, x_2, ..., x_n)$ and $Y = (y_1, y_2, ..., y_n)$, where x_i and y_i represent vectors in p dimensional space. An

orthogonal matrix R and a vector t are needed to be found, so that:

$$M^2 = \sum_{i=1}^{n} ||x_i - Ry_i - t||^2 \tag{2.39}$$

It can be shown that finding the translation t and rotation the matrix R can be considered separately. Translation can easily be found if each configuration is centered. If the rotation is already known, then the translation can be found. Let us denote $z_i = Ry_i + t$. Then,

$$\sum_{i=1}^{n} ||x_i - z_i||^2 = \sum_{i=1}^{n} ||x_i - \overline{x}||^2 + \sum_{i=1}^{n} ||z_i - \overline{z}||^2 + n||\overline{x} - \overline{z}||^2 \tag{2.40}$$

Since only the third term depends on the translation vector, this function is minimized when the third term is equal to 0. Thus:

$$t = \bar{x} - R\bar{y} \tag{2.41}$$

The first step in Procrustes analysis is to center matrices X and Y and calculate the mean values of the columns of them. Once mean of each column is subtracted from the related column, the remaining problem is to find the orthogonal matrix (matrix of rotation or inverse). It can be written as:

$$\begin{aligned} M^2 &= \sum_{i=1}^{n} ||x_i - Ry_i|| = tr((X - YR)(X - YR)^T) \\ M^2 &= tr(XX^T) + tr(YRR^TY) - 2tr(X^TYR) \\ M^2 &= tr(XX^T) + tr(YY^T) - 2tr(X^TYR) \end{aligned} \tag{2.42}$$

where tr is the trace operator, defined as:

$$tr(A) = a_{1,1} + a_{2,2} + ... + a_{n,n} = \sum_{i=1}^{n} a_{i,i} \tag{2.43}$$

and circular permutation of matrices is valid under trace operator. R is an orthogonal matrix:

$$RR^T = I \tag{2.44}$$

Since in the expression of M^2 only the last term is dependent on R, the problem reduces to constrained maximization of $tr(X^TYR)$. It can be computed using Lagrange multipliers technique. If a symmetric matrix of constraints is defined by $\frac{1}{2}A$, then V is needed to be maximized, where V is defined as:

$$V = tr(X^TYR - \frac{1}{2}A(RR^T - I)) \tag{2.45}$$

If the first derivatives of this expression are computed with respect to matrix R and equated to 0, then,

$$Y^TX = AR \tag{2.46}$$

R is an orthogonal matrix, therefore,

$$\begin{aligned} &Y^TX = AR \;\rightarrow\; Y^TXX^TY \\ &Y^TX = ARR^TA \;\rightarrow\; (UDV^T)(VDU^T) \\ &Y^TX = A^2 \;\rightarrow\; UD^2U^T \\ &Y^TX = A^2 \;\rightarrow\; A \\ &Y^TX = UDU^T \end{aligned} \tag{2.47}$$

and

$$Y^TX = AR \;\rightarrow\; UDV^T = UDU^TR \;\rightarrow\; UV^T = R \tag{2.48}$$

It gives the solution for the rotation (orthogonal) matrix. Then, least-squares distance

between two configurations can be calculated as follows:

$$\begin{aligned} M_0^2 &= tr(XX^T) + tr(YY^T) - 2tr(X^TYR) \\ M_0^2 &= tr(XX^T) + tr(YY^T) - 2tr(VDU^TUV^T) \\ M_0^2 &= tr(XX^T) + tr(YY^T) - 2tr(VDV^T) \\ M_0^2 &= tr(XX^T) + tr(YY^T) - 2tr(D) \end{aligned} \tag{2.49}$$

Thus the expressions for rotation matrix and the distances between configurations after matching are found. Note that to find the distance between configurations it is not necessary to rotate one of them. Lastly, following expression shows that it is even not necessary to do SVD to find distance between configurations:

$$M_0^2 = \ tr(X^TX) + tr(Y^TY) - 2tr((X^{\mathrm{T}}YY^TX)^{1/2}) \ . \tag{2.50}$$

2.3.3. Average Nose Model Generation

Average Nose Model (ANM) is a common 3D nose model, which is composed by noses in the gallery. For ANM-based registration, first an average nose model is needed to be constructed. An Average Face Model (AFM) generation method is suggested by Irfanoğlu *et al.* in [53]. In this method, the face data with minimum number of points is used as the candidate base model. Then,the fiducial points are manually localized on the training face models. 10 fiducial points are used by the proposed method. After landmark localization, an average consensus model is computed by using Procrustes analysis as described in Section 2.3.2. Then Thin-Plate Spline (TPS) algorithm is used to warp training faces on to the average consensus model. Finally the AFM is constructed by trimmed averaging of the training faces.

A modified version of the explained AFM generation method is used to generate an ANM in this thesis. First, five fiducial points on the training noses are manually localized. The selected fiducial points, namely, nose saddle left, nose saddle right, left nose peak, nose tip and right nose peak, are shown in Figure 2.38.

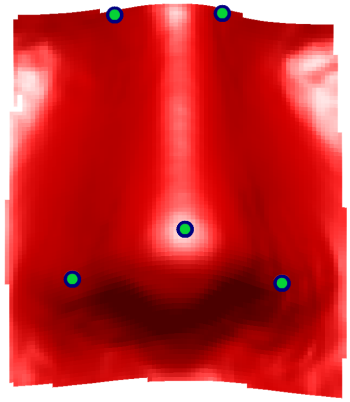


Figure 2.38. The selected fiducial points on an example ANM.

Mean values of the selected fiducial points are calculated to use as the landmarks of the ANM. Then, all training faces are aligned by using Procrustes analysis with respect to the mean values of the fiducial points. After the alignment process, a regular x-y grid is generated which comprises all of the training noses and it is placed on the given noses. The depth values corresponding to each vertex are re-sampled. As a result, all of the training noses have corresponding x and y values. Then, mean of depth values of the aligned training faces are calculated. In averaging, only points which are valid for all of the training noses, are used. Therefore, ANM is also cropped in the averaging procedure. Lastly, all training faces in gallery are translated so that the nose tip will be the origin $(0, 0, 0)$ of the nose model.

2.4. Part-Based 3D Face Recognition

Recognition step of this work is based on average of point-to-point distances of all points of a test face for all gallery faces. The similarity measure used for comparison as described above is called the Point Set Difference (PSD). Before classification, a regular x-y grid which is used in ANM generation, is placed on all of the test noses in the probe set. Additionally, the depth values corresponding to each vertex are re-sampled. In classification, if $\mathcal{N}$ and $\mathcal{N}'$ are the re-sampled depth maps of a registered training and

test nose, respectively, a mean error function, E_{PSD} can be formulated as:

$$E_{PSD} = \frac{1}{G_h \times G_h} \sum_{i=1}^{G_h} \sum_{j=1}^{G_w} w_{i,j} \times \left| \mathcal{N}_{i,j} - \mathcal{N}'_{i,j} \right| \tag{2.51}$$

where, G_h and G_w indicate height and width of the re-sampled depth maps, respectively. $w_{i,j}$ denotes the intersection of valid points in $\mathcal{N}$ and $\mathcal{N}'$. $w_{i,j}$ is zero for invalid points and one otherwise. Rank-one classification method is used on E_{PSD}. Gallery sample with the minimum error is selected as the recognized subject.

