



Automated Facial Anthropometry Over 3D Face Surface Textured Meshes

Augusto Enrique Salazar Jiménez

Universidad Nacional de Colombia
Facultad de Ingeniería y Arquitectura
Departamento de Electricidad, Electrónica y Computación
Manizales, Caldas, Colombia

2014



Antropometría Facial Automatizada sobre Modelos 3D Texturizados

Augusto Enrique Salazar Jiménez

Universidad Nacional de Colombia
Facultad de Ingeniería y Arquitectura
Departamento de Electricidad, Electrónica y Computación
Manizales, Caldas, Colombia
2014

Automated Facial Anthropometry Over 3D Face Surface Textured Meshes

Augusto Enrique Salazar Jiménez

A thesis submitted for the degree of:
PhD in Engineering

Advisor:
Flavio Prieto, PhD.

Research group:
Perception and Intelligent Control

Universidad Nacional de Colombia
Facultad de Ingeniería y Arquitectura
Departamento de Electricidad, Electrónica y Computación
Manizales, Caldas, Colombia
2014

*A mi madre y a quienes
cuidan de mí desde la distancia*

Acknowledgments

First, I want to thank Professor Flavio Prieto for his mentorship through all the years, his infinite patience and trust, and the opportunity to work in the laboratory of Professor Chang Shu. To Chang I owe much of my professional growth during the last years and he was always willing to give me advice. He gave me a place in his team, where I had the fortune to meet professionals whose discussions refined the course of my ideas and work. I thank the Professors Pedro Vizcaya and Jean Pierre Charalambos, as well as Flavio and Chang, for the time they spent to review the proposal, for the detailed reading of the thesis and their attention during the thesis defense. Their comments and suggestions were fundamental in raising the quality of this work.

My acknowledgments also go to Stefanie Wuhler for all her teaching, guidance, kindness and incredible talent to find solutions; to Timo for his cooperation, friendship and for making my life more fun within the group; to Alan for his feedback and advice; and to all the people I met in Canada and Germany who in one way or another were part of my job, especially Pengcheng for his kindness and Jonathan for the discussions.

My respect and gratitude for their kindness go to Alexander Cerón, Marco Jinete and Hernán Felipe García who collaborated in various stages of this work.

Many thanks to three generations of the research group of *Percepción y Control Inteligente*, who have been witnesses and participants in my entire career as a researcher.

To my mother, my sister and my father, unconditional companions in this adventure of life: Thank you for your support and good example, and that you have never given up on me. I am very fortunate that you are part of my life.

To my Maris and Carlitos: Thanks for showing me other points of view every day, from which life looks much better. I hold you in my heart.

To my wife Anna and the baby coming, who have been my great support in the last part of this way: You injected the motivation into my life, which gave me the needed strength. My love, thanks for your patience and help.

Last but not least, I want to acknowledge the program for Scholarships - *Doctorados Nacionales de COLCIENCIAS*, the *Universidad Nacional de Colombia*, the National Research Council of Canada and the Cluster of Excellence on Multimodal Computing and Interaction, UdS, Germany, without whose funding this work would not have been possible.

Agradecimientos

Primero que todo agradezco a Flavio por el acompañamiento, su paciencia infinita y por confiar en mí para enviarme al laboratorio del profesor Chang Shu. A Chang le debo gran parte de mi crecimiento profesional de los últimos años, pues me dió un puesto en su equipo de trabajo, donde tuve la fortuna de conocer profesionales cuyas discusiones refinaron el rumbo de mis ideas y siempre tuvieron muy buena disposición para atenderme. A los profesores Pedro Vizcaya y Jean Pierre Charalambos por el tiempo que dedicaron a la revisión de la propuesta, la lectura detallada de la tesis y la atención prestada en la sustentación; sus comentarios y sugerencias fueron fundamentales para la elevar la calidad de este trabajo.

A Stefanie por todas sus enseñanzas, la guianza, su amabilidad e increíble manera de encontrar soluciones. A Timo por su colaboración, amistad y por hacer mi vida más amena dentro del grupo *Non-Rigid Shape Analysis*. A Alan por la realimentación recibida. A las demás personas que conocí en Canadá y Alemania que de una u otra manera fueron parte de mi trabajo, en especial a Pengcheng por su amabilidad y a Jonathan por su tiempo para las discusiones.

A Alexander Cerón, Marco Jinete y Hernán Felipe García quienes colaboraron en distintas fases de este trabajo, a ellos mi respeto y gratitud por su amabilidad.

A tres generaciones del grupo de Percepción y Control Inteligente de la Universidad Nacional de Colombia - Sede Manizales, quienes han sido testigos y participes de toda mi carrera como investigador.

A mi madre, mi hermana y mi padre, compañeros incondicionales en esta aventura de la vida. Gracias por el apoyo, el ejemplo y por nunca perder la esperanza en mí. Soy muy afortunado de que hagan parte de mi vida.

A mi Maris y Carlitos por siempre mostrarme otras perspectivas desde las cuales la vida se ve mucho mejor. Los llevo en mi corazón.

A mi esposa Anna y el Bebé que viene en camino, quienes han sido mi gran apoyo en la última parte del camino. La motivación que le han inyectado a mi vida, me dio la fuerza que me faltaba. Mor mio gracias por tu paciencia y tu gran ayuda.

Finalmente, al programa de Becas - Doctorados Nacionales de Colciencias, la Universidad Nacional de Colombia, el *National Research Council of Canada* y el *Cluster of Excellence on Multimodal Computing and Interaction*, sin cuyo financiamiento, éste trabajo no hubiera sido posible.

Abstract

The automation of human face measurement means facing major technical and technological challenges. The use of 3D scanning technology is widely accepted in the scientific community and it offers the possibility of developing non-invasive measurement techniques. However, the selection of the points that form the basis of the measurements is a task that still requires human intervention. This work introduces digital image processing methods for automatic localization of facial features. The first goal was to examine different ways to represent 3D shapes and to evaluate whether these could be used as representative features of facial attributes, in order to locate them automatically. Based on the above, a non-rigid registration procedure was developed to estimate dense point-to-point correspondence between two surfaces. The method is able to register 3D models of faces in the presence of facial expressions. Finally, a method that uses both shape and appearance information of the surface, was designed for automatic localization of a set of facial features that are the basis for determining anthropometric ratios, which are widely used in fields such as ergonomics, forensics, surgical planning, among others.

Keywords— 3D shape descriptors, automatic landmark prediction, non-rigid 3D registration, facial expression-invariant, face anthropometry

Resumen

La automatización de la medición del rostro humano implica afrontar grandes desafíos técnicos y tecnológicos. Una alternativa de solución que ha encontrado gran aceptación dentro de la comunidad científica, corresponde a la utilización de tecnología de digitalización 3D con lo cual ha sido posible el desarrollo de técnicas de medición no invasivas. Sin embargo, la selección de los puntos que son la base de las mediciones es una tarea que aún requiere de la intervención humana. En este trabajo se presentan métodos de procesamiento digital de imágenes para la localización automática de características faciales. Lo primero que se hizo fue estudiar diversas formas de representar la forma en 3D y cómo estas podían contribuir como características representativas de los atributos faciales con el fin de poder ubicarlos automáticamente. Con base en lo anterior, se desarrolló un método para la estimación de correspondencia densa entre dos superficies a partir de un procedimiento de registro no rígido, el cual se enfocó a modelos de rostros 3D en presencia de expresiones faciales. Por último, se plantea un método, que utiliza tanto información de la forma como de la apariencia de las superficies, para la localización automática de un conjunto de características faciales que son la base para determinar índices antropométricos ampliamente utilizados en campos tales como la ergonomía, ciencias forenses, planeación quirúrgica, entre otros.

Palabras clave— descriptores de forma 3D, estimación de puntos característicos, registro 3D no rígido, invarianza a las expresiones faciales, antropometría facial

Contents

List of Figures	xxi
List of Tables	xxiv
1 Introduction	1
2 Facial features detection techniques: an Anthropometric perspective	5
2.1 Pixel or Feature Level	6
2.2 Eigen-X	7
2.3 Deformable Models	8
2.3.1 Based on energy functions	8
2.3.2 Active Shape (ASM) and Appearance Models (AAM)	9
2.4 3D Approaches	11
2.5 Multimodal Approaches	12
2.6 Summary	13
3 3D Shape Descriptors for Facial Features Detection	15
3.1 3D Shape Descriptors	16
3.2 Discriminant analysis	18
3.3 Experimental Setup	19
3.3.1 3D Face Template	19
3.3.2 Databases	20
3.3.3 Tests	21
3.4 Experimental Results	22
3.4.1 Global Relevance	22
3.4.2 Local relevance	23
3.4.3 Tests with facial expressions	25
3.5 Conclusions	27

4	Non-rigid registration of Faces	29
4.1	Related Work	30
4.1.1	Finding Landmarks on Face Models	31
4.1.2	Correspondence Computation	33
4.1.3	Use of Blendshape Models	35
4.2	Landmark Prediction	35
4.2.1	Learning	36
4.2.2	Prediction with Belief Propagation	37
4.2.3	Restricting the search region	37
4.2.4	Classification of Vertices	38
4.2.5	Refining the Nose Landmarks	40
4.2.6	Aligning Landmark Graph to Scan	41
4.3	Registration	43
4.3.1	Affine Alignment	44
4.3.2	Expression Fitting	44
4.3.3	Shape Fitting	46
4.4	Experiments and results	48
4.4.1	Database	48
4.4.2	Landmark prediction accuracy	48
4.4.3	Registration	51
4.4.4	Comparison to 3D Morphable Model	55
4.4.5	Application	60
4.4.6	Limitations	60
4.5	Conclusions	61
5	3D Anthropometry of the Face	63
5.1	Related Work	63
5.2	Non-contact Face Anthropometry	65
5.2.1	Nose and Eyes Landmarks Detection	65
5.2.2	Mouth Landmarks Detection	66
5.3	Experimental Setup	67
5.3.1	Data Collection	69
5.3.2	Anthropometric Dimensions	70
5.3.3	Semantic Segmentation	72
5.3.4	Mouth Contour Extraction	75
5.4	Experimental Results	77
5.4.1	Automatic Landmarks Location Accuracy	77
5.4.2	Segmentation	78

5.4.3	Dimensions Magnitude	82
5.4.4	Automatic Mouth Contour Location	82
5.5	Conclusions	86
6	Conclusion and Further Work	89
A	Anthropometric Dimensions Description	91
	References	105

List of Figures

3.1	Parameters of the SPIN image.	17
3.2	Circles used to compute the Finger Print descriptor. Red and green circles correspond to the Geodesic and Euclidean circles respectively.	18
3.3	(a) Facial regions. (b) Face Template	20
3.4	Snapshots of models from the databases. DB_s (a,b). 3DImDB (c,d). <i>Human Face</i> (e,f).	21
3.5	Regions considered for each local test	21
3.6	Models from HumanFace database. Original data (a-d). Bending-invariant canonical forms (e-h).	28
4.1	Overview of the fully automatic expression-invariant face correspondence approach.	30
4.2	Face model with landmarks. Locations and landmark graph structure.	36
4.3	PCA-based clustering. Left: Landmarks on a face model. Upper Right: Initial clusters formed with all the samples. Lower Right: Final cluster after removing the samples beyond a 1.5 standard deviations from the cluster medoid. Minimum volume enclosing ellipsoids (3D and upper views).	38
4.4	Example of vertex labeling result. (A) Notice how the points on the nose tip region are correctly labeled. (B) Some vertices are assigned to two classes. This situation is because of the left-right symmetry of the features. (C) Points located far from the region of interest are discarded.	39
4.5	<i>Umbilics</i> of different 3D facial models of the same subject performing different expressions. Notice how the <i>umbilics</i> are distributed all over the surface, and in most of the cases umbilics are present at the locations of salient facial features.	40
4.6	Framework of the proposed initial alignment method.	42

4.7	Registration procedure. First, the template and the scan are aligned using the predicted landmarks. Second, the expression is fitted using a blendshape model. Finally, an energy-based surface fitting method is used to fit the shape. At the end, the overlap between the scan and the template is maximized and a point-to-point correspondence for the face shapes in different expressions is obtained.	43
4.8	Left: template rest pose A_0 and a set of blendshapes A_i . Right: examples of models generated as linear combinations of blendshapes.	45
4.9	Regions used in the expression fitting procedure.	45
4.10	Characteristics of the BU-3DFE database.	49
4.11	Examples of the landmark prediction results. Red and green spheres correspond to the manually placed and predicted landmarks, respectively. First row: female subjects; Second row: male subjects.	52
4.12	Error at landmark points not used for registration. Left: set of points. Right: summary of errors.	53
4.13	Cumulative distribution of the number of models where the error at all the landmark points not used for registration is below a threshold. Example of registration results (left and right). Error distribution (center).	54
4.14	Examples of registration results. The input, fitted expression, error mapped, and texture mapped models are provided for each example.	56
4.15	Examples of fitting to models of the same subject performing an expression in different levels. Fear (first three rows). Surprise (last three rows). For each example, first, second, and third rows are the input, output, and textured models.	57
4.16	Comparison of shape distance of 3DMM fitting and the results of the proposed method. Top to bottom: input scan, 3DMM fitting, proposed method result.	58
4.17	Distance between the surface of the template P and the surface of input model F . Histograms and the false color visualization (different views) of the magnitude of the mean and standard deviation of the distance.	59
4.18	Real models used to compute the multilinear model (shown in boxes) and synthetic models generated from the multilinear model.	61
4.19	Incorrect shape fitting. The differences in topology of the input and template meshes cause incorrect expression and shape fitting.	61
4.20	Challenging test scenario. Mapped error models correspond to the fitting result. Test was carried out over one model of the Bosphorus database.	62
5.1	Overview of the fully automatic face anthropometry approach	65
5.2	Framework of the proposed multimodal landmark detection method.	68
5.3	Set of landmarks	69

5.4	Example of model from the DB_H database. (a) Textured and (b) Geometry. (c) Points (blue) used to compute the absolute (red) position of the landmarks. (d) and (e) show the location of the anatomical (blue) and physical (red) landmarks computed from the manually annotated points.	70
5.5	Plaster model. (a) Picture and (b) 3D Geometry.	71
5.6	Set of dimensions	71
5.7	Face template used for the semantic segmentation of the face.	73
5.8	Curves used to define the external contour of the lips.	76
5.9	Parametric functions used to define a external lips contour template.	76
5.10	Examples of the landmark detection results. Red and green spheres correspond to the manually placed and predicted landmarks, respectively. (a)-(c): female subjects; (d)-(f): male subjects.	78
5.11	Evaluation of the segmentation for the tests T_1 to T_4	79
5.12	Results of the semantic segmentation.	80
5.13	Evaluation of the segmentation for each region individually.	81
5.14	Mouth contour detection results. Red: Ground truth. Green: Detected.	85
5.15	Lips contour location error. Mean (left) and Standard deviation (right).	85

List of Tables

3.1	Fisher's coefficients for the Global test. R indicates the ranking.	22
3.2	Fisher's coefficients for the Eyes test.	23
3.3	Fisher's coefficients for the Nose test.	24
3.4	Fisher's coefficients for the Cheeks test.	24
3.5	Fisher's coefficients for the Mouth test.	25
3.6	Fisher's coefficients for the Chin test.	26
3.7	Fisher's coefficients for the Local test over <i>Human Face</i> neutral group. Gray columns correspond to the coefficients of the left side of the face.	26
3.8	Fisher's coefficients for the Local test over <i>Human Face</i> smile open mouth group. Gray columns correspond to the coefficients of the left side.	27
4.1	Error of landmark prediction with training set T_n . $T < 10$, $T < 20$, and $T < 30$ correspond to the detection rates with a tolerance of 10mm, 20mm and 30mm, respectively.	50
4.2	Error of landmark prediction with training set T_e . $T < 10$, $T < 20$, and $T < 30$ correspond to the detection rates with a tolerance of 10mm, 20mm and 30mm, respectively.	50
4.3	Comparison of mean errors of the proposed method and two different approaches.	51
5.1	Discrepancy between real and 3D measurements.	72
5.2	Error of landmark detection used the multimodal approach. $T < 5$ corresponds to the detection rates with a tolerance of 5mm.	77
5.3	Set of evaluated regions.	82
5.4	Error at landmarks points not used for the initial alignment. Gray column corresponds to the landmarks located at the left side of the face and the other column to the landmarks located at the right. N.A. means that the point is located on the center line of the face.	83

5.5	Error of the magnitude of the dimensions.	84
5.6	Error of the surface measurements of the mouth contour.	86

Introduction

Anthropometry is the science and practice of human body measurement. Its main aim is the characterization and description of human body morphology's variation [1]. Face anthropometry is one of the most widely studied branches of anthropometry, which provides objective ways to assess the morphology and to detect changes derived from aging. In particular, face anthropometry is used as a base to diagnose acquired and/or genetic malformations [2, 3], to plan and evaluate surgeries [4], to study normal and abnormal growth [5, 6], or to determine results in different stages of treatments [7, 8], among others.