Two different types of gallery set are used for E_{PSD} computation. The first one is composed of automatically cropped and the registered nose models. The other one includes enlarged nose models. The enlarged nose models are constructed by dilation with 10 pixels of the automatically cropped nose regions' bounding boxes. The results of the recognition tests will be given in Chapter 3.

3. EXPERIMENTS AND RESULTS

3.1. Bosphorus 3D Face Database

Bosphorus 3D face database (BOS) [9, 11] is used for the experiments indicated in this chapter about the proposed algorithms. This database is recently collected by Boğaziçi University Multimedia Group (BUMM) and will be open for academic usage soon. BOS is composed of 3D facial surface point clouds and correlated 2D texture images of the subjects. It is the first database which has so many pose, expression and occlusion variations with manually localized 24 fiducial points. Manually labelled fiducial points are shown in Figure 3.1. The database consists of 3,396 facial scans of 81 subjects. There are 51 men and 30 women in the database. Majority of the subjects are Caucasian and aged between 25 and 35. The images are captured using Inspeck Mega Capturor II. The subject variation in this database incorporates various expressions,

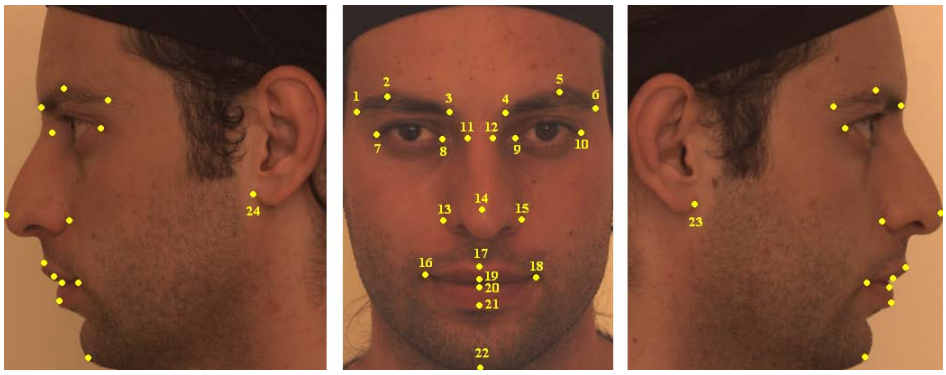


Figure 3.1. Manually labelled feature points on BOS.

poses, but also realistic occlusions like eye glasses, hair tassel and eye rubbing. The pose variations take rotations systematically into account. The facial expressions include six universal expressions (happiness, surprise, fear, sadness, anger and disgust), as well as expressions based on facial action units (AU) of the Facial Action Coding System (FACS) [54]. This group is composed of three sets: i) 20 lower face AUs (see Figure 3.2), ii) five upper face AUs (see Figure 3.3) and iii) three AU combinations (see Figure 3.4). Subject variations containing facial expressions, rotations, and occlusions are shown in

Figure 3.5, Figure 3.7 and Figure 3.6 respectively. For more realistic expressions, professional actors and actresses have been enrolled in the Bosphorus database.

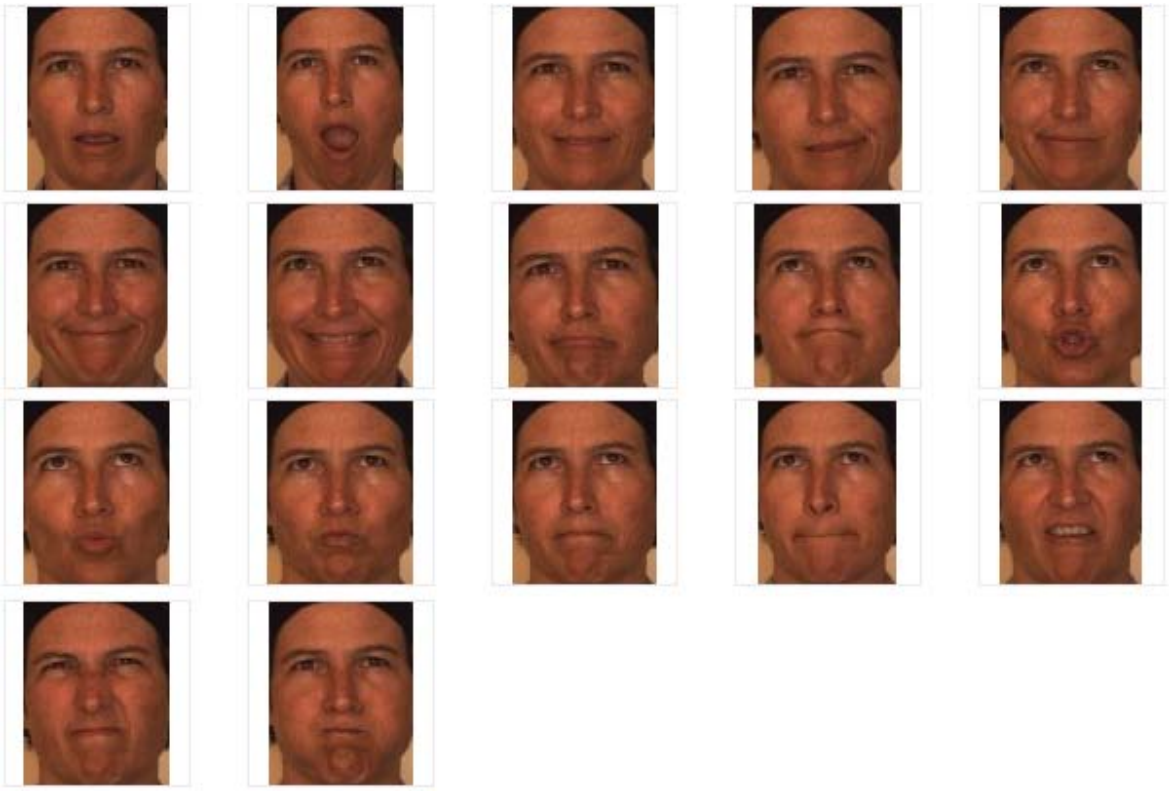


Figure 3.2. Samples from lower face action units.

Figure 3.3. Samples from upper face action units.

Figure 3.4. Some action unit combinations.

(a) (b) (c)

(d) (e) (f)

Figure 3.5. Emotional expressions: (a) Happiness , (b) surprise, (c) fear, (d) sadness, (e) angry, and (f) disgust.

Our database has two versions as v.1 (BOSv1) and v.2 (BOSv2). BOSv1 includes 34 subjects with only 10 expressions, 13 poses, four occlusions and four neutral faces, thus resulting in a total of 31 scans per subject. BOSv2 is designed for both expression understanding and face recognition schemas. In version 2, there are 47 people with 53 different face scans per subject. Each scan is purposed for covering one pose and/or one expression type, and most of the subjects have only one neutral face, though some of them have two. Totally there are 34 expressions, 13 poses, four occlusions and one/two neutral faces. 30 of these 47 subjects are professional actors/actresses.

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

Figure 3.6. Occlusions: (a) Eye glasses, (b) hair, (c) eye occlusion, (d) mouth occlusion.

Figure 3.7. Poses: (a) Neutral pose, (b) yaw rotations of -10°, (c) -20°, (d) -30°, (e) -45°, (f) -90°, (g) +45°, and (h) +90°, respectively; pitch rotations of (i) strong upwards, (j) slight upwards, (k) slight downwards, (l) strong downwards; (m) bottom-right, and (n) upper right.

3.2. Landmark Localization

Two types of curvature based heuristic nose tip estimators are proposed in this thesis. The first one is a part of the segmentation algorithm which is explained in Section 2.1 and denoted as MC1. The second method is denoted as MC2 and explained in Section 2.2. As another approach, a simple nose tip heuristic from the literature [55] is used and denoted as MD. In MD, the point with the highest depth value is selected as the tip of the nose.

Figure 3.8 and Figure 3.9 show the accuracies of with MC2 and MD on BOSv1 and BOSv2 datasets. The figures display the localization success rate against the acceptance distance as a percentage of inter-eye distance. The resulting curve is like a ROC curve: performance gets better as the threshold is increased. The results show that the nose tip is localized with good accuracy even under strong rotations and occlusions. The simple nose tip heuristic, MD cannot operate under these conditions. However, MC2 usage has a risk because of its search methodology. It estimates labelled candidate regions for the nose. Hence, if there is a wrong estimation in the nose region selection, localized nose tip is far away from the real location. That is why MC2 is not used in the overall segmentation system.

MC1 has not been tested for pure yaw rotations greater than 30 degrees because the faces with rotation angles greater than 30 degrees are discarded in our recognition experiments. The occlusion conditions also cannot be included in MC1 because this approach is sensitive to noise on the range data. In other words, curvature-based vertical and horizontal protrusion scans in MC1 are directly affected by the partial occlusions on the face. Despite of all, MC1 is more robust than MC2 as seen in Figure 3.10 and Figure 3.11. The main factor about its robustness and performance is that MC1 has a pose correction step and the estimated nose tip is updated several times along the nose segmentation process.

Table 3.1. Dataset details of landmark experiments with MC2 and MD.

Experiment	Test Type	#of Samples	
		BOSv1	BOSv2
EXP-1	Neutrals and Facial Expressions	475	1576
EXP-2	-Yaw Rotations (10,20,30,45 degrees)	135	188
EXP-3	-90 degrees Yaw Rotation	34	47
EXP-4	+45 degrees Yaw Rotation	34	47
EXP-5	+90 degrees Yaw Rotation	34	47
EXP-6	Strong and Slight Upwards	68	94
EXP-7	Strong and Slight Downwards	68	94
EXP-8	+20 degrees Pitch and +45 Yaw	34	47
EXP-9	-20 degrees Pitch and +45 Yaw	34	46
EXP-10	Occlusion of Eye with Hand	34	47
EXP-11	Occlusion of Mouth with Hand	34	47
EXP-12	Eye Glasses Occlusion	33	47
EXP-13	Hair Occlusion	23	29

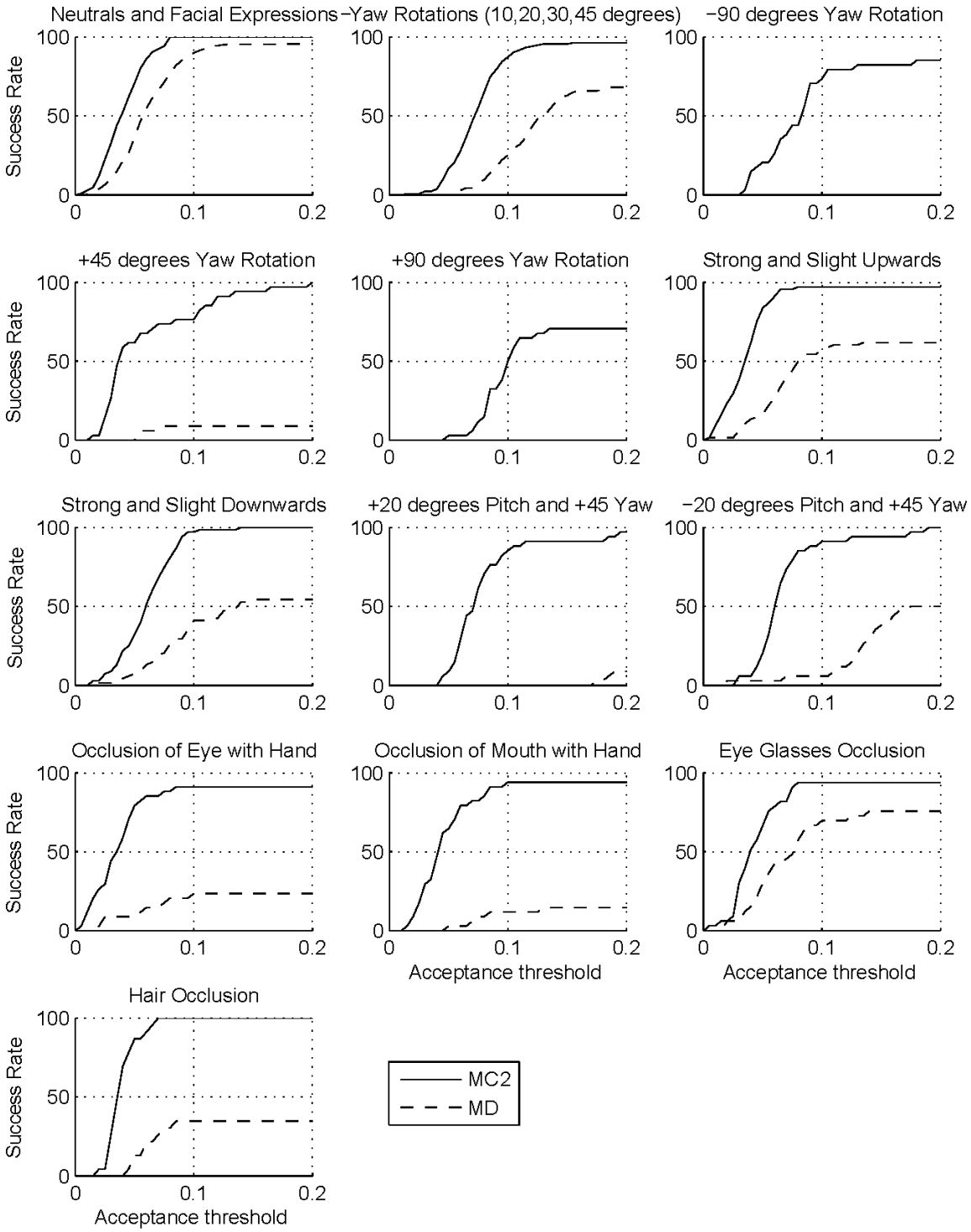


Figure 3.8. The success rate versus the acceptance distance (as a percentage of inter-eye distance): MC2 and MD on BOSv1 dataset.