In order to obtain reliable measurements, detailed knowledge of the accurate location of landmarks on the head and face surface is crucial. Landmarks are classified as *osseous* when they are located on the surface of the underlying bone and *soft* if they are on the skin surface. Even on a normal face, accurate identification of landmarks requires some experience. Traditional anthropometry is done manually with the help of a set of devices such as calipers and measure tapes. This has sociological, logistical and technical disadvantages, such as time-consuming measurement procedures; dependence on the researcher's abilities to produce consistent and accurate measurements; as well as unwanted physical contact between subject and researcher. Thus, there are several limitations when there is a need of collecting information of a wide set of information from an extensive group of subjects [9].

Advances in imaging technology allowed the emergence of no-contact anthropometry. The new acquisition methods allow to overcome many drawbacks of traditional anthropometry, e.g., with 3D laser technology it is possible to acquire an entire database of surface models of human body in a relatively short time [10]. Different researchers have analyzed and demonstrated the reliability of measures taken on 3D models [9, 11]. Despite the advantages of non-invasive anthropometry, the task of selecting the points that are the base to compute

the anthropometric ratios, still needs the human intervention.

In order to be processed, a face can be modeled as a simple 2D pattern, a parametrized vector, or as a complex set of 3D points with polygonal meshes or parameters for each degree of freedom or variation [12]. For each of these representations or combinations of them, many techniques have been developed. One of the fields that received most attention is the face identification, but also face modeling, synthesis, and identification of expressions have been of great interest. However, the performance of several approaches relies in the proper location of facial landmarks. Therefore, the development of techniques that minimize the human intervention in the task of facial feature location will positively affect the range of applications in the medical and the computer graphics communities.

In this thesis the problem of the automatic locations of facial features such as landmarks, regions, and contours, is addressed. The first stage consists in the selection of a proper descriptor or set of descriptors that is discriminative enough to identify the facial features. Next, a non-rigid registration approach is used to register and input scan with a 3D face template, which carries information of face anatomy, allowing the segmentation of the face in regions with a semantic meaning. Finally, the shape and texture information are used to derive a method for automatic face anthropometry. The methods proposed in this work are robust to the non-linear local variation due to facial expression. Also, they allow the detection of a considerable set of landmarks, anatomical regions, and the mouth contour. In addition, all these procedures are performed in a fully automatic way.

Next, the organization of the document and the publications derived from the findings of this thesis are listed. In particular,

Chapter 2 Presents an analysis of several facial features detection techniques as if they will be used for automatic face anthropometry. The discussed techniques belong to the most used methods for facial features detection.

Chapter 3 Exhibits a study of the behavior of a 3D shape descriptors set computed on the surface of 3D face models. The relevance of the descriptors was determined using the Fisher's coefficient. Two different tests were performed in order to determine the global and local relevance of the set of descriptors.

The findings of this study were presented at the International Symposium on Visual Computing [13] and accepted to International Journal of Signal And Imaging Systems Engineering [14].

Chapter 4 Introduces a non-rigid registration method to compute point-to-point correspondences among a set of human face scans in a fully automatic way. The complete automation of the procedure is accomplished with the inclusion of a landmark prediction method that learns local properties and spatial relationships between the landmarks and performs statistical inference over the trained model. The method showed to be robust in presence of several kind of facial expressions. A consistent correspondence was found for most of the tested models.

This work was accepted to Machine Vision and Applications [15] and used to register the database that was used in the work accepted to Computer Vision and Image Understanding [16].

Chapter 5 Presents a method that uses both shape and appearance information of surfaces for automatic location of a set of facial features. The method is evaluated as a tool for automatic face anthropometry. The potential of locating osseous and soft landmarks is assessed. Also, the ability of the method to segment the model into several semantic regions is tested. Finally, an evaluation of the consistency of the extracted 3D mouth contour is presented.

At the end of each chapter, the conclusions are stated. Chapter 6 summarizes the main conclusion of this thesis and gives ideas for future work.

Facial features detection techniques: an Anthropometric perspective

The issue of automatic detection of facial features has been widely discussed in the image processing and pattern recognition literature. This field has several practical applications such as identification, location and tracking of people, expressions recognition, 3D pose estimation, codification and image reconstruction, and recently, non-invasive anthropometry.

Anthropometry allows establishing the spatial correspondence between relative points in human body structures and their geometric variation of their relative location; this characteristic serves as a base to derive anthropometric ratios. As an example, a nasal index can be defined as a ratio between nasal wide and height. Farkas [17] carried out a thorough review of basic anthropometric craniofacial ratios that serve as a reference in innumerable studies.

Regarding to automatic feature detection, the use of a serial search methodology is popular in the literature; that is, initially a coarse detection of features is carried out and then another search over the detected region is done in order to refine the feature location. The main obstacles to overcome during feature detection include: the variability of subject's appearance, pose, facial expressions, presence or absence of structural components (beard, moustache, glasses, etc.) and the lighting conditions. In general, the criteria of technical quality of a technique are related to the ability to overcome the mentioned obstacles. However, for anthropometry, the main criterion is the accuracy in locating: points, contours, and regions, which are useful to determine the anthropometric ratios.

The aim of this chapter is to analyze different characteristics of several facial features detection techniques as if they will be used for automatic face anthropometry. The discussed

techniques belong to the most used methods for facial features detection. For each technique a small review about its origin, its mathematical foundation, its pros and cons, and a description of several works that uses such techniques for facial features detection, is presented. In addition, a discussion of the utility of the techniques for anthropometric applications is stated.

This chapter is divided as follows: Pixel or feature level techniques are described in Section 2.1. Section 2.2 reviews the Eigen-X approaches. The deformable models are reported in Section 2.3. Section 2.4 depicts the approaches that are mainly based in the analysis of 3D data. Section 2.5 gives some ideas about the approaches that use hybrid 2D-3D information as input of the facial feature detection procedure. Finally, a summary of the chapter is presented in Section 2.6.

2.1 Pixel or Feature Level

Pixel or feature level was one of the first techniques used to detect facial features. In general, the technique is based on a geometric ratio such as position or width of different attributes. Usually the attributes are taken from integral projections (*IP*) of the original image. *IP* results from defining a vertical and a horizontal *IP* starting from an image $I(x, y)$, in a rectangle $[x_1, x_2] \times [y_1, y_2]$, which can be defined as follows:

$$v(x) = \sum_{y=y_1}^{y_2} I(x, y), \quad h(y) = \sum_{x=x_1}^{x_2} I(x, y).$$

The horizontal *IP* can be used to extract the face left and right boundaries as well as the nose; the vertical *IP* can be used to extract the eyes, mouth and the nose base. Peaks and valleys of *IP* are analyzed against a threshold to detect and extract the attribute's position [18]. Kanade satisfactorily used this technique, in [19], in his pioneer work on face recognition [19]. The technique can also be used in colored images [20]. In general, methods based on color information are more resistant to lighting variations; a description and a list of techniques for skin modeling can be found in [20].

Facial features detected by this technique in an initial stage are face [21, 22], eyes, nose and mouth [18, 23, 24], eyebrows, chin and ears [25, 26]; in a fine detection stage, this technique detects pupil center, nostrils, eye and mouth corners, [21, 26]. By adjusting functions on the binary image, shapes of eyes, nose, mouth and chin are detected [27]. This technique detects a small number of points so the measurements that can be obtained are limited. In addition, since the framework is a 2D field, the dimensions are linear or flat with possible

correction factors to determine the length of the face [23, 26].

The assessment of the measurements obtained with this technique are generally done by visual inspection [23, 26], or expressed as a percentage of the size of the analyzed region [21, 22]. Nevertheless, some works report results in millimeters or in pixels with up to 2mm accuracy [28].

This technique is sensitive to lighting change, uneven background and head pose [18, 28, 25]; in addition, the presence of accessories and facial hair reduces the performance significantly; to deal with these limitations, heuristics are introduced with restrictions subject to the face morphology [26]. The process is clearly affected by the image size, as every pixel has to be evaluated. Processing time has been considerably reduced; the first system used to take up to 10 minutes to process an image [19], currently, attributes can be detected in real time [26, 22].

2.2 Eigen-X

Among the techniques based on appearance, the Principal Component Analysis (PCA) is widely used to extract facial features. Kirby and Sirovich state that any human face -either partial or complete- can be described in an optimal coordinate system [29]. In this approach, an image is transformed into a small set of attributes called eigenfaces. Eigenfaces are normalized eigenvectors from the covariance matrix of the training set [30]. The corresponding eigenvalues $\{\lambda_i\}_{i=1}^M$ are directly correlated to the variability ranges of the projections $\{y_i\}_{i=1}^M$.

Eigenfaces approach assumes that an image x out of a $\{x_i\}_{i=1}^{N_t}$ training dataset can be approached by a linear combination of few $\{f_i\}_{i=1}^N$ eigenfaces, that is:

$$x \approx \bar{x} + \sum_{i=1}^M y_i f_i,$$

where, \bar{x} is the mean value of the training dataset and $\{y_i\}_{i=1}^M$ contains the projection of the normalized facial image on the first eigenfaces. The eigenfaces concept can be extended to eigenfeatures [31], such as eigeneye, eigenmouth [32] and eigennose. Other approach is the Eigen-harmonics Faces [33], with which face recognition can be done at different lighting conditions.

These methods are special at coarse face, eyes and mouth detection [34, 35, 36, 37], as

well as eyebrows and nose detection [36]. They can also detect specific points such as the eye-center, nostrils, nose tip and mouth corner [34, 38], and facial regions [39]. As the regions are defined by lines, such divisions are not enough to make an adequate anatomical characterization of faces. Besides, since input images are normally 2D, measurements are linear [35, 27, 36]. Nevertheless, it is possible to carry out a greater number of measurements [39]. In addition, the training is demanding since the points are labeled manually.

The evaluation of the detected features is carried out objectively [34, 35, 40, 38] only in some cases visual inspection is employed [27]. Nevertheless, reported values are in pixels without the equivalence in millimeters. This technique allows an accuracy of up to 1 pixel in images where most area is occupied by the face.

As databases used in the training that include acquisition variables, the algorithms are robust to different face orientations, lighting and scale [39, 34] and facial expressions [35]. Nevertheless, these methods are sensitive to background variations [36]. In addition, heuristics are used as a tool to increase the algorithm performance. Processing time is not an issue for these approaches as they limit themselves to evaluation based on a classifier or to the comparison with a template, which can be done in milliseconds [27]. However, algorithm training takes long due to the manual labeling of the points.

2.3 Deformable Models

Deformable models have been widely used in the last decades in the image interpretation field, especially in those that have highly variable structures such as faces. Deformable models can be classified in different ways; here they will be classified based on energy functions, Active Shape models and Active Appearance models.

2.3.1 Based on energy functions

These models are energy minimization curves, which are constrained by internal forces of continuity and guided to features by external forces. In many areas they are used as tools for edge detection, motion tracking, stereo matching and more generally, to solve problems which require the fitting of a model to an object in an image by minimizing the energy [41]. The model known as snakes (active contours), was introduced in [42, 43] and unlike the models presented in [44, 45, 46, 47], the snakes were adopted due to the unified treatment of the optimization process, allowing to take advantage of standard numerical techniques for the treatment of partial derivatives equations [48]. From the physical point of view, a snake is a

set of control points, called snaxels, which are connected to each other. Each snaxel has an associated energy that rises or falls depending on the forces acting on it.

Let $\mathcal{C}(p) : [0, 1] \rightarrow \mathbb{R}^2$ a parametric curve, and $\mathbf{I} : [0, a] \times [0, b] \rightarrow \mathbb{R}^+$ an image where object boundaries need to be detected. It is possible to associate in the curve \mathcal{C} an energy given by [42]:

$$E(\mathcal{C}) = \alpha \int_0^1 |\mathcal{C}'(\tau)|^2 d\tau + \beta \int_0^1 |\mathcal{C}''(\tau)|^2 d\tau - \lambda \int_0^1 |\nabla \mathbf{I}(\mathcal{C}(\tau))| d\tau, \quad (2.1)$$

where, α , β and λ are positive constants (α and β , determine the elasticity and stiffness of the curve). The first two terms control the smoothness of the contour to be detected (internal energy) and the third term is the responsible of the attraction of the contour towards the object (external power).

These models had been used to the coarse detection of face, eyebrows, eyes, nose, mouth, and in some cases the ears [49, 50, 51]. In fine detection, the developments have been focused on the detection of the eyebrow and lip shape [52, 53]; the iris, the nose; chin and face edge [49, 54]; and even the cheek and ear shape [55, 51]. The input information corresponds to 2D [54, 53] and/or 3D images [52, 49, 50, 51]. Therefore, other linear, circular and face measurements can be obtained.

In general, feature detection by these techniques is evaluated by visual inspection or by detection percentage [49, 54]. This is due to the fact that algorithms can converge freely or restrictively by prior knowledge of the morphology of the target object. Therefore a manual labeling of the coordinates of convergence is not required. Besides, since the edge is made of many points, the quantitative evaluation is a demanding task. Some works reported results measured in millimeters [53] that serve as a base to infer that features can be detected with an accuracy level of up to 0.8 mm.

This technique is sensitive to initialization, noise, low lighting conditions, among others. The method is also sensitive to head pose, especially when there are self-occlusions or a wide perspective angle view [51, 53]. Deformable models are robust against significant appearance and shape variations resulting from facial expressions [50].

2.3.2 Active Shape (ASM) and Appearance Models (AAM)

This statistical approach is used in shape modeling and features extraction. These models represent a target structure by means of a statistical model of the shape obtained from the

training. In this sense, these models are very related with the Eigen-X approaches. However, due to the fitting procedure of either an ASM or AMM includes different energy terms, it was decided to classify them into a different category. Cootes et al. [56, 57] introduced this method and along of the years it has been improved by other authors. In the original version, the initial set of points is obtained from the average shape, which is derived from the training information and its accuracy depends on the amount of variability included in the training set. Furthermore, local structure of points is represented by changes in the intensity values of pixels along a profile line that passes through the points. This based on the assumption that usually the facial features are located on the strong edges [58].

In the ASM technique, the position of n points named as landmarks is selected by an expert, over a set of training images. This set of points is represented by a vector $\mathbf{X} = (x_1, y_1, \dots, x_n, y_n)^T$, where x_i and y_i are the coordinates of the i – th landmark. By analysis of shape variations in the training vector, a model to represent such variations is build:

$$\mathbf{X} \approx \bar{\mathbf{X}} + \mathbf{P}\mathbf{b}. \quad (2.2)$$

The vector $\bar{\mathbf{X}}$ includes the average value of the coordinates of the n points. \mathbf{P} is a matrix with the first t eigenvectors of the covariance matrix, and \mathbf{b} is a vector to define the model parameters. The variance of the i – th parameter, P_i , along the data set is given by the corresponding eigenvalue λ_i .

In this approach, initialization is important. When the initialization is poor, the searching process can fail or can be very slow. Therefore, a good initialization can help to find the optimal solution in less iterations. For this purpose, the AAM, which includes the appearance of the target pattern into the model, has shown to be less sensitive to poor initialization, improving the overall quality of the fitting.

These models have been used for coarse detection of facial features such as eyes, nose and mouth by analyzing gradient information [59]; or the estimation of the head pose by using genetic algorithms [60]. The model introduced in [56] or its extensions are generally used for fine detection of features. Features detected include face edge, eyebrows, eyes, nose, and mouth [58, 61, 62, 63, 64].

In some works accuracy evaluation is done visually [58], but it is general done objectively [58, 61, 62, 63]. Here, contrary to the Eigen-X approaches, it is necessary to label a more extensive set of points. Therefore, training is more demanding than in the previously mentioned techniques. Facial features can be detected with an accuracy level of up to 1.9

pixels [62], less than for Eigen-X.

Robustness in these techniques as in Eigen-X approaches depend on the properties of the training database. The greatest variations are found in the head pose and the non-homogenous background [58, 61, 62]. Training different models for different poses reduces these variations but increases the number of images that have to be labeled manually. Model initialization is very important, when it is poor, the search process may fail or be too slow [64].

2.4 3D Approaches

In spite of the amount of effort to develop robust recognition systems, there are still problems with respect to providing the systems with invariance to illumination, pose and facial expressions. This can be due, largely, to the limited input information. The system of human senses provides a wide information set, which has been attempted to be adapted to imaging work either grayscale or color ones. It is suspected that problems in limited information is due to the drawback in face characterization, which involves the analysis of face shape and curvature. 3D sensors provide information of such resolution and accuracy which allow (with an appropriate noise treatment) to make accurate calculations of face curvature [65].

Common approaches of this kind of techniques are based on the analysis of the sign of Mean and Gaussian Curvatures, and other descriptors that are a variation of them (see Chapter 3 for a more detailed description). Moreover, curvature as a property of the local surface has the quality of being point-of-view invariant. In [66] five methods of curvature estimation are evaluated and classified in Analytic and Discrete estimation. Analytic estimation first adjusts a local surface around a point and uses the parameters of the surface equation to determine the curvature value. Instead of adjusting the surface, the discrete approaches estimate numerically either the curvature or the derivatives [51].

Most systems based on these techniques have been oriented to face identification [67, 68]. Nevertheless, their use in anthropometry has gained relevance, as in addition to the linear measurements, it is possible to render detailed characterization of surface morphology. In coarse detection, nose, eyes and mouth are detected in [69, 70, 71]. Since the segmented region may contain the feature totally or partially, the detected regions boundaries are abstract; nevertheless, it also contains regions that do not belong there, therefore detection quality cannot be determined easily. Fine detection is focused on eye corners and nose tip [72, 73, 74, 75]. Nevertheless, it can also detect extreme points such as nose base, low chin and mouth corners [76, 71], face mid-line [69], and different anthropometric landmarks [77, 78]. The use of

3D information allows linear, angular, circumferences and surface length dimensions, which have a great potential for being used in face anthropometry.

There is balance in the evaluation methodologies used; some works are evaluated by visual inspection [76, 69], others by detection percentage [74] and others quantify the evaluation in millimeters or even in degrees [72, 77, 78, 62].

Systems based on this technique are robust to lighting and pose variations as the curvature, a property to the local surface, is invariant to perspective, lighting or color [78]. Nevertheless, this technique is weak against facial hair, artifacts and self-occlusion [72]. Feature detection can be carried out in times between 0.33 and 20 seconds [76, 69, 71].

2.5 Multimodal Approaches

This category corresponds to the approaches that take advantage of using several sources of information. These techniques are reference in the literature as multimodal approaches. The key idea to combine methods that complement each other in order to complete a task that could not be done using each method individually.

In the case of the multimodal approaches, the coarse detection is focused on the eyes, nose and mouth regions; furthermore, it is possible to detect the eyebrows, ears and face contour as well. It is also possible to split the face into several patches [79] but separation boundaries are rectangular and do not have a semantic meaning. On the other hand, fine detection of the eyes and mouth corners, nose tip, nostrils, eyebrows and chin, also can be performed.

Since 2D and/or 3D information is used in these approaches, all kind of measurements can be obtained. Although the amount of detected points is low, some works have combined deformable models with stereo vision techniques, which enable obtaining a large amount of points and extracting 3D contours.

Thanks to the fact that several methods and different type of information are combined, computational time is lower than for those approaches that only depth or 3D information. Therefore, it is good to take into account that including different kinds of information and different algorithms for the different features detection, will contribute to an efficient work.

2.6 Summary

Studying human face morphology has generated a set of needs aimed to achieve a more detailed structure description; image-processing techniques satisfied many of these needs. Nevertheless, developing systems invariant to conditions such as lighting, pose, facial expressions, presence of artifacts (beard, glasses, jewelry, etc.) is still complex. Due to that, and in order to obtain adequate results, work conditions are generally restricted. Among the herein referenced techniques, the ones using both 3D and texture information show great potential to be included as part of a robust facial feature detection system.

Based on the above, the type of measurements is generally limited to linear and angular distances leaving out superficial measurements. Evidently, the use of 3D imaging allows results that are more accurate in addition to obtaining descriptions of the surface that are more detailed. Nevertheless, currently developed systems still do not show reliable results, therefore, there is still a lot to do in order to increase the performance in the analysis of facial images and 3D models.

Bearing in mind current developments, an efficient anthropometry system should use techniques that complement each other. For instance, in order to reduce the search area, an algorithm that carries out an initial search on the texture of a 3D, could be used. Afterwards, to achieve fine detection, an algorithm that use surface information could be very effective.

3D Shape Descriptors for Facial Features Detection

As the aim of this work is to develop a method to distinguish the main structures of the human face, the first step consists in the selection of a proper feature or set of features that allow the correct identification of the structures that should be analyzed. In 2D imaging, color and texture attributes are the basis for the selection of the regions of interest, followed by refined features extraction procedures, that locate corners and contours that correspond with facial features like eyebrows, mouth and eye corners, nose tip, lips contour, among others. However, for a proper description of the facial structures it is necessary to include the third dimension. Several approaches have been proposed for 3D facial feature detection, most of them use representation models based on geometrical shape descriptors and establish relations of similarity by means of a process of feature matching [80]. A shape descriptor must have the following strengths: discriminant capacity, quick to compute, concise to store, pose independent and efficient to match. Curvature-based shape descriptors may be the ones that better fit the mentioned requirements. However, thinking of facial features detection, accuracy becomes one of the most important requirements. Therefore, descriptors which offer a better characterization of the surface to a high computational and/or memory cost, also must be considered.

The complexity of the face surface implies that a single descriptor is unable to represent the entire surface. Therefore, in addition to evaluating the conditions outlined in the previous paragraph, it is of crucial interest to analyze the behavior of a descriptor on the face surface. This chapter introduces a study of the behavior of a 3D shape descriptors set computed on the surface of 3D face models. Instead of defining clusters of vertices based on the value of a given primitive surface feature [81], a face template composed by 28 anatomical regions was used to segment the models and to extract the location of different landmarks and fiducial

points. The set of shape descriptors considered in the study includes Minimum, Maximum, Mean, and Gaussian curvatures, Shape Index [82], Curvedness, SPIN images [83], and Finger Prints [84]. Fisher's coefficient [85] was used to establish the relevance of the descriptors in a global and local way.

Facial expressions add variations that must be considered as they are important to understand the facial dynamics. The proposed study was carried out over three data sets that include neutral pose and faces showing expression. The study was extended in order to verify, if the bending-invariant canonical form of the surface eases the identification of facial features in the presence of facial expressions. Therefore, the descriptors were computed directly from the surface and from its bending-invariant canonical form [86].

The rest of this chapter is structured as follows. The set of 3D shape descriptors used in the study is described in Section 3.1. Section 3.2 is devoted to the Fisher's discriminant analysis. Experimental setup and results are reported in Sections 3.3 and 3.4, respectively. Finally, the findings of the study presented in this chapter are summarized in Section 3.5.

3.1 3D Shape Descriptors

First, the set of curvature-based shape descriptor is described. The curvature in facial features detection has been widely used, especially when 3D face geometry information is available. The curvature in the plane is defined as $\kappa = \frac{d\alpha}{ds}$, where α is the angle formed by the vector tangent to the curve (to the direction of displacement) with a fixed direction. Some curvatures can be associated to a surface S in \mathbb{R}^3 : principal curvatures k_1 and k_2 , Mean curvature

$$H = \frac{k_1 + k_2}{2},$$

and Gaussian curvature,

$$K = k_1 k_2,$$

By analyzing the different curvature values it is possible to detect mouth and eyes region [78], to segment a face model as input of a recognition system [87], or to detect several landmarks on the facial surface [69]. Other approaches combine the curvature information along with other features obtained from 2D information [88] and/or *a priori* knowledge of the face geometry [70].

Based on the principal curvature values, other descriptors such as the Shape Index have

been proposed [82]. For a point p on a surface, the Shape Index is defined as

$$S_I(p) = \frac{1}{2} - \frac{1}{\pi} \tan^{-1} \frac{k_1(p) + k_2(p)}{k_1(p) - k_2(p)},$$

S_I has been used to locate facial features in [72, 76, 89].

One more descriptor is the Curvedness defined as

$$R(p) = \sqrt{(k_1^2(p) + k_2^2(p)) / 2},$$

R represents the amount of curvature in a region, enabling the perception of the variation in the shape scale of the objects. The Curvedness is useful in defining the criteria for automatic segmentation of triangular meshes [90].

Other approaches such as SPIN images (SI), describe the surface by means of images related with each oriented point of the surface. An oriented point defines a five degrees of freedom basis (p, n) using the tangent plane P through p oriented perpendicularly to n and the line L through p parallel to n (see Figure 3.1). Parameters of the SPIN image are represented in the *spin-map* S_O (see Equation 3.1) as the function that projects 3D points x to the 2D coordinates of a particular basis (p, n) corresponding to the oriented point O . (for details see [83]).

$$S_O : \mathbb{R}^3 \rightarrow \mathbb{R}^2$$

$$S_O(x) \rightarrow (\alpha, \beta) = \left(\sqrt{\|x - p\|^2 - (n \cdot (x - p))^2}, n \cdot (x - p) \right). \quad (3.1)$$

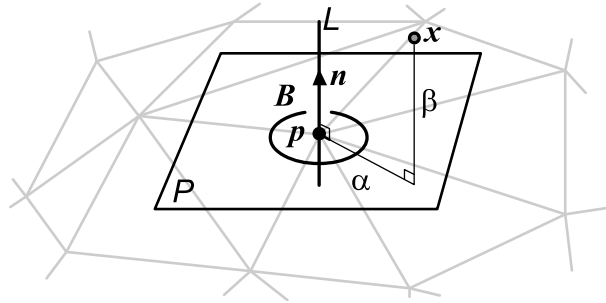


Figure 3.1: Parameters of the SPIN image.

For facial features detection, SPIN images have been combined with methods such as Support Vector Machines [74] and Markov Random Fields [77].

On the other hand, Bronstein *et al.* [86] introduced the bending-invariant canonical form

(BICF) for 3D face recognition in presence of facial expression. The BICF is computed as the embedding of the intrinsic geometry of the face surface to \mathbb{R}^3 . To compute this embedding, a least-squares multi-dimensional scaling [91], with geodesic distances between vertices as dissimilarities, is performed. Wuhler *et al.* [92] combined BICF with a descriptor called *Finger Print (FP)* [84] to predict landmarks on 3D human scans in varying poses. The descriptor uses a measure related to the area of a geodesic circle centered at the point to be characterized. The descriptor at a point p_k ($k = 1, 2, \dots, N$, N is the number of vertices in the model) is obtained by computing the distortion of the geodesic disks with respect to Euclidean disks of the same radius. More specifically, the distortion of the area $A(c)$ of the geodesic disk c of radius r centered at p_k is computed as $d(r) = A(c)/(\pi r^2)$; surface descriptor is a vector of distortions obtained by varying the radius of the geodesic circle (see Figure 3.2).

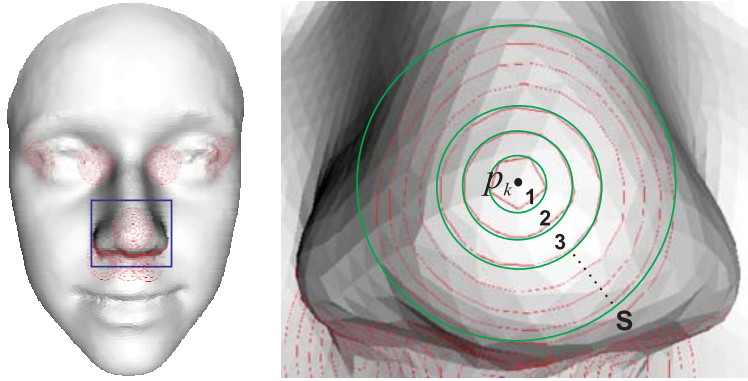


Figure 3.2: Circles used to compute the Finger Print descriptor. Red and green circles correspond to the Geodesic and Euclidean circles respectively.

3.2 Discriminant analysis

Despite algorithms like Principal Components Analysis, which find components of a set of features useful for data representation, they do not allow to find the features that have the most relevant information, which is an important step before performing a classification or recognition process. Fisher's discriminant analysis finds the features that carry the most relevant information by projecting the data in a space with less overlapping within classes.

Consider a data set with d dimensions and n measures x_1, \dots, x_n which is composed of l classes C_i . Fisher's linear discriminant is performed with two (sub sets or) classes with N_1 and N_2 elements obtaining a criterion of separation.

Each class has mean $m_i = \frac{1}{N_i} \sum_{x \in C_i} x$. The data are projected in a new space $y = w^T x$.

These two projected classes have means μ_1 and μ_2 respectively by using $\mu_i = \frac{1}{N_i} \sum_{y \in C_i} \mathbf{y}$.

The separation between classes is obtained by finding a \mathbf{w} that maximizes $m_2 - m_1 = \mathbf{w}^T(\mu_2 - \mu_1)$ [85], where $m_i = \mathbf{w}^T \mu_i$. The within-class variance of the transformed data from the class C_i is obtained from $\sigma_k = \sum_{n \in C_i} (y_i - m_k)^2$.

The total within-class variance for the whole data set is defined as $\sigma_i^2 + \sigma_j^2$. Fisher's criterion is obtained as the ratio of the between-class variance to the within-class variance by using the Equation 3.2.

$$F_{ij} = \frac{(\mu_i - \mu_j)^2}{\sigma_i^2 + \sigma_j^2}. \quad (3.2)$$

Finally, Fisher's coefficient is computed as the mean of all combinations of the Equation 3.2 evaluated for each feature.

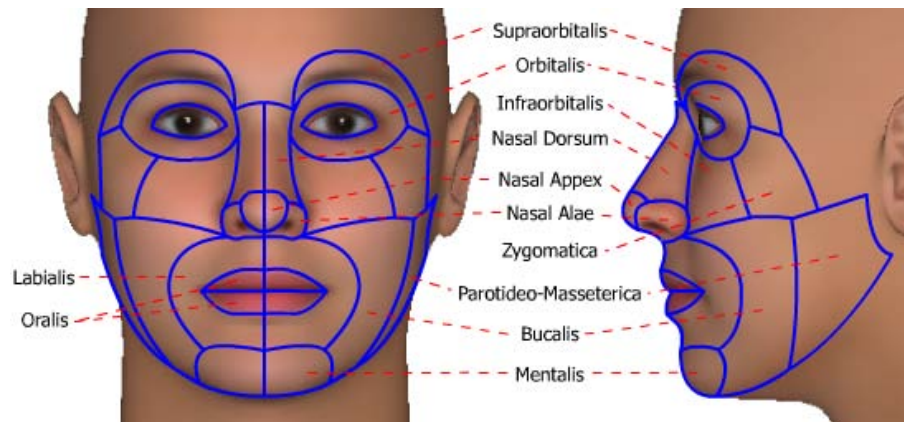
3.3 Experimental Setup

Although several shape descriptors have been proposed, just a few works studied the usefulness of the descriptors for segmenting 3D face models taking into account anatomical information. Moreover, the set of points detected is restricted and the segmentation of regions is not adequate [81], limiting their application in tasks of automatic characterization of the face morphology. Later in this chapter the importance and usefulness of 3D facial geometric shapes (curvature-based), is studied. In comparison with the work by Wang *et al.* [93], the analysis also includes a large set of shape representations such a SPIN images, Finger Prints, and its combination with the BICF. In addition, a more detailed template of the face was developed, and a large set of points was considered.

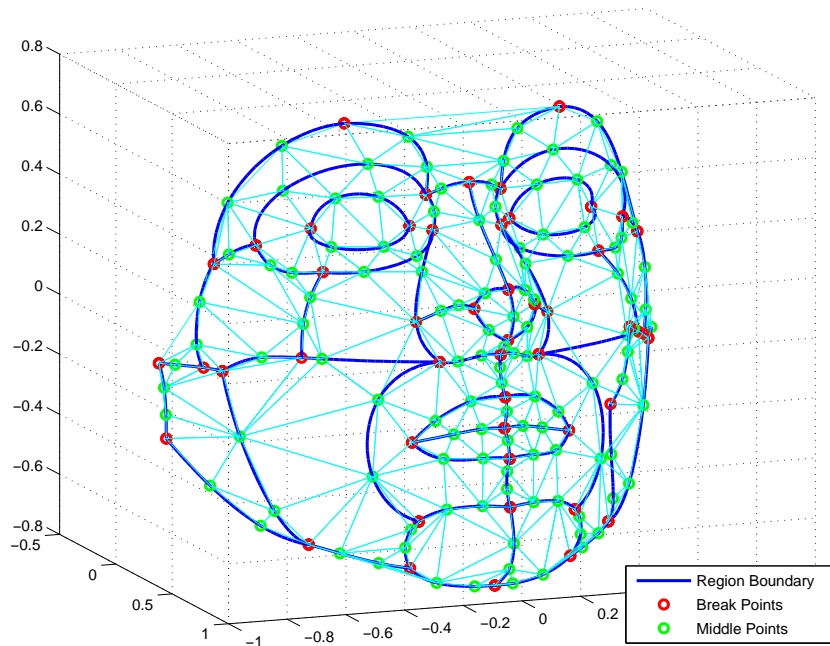
3.3.1 3D Face Template

Each 3D facial model has to be segmented in 28 regions (see Figure 3.3a), which correspond to anatomical regions of the facial soft tissue, which are used to describe an injury in forensics and/or to plan a surgery in many other medical contexts. A region boundary is defined by using Bezier curves, each one with two break points and two control points. Since the control points do not lie on the curve, they are not included in the template; instead, points (middle points) on the curve, which are equally spaced from the break points are defined. The entire template is composed of 68 region boundaries, 46 break points, 136 middle points and 338 triangles. Both break and middle points are the vertices of the 3D face template (see Figure

3.3b).