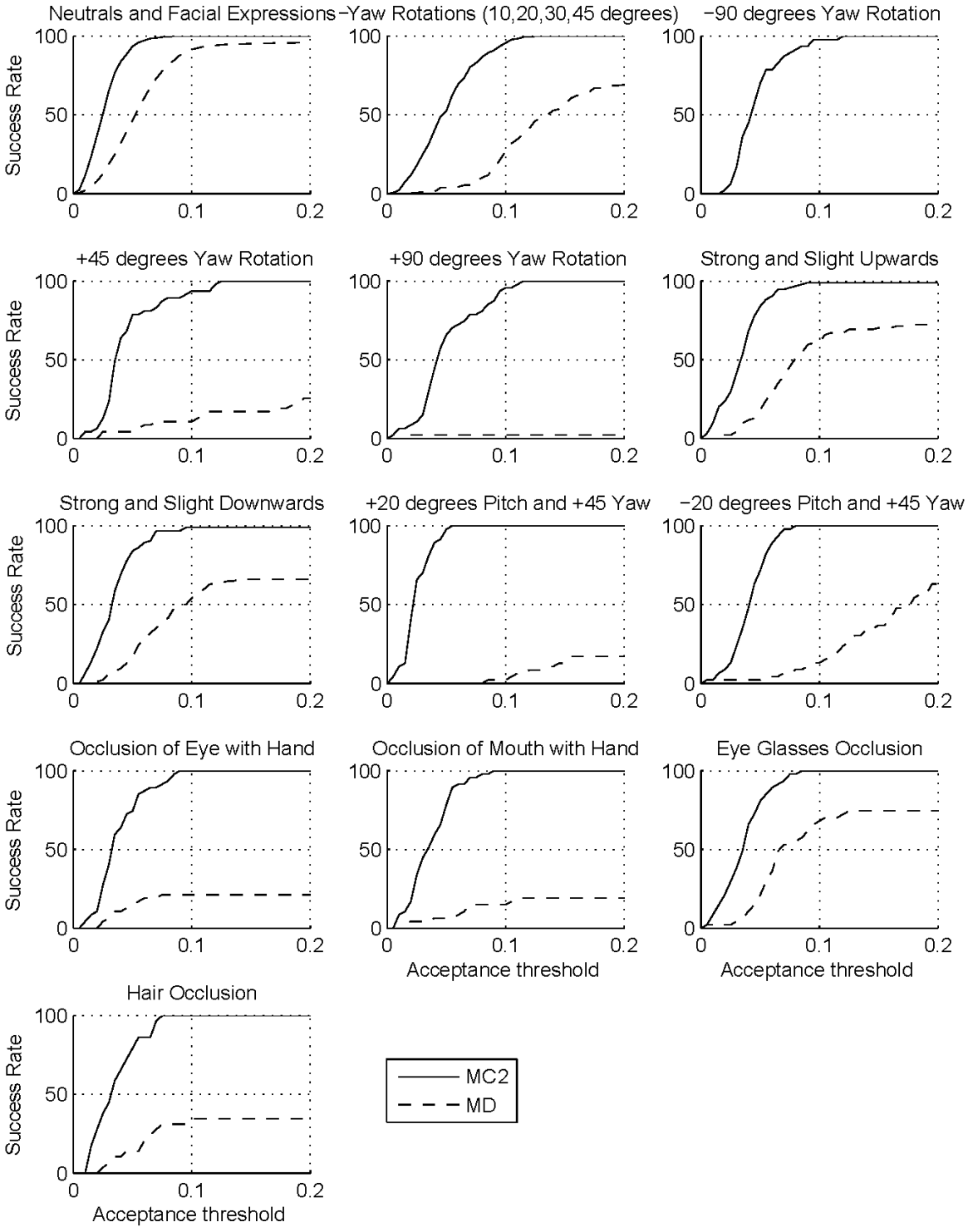


Figure 3.9. The success rate versus the acceptance distance (as a percentage of inter-eye distance): MC2 and MD on on BOSv2 dataset.

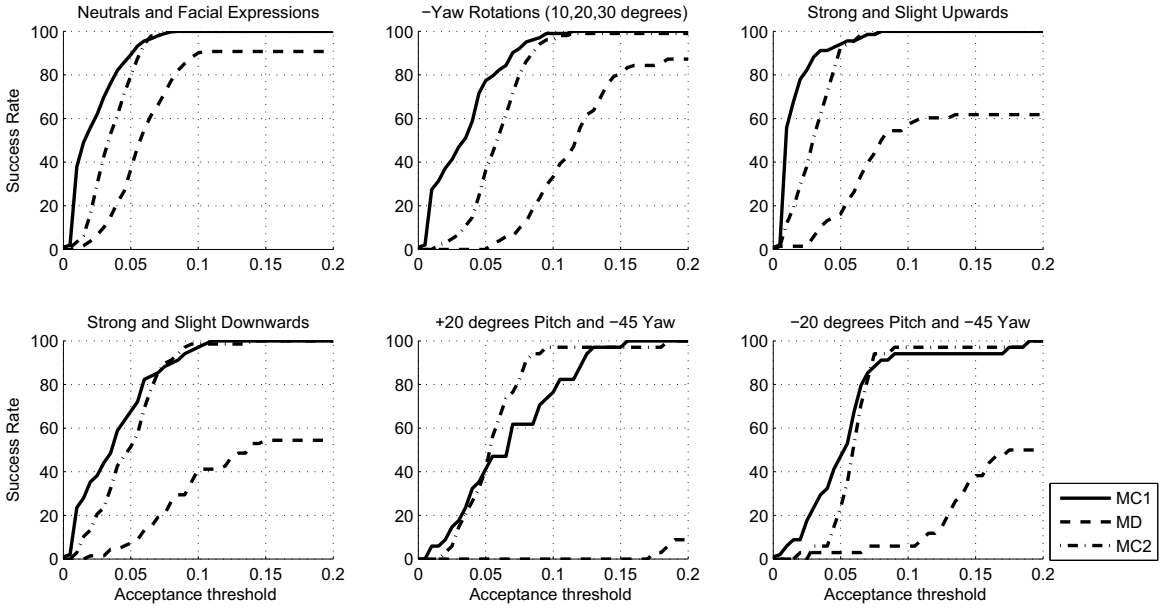


Figure 3.10. The success rate versus the acceptance distance (as a percentage of inter-eye distance): MC1, MC2, and MD on BOSv1 dataset.