(a)



(b)

Figure 3.3: (a) Facial regions. (b) Face Template

3.3.2 Databases

Three databases were used: synthetic face models database (DB_s), range-scans face database (3DImDB), and facial expression database *Human Face*. The database DB_s contains 20 models of different characters, 10 female and 10 male characters. Database 3DImDB were captured

from 10 subjects using a Minolta Vivid 9i 3D digitizer. *Human Face* database ¹ contains 15 expressions of the same face showing different facial expressions. Figure 3.4 shows models of each one of the databases used in the study.

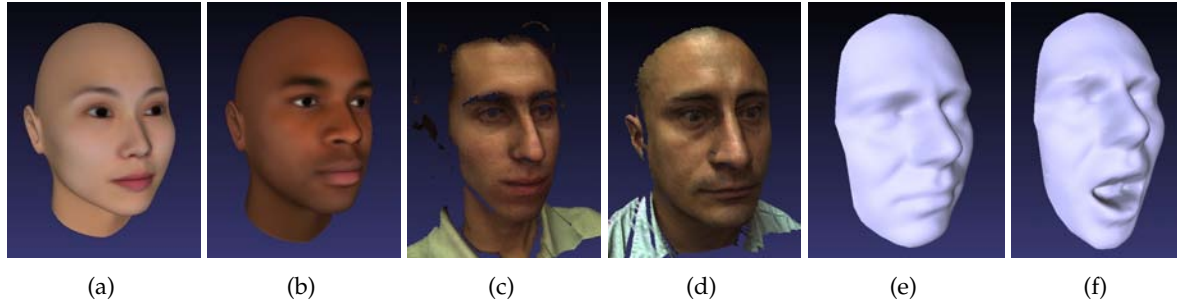


Figure 3.4: Snapshots of models from the databases. DBs (a,b). 3DImDB (c,d). *Human Face* (e,f).

3.3.3 Tests

Two kinds of tests were designed. The first one was meant to establish which descriptor is the most representative over all the set of points (global relevance). The second test is composed of several tests depending on the facial region where the points are located (local relevance). Global relevance is computed based on the Fisher analysis [85]. Eight Fisher's coefficients are estimated, one for each shape descriptor. This test is called Global. Local relevance is estimated in five different facial regions nose, mouth, chin, eyes, and cheeks. This test is called Local. For both region and region boundaries, a set of Fisher's coefficients is obtained. Figure 3.5 shows which regions are considered in each local test.

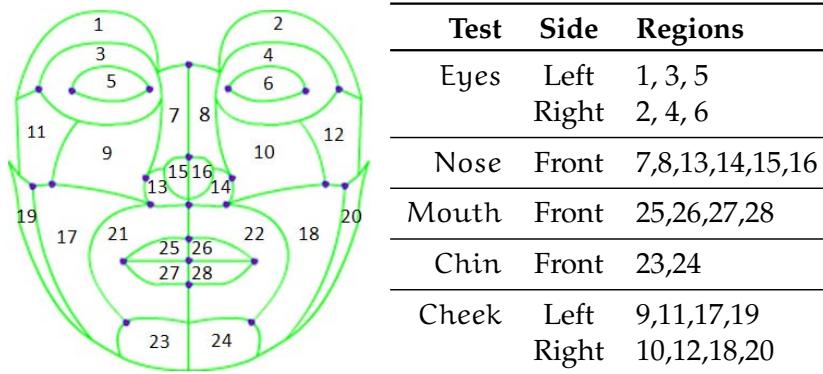


Figure 3.5: Regions considered for each local test

¹http://tosca.cs.technion.ac.il/book/resources_data.html

Additionally, in order to see the relevance of the descriptors in presence of facial expressions, the *Human Face* models were grouped into four categories: neutral, open mouth, smile and smile open mouth. For each group, a Local test was performed. The shape descriptors were computed directly from the surface and from its BICF.

3.4 Experimental Results

3.4.1 Global Relevance

Here, the relevance between the mean of the shape descriptors computed on the regions and contour regions of the 28 regions of the face (see Table 3.1), was computed.

D	DBs				3DImDB			
	Contours Value	R	Regions Value	R	Contours Value	R	Regions Value	R
k_1	3,69	8	9,89	7	0,67	6	1,24	6
k_2	10,68	1	39,87	2	1,03	3	1,76	4
H	5,94	5	18,27	3	1,01	4	2,25	3
K	5,20	6	13,83	5	0,56	7	0,69	7
S_I	6,52	4	14,76	4	1,55	2	3,93	2
R	8,27	2	41,02	1	0,79	5	1,45	5
SI	8,12	3	8,66	8	0,36	8	0,29	8
FP	4,96	7	10,15	6	2,12	1	4,57	1

Table 3.1: Fisher's coefficients for the Global test. **R** indicates the ranking.

Results of the Global test show that for synthetic data the relevance of the descriptors is different depending on the data set analyzed. For synthetic data, the curvature-based shape descriptors k_2 and R have the best values of the Fisher's coefficient. Regarding to the range-scan data set, the FP and S_I descriptors were the best in all cases but the values of the coefficients were lower than the ones for synthetic data. Due to the 3D models from range-scan data are not generated with regular distances between its vertices, which generates a concentration of descriptors values in dense areas, it is necessary to analyze a larger data set in order to catch this non-linear behavior. Unfortunately, a larger manually segmented data set was not available at the time this study was performed. However, with the techniques that will be described in Chapters 4 and 5, the segmentation procedure could be done automatically,

therefore, a better ground truth will be available in the future.

3.4.2 Local relevance

3.4.2.1 Eye area test

In this test, the face side variable was included. The results are shown in Table 3.2 (subscripts L and R correspond to the left and right eyes respectively). For the 3DImDB database, a remarkable difference between the rankings of the contours of both sides was observed. This is because the contours of the eye area present great shape variation, they can be asymmetric, and the points are not enough or could be located in different places on both sides of the face. In addition, the ranking of contours and regions differed a lot. This implies that contours and regions should be analyzed separately. Regarding to the regions, the results shows that both side regions share almost the same ranking, this proves that in regions with high detail, in order to describe the shape properly, a big amount of points is required.

D	DBs				3DImDB _L				3DImDB _R			
	Contour Value	R	Region Value	R	Contour Value	R	Region Value	R	Contour Value	R	Region Value	R
k ₁	6,61	4	12,75	7	1,54	2	1,86	2	0,18	6	1,63	2
k ₂	14,12	2	107,79	1	0,43	7	1,12	4	0,70	3	1,17	5
H	3,13	6	35,85	4	0,48	6	0,24	7	0,46	5	0,51	7
K	1,50	8	35,90	3	0,56	5	0,46	6	0,03	7	0,64	6
S _I	3,17	5	18,65	6	0,94	3	1,70	3	0,55	4	1,21	4
R	12,00	3	54,95	2	0,88	4	2,25	1	0,75	2	1,84	1
SI	1,56	7	1,33	8	0,01	8	0,01	8	0,01	8	0,01	8
FP	15,18	1	22,50	5	2,15	1	1,05	5	0,94	1	1,27	3

Table 3.2: Fisher's coefficients for the Eyes test.

3.4.2.2 Nose area test

The values obtained in this test were higher than the ones of the other regions (see Table 3.3). Results were similar for both data sets. Despite the fact that the nose varies considerably from one subject to another, shape variation occurs in a low scale, then, the differences between the resolution of the data sets prevent that subtle variations could be compared between them.

D	DBs				3DImDB			
	Contour Value	R	Region Value	R	Contour Value	R	Region Value	R
k_1	18,619	3	28,706	4	2,056	3	4,129	3
k_2	4,089	7	19,761	5	1,067	5	1,659	6
H	11,773	4	29,135	3	1,879	4	2,347	5
K	7,017	6	11,266	7	0,649	6	0,553	7
S_I	25,129	2	42,749	2	7,218	2	9,549	2
R	8,975	5	11,805	6	0,436	7	2,409	4
SI	0,186	8	0,182	8	0,001	8	0,003	8
FP	28,360	1	50,480	1	9,528	1	10,826	1

Table 3.3: Fisher's coefficients for the Nose test.

3.4.2.3 Cheeks area test

Despite the surfaces in the cheeks are soft, values of the Fisher's indexes were high. Contrary to the eyes test, contours of both sides shared almost the same ranking, and the ranking of the regions are different. The difference in the values shows that the asymmetry affects the capacity of representation of the descriptors. In this case, variations of shape occurs in a great scale. Therefore, for real data, the descriptors were able to characterize the morphology changes in both sides of the face, which was not possible in the nose region.

D	DBs				3DImDB_L				3DImDB_R			
	Contour Value	R	Region Value	R	Contour Value	R	Region Value	R	Contour Value	R	Region Value	R
k_1	13,61	4	18,29	2	1,68	2	3,26	1	3,45	2	13,57	1
k_2	0,72	8	0,91	8	0,22	7	0,21	7	0,01	8	0,01	8
H	17,57	2	16,90	4	1,40	3	1,47	3	3,01	3	2,96	3
K	0,81	7	0,93	7	0,91	6	1,02	5	1,48	4	0,78	6
S_I	37,34	1	29,04	1	3,51	1	2,57	2	4,54	1	7,15	2
R	10,57	5	17,53	3	1,40	4	1,18	4	1,44	5	1,06	5
SI	15,71	3	15,78	5	0,11	8	0,10	8	0,11	7	0,10	7
FP	9,81	6	7,55	6	1,03	5	0,97	6	1,13	6	2,58	4

Table 3.4: Fisher's coefficients for the Cheeks test.

3.4.2.4 Mouth area test

The values obtained in this test were the lowest (see Table 3.5). A great difference between the ranking of the contours and regions was obtained, this is because the contours are located in the place where the surface changes its orientation, and then, the nature of the information from the vertices of the region and contours is very different. Another reason is that the changes in the surface of the mouth area are soft (except in the mouth corners) making the task of characterization more difficult than in other regions of the face.

D	DBs				3DImDB			
	Contour Value	R	Region Value	R	Contour Value	R	Region Value	R
k_1	0,187	5	0,077	6	0,014	6	0,072	2
k_2	0,225	4	0,143	4	0,027	4	0,019	6
H	0,141	7	0,140	5	0,010	7	0,027	5
K	0,145	6	0,071	7	0,056	2	0,051	3
S_I	0,329	2	1,331	1	0,019	5	0,005	7
R	0,276	3	0,151	3	0,043	3	0,041	4
SI	0,042	8	0,043	8	0,001	8	0,001	8
FP	0,439	1	0,372	2	0,071	1	0,095	1

Table 3.5: Fisher's coefficients for the Mouth test.

3.4.2.5 Chin area test

Table 3.6 shows the results of this test. As in the mouth test, the situation regarding to the values and rankings was similar, which shows that for the kind of surfaces present in the mouth and chin areas, the shape descriptors considered in this study are not able to describe the morphology properly.

3.4.3 Tests with facial expressions

Results were similar to the ones obtained with the range-scan data set. Table 3.7 shows the values of the Fisher's coefficients for the test using the models of the neutral group. T_i corresponds to each local test, subscript $i = 1, 2, \dots, 5$ corresponds to the nose, mouth, chin, eyes and cheeks regions, respectively.

D	DBs				3DImDB			
	Contour Value	R	Region Value	R	Contour Value	R	Region Value	R
k_1	0,026	4	0,005	5	0,001	6	0,346	3
k_2	0,066	2	0,047	1	0,202	2	0,348	2
H	0,077	1	0,027	3	0,276	1	0,047	7
K	0,010	7	0,005	6	0,024	5	0,983	1
S_I	0,021	5	0,001	8	0,001	7	0,107	5
R	0,031	3	0,039	2	0,049	4	0,100	6
SI	0,004	8	0,004	7	0,001	8	0,001	8
FP	0,020	6	0,019	4	0,116	3	0,175	4

Table 3.6: Fisher's coefficients for the Chin test.

D	T₁	T₂	T₃	T₄		T₅	
	Ind/R	Ind/R	Ind/R	Ind/R	Ind/R	Ind/R	Ind/R
k_1	2,145 / 3	0,019 / 6	0,001 / 6	1,42 / 2	0,16 / 6	1,56 / 2	3,38 / 2
k_2	1,167 / 5	0,031 / 4	0,213 / 2	0,38 / 7	0,67 / 3	0,31 / 7	0,01 / 8
H	1,834 / 4	0,016 / 7	0,256 / 1	0,40 / 6	0,51 / 5	1,52 / 3	2,84 / 3
K	0,678 / 7	0,067 / 2	0,019 / 5	0,49 / 5	0,04 / 7	0,98 / 5	1,57 / 4
S_I	7,521 / 2	0,023 / 5	0,001 / 7	0,88 / 3	0,63 / 4	3,24 / 1	4,21 / 1
R	0,532 / 6	0,058 / 3	0,052 / 4	0,79 / 4	0,69 / 2	1,23 / 4	1,01 / 6
SI	0,001 / 8	0,001 / 8	0,001 / 8	0,01 / 8	0,01 / 8	0,10 / 8	0,10 / 7
FP	9,876 / 1	0,089 / 1	0,097 / 3	1,89 / 1	0,79 / 1	0,96 / 6	1,02 / 5

Table 3.7: Fisher's coefficients for the Local test over *Human Face* neutral group. Gray columns correspond to the coefficients of the left side of the face.

In comparison with the values of the descriptors of the local tests with the range-scan data set, there were not significant differences. The main difference is in the rankings of the shape descriptors of the mouth and cheeks region, which are the ones with more changes, because of the facial expression (See Tables 3.7 and 3.8). Notice that in all cases the values of the descriptors were small. This situation demonstrates the need for a mixed representation, which includes the 3D shape descriptors as well as the geometric relationship between the characteristic points, which describe each one of the face regions.

D	T₁		T₂		T₃		T₄		T₅	
	Ind/R	Ind/R	Ind/R	Ind/R	Ind/R	Ind/R	Ind/R	Ind/R	Ind/R	Ind/R
k ₁	1,937 / 3	0,021 / 7	0,001 / 6	1,56 / 2	0,25 / 6	1,01 / 6	1,22 / 6			
k ₂	0,093 / 7	0,049 / 3	0,319 / 1	0,48 / 7	0,82 / 2	0,42 / 7	0,01 / 8			
H	1,494 / 4	0,023 / 6	0,301 / 2	0,52 / 5	0,63 / 5	1,03 / 5	1,75 / 4			
K	0,492 / 5	0,059 / 2	0,021 / 5	0,51 / 6	0,03 / 7	1,62 / 2	2,91 / 3			
S _I	5,285 / 2	0,078 / 1	0,001 / 7	0,92 / 3	0,72 / 4	1,48 / 4	1,27 / 5			
R	0,382 / 6	0,041 / 4	0,049 / 4	0,85 / 4	0,76 / 3	3,59 / 1	4,67 / 1			
SI	0,001 / 8	0,001 / 8	0,001 / 8	0,01 / 8	0,01 / 8	0,11 / 8	0,10 / 7			
FP	7,634 / 1	0,035 / 5	0,088 / 3	2,01 / 1	0,95 / 1	1,61 / 3	3,66 / 2			

Table 3.8: Fisher’s coefficients for the Local test over *Human Face* smile open mouth group. Gray columns correspond to the coefficients of the left side.

3.4.3.1 Tests with Bending-Invariant Canonical Forms

Except for the nose region, the values of the coefficients of the 3D shape descriptors computed over the BICF of the models were close to zero. Figure 3.6 illustrates how the BICF removes most of the variations due to facial expressions. Only the region of the nose is easy to identify. This representation could be useful to model spatial relationships between fiducial points of the face and to develop a hierarchical search strategy for facial features detection.

3.5 Conclusions

An analysis of the relevance of nine 3D shape descriptors on points, contours, regions, areas and sides of the face was carried out. Based on the Fisher’s coefficient, it was shown how the morphology of the face surface influences the capacity of representation of the descriptors. From this, it was determined which descriptors, whether curvature-based, point of view-based or local information-based, are more appropriate to characterize each of the major areas that define the face.

In tests, where the side of the face was included as a variable, it was shown that the descriptors are suitable to capture changes due to the natural asymmetry of the face. However, in order to characterize such a variation, an analysis should be performed over a comprehensive database. In addition, it should be noted that depending on the area to be assessed, an analysis at a higher or lower resolution (e.g., regions of the nose and mouth) is required.