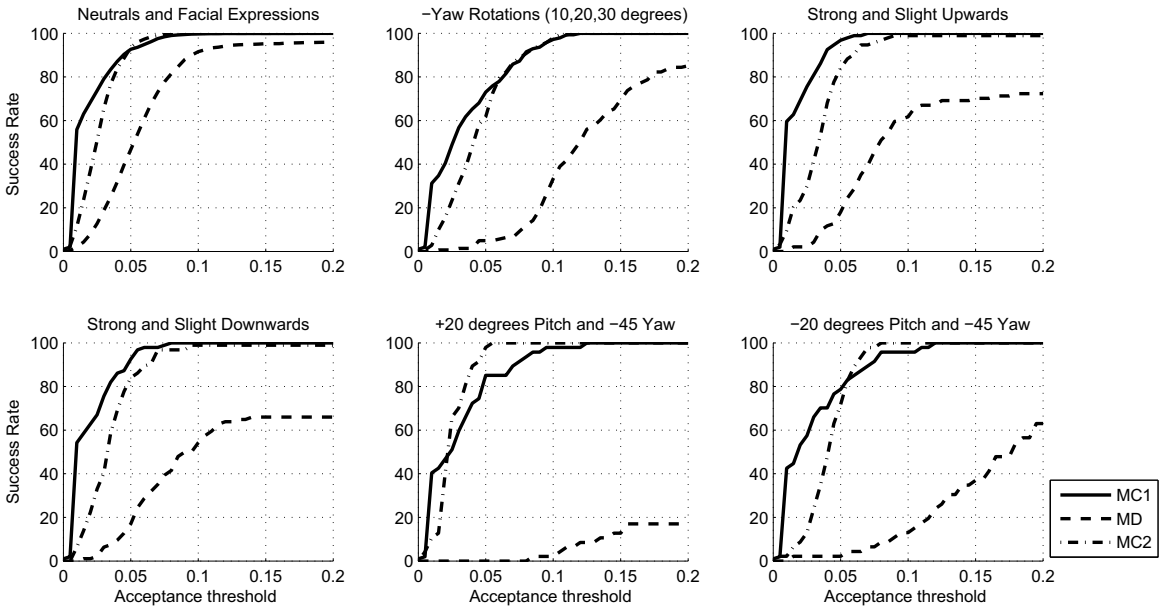


Figure 3.11. The success rate versus the acceptance distance (as a percentage of inter-eye distance): MC1, MC2, and MD on BOSv2 dataset.

3.3. Nose Segmentation

Nose segmentation performance cannot be evaluated numerically as there is no ground truth. Instead, we evaluate the success of the segmentation by the performance of the succeeding recognition step. In the following figures (Figure 3.12 to Figure 3.16), segmented nose examples are shown and overall recognition results are given in Section 3.4.

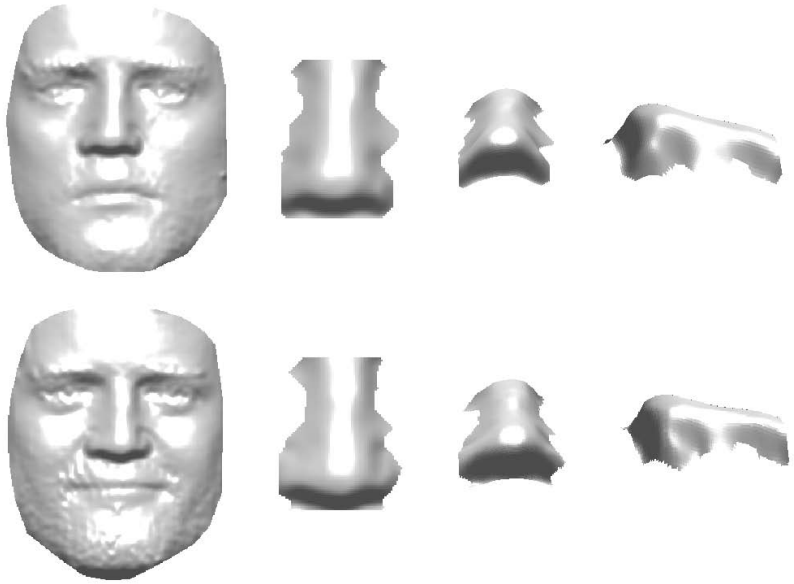


Figure 3.12. Examples of segmented nose region under frontal expressions: Whole face and different views of the segmented nose are shown, respectively.

Figure 3.13. Example of segmented nose region under -20 degrees of yaw rotation: Whole face and different views of the segmented nose are shown, respectively.

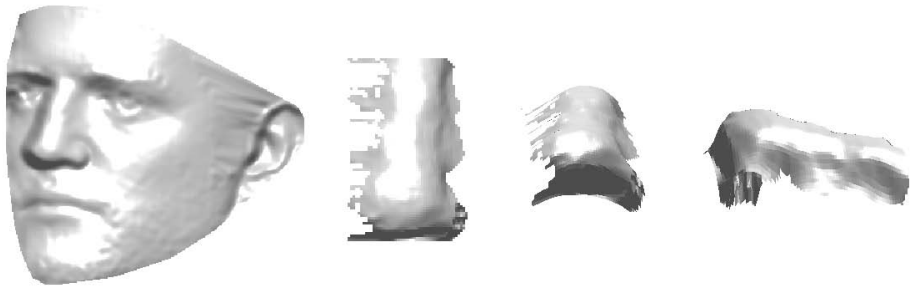


Figure 3.14. Example of segmented nose region under −30 degrees of yaw rotation: Whole face and different views of the segmented nose are shown, respectively.

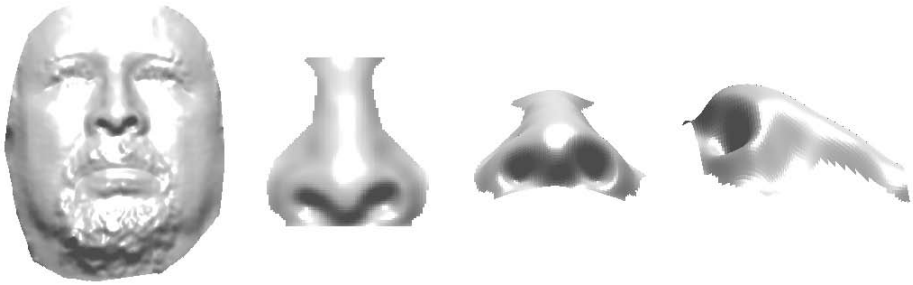


Figure 3.15. Example of segmented nose region under strong upwards rotation: Whole face and different views of the segmented nose are shown, respectively.

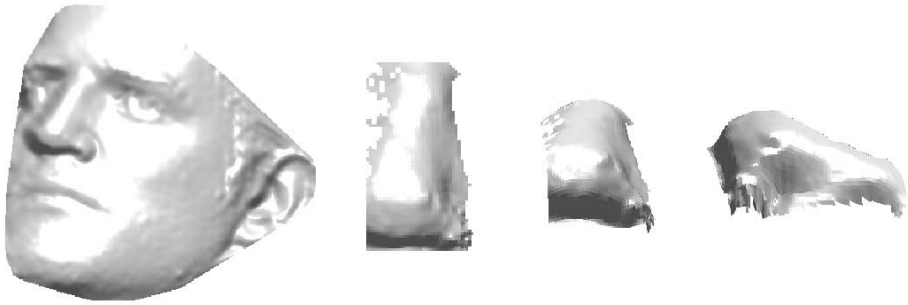


Figure 3.16. Example of segmented nose region under −45 degrees of yaw and +20 degrees of pitch rotations combination: Whole face and different views of the segmented nose are shown, respectively.

3.4. Part-Based 3D Face Recognition

Both versions of BOS are used in the experiments. For each set, the gallery is composed of one neutral scan per subject. The remaining faces are divided into two groups as frontal scans and poses with rotation. The remaining frontal scans form the frontal probe set, and the remaining poses with rotation form the rotation probe set. The poses with rotation of ±45 and ±90 degrees are discarded in the rotation probe set. Number of samples in the gallery and probe sets for each version are given in Table 3.2.

Table 3.2. The distribution of faces into gallery and probe sets for BOSv1 and BOSv2.

Version	Type	Gallery	Frontal Probe Set	Rotation Probe Set
BOSv1	Neutral	34	102	306
	Expression	-	339	-
	Total	34	441	306
BOSv2	Neutral	47	19	423
	Expression	-	1508	-
	Total	47	1527	423

3.4.1. Coarse Registration Approaches for ICP

Two different coarse registration approaches are implemented for this experiment. In the faster approach (single-pass), 22 regional landmarks of a face, labelled as 1 to 22 in Figure 3.1, are aligned to the regional landmarks of the Average Face Model via Procrustes analysis. In the second approach (two-pass), prior to registration with the nose ARM, a face is densely aligned to the AFM by ICP. This method, together with ARM-based dense registration, includes two ICP alignments. The overall registration accuracy is intended to be improved by utilizing the two-pass approach. In this experiment, only frontal samples of BOSv1 and BOSv2 are used because all of the 22 landmarks have not been manually localized for rotation variations. PSD is used as the

similarity measure for this experiment. The recognition results for the two different coarse alignment approaches using the nose ARM are given in Table 3.3. It should be noted that both coarse alignment methods use the whole facial information, either only the landmark locations or the whole facial surface. On the contrary, when densely registering a face with an ARM, only the nose region is considered. As the results exhibit, no improvement is achieved by the two-pass registration in BOSv1. The two-pass approach introduces better results in BOSv2, because of the increased expression variations.

Table 3.3. Comparison of coarse alignment approaches on the nose ARM: Recognition rate percentages.

Total Frontal Probe Set	Single-pass	Two-pass
BOSv1	88.89	89.12
BOSv2	82.45	84.15

3.4.2. Effect of the Number of Landmarks Used on Coarse Registration

To explore the effect of the number of landmarks used in coarse registration, single-pass coarse registration tests have been done with different number of landmarks on BOSv1. Only the nose region is used in these experiments. Coarse registration is done with Procrustes analysis by using related landmark points which are indicated in Table 3.4 and fine registration is processed by using nose ARM via ICP. As shown in Table 3.5 (see Figure 3.4), using all landmarks on the facial surface gives better results in coarse registration rather than using related landmarks only. Test by using only mouth corners, eyes' corners and the nose tip has the same success rate with 22 landmarks used but including chin landmark degrades the performance of registration because the mouth region is strongly affected by facial expressions.

Table 3.4. Number of landmarks used in different landmark selection experiments and their labels.

Number of Landmarks	Labels
5	11-15
7	7-10, 14, 16, 18
8	7-10, 14, 16, 18, 22
22	1-22

Table 3.5. Recognition rates for different subsets of landmarks.

BOSv1	Number of Landmarks			
	5	7	8	22
Frontal Neutral Probe Set	92.16	100.00	100.00	100.00
Frontal Expression Probe Set	81.12	85.84	81.71	85.84
Total Frontal Probe Set	83.67	89.12	85.94	89.12

3.4.3. Recognition Tests

The automatically cropped and registered nose samples are used in tests. The results are computed by using rank-one recognition as indicated in Section 2.4. To show the accuracy of the method we proposed, we compute the registration by not only one-to-all ICP but also ANM-based one-to-one ICP. One-to-all method is called RN1 and the ANM-based method is shown by RN2. Part-based recognition success rates on the total frontal probe set and the rotated probe set, are given in Table 3.6 and Table 3.7, respectively. Two different types of gallery sets are used in frontal tests of one-to-all methods as indicated in Section 2.4. The first one is composed of automatically cropped and registered nose models, denoted as *Original*. The other one includes enlarged nose models which are constructed by dilation with 10 pixels of the

automatically cropped nose regions' bounding boxes. Gallery of enlarged nose models is denoted as *Enlarged*.

The method which is suggested in [22] is also implemented and tested for two different settings as baseline results. In this method, the nose is cropped by a sphere with a radius r. The center of the sphere is defined by offsets with respect to the tip of the nose, namely x-axis offset (x_o) and y-axis offset (y_o). Nose cropping parameters have been selected as the values used in [22] which provide the best results in terms of rank-one recognition rate and verification with the true accept rate at 0.1 per cent false acceptance rate in FRGC v2. These tests are denoted as RF1 and RF2, respectively. The related parameters are set as $x_o = 0$,$y_o = +30$, $r = 40$ millimeters for RF1 and $x_o = 0$,$y_o = +20$ and $r = 35$ millimeters for RF2. Nose tip coordinates which are used in this method, are provided by MC2. One-to-all ICP is used for the registration of the baseline method.

Table 3.6. The part-based recognition success rates for the frontal poses with expression in BOS (* denotes the highest recognition rates).

Method	Gallery Type	BOSv1	BOSv2
RN1	Enlarged	92.97	88.41
	Original	94.10*	89.19*
RN2	Original	93.88	89.19*
RF1	Enlarged	73.71	65.62
	Original	89.57	80.48
RF2	Enlarged	85.94	78.26
	Original	91.84	81.79

Test results indicate that proposed methods, RN1 and RN2 with *Original* gallery type, provide the highest recognition rates for the frontal expressions in BOSv1, 94.10 per cent and in BOSv2, 89.19 per cent. The nose region which is segmented by using its natural boundaries is more discriminative than the point set of the manually cropped

Table 3.7. The part-based recognition success rates for the poses with rotation in BOS (* denotes the highest recognition rates).

Method	Gallery Type	BOSv1	BOSv2
RN1	Original	79.41*	60.99
RN2	Original	73.20	62.55*

coarse nose region. Therefore, neither the nose ARM-based method in Section 3.4.1 nor the method which is proposed by Faltemier *et al.*, RF1 and RF2, can reach the accuracies of RN1 and RN2 which are based on our full automatic nose segmentation system. As explained above, RF1, RF2 and the nose-ARM based methods use nose models which also comprise an extra portion of the facial surface outside of the nose region. Because of that the recognition success rates of the mentioned methods are degraded. If we compare the results for *Original* and *Enlarged* types of gallery set usages, it can be seen that Enlarged gallery set usage degrades the accuracy because of the same reason.