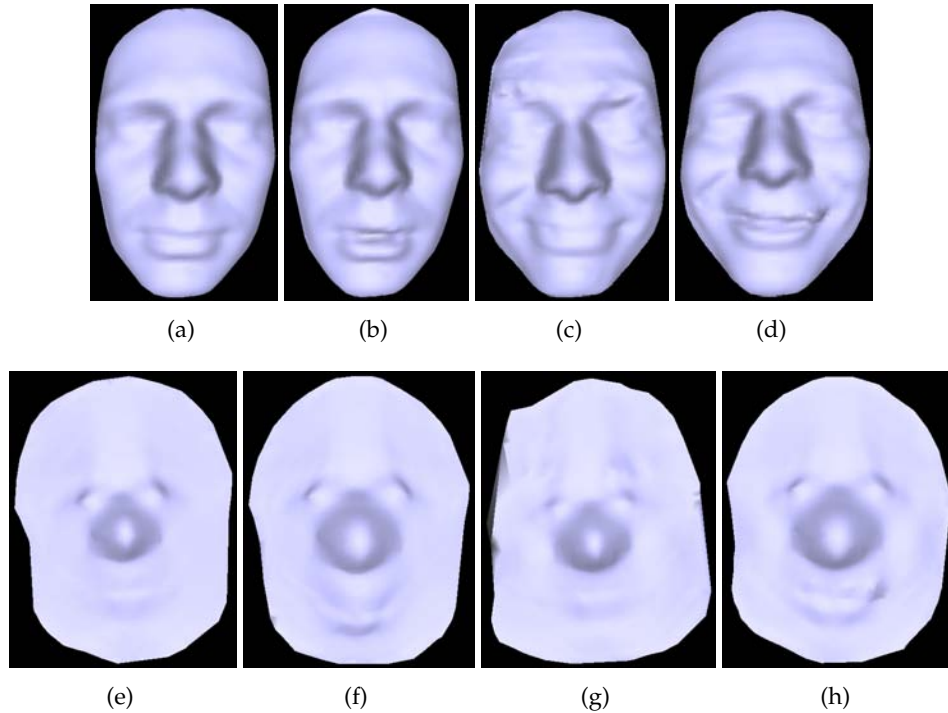


Figure 3.6: Models from HumanFace database. Original data (a-d). Bending-invariant canonical forms (e-h).

In the regions belonging to the mouth and chin, the studied descriptors did not show a difference in their levels of relevance, therefore, it is necessary to conduct a study that includes other kind of descriptors. Anyway, it should be noted that in order to identify those regions, the search strategy should consider the low variability of the surfaces.

It was showed how the bending-invariant canonical forms are able to remove the majority of the variation due to facial expression, but this feature complicates the location of landmarks based on the analysis of the shape descriptors values. As most of the regions of the face becomes flat, the variation of the descriptors values is subtle. Thinking in facial features extraction, the bending-invariant canonical forms are useful to identify the nose region with high accuracy, for the others face regions, it is necessary to model the geometric relationships between the different fiducial points.

Non-rigid registration of Faces

After analyzing the behavior of several 3D shape descriptors over the face geometry, the next step is to develop a method that allows the automatic identification of the different facial structures where the shape descriptors were analyzed. The non-rigid registration of surfaces is a power tool that allows the dense point-to-point correspondence between surfaces. Therefore, the registration of a face template to a 3D triangle mesh of the face of a subject, allows the automatic detection of as many facial features as the intrinsic geometry of the template carries. In addition, other attributes as the identity variations and expression dynamics could be also be analyzed. Therefore, a method developed for the non-rigid registration of faces should be able to work in presence of: noise, different initializations, variations of pose, identity and expressions changes, among others.

In this chapter the problem of computing point-to-point correspondences among a set of human face scans with varying expressions in a fully automatic way, is considered. As a result a raw 3d model will be parameterized in such a way that likewise anatomical parts correspond with the ones of a face template, where the locations of landmarks and anatomical regions are known. Facial expression affects the geometry of the human face and therefore is important for facial shape analysis. Computing accurate point-to-point correspondences for a set of face shapes in varying expressions is a challenging task because the face shape varies across the database and each subject has its own way to perform facial expressions. The problem is further complicated by incomplete and noisy data in the scans.

This chapter describes a novel technique to compute correspondences between a set of facial scans with varying expressions that does not require the scans to be spatially aligned. The correspondence computation procedure uses a template model P as prior knowledge on the geometry of the face shapes. Unlike Xi and Shu [94], the aim is to find correspondences for faces with varying expressions. Hence, it is not enough to have a template model that

captures the face shape of a generic model, but also the expressions of a generic model need to be captured. To achieve this, P is modeled as a blendshape model as in Li et al. [95]. In a blendshape model, expressions are modeled as a linear combination of a set of basic expressions. Hence, blendshape models are both simple and effective to model facial expressions.

This approach proceeds as follows. First, a database of human face scans with manually placed landmark positions is used to learn local properties and spatial relationships between the landmarks using a Markov network. Given an input scan F without manually placed landmarks, primary the landmark positions are predicted on F by carrying out statistical inference over the trained Markov network. Sections 4.2.1 and 4.2.2 discuss this step. In order to perform statistical inference, the search region for each landmark needs to be restricted. This is detailed in Sections 4.2.3 to 4.2.6. The predicted landmarks are used to align P to F . In order to fit the expression of P to the expression of F , the template is aligned to the scan as outlined in Section 4.3.1 and the weights of the generic blendshape model are optimized as discussed in Section 4.3.2. Finally, the shape of P is changed to fit the shape of F as outlined in Section 4.3.3. Figure 4.1 shows an overview of the method.

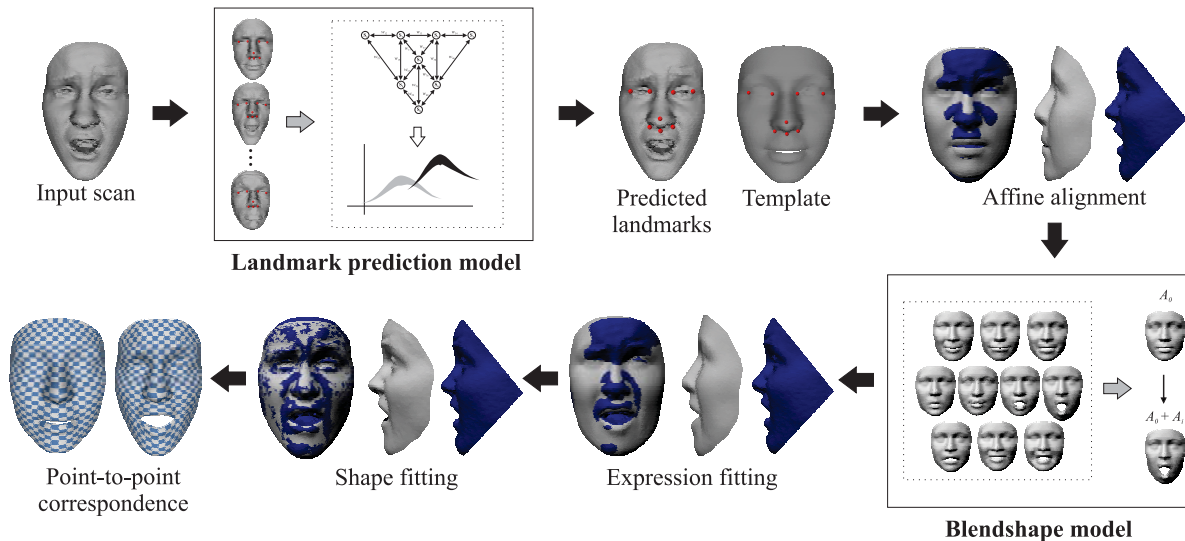


Figure 4.1: Overview of the fully automatic expression-invariant face correspondence approach.

4.1 Related Work

This section reviews literature in face shape analysis related to finding landmarks on face models, computing correspondences between three-dimensional shapes, and using blendshape models for facial animation.

4.1.1 Finding Landmarks on Face Models

Traditionally, facial features are detected in 2D images. In this setting, facial feature detection can be achieved in an unsupervised (see for instance [96, 97]), semi-supervised (see for instance [98]) or supervised (see for instance [99]) manner. Unsupervised methods do not use prior information about the geometry of target object. However, these methods only estimate a global affine transformation between the source and the target object. On the other hand, semi-supervised and supervised methods estimate a shape deformation described by a set of landmarks, which provide more accurate and consistent results. To incorporate prior knowledge about landmark locations, it often suffices to annotate only a few examples manually [98].

Recent developments on 3D data acquisition have allowed to overcome the problems attached to the 3D technologies. However, only a few approaches consider 3D landmark detection, while accounting for expression and pose variations [100]. It is well-studied that facial landmarks play an important role in applications, such as face or expression recognition [101].

Ben Azouz et al. [77] propose a method to find correspondences by automatically predicting marker positions on 3D models of a human body. The method encodes the statistics of a surface descriptor and geometric properties at the locations of manually placed landmarks in a Markov network. This method works only for models with slight variation of posture.

Mehryar et al. [100] introduce an algorithm to automatically detect eyes, nose, and mouth on 3D faces. The algorithm correctly detects the landmarks in the presence of pose, facial expression and occlusion variations. This method is useful as initial alignment but not for an accurate registration.

Berreti et al. [102] combine principal curvatures analysis, edge detector and SIFT descriptors to find 9 landmarks on the eyes nose and mouth regions in range images. The landmarks are properly detected in the presence of facial expressions but the method relies in anthropometric facial proportions to define the search regions and assumes that the face is upright oriented.

Creusot et al. [103] present a method to localize a set of 13 facial landmark points under large pose variation or when occlusion is present. Their method learns the properties of a set of descriptors computed at the landmark locations and encodes both local information and spatial relationships into a graph. The method works well for neutral pose. However, in the presence of expression variation, the accuracy decreases considerably.

Segundo et al. [71] develop a method for face segmentation and landmark detection in range images. The landmark detection method combines surface curvature information and depth relief curve analysis to find five landmarks located on the nose and eye regions. The landmarks are properly detected in the presence of facial expressions and hair occlusions, but the method relies on a specific acquisition setup.

Perakis et al. [104, 105] present a method to detect landmarks under large pose variations using an Active Landmark Model (ALM), which is a statistical shape model learned from eight manually annotated landmarks. Using a combination of the Shape Index descriptor and Spin Images, the search space for the fitting of the ALM is defined. The final set of landmarks is defined by selecting the set of candidates that satisfies the geometric restrictions encoded in the ALM. The experiments show that the method works in the presence of facial expressions and pose variation up to 80 degrees around the y-axis.

Nair and Cavallaro [106] use a point distribution model to estimate the location of 49 landmarks on the eyebrow, eye and nose regions. The method works well in the presence of expressions and noisy data. However the error in the localization of landmarks is quite high (a comparison of the results is provided in Section 4.4.2).

Lu and Jain [107] present a multimodal approach for facial feature extraction. The nose tip is located using only the 3D information, and the eyes and mouth corners are extracted using 2D and 3D data. As their focus is handling changes in head pose and lighting conditions, variations due to facial expressions are not considered in their experiments. This multimodal approach is used by Lu et al. [108] as part of a system for face recognition in the presence of pose and expression variation (only smiling expression variations are included in the test data). The authors claim that the expression changes decrease the accuracy of the system. However, quantitative results of the landmark detection are not provided. In addition, the requirement of the texture data is a limitation of the multimodal approaches because sometimes such information is not available.

As here the aim is to obtain accurate point-to-point correspondences, a landmark prediction method based on the approach of Ben Azouz et al. [77], was derived. The surface descriptor which was used is able to catch the local geometry properly [84] and, by combining it with a canonical representation [109], this new approach is able to detect landmarks in the presence of facial expressions. A machine learning-based approach was selected to avoid classic assumptions such as: the nose tip is the closest point to the camera [110], the inner-

corners of the eyes and the tip of the nose are the most salient points [71], the 3D face scan is in a frontal upright canonical pose [102], among others. The advantage is that learning-based approaches can easily be extended to other contexts.

4.1.2 Correspondence Computation

Several methods have been proposed to solve the problem of establishing a meaningful correspondence between shapes. Here, the focus lies on computing correspondences between human face shapes. Methods that do not assume templates usually have the problem that some points are not registered accurately. To remedy this, a template model is assumed. In the following, only approaches that use template models are reviewed (for details about methods for correspondence computation see the survey of van Kaick et al. [111]).

Passalis et al. [112] proposed a 3D face recognition method that uses facial symmetry to handle pose variation and missing data. A template is fitted to the shape of the input model as follows: an Annotated Face Model [113] is iteratively deformed towards the input using automatically predicted landmarks and an algorithm based on Simulated Annealing. When dealing with facial expressions, the performance of the recognition system decreases. This is due to an incorrect registration of the mouth region. Mpiperis et al. [114] propose a method that supports both 3D face recognition and expression recognition. A template model is fitted to the shape of the input model using an elastic deformation model. Both works do not show direct evaluations of the fully-automatic registration methods as this is not the main part of these works.

Guo et al. [115] propose a multimodal approach to automatically compute correspondences between 3D face models. The approach predicts 17 landmarks using a PCA-based method and uses these features to deform a template to the input model using a thin-plate spline. Although the registration results are shown to be accurate, the method cannot compute correspondences in the presence of expression variation.

Huang et al. [116] recently presented an approach to register 3D facial models in the presence of facial expressions. They first detect a set of landmarks using texture information with the help of an active appearance model. These points are used in an iterative fitting procedure, which combines displacement mapping, point-to-surface mapping, and a regional blending algorithm to fit a template to the 3D surface. The fitting accuracy of this method is evaluated on manually selected landmarks, and a high fitting accuracy is presented, thereby demonstrating that the combined use of geometry and texture leads to good results. In contrast, the method presented in this chapter is purely geometry-based, and could therefore in

principle also be applied to 3D data of faces without reliable texture information.

Statistical learning-based approaches have been effectively used to model facial variations oriented to both the synthesis and recognition of faces. Blanz and Vetter [117] developed a 3D morphable model (3DMM) for the synthesis of 3D faces from photographs. As the registration is specific to the scanning setup, rigid alignment of the scans is assumed. Lu and Jain [118] present an approach to perform face recognition using 3D face scans. The approach builds a 3DMM for each subject in the database. When a test image becomes available, the approach matches the scan to a specific individual using the learned 3DMM. Unlike the here presented method, their training data is parameterized using manually placed landmarks and the test scans are parameterized using individual-specific deformation models. Basso et al. [119] extend the method of Blanz and Vetter [117] to register 3D scans of faces with arbitrary identity and expression. The rigid alignment of the scans is also assumed for registration. To avoid the use of texture information, Amberg et al. [120] present a method to fit a 3DMM to 3D face scans using only shape information. They demonstrate the performance of the method in the presence of expression variation, occlusion and missing data, but do not conduct extensive evaluations of the registration.

Registration methods based on iteratively deforming a template to the data are an alternative to statistical learning-based approaches. Allen et al. [121] present an approach to parameterize a set of 3D scans of human body shapes in similar posture. To fit the template to each scan, the method proceeds by using a non-rigid iterative closest point (ICP) framework coupled with a set of manually placed marker positions. Xi and Shu [94] extend the method of Allen et al. [121] to deform a template model to a head scan. The shape fitting is carried out as in Allen et al. [121] but uses radial basis functions to speed up the deformation process. Unlike the method presented here, this only allows for neutral expressions and uses manually placed markers to align the template to a head scan. Wuhrer et al. [122] propose a method to deform a template model to a human body scan in arbitrary posture. The method works in two stages: posture and shape fitting. Posture fitting relies on the location of different landmarks, which are predicted in a fully automatic way using a statistical model of landmark positions learned from a population. The method described in this chapter can be viewed as an extension of this approach, but instead of fitting the posture, the expression is fitted using blendshapes (see Section 4.1.3).

Methods that compute a correspondence between two surfaces by embedding the intrinsic geometry of one surface into the other one by using Generalized Multi-Dimensional Scaling (GMDS) [123] are another alternative to deal with variations due to facial expres-

sions [124]. The performance of these methods has been demonstrated for face recognition. As GMDS methods do not take care that close-by points on one surface map to close-by points on the other, the results are often spatially inconsistent. This prevents such methods from being used for shape analysis.

4.1.3 Use of Blendshape Models

Modeling expressions using blendshape models is an alternative to approaches based on statistical models where a comprehensive database annotation process has to be carried out to extract variational information. In a blendshape model, movements of the different facial regions are assumed to be independent. Any expression is then modeled as a linear combination of the differences between a set of basic expressions, called *blendshapes*, and a neutral expression. That is, to produce an expression, the displacements causing the movement are linearly combined. Using a representative set of blendshapes, this simple model is effective to model facial expressions.