According to the tests' results of the poses with rotations, the best rates achieved by RN1 on BOSv1 and RN2 on BOSv2 are 79.41 per cent and 62.55 per cent, respectively. The recognition rates are obviously decreased with respect to frontal poses by rotation variations. However the performance of the system is accurate enough for rotation conditions. It is needed to be noted that a very small part of a nose can be extracted under rotation variations.

Note that there is not a big gap between the recognition performances of the proposed one-to-one and one-to-all methods despite very high computation complexity of the one-to-all method. The run-time computation complexities of one-to-one ANM-based method and one-to-all method are directly proportional with N_P and, $N_P \times N_G$, respectively; where, N_G denotes the number of samples in the gallery and N_P denotes the number of samples in the probe set.

4. CONCLUSIONS

A fully automatic part-based 3D face recognition system has been proposed in this thesis. Curvature-based face representation and part-based registration methods have been used to cope with rotations and facial expressions in the developed system. Our system initially finds the tip of the nose accurately, estimates the yaw rotation of the head pose and corrects it, then crops the nose region by the natural boundaries and lastly, matches 3D nose models in order to recognize a subject. The overall system is divided into three main modules: 1) Landmark localization, pose correction and nose segmentation, 2) registration, and 3) identification. The proposed algorithms and the baseline approaches for these modules are compared. Conclusions are given under the following categories, based on the findings in this thesis:

Landmark Localization Task: The automatic landmarking systems we have proposed in Section 2.1 and in Section 2.2 are curvature based heuristic methods, abbreviated as MC1 and MC2, respectively. The proposed algorithms estimate the regions with high protrusion and select the most probable location based on curvature values. Principle normal curvatures are used since they are invariant to rotation and translation. Additionally, MC2 uses a voting algorithm to select the nose region from candidate regions.

MC1 is more robust than the second approach, MC2, despite the applicability of MC2 to occlusion conditions and strong yaw rotations such as ± 45 and ± 90 degrees. It needs to be noted that the nose segmentation algorithm relies on the nose tip location found by the MC1 algorithm. Therefore it includes pose correction and several corrections of localization all through the segmentation procedure. The results show that the nose tip is localized with good accuracy even under strong rotations and occlusions by MC2. However, MC2 usage has a risk because of its search methodology. It estimates several candidates to localize the nose tip. Hence, if there is a wrong estimation, localized nose tip is far away from the real location. That is why MC2 is not used in the overall segmentation system.

Pose Estimation and Correction Task: Estimation and correction of yaw rotations of poses is one of the major features of our system. This way, curvature based feature maps are updated and the nose region is revealed. After yaw correction of poses, a coarse estimation of pitch rotation is also done for the estimation of upper and lower limits of the nose region. Note that the pose correction provides accurate nose segmentation under pose variations and this affects the recognition performance extremely in our system.

Nose Segmentation Task: There are several nose segmentation methods suggested in the literature, however we decide and propose to crop nose region by its natural boundaries via the facial anatomical features. As seen in the results in Section 3.4, the recognition rates are obviously increased by using natural regions of the noses as opposed to enlarged nose regions. This finding can be explained by two aspects. The first one is that the natural cropping provides much more discriminative features than the pure point set of a geometrically cropped nose region. In the second point of view, nose is the most stable part of the facial surface due to facial expressions and an extra portion of the facial surface outside of the nose region degrades the overall recognition performance.

Registration Task: Efficient normalization, alignment and registration of facial components play an important role due to the similarities between faces for the face recognition task. Most of the proposed systems register gallery and probe faces at the identification phase by the one-to-all technique. However, time complexities of this type of approaches are extremely high. In order to get rid of this disadvantage, we use an Average Nose Model. Our approach registers the gallery nose models with the ANM before the fine registration step and only one registration with the probe nose is computed at the recognition step. According to our findings ANM-based one-to-one registration method provides sufficiently good recognition performance in a very fast manner.

Limitations of the System: Heuristics used in this thesis are sufficiently feasible for facial surfaces but the proposed nose segmentation system is not applicable to

higher yaw rotations more than 45 degrees and partial occlusions on faces. High raw rotation conditions can be dealt with a more sophisticated pose estimation system which provides higher accuracy. Under facial occlusions, a pre-filtering phase is needed to discard surface irregularities.

In brief, we show that high recognition performances can be gained under facial expression and rotation variations by part-based 3D face recognition systems. In this work, up to 94.10 per cent of recognition success rate on facial expressions and up to 79.41 per cent of recognition success rate on rotation poses are reported by using only the nose region. The importance of the components segmentation with their natural-boundaries is also shown in this study.

In this thesis, we have focused only on nose landmark localization and nose segmentation. Although the nose is definitely the most invariant and discriminative part of the face, other parts also play an important role in recognition. For example, human subjects attribute more importance to the eyes. Localization of other landmarks and their segmentation would definitely add more information. Their segmentation and fusion of information from different parts is outside the scope of this thesis and are subjects of further research. Additionally, scale invariant features [56] can be used to detect facial landmark points and segmentation can be carried out by using watershed segmentation [57] method.

REFERENCES

1. Hamouz, M., J. R. Tena, J. Kittler, A. Hilton and J. Illingworth, *3D assisted face recognition: A survey*, chap. 1, p. 323, Springer, 2007.

2. Scheenstra, A., A. Ruifrok and R. C. Veltkamp, "A survey of 3D face recognition methods", Vol. 3546, pp. 891–899, 2005.

3. Bowyer, K. W., K. Chang and P. Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D+2D face recognition", *Computer Vision and Image Understanding*, Vol. 101, pp. 1–15, 2006.

4. Zhao, W., R. Chellappa, P. J. Phillips and A. Rosenfeld, "Face recognition: A literature survey", *ACM Computing Surveys*, Vol. 35, No. 4, p. 399458, 2003.

5. Weyrauch, B., J. Huang, B. Heisele and V. Blanz, "Component-based face recognition with 3D morphable models", *First IEEE Workshop on Face Processing in Video.*, 2004.

6. Beumier, C. and M. Acheroy, "Automatic 3D face authentication", *Image and Vision Computing*, Vol. 18, No. 4, pp. 315–321, 2000.

7. Moreno, A. B. and Á. Sánchez, "GavabDB: A 3D face database", *2nd COST275 Workshop on Biometrics on the Internet*, 2004.

8. Messer, K., J. Matas, J. Kittler, J. Luettin and G. Maitre, "XM2VTSDB: The extended M2VTS database", *2nd Inf. Conf. on Audio and Video-based Biometric Person Authentication*, 1999.

9. Savran, A., O. Çeliktutan, A. Akyol, J. Trojanova, H. Dibeklioğlu, S. Esenlik, N. Bozkurt, C. Demirkır, E. Akagündüz, K. Çalıskan, N. Alyüz, B. Sankur, İ. Ulusoy, L. Akarun and T. Sezgin, "3D face recognition performance under adversarial

conditions", *eNTERFACE07 Workshop on Multimodal Interfaces*, 2007.

10. Phillips, P. J., P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min and W. J. Worek, "Overview of the face recognition grand challenge", *IEEE Conf. Computer Vision and Pattern Recognition*, Vol. 1, pp. 947–954, 2005.

11. Savran, A., N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur and L. Akarun, "Bosphorus database for 3d face analysis", *The First COST 2101 Workshop on Biometrics and Identity Management (BIOID)*, 2008.

12. Gordon, G., "Face recognition based on depth and curvature features", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 108–110, 1992.

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

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

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

16. Alyüz, N., B. Gökberk, H. Dibeklioğlu and L. Akarun, "Component-based registration with curvature descriptors for expression insensitive 3D face recognition", *International Conference on Automatic Face and Gesture Recognition (accepted)*, 2008.

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

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

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

20. Besl, P. and N. McKay, "A method for registration of 3-D shapes", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, pp. 239–256, 1992.

21. Faltemier, T., K. W. Bowyer and P. J. Flynn, "3D face recognition with region committee voting", *Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, pp. 318–325, 2006.

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

23. Mian, A. S., M. Bennamoun and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 11, pp. 1927–1943, 2007.

24. Wang, Y., C. Chua and Y. Ho, "Facial feature detection and face recognition from 2D and 3D images", *Pattern Recognition Letters*, Vol. 23, No. 10, pp. 1191–1202, 2002.

25. Boehnen, C. and T. Russ, "A fast multi-modal approach to facial feature detection", *WACV05*, Vol. 1, pp. 135–142, 2005.

26. Lu, X., D. Colbry and A. Jain, "Three-dimensional model based face recognition", *ICPR*, Vol. 1, pp. 362–366, 2004.

27. Lu, X. G. and A. K. Jain, *Multimodal facial feature extraction for automatic 3D face recognition*, Tech. rep., Departmnet of Computer Science, Michigan State

University, 2005.

28. Lu, X. G. and A. K. Jain, "Automatic feature extraction for multiview 3d face recognition", *7th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 585–590, 2006.

29. Cootes, T., G. Edwards and C. Taylor, "Active appearance models", *IEEE Trans. PAAMIJ*, Vol. 23, No. 6, pp. 681–685, 2001.

30. Xiao, J., S. Baker, I. Matthews and T. Kanade, "Real-time combined 2D+3D active appearance models", *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 535–542, 2004.

31. Senior, A. W., "Face and feature finding for a face recognition system", *Audio and Video-based Biometric Person Authentication*, pp. 154–159, 1999.

32. Akagündüz, E. and İ. Ulusoy, "3D face representation using scale and transform invariant features", *16th IEEE Signal Processing, Communication and Applications Conference*, 2008.

33. Salah, A. A., N. Alyüz and L. Akarun, "Registration of three-dimensional face scans with average face models", *Journal of Electronic Imaging*, Vol. 17, p. 011006, 2008.

34. Ye, Y. S. and Y. Wu, "ICA/NFL local face recognition", *Journal of Image and Graphics*, Vol. 10, No. 4, pp. 468–472, 2005.

35. Samaria, F., *Face recognition using hidden Markov models*, Ph.D. thesis, University of Cambridge, 1994.

36. Su, G. D., C. P. Zhang, R. Ding and C. Du, "MMP-PCA face recognition method", *Electronics Letters*, Vol. 38, No. 25, pp. 1654–1656, 2002.

37. Sun, X., B. Liu, and B. Y. Liu, "Face recognition based on block-PCA", *Computer Engineering and Applications*, Vol. 41, pp. 80–82, 2005.

38. Lengagne, R., J. Tarel and . Monga, "From 2D images to 3D face geometry", *2nd international Conference on Automatic Face and Gesture Recognition*, pp. 301–316, 1996.

39. Gu, C. L., B. C. Yin, Y. L. Hu and S. Q. Cheng, "Resampling based method for pixel-wise correspondence between 3D faces", *International Conference on Information Technology: Coding and Computing*, Vol. 1, pp. 614–619, 2004.

40. Gong, X. and G. Wang, "Automatic 3D face segmentation based on facial feature extraction", pp. 1154–1159, 2006.

41. Eraslan, A. H., "ILEFIS (Integrated law enforcement face identification system) project", *NIJ-OST Biometrics in Criminal Justice Conference*, 1999.

42. Eraslan, A. H., "ILEFIS (Integrated law enforcement face identification system)", *Biometrics Summit Conference*, 2000.

43. Eraslan, A. H., "3D universal face-identification technology: Knowledge-based composite-photogrammetry", *Biometrics Consortium Conference*, 2004.

44. Weyrauch, B., B. Heisele, J. Huang and V. Blanz, "Component-based face recognition with 3D morphable models", *Computer Vision and Pattern Recognition Workshop*, pp. 85–85, 2004.

45. Goldfeather, J. and V. Interrante, "A novel cubic-order algorithm for approximating principal direction vectors", *ACM Trans. Graph.*, Vol. 23, No. 1, pp. 45–63, 2004.

46. Dorai, C. and A. K. Jain, "COSMOS - A representation scheme for 3D free-form objects", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 10, pp. 1115–1130, 1997.

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

48. Hesher, C., A. Srivastava and G. Erlebacher, "A novel technique for face recognition using range imaging", *7th Int. Symposium on Signal Processing and Its Applications*, pp. 201–204, 2003.

49. Lee, Y., K. Park, J. Shim and T. Yi, "3D face recognition using statistical multiple features for the local depth information", *16th Inf. Conf. on Vision Interface*, 2003.

50. Malassiotis, S. and M. G. Strintzis, "Pose and illumination compensation for 3D face recognition", *Inf. Conf. on Image Processing*, 2004.

51. Lao, S., Y. Sumi, M. Kawade and F. Tomita, "3D template matching for pose invariant face recognition using 3D facial model built with iso-luminance line based stereo vision", *Inf. Conf. on Pattern Recognition*, Vol. 2, pp. 911–916, 2000.

52. Arun, K. S., T. S. Huang and S. D. Blostein, "Least-squares fitting of two 3-D point sets", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9, No. 5, pp. 698–700, 1987.

53. İrfanoğlu, M. O., B. Gökberk and L. Akarun, "3D shape-based face recognition using automatically registered facial surfaces", *International Conference on Pattern Recognition (ICPR'04)*, Vol. 4, pp. 183–186, 2004.

54. Ekman, P. and W. V. Friesen, *The facial action coding system: A technique for the measurement of facial movement*, Consulting Psychologists Press, 1978.

55. Hayes, C. D., Jr., L. J. Hayes and L. Y. Johnson, "System and method for creating a digitized likeness of persons", US Patent 6283858, September 2001.

56. Lowe, D. G., "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, Vol. 60, No. 2, pp. 91–110, November 2004.

57. Mangan, A. P. and R. T. Whitaker, "Partitioning 3D surface meshes using watershed segmentation", *IEEE Transactions on Visualization and Computer Graphics*, Vol. 5, No. 4, pp. 308–321, Oct-Dec 1999.