Li et al. [95] propose a method to transfer the expression of a subject to an animated character. Their framework allows to create optimal blendshapes from a set of example poses of a digital face model automatically. Weise et al. [125] present a framework for real-time 3D facial animation. The method tracks the rigid and non-rigid motion of the user's face accurately. They incorporate the expression transfer approach of Li et al. [95] in order to find much of the variation from the example expressions. The registration stage requires offline training where a generic template is fitted to the face of a specific subject. To obtain the results, manual marking of features has to be carried out.

Because of the advantages of modeling expression using linear blendshapes, here they are used to aid in shape matching and only a blending weight per expression is optimized. This reduces the dimensionality of the optimization space drastically. Since the used database of blendshapes is small, the expression fitting stage of the proposed algorithm is efficient and helps to improve the results significantly.

4.2 Landmark Prediction

This section outlines how to predict a set of landmark positions on a face scan. To establish the correspondences across the whole database, a template is fitted to each model. The fitting process begins with the extraction of the locations of eight landmarks shown as red spheres in Figure 4.2. The locations of the landmarks were selected based on the fact that in the

presence of facial expressions, the corners of the eyes, and the base and tip of the nose do not move drastically. Each landmark is located automatically on the face surface by means of a Markov network following the procedure proposed by Ben Azouz et al. [77]. The network learns the statistics of a property of the surface around each landmark and the structure of the connections shown in Figure 4.2.

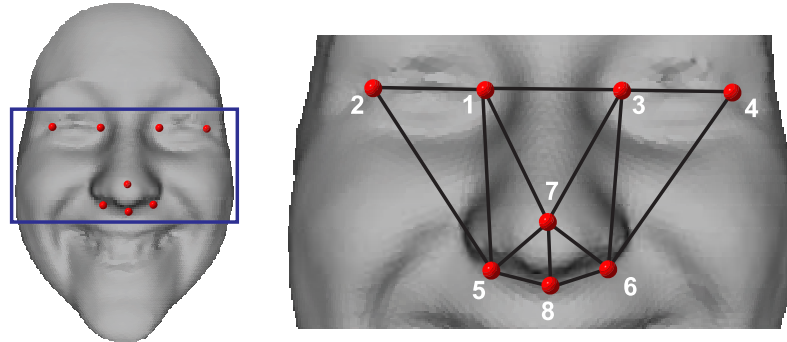


Figure 4.2: Face model with landmarks. Locations and landmark graph structure.

4.2.1 Learning

Two important aspects have to be defined for the training of the Markov network. First, each landmark l_i ($i = 1, 2, \dots, L$), represented by a network node, is described using a node potential ϕ_i . The surface descriptor Finger Print *FP* (described in Section 3.1) is used. In addition to the reasons exposed in the previous chapter, the *FP* is used as potential because it is isometry-invariant. Hence, in scenarios where the surface undergoes changes that preserve isometry, *FP* is effective to encode the surface information of an object.

Second, a link between landmarks l_i and l_j , represented by a network edge, is described using an edge potential $\psi_{i,j}$. Although the locations of the landmarks were selected based on the observations that nose and eye regions do not change much in the presence of expressions, some distortions along the edges of the Markov network may occur. To minimize the effects of the face movements, the canonical form [109] of each model is computed and the edge potential is defined as the relative position of landmark l_i with respect to landmark l_j in the canonical form space.

The Markov network training process learns the distributions of both node and edge potentials for each individual node and edge of the network, respectively. In this case Gaussian distributions for both the node and edge descriptors are assumed, and the distributions are learned using maximum likelihood estimation. This commonly used distribution was chosen

to derive an efficient algorithm that is easy to implement. While this distribution may not be satisfied in practice, experimentally was found that using this simplified assumption yields satisfactory results.

4.2.2 Prediction with Belief Propagation

The estimation of the location of landmarks on a test model is carried out by using probabilistic inference over the Markov network. The aim is to find landmark locations l_i , such that the joint probability

$$p(l_1, \dots, l_L) = \frac{1}{Z} \prod_i \phi_i(l_i) \prod_{i,j} \psi_{i,j}(l_i, l_j) \quad (4.1)$$

is maximized, where Z is a normalizing factor. In practice, an approximate solution using the loopy belief propagation algorithm [126] was found. This algorithm requires a set of possible labels for each node. This means that a number of candidate locations for each landmark has to be provided.

Wuhrer et al. [122] use canonical forms to learn the average locations of the landmarks, but because of the flipping-invariant property of the canonical forms, it is necessary to compute eight different alignments and select the one that leads to the minimum distance between the scan and the deformed template. To remedy this, a method to restrict the search space based on a rough template alignment is introduced in this chapter. Thus, only one fitting process has to be computed, reducing the computing cost by a factor of eight.

4.2.3 Restricting the search region

There are two reasons to reduce the search space for the landmarks: to increase the efficiency of the landmark prediction and to eliminate the ambiguity caused by the facial symmetry. Here, the problem of restricting the search region for the landmarks is treated as a 3D face pose estimation problem. In this case, the estimated pose does not have to be so accurate since the Markov network refines the position of the landmarks, but it has to be accurate enough to identify the left and right sides of the face. The proposed face pose estimation method finds four landmarks located on the nose region and extracts the information of the face symmetry planes by using a template of the landmark graph. Once the nose landmarks are labeled, the final position of the entire set of landmarks is obtained by transforming the template to the coordinate system of the test model. Figure 4.6 shows the main steps of the proposed search space restriction method.

4.2.4 Classification of Vertices

Before explaining the rough template alignment procedure, a method to classify a vertex of a 3D model into a specific class is introduced. The classes correspond to the nodes of the Markov network and the 3D model corresponds to a 3D face model. The decision rules are derived from a clustering procedure over the Principal Components Analysis (PCA) projections of a surface feature and a pre-selection method based on the surface primitives.

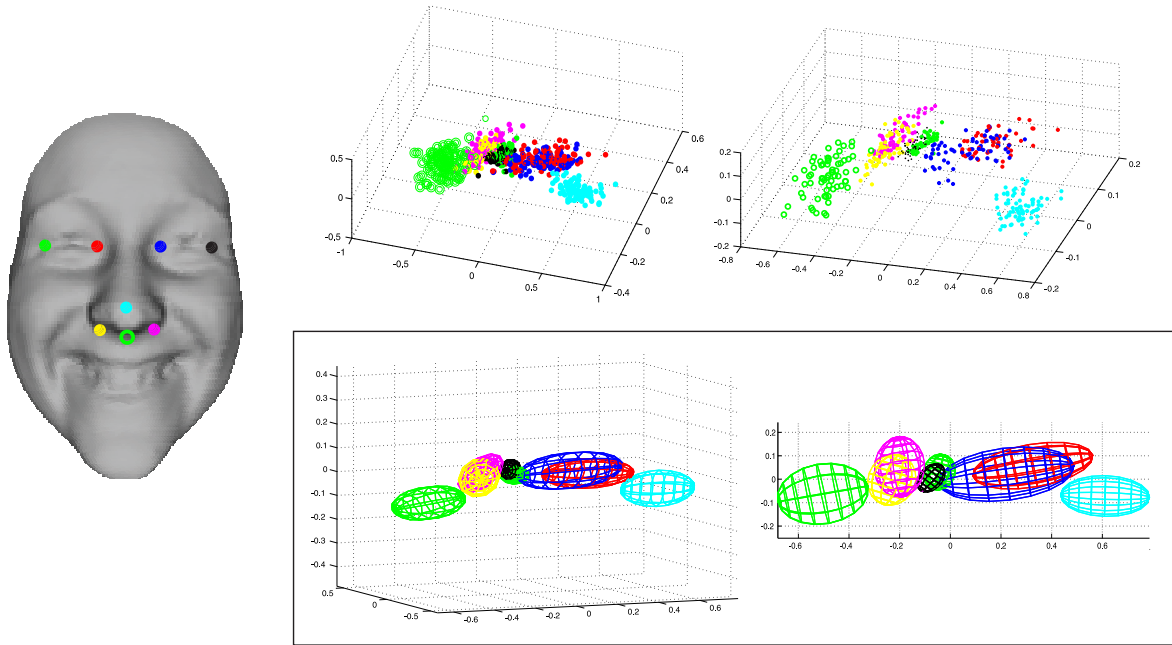


Figure 4.3: PCA-based clustering. Left: Landmarks on a face model. Upper Right: Initial clusters formed with all the samples. Lower Right: Final cluster after removing the samples beyond a 1.5 standard deviations from the cluster medoid. Minimum volume enclosing ellipsoids (3D and upper views).

As the value of the FP descriptor at each landmark l_i was computed during the Markov network training process, the distributions of the surface descriptors can be modeled and used to classify a vertex v_k on the face surface into a class i (each landmark corresponds to a class). PCA is a useful tool to compress a high-dimensional space into a linear low-dimensional space. When the space corresponds to a multidimensional feature space, sometimes, depending on the distinctiveness of the features, it is possible that elements of the same class form clusters in the PCA space. Here, the FP descriptor can be viewed as S -dimensional vector and PCA is used to reduce the dimensionality to D . In this case, $D = 3$ was chosen. Figure 4.3 shows the results of applying PCA to the data from the subjects in neutral and performing six expressions (for information about the database, see Section 4.4.1).

Although samples of the same class tend to form groups in the PCA space, some groups

overlap due to symmetric landmarks. In order to improve the separation between classes, a new cluster is defined, denoted as *M-cluster*, by removing the samples which are farther than M ($M \in \mathbb{R}^+$) times the standard deviation from the cluster medoid. Medoids are representative objects of a cluster whose average dissimilarity to all the objects in the cluster is minimal [127]. For instance, Figure 4.3 shows the *M-clusters* formed by setting $M = 1.5$. With this value, the clusters corresponding to the landmarks nose tip and subnasal (points 7 and 8 in Figure 4.2) do not overlap any of the clusters. In Sections 4.2.5 and 4.2.6 will be shown that with a good separation between these two classes, a proper landmarks prediction can be obtained.

A rule E_i is derived for a class i based on a clustering procedure. The rule E_i is defined as the minimum volume enclosing ellipsoid of a *M-cluster* _{i} (see Figure 4.3). E_i is obtained from the representation of the ellipsoid in the center form as $(p_k - C_i)^T A (p_k - C_i) \leq 1$, where C_i corresponds to the center of the ellipsoid corresponding to class i and A is the 3×3 matrix of the ellipse equation. When a new point p_k becomes available, each E_i is evaluated in order to see if the point satisfies the equation. As some *M-clusters* are overlapping, it is possible that more than one label be assigned to the same p_k . Similarly, it is possible that p_k is not assigned to any class because the point lies in a region that is not of interest. Figure 4.4 shows an example of the vertex classification results obtained using the proposed method.

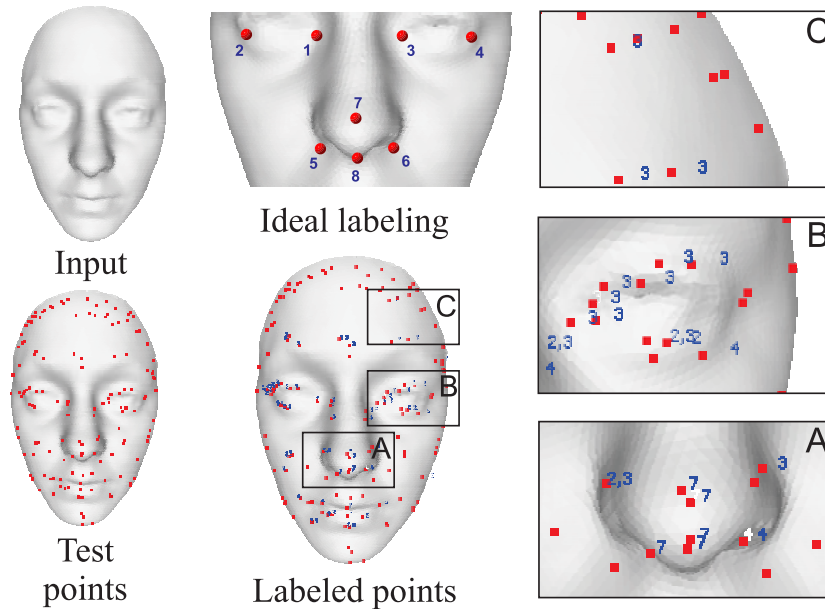


Figure 4.4: Example of vertex labeling result. (A) Notice how the points on the nose tip region are correctly labeled. (B) Some vertices are assigned to two classes. This situation is because of the left-right symmetry of the features. (C) Points located far from the region of interest are discarded.

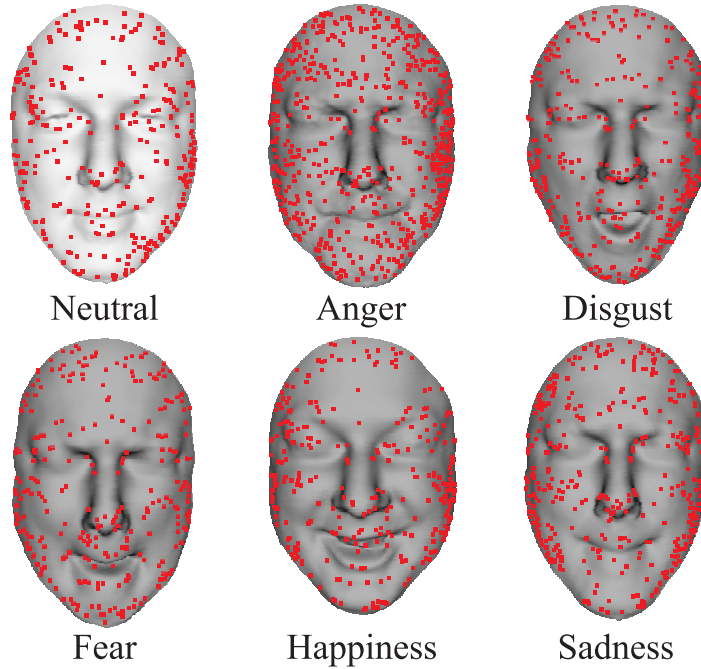


Figure 4.5: *Umbilics* of different 3D facial models of the same subject performing different expressions. Notice how the *umbilics* are distributed all over the surface, and in most of the cases umbilics are present at the locations of salient facial features.

It is not efficient to compute the descriptor value and its projection to PCA space for all the vertices of the mesh. To reduce the search space, samples were computed on the surface using a curvature-based descriptor. More precisely, all surface *umbilics* [128] were used as samples, umbilics are the points on the surface where the principal curvatures are identical (that is, $k_1 = k_2$). This sampling approach was chosen because it can be observed experimentally that most landmark positions are located close to a umbilic, as shown in Figure 4.5.

4.2.5 Refining the Nose Landmarks

This section describes the procedure to select candidates for four points on the nose area, which are used as initial guess of the landmarks: right subalare, left subalare, nose tip, and subnasal, which are labeled as 5, 6, 7 and 8, respectively (see Figure 4.2). Following the classification procedure described in Section 4.2.4, for each umbilic of the input scan F , the FP descriptor is computed, projected into PCA space, and labeled (in the following this procedure will be referenced as *FPPCA*). The result is a set of candidates for each landmark class (see first row of Figure 4.6). To find an initial position of landmark l_i , the points in the neighborhoods of umbilics that were labeled l_i were considered.

The search starts in the nose tip region. The point selected as starting point, it is the vertex v of F that corresponds to the umbilic that after *FPPCA* is the closest point to the medoid of the cluster of points labeled as nose tip. The new search space corresponds to the set of vertices v_k within the geodesic circle of radius r centered at v . For each v_k , *FPPCA* is applied. In this step, only points v_k that are either labeled as nose tip or subnasal were considered. This procedure is depicted in the second row of Figure 4.6.

Next, the positions of the right and left subalare are refined. The refinement starts from the point v closest to the medoid of all points that were labeled as subnasal in the previous step. The algorithm proceeds by classifying points v_k in a geodesic neighborhood of radius r of v using *FPPCA*. In this step, only the points v_k that are labeled as right or left subalare are considered. Since the *M-clusters* of these two classes strongly overlap, most of the labeled points are assigned to two classes and the non-relevant points are discarded (see third row of Figure 4.6). Since the labeled vertices are distributed over both sides of the nose, this set of vertices is split up into two sets by performing a k – means clustering with $k = 2$. The two new sets of vertices still have both labels, and the point closest to the medoid of each cluster is selected as a possible candidate (see fourth row of Figure 4.6). It remains to determine which of these points corresponds to the right subalare, and which one to the left.

4.2.6 Aligning Landmark Graph to Scan

So far, four points on the nose region have been selected and labeled. Due to the face symmetry, two of the points have the same labels. To solve this problem, a template P_a of the upper part of the face with the same structure as the landmark graph (see Figure 4.2) is roughly aligned to the input scan F . This helps also to estimate the initial guess of the remaining landmarks: right inner eye corner, right outer eye corner, left inner eye corner, and left outer eye corner, which are labeled as 1, 2, 3 and 4, respectively.

Here, a rigid alignment \mathbf{T} that best aligns the point set v_a from P_a with the point set v_b from F is computed. The point set v_a corresponds to the points labeled 5 to 8 of P_a , and v_b corresponds to the four points on the nose region of F . As the labels 5 and 6 of the points in v_b are unknown, there are two possible configurations for the alignment. As a result two linear transformations \mathbf{T}_1 and \mathbf{T}_2 are obtained. In order to select the transformation that produces a valid result, the transformed point sets $P_1 = \mathbf{T}_1 P_a$, and $P_2 = \mathbf{T}_2 P_a$, are computed. One of the transformations produces a vertical “flip” of the template, resulting in a wrong estimation of the coordinates of the points in the eye region. Therefore, the point set P_i that minimizes the sum of Euclidean distances to closest points on F is the correct transformation. This procedure is depicted in the fifth row of Figure 4.6.

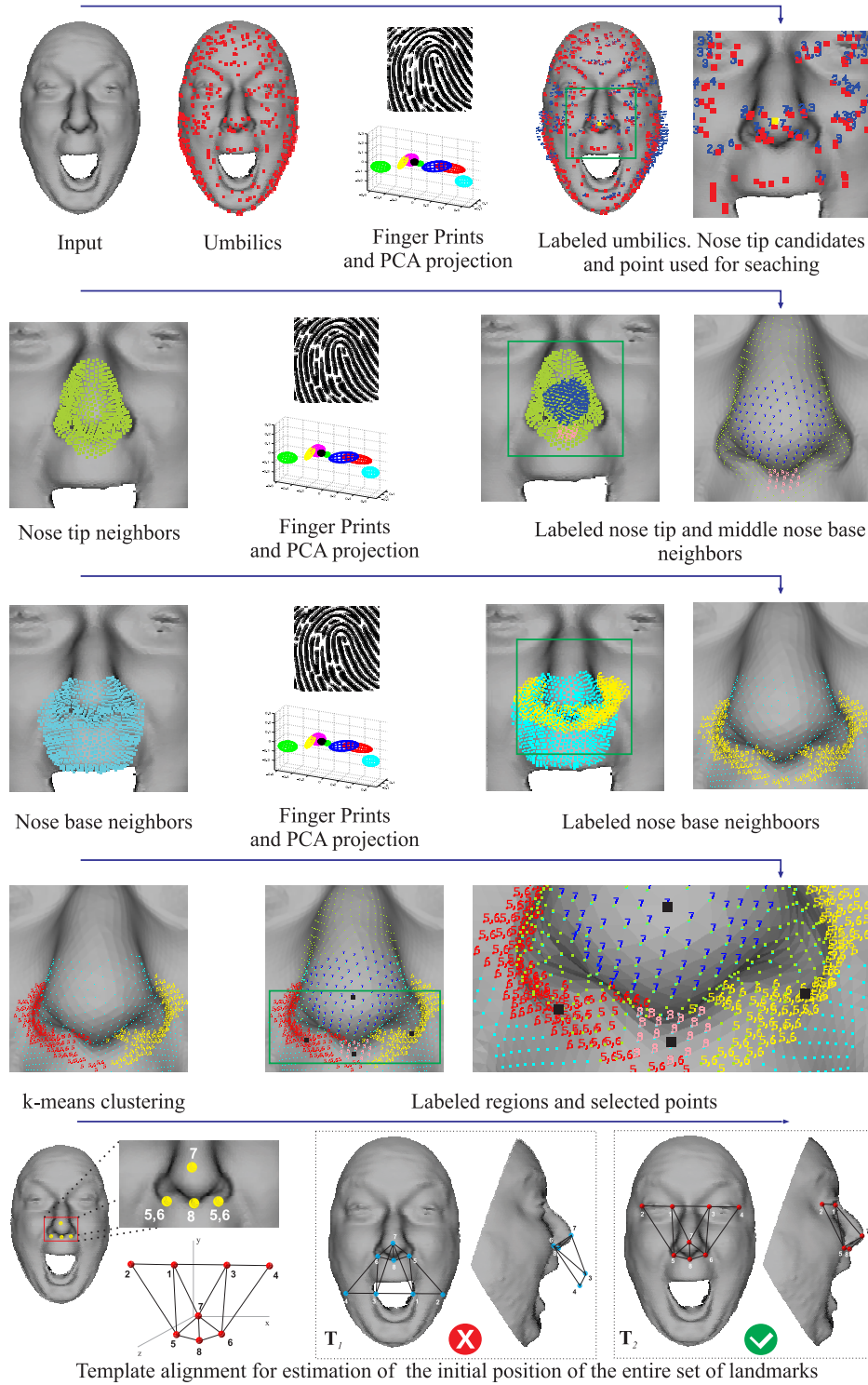


Figure 4.6: Framework of the proposed initial alignment method.

The locations of the transformed template vertices are used to define the search space region on which statistical inference is performed, as discussed in Section 4.2.2. The regions are defined as all points within distance r from the transformed points P_i .

4.3 Registration

This section describes how a template is fitted to a 3D scan of the face. The input scan corresponds to a face of a subject performing a facial expression. Fitting a template to this scan is challenging because the facial geometry has large variations due to different face shapes and facial muscle movements. Here a registration method is proposed. The expression and the shape are fitted separately in order to handle the complexity of the problem. Figure 4.7 shows an overview of the proposed method.

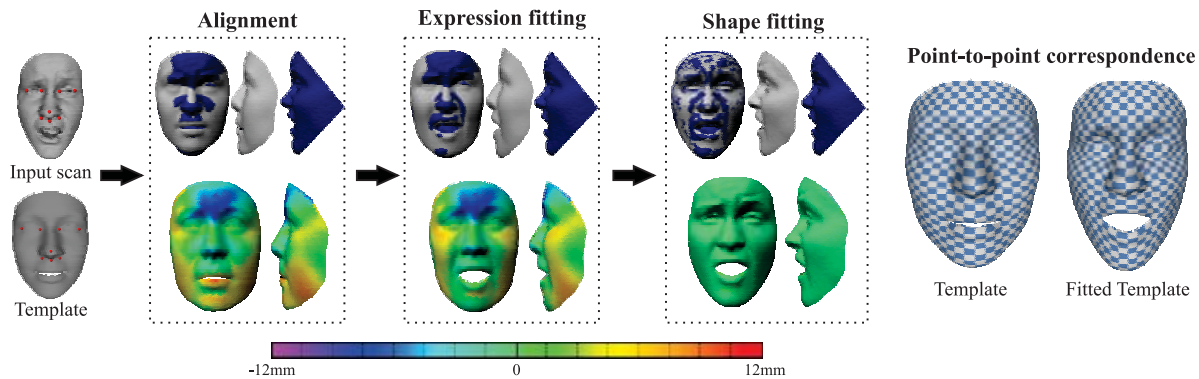


Figure 4.7: Registration procedure. First, the template and the scan are aligned using the predicted landmarks. Second, the expression is fitted using a blendshape model. Finally, an energy-based surface fitting method is used to fit the shape. At the end, the overlap between the scan and the template is maximized and a point-to-point correspondence for the face shapes in different expressions is obtained.

In this case, the facial expression fitting problem is addressed as a facial rigging problem. In facial rigging, a facial expression is produced by changing a set of parameters associated with the different regions of the face modeled using blendshapes. Conceptually, to generate a facial shape from a 3D rest pose face template, just a set of vertices is moved to a new location, e.g., lift an eyebrow or open the mouth (see Figure 4.8). In this sense and similar to the approach proposed by Li et al. [95], a facial expression is modeled as a linear combination of facial blendshapes (denoted by A_i), which are expressed as vectors of displacements from the rest pose (denoted by A_0).

4.3.1 Affine Alignment

To solve the fitting problem, the template A_0 in neutral pose is aligned to a scan F as follows. Both A_0 and F contain a set of landmarks denoted by \bar{l}_i and l_i , respectively. The landmarks l_i were predicted using the method described in Section 4.2. The alignment is carried out by finding a 3×4 transformation matrix \mathbf{T}_A that minimizes the energy

$$E_{\text{land}} = \sum_{i=1}^L (\mathbf{T}_A \bar{l}_i - l_i)^2, \quad (4.2)$$

with respect to the 12 parameters in \mathbf{T}_A using a quasi-Newton approach starting from \mathbf{T}_A as identity matrix.

4.3.2 Expression Fitting

The aim of this step is to model expression variations using a small number of basis shapes. An expression can be generated using a small number of parameters as

$$P(\alpha_i) = A_0 + \sum_{i=1}^j \alpha_i A_i, \quad (4.3)$$

where A_0 corresponds to the rest pose, $A_i, i > 0$ correspond to the blendshape displacements, and α_i ($0 \leq \alpha_i \leq 1$) are the blending weights of expression $P(\alpha_i)$. For each blendshape A_i , Figure 4.8 shows the corresponding expressions. The 3D models used in both the creation of A_0 and the generation of A_i were obtained using a commercial software. Notice that mostly mouth displacements are considered. As the expressions are generated as a linear combination of displacements, to avoid exaggerated undesired expressions, it is important that no two blendshapes add the same kind of displacement. By using a blendshape model, the facial expression fitting problem is transformed into an optimization problem, where the value of each α_i has to be estimated.

Recall that A_0 and F are affinely aligned. The α_i that best match the expression of F , it is found by dividing $P(\alpha_i)$ into three regions: chin, mouth, and remaining face (as shown in Figure 4.9). The division is motivated by the fact that the chin and lip regions vary drastically from one expression to another (mostly in terms of displacements). Thus it is desirable to inspect the quality of the fitting in each of these regions separately by assigning higher weights to points in these regions than to points in the remaining face.

To fit the expression, the energy

$$E_{\text{expr}} = \sum_r \omega_r \langle (\text{NN}(p_r(\alpha_i)) - p_r(\alpha_i)), \bar{n}(\text{NN}(p_r(\alpha_i))) \rangle^2, \quad (4.4)$$

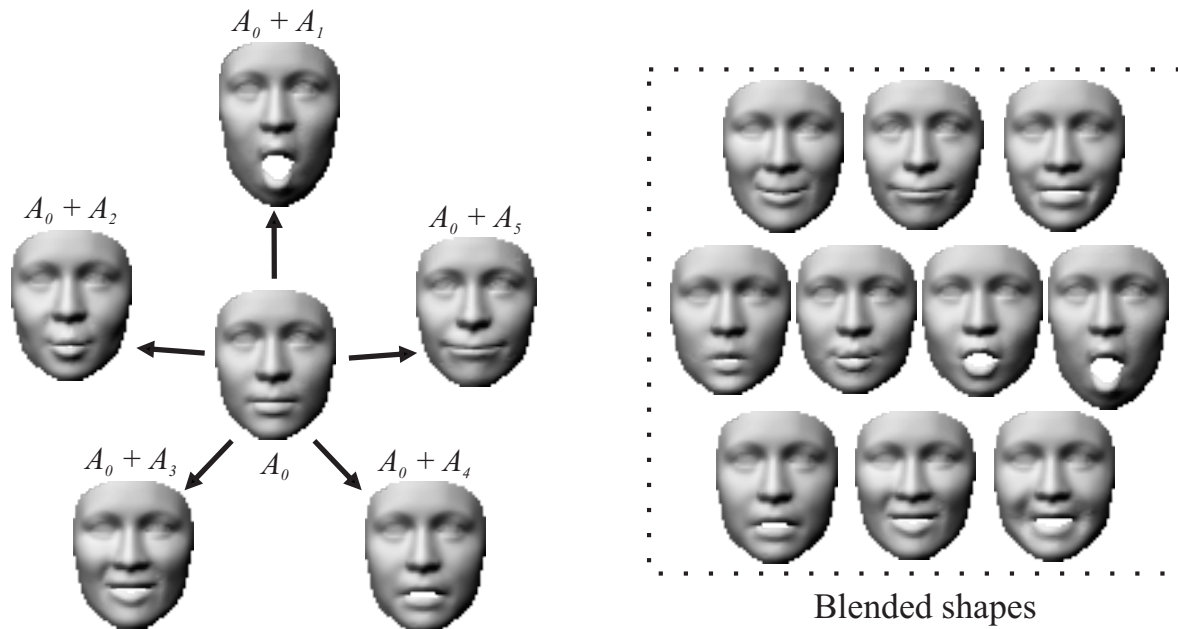


Figure 4.8: Left: template rest pose A_0 and a set of blendshapes A_i . Right: examples of models generated as linear combinations of blendshapes.

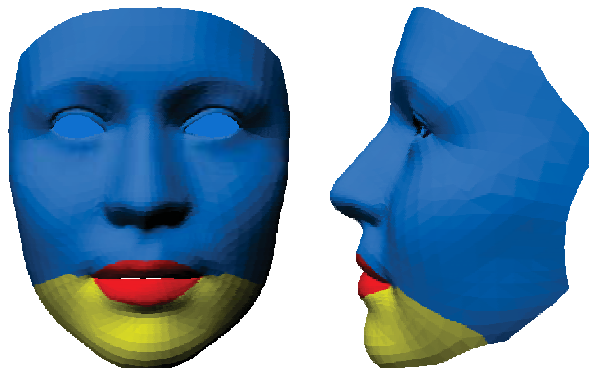


Figure 4.9: Regions used in the expression fitting procedure.

is used. $p_r(\alpha_i)$ are the vertices of $P(\alpha_i)$, $NN(p_r(\alpha_i))$ indicates the nearest neighbor point of $p_r(\alpha_i)$ on F , $\vec{n}(NN(p_r(\alpha_i)))$ is the unit outer normal vector of $NN(p_r(\alpha_i))$, $\langle \cdot, \cdot \rangle$ denotes the dot product of two vectors, and ω_r is a weight associated with $p_r(\alpha_i)$. The energy pulls each vertex of the template to the nearest point on the tangent plane of its nearest neighbor on F . The weight ω_r is used for two purposes: to give different weight to the mouth, chin, and remaining regions of the model, and to make the method more robust to both the presence of outliers and mis-oriented surfaces. To achieve the first goal, ω_r is set to either ω_{mouth} , ω_{chin} , or $\omega_{\text{remaining}}$, depending on the region containing $p_r(\alpha_i)$. To achieve the second goal, only the nearest neighbor is considered if the angle between the outer normal vectors of $p_i(\alpha_i)$ and $NN(p_r(\alpha_i))$ is small. Specifically, $\omega_{\text{remaining}}$ is set to zero if the angle is larger than φ . To force the fit to be exact, ω_{chin} and ω_{mouth} are set to zero if the angle is larger than $\varphi/2$. The expression is fitted by minimizing Eq. 4.4 with respect to the blending weights α_i . In the experiments performed here, φ was set to 80 degrees.

The minimization of E_{expr} is carried out in two stages. In the first stage, it is inspected if some movement occurs in the chin. Once the position of the chin is known, to refine the match with the expression of the input model, it is necessary to inspect the position of the lips. Based on this, the expression fitting procedure proceeds as follows: First, the weight ω_{mouth} is set to zero, thus the minimization is only guided by vertices that are not in the mouth region. In this step $\omega_{\text{remaining}}$ is set to one and ω_{chin} is defined as $1 - (V_{\text{chin}}^{\text{valid}}/V_{\text{chin}})$, where V_{chin} is the number of vertices in the chin region and $V_{\text{chin}}^{\text{valid}}$ is the number of valid nearest neighbors in this region. The second step begins when at least 80% of the vertices in the chin region have valid nearest neighbors. At this time, ω_{mouth} is set to $1 - (V_{\text{mouth}}^{\text{valid}}/V_{\text{mouth}})$, where V_{mouth} is the number of vertices in the mouth region and $V_{\text{mouth}}^{\text{valid}}$ is the number of valid nearest neighbors in this region. The minimization process ends when at least 60% of the vertices in the mouth region have valid nearest neighbors. This weight variation scheme ensures that the chin and mouth regions of $P(\alpha_i)$ match the expression of F . The threshold values for ω_{chin} and ω_{mouth} were chosen based on experimental observations.

This step fits the expression of the template to the expression of the scan. However, since the deformations are modeled by a small number of parameters, the deformation during this step is restricted, and fine shape details cannot be modeled by this step.

4.3.3 Shape Fitting

To find a more accurate local fitting, the shape of $P(\alpha_i)$ is fitted to the shape of F . For ease of notation, $P = P(\alpha_i)$ is used in the following.

The shape fitting is, again, treated as an optimization problem similar to the method proposed by Allen et al. [121] and extended by Li et al. [129]. The goal is to find a set of 3×4 transformation matrices \mathbf{T}_i for each vertex p_i of P such that p_i is moved to the new location $\tilde{p}_i = \mathbf{T}_i p_i$ to fit the shape of F . The transformed version of P is denoted \tilde{P} . The transformation matrices \mathbf{T}_i are obtained by minimizing an energy function, which is a weighted sum of three energy terms.

The first term is the data term

$$E_{\text{data}} = \sum_i \omega_i \langle (\text{NN}(\tilde{p}_i) - \tilde{p}_i), \vec{n}(\text{NN}(\tilde{p}_i))) \rangle^2, \quad (4.5)$$

where $\text{NN}(\tilde{p}_i)$ indicates the nearest neighbor of \tilde{p}_i on F , and $\vec{n}(\text{NN}(\tilde{p}_i))$ is the normalized outer normal of $\text{NN}(\tilde{p}_i)$. The weight ω_i is set to one if the angle between the outer normal vectors of \tilde{p}_i and its nearest neighbor is at most 80 degrees, and to zero otherwise. The data term ensures that the template is deformed to resemble the input scan.

The second energy is a regularization term that encourages smooth transformations between neighboring vertices of the mesh. This energy is called the regularization energy E_{reg} and it is defined as

$$E_{\text{reg}} = \sum_{(i,j) \in E(\tilde{P})} (\mathbf{T}_i - \mathbf{T}_j)^2, \quad (4.6)$$

where $E(\tilde{P})$ is the set of edges of \tilde{P} . This term prevents adjacent parts of P from being mapped to disparate parts of F , and also encourages similarly-shaped features to be mapped to each other [121].

The final energy term encourages the transformation matrices to be rigid. The rigid energy E_{rigid} , which measures the deviation of the column vectors of \mathbf{T}_i from orthogonality and unit length, is defined as

$$E_{\text{rigid}} = \sum_{i=1}^r \left(\left((\mathbf{a}_1^i)^T \mathbf{a}_2^i \right)^2 + \left((\mathbf{a}_1^i)^T \mathbf{a}_3^i \right)^2 + \left((\mathbf{a}_2^i)^T \mathbf{a}_3^i \right)^2 + \left(1 - (\mathbf{a}_1^i)^T \mathbf{a}_1^i \right)^2 + \left(1 - (\mathbf{a}_2^i)^T \mathbf{a}_2^i \right)^2 + \left(1 - (\mathbf{a}_3^i)^T \mathbf{a}_3^i \right)^2 \right), \quad (4.7)$$

where $\mathbf{a}_1^i, \mathbf{a}_2^i, \mathbf{a}_3^i$ are the first three columns vectors of \mathbf{T}_i .

The energy terms described above are combined in the weighted sum

$$E_{\text{shape}} = \omega_{\text{data}} E_{\text{data}} + \omega_{\text{reg}} E_{\text{reg}} + \omega_{\text{rigid}} E_{\text{rigid}}. \quad (4.8)$$

The shape is fitted by minimizing E_{shape} with respect to the parameters \mathbf{T}_i . To encourage smooth and rigid transformations, the weights are set as follows: $\omega_{\text{data}} = 1$, $\omega_{\text{reg}}^0 = 20000$, and $\omega_{\text{rigid}}^0 = 10$. Similar to Li et al. [129], whenever the energy change is negligible, the weights are relaxed as $\omega_{\text{reg}}^t = 0.5\omega_{\text{reg}}^{t-1}$ and $\omega_{\text{rigid}}^t = 0.5\omega_{\text{rigid}}^{t-1}$ to give more weight to the data term. This allows the template to deform towards the scan. The algorithm iterates until the relative change in energy $(E_{\text{shape}}^{i-1} - E_{\text{shape}}^i)/E_{\text{shape}}^{i-1}$, where i is the iteration number, is less than 0.0001. For each set of weights, a quasi-Newton approach [130] was used to solve the optimization problem, and at most 1000 iterations are performed.

As the template only includes the shape of the face and the template can be free deformed during the shape fitting, in both expression and shape fitting procedures, the boundary points of the input model are ignored to prevent that the fitting results include noise shapes from the hair or ears of the input model.

4.4 Experiments and results

4.4.1 Database

The database BU-3DFE [131] was used for all of the experiments. This database consists of 3D face models from 100 subjects (56 Females and 44 Males) in neutral pose and with the following facial expressions: *surprise, happiness, disgust, sadness, anger* and *fear*. There are four scans of each facial expression, corresponding to different levels of intensity from *low* to *highest*. As a file containing the raw data of each scan is also available, there are a total of 50 files per subject, 25 raw scans and 25 corresponding to the cropped faces. Figure 4.10 shows snapshots of different scans from the BU-3DFE database. In the experiments, a subset of 700 3D models corresponding to the cropped faces of the subjects performing the expressions in the highest level was used.

4.4.2 Landmark prediction accuracy

Two different subsets of models of 50 subjects (25 females and 25 males) to train the landmark prediction model were used. First, the subset T_n consisting of 50 models of subjects in neutral pose was used as training set. Second, the subset T_e consisting of 350 models of the same 50

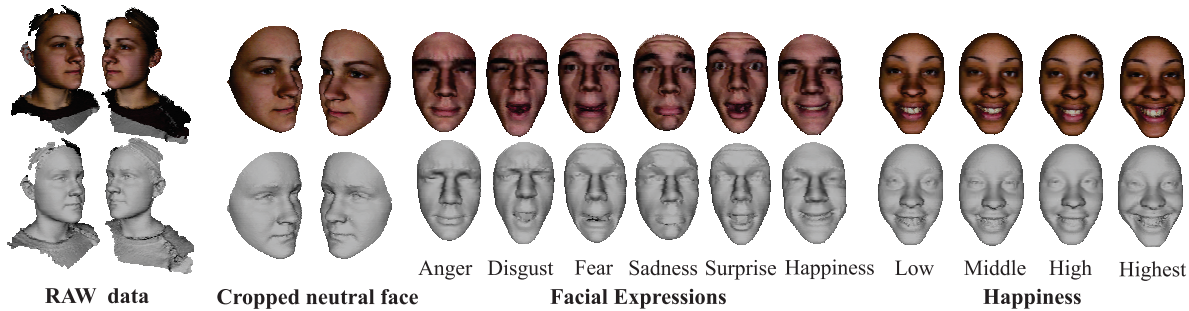


Figure 4.10: Characteristics of the BU-3DFE database.

subjects in neutral pose and performing six different facial expressions was used as training set. As T_n covers the shape variability and T_e covers both shape and expressions variability, this enables the evaluation of the influence of the variabilities considered in the training sets. The accuracy of the landmark prediction algorithm is evaluated over the remaining 50 subjects of the database (31 females and 19 males). The test database corresponds to 350 models of subjects in both neutral pose and when performing six different facial expressions.

To evaluate the accuracy of the landmark prediction algorithm, the error of the Euclidean distance between a manually located landmark l_i and its corresponding estimation \hat{l}_i was computed. The mean, the standard deviation, and the maximum of the error were computed. Also the detection rates were computed by counting the percentage of test models where the landmark \hat{l}_i was predicted with an error below 10mm ($T < 10$), 20mm ($T < 20$), and 30mm ($T < 30$). Tables 4.1 and 4.2 show the results of the evaluation for the test with T_n and T_e as training databases, respectively.

The best landmark prediction results were obtained when T_e is used for training. In both experiments, the landmarks located in the nose region are better predicted than the ones located in the eye region. The tip of the nose is predicted with the lowest error and the outer corners of the eyes are predicted with the highest error. One of the reasons that the outer corners of the eyes are not predicted as well as the other landmarks is that the initial position is found based on the alignment of the landmark template (see Figure 4.6). This adds an estimation error that is reflected in the values of the standard deviation. The values of the detection rates show the improvement in accuracy of the landmark prediction when T_e is used as training set. This indicates that for the configuration of the landmark prediction model described in this chapter, the variations due to both shape and expression have to be considered.

These results of landmark prediction were compared with two approaches where the BU-3DFE database is also used for testing. Segundo et al. [71] used 2500 range images obtained

Landmark	Mean \pm Std [mm]	Max. [mm]	T < 10 [%]	T < 20 [%]	T < 30 [%]
Right inner eye corner	10.35 \pm 6.13	33.93	53.71	87.14	92.57
Right outer eye corner	11.79 \pm 7.77	34.73	27.71	85.71	93.43
Left inner eye corner	11.63 \pm 6.82	34.16	44.57	86.57	94.00
Left outer eye corner	12.57 \pm 7.23	34.29	31.43	89.14	95.71
Right subalare	9.96 \pm 6.59	33.49	66.00	86.86	98.00
Left subalare	10.93 \pm 6.87	34.15	55.14	87.43	94.29
Nose tip	7.42 \pm 5.64	32.03	82.57	92.00	96.86
Subnasal	7.12 \pm 5.87	33.75	84.57	87.43	95.43

Table 4.1: Error of landmark prediction with training set T_n . T < 10, T < 20, and T < 30 correspond to the detection rates with a tolerance of 10mm, 20mm and 30mm, respectively.

Landmark	Mean \pm Std [mm]	Max. [mm]	T < 10 [%]	T < 20 [%]	T < 30 [%]
Right inner eye corner	6.14 \pm 4.54	34.39	80.86	95.14	97.43
Right outer eye corner	8.49 \pm 6.12	34.54	62.29	95.14	97.71
Left inner eye corner	6.75 \pm 4.21	33.75	84.00	96.57	98.29
Left outer eye corner	9.63 \pm 5.82	34.63	63.14	93.43	98.86
Right subalare	7.17 \pm 3.3	32.23	85.43	95.14	97.43
Left subalare	6.47 \pm 3.07	32.3	89.71	96.86	97.43
Nose tip	5.87 \pm 2.7	29.91	93.71	97.43	100
Subnasal	5.57 \pm 2.03	30.26	95.43	98.29	99.71

Table 4.2: Error of landmark prediction with training set T_e . T < 10, T < 20, and T < 30 correspond to the detection rates with a tolerance of 10mm, 20mm and 30mm, respectively.

Landmark	[71] [mm]	[106] [mm]	Proposed Method [mm]
Right inner eye corner	6.33	20.46	6.14
Right outer eye corner	N.A.	12.11	8.49
Left inner eye corner	6.33	19.38	6.75
Left outer eye corner	N.A.	11.89	9.63
Right subalare	6.49	N.A.	7.17
Left subalare	6.66	N.A.	6.47
Nose tip	1.87	8.83	5.87
Subnasal	N.A.	N.A.	5.57

Table 4.3: Comparison of mean errors of the proposed method and two different approaches.

from the raw data, and Nair and Cavallaro [106] used 2350 of the 2500 3D cropped face models available. Table 4.3 shows the mean of the error of the landmark prediction. For all the landmarks, the here described approach outperforms the approach of Nair and Cavallaro [106]. Compared to Segundo et al. [71], for all the landmarks but the nose tip the mean error is similar. Recall however that Segundo et al. [71] use a more challenging dataset for testing.

Although the obtained landmark prediction error appears to be high, it is still possible to obtain a proper point-to-point correspondence since the landmarks are only used to align the template to the scan. Afterwards, a non-rigid iterative closest point framework is used to deform the expression and shape of the template. Figure 4.11 shows some examples of the landmark prediction results over models of subjects with different facial shapes and performing different expressions.

In the following T_e is used as training data set. Furthermore, only the models where all landmarks are predicted within 30mm of the ground truth (332 of the 350 models) are considered.

4.4.3 Registration

The proposed dense point-to-point correspondence algorithm was tested on 332 models.

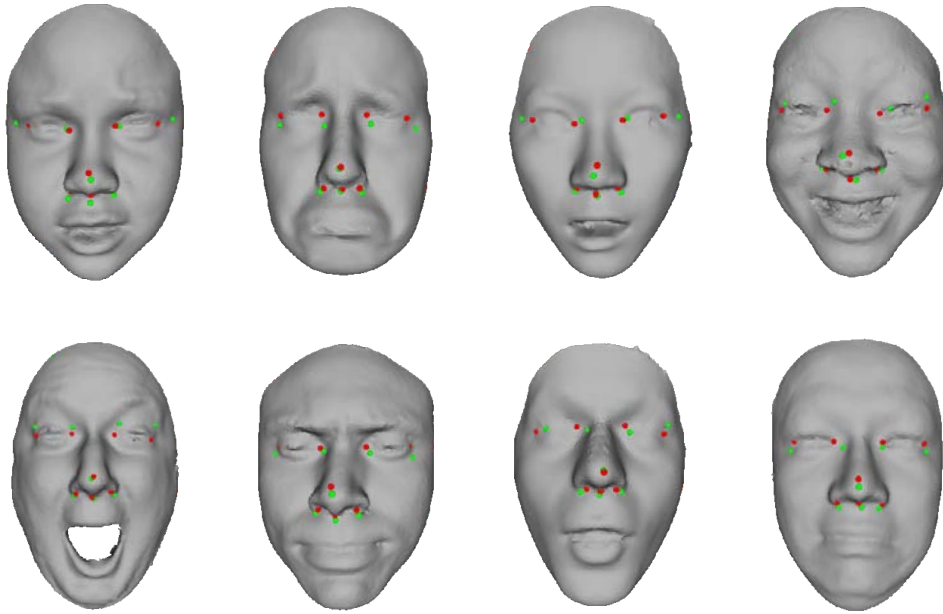


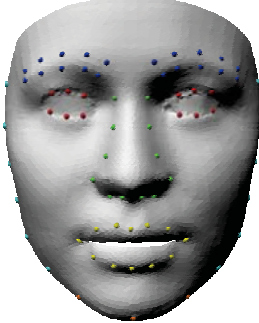
Figure 4.11: Examples of the landmark prediction results. Red and green spheres correspond to the manually placed and predicted landmarks, respectively. First row: female subjects; Second row: male subjects.

4.4.3.1 Landmark Fitting Accuracy

To evaluate the accuracy of the registration, the error in the location of manually placed landmark points present in the BU-3DFE database that are not considered for the alignment is computed. The error corresponds to the Euclidean distance between a manually placed point and its corresponding location after registration. The set of points considered for the evaluation (see Figure 4.12) includes 20 points on the eyebrows (10 left, 10 right), 12 points on the eye contours (6 left, 6 right), 12 points in the nose region, 12 points on the outer contour of the lips, 3 points on the chin, and 12 points on the face contour (6 left, 6 right).

Table 4.12 shows the mean, the standard deviation, and the maximum of the error, as well as the detection rates. In this case, the mean and the standard deviation were computed over all points in a region and over all 332 models used for correspondence computation. Furthermore the detection rates are computed by counting the percentage of test models where all the points belonging to the same region were predicted with an error below 10mm ($T < 10$), 20mm ($T < 20$), and 30mm ($T < 30$).

The points on the eye contour and the nose region were found with lower mean error and variation than the points on the mouth, chin, and eyebrows regions. This situation is expected because the movements in the eyebrows and mouth are more pronounced than in



Points	Mean \pm Std [mm]	Max. [mm]	T < 10 [%]	T < 20 [%]	T < 30 [%]
Left Eyebrow	6.28 \pm 3.30	25.36	52.87	98.79	100
Right Eyebrow	6.75 \pm 3.51	23.59	45.62	98.19	100
Left Eye	3.25 \pm 1.84	12.53	98.19	100	100
Right Eye	3.81 \pm 2.06	12.24	96.07	100	100
Nose	3.96 \pm 2.22	16.97	87.61	100	100
Mouth	5.69 \pm 4.45	45.36	52.57	94.26	98.79
Chin	7.22 \pm 4.73	33.80	58.01	95.47	99.39
L. Face	18.48 \pm 8.52	52.17	0.60	22.36	64.05
R. Face	17.36 \pm 9.17	58.36	0.30	22.96	60.12

Figure 4.12: Error at landmark points not used for registration. Left: set of points. Right: summary of errors.

the other areas. The big difference between the error on the face contour points with respect to the other regions is mainly because of there are no strong anatomical attributes that help to define the face contour, which results in highly inconsistent manually placed markers across the database.

The next point which will be discussed is the quality of the results after the final shape fitting step. Figure 4.13 shows the cumulative distribution of the number of models where the error at the landmark points not used for registration is below a threshold (due to noise, the set of ground truth points on the face contour was not included). Note that even when the error at some points is slightly high, it was found that both the face regions and the surface geometry of the input models are consistently matched with their counterparts in the deformed template.

4.4.3.2 Surface Fitting Accuracy

To evaluate the accuracy of the fitting, the Modified Hausdorff Distance (*MHD*) is computed. The *MHD* is a metric for shape comparison that measures the degree of mismatch between two points sets. Therefore, it is useful to demonstrate the quality of a registration algorithm [112]. The *MHD* is defined as [132]:

$$\text{MHD}(P, F) = \frac{1}{N_p} \sum_{i=1}^{N_p} \min_{f_j \in F} |p_i, f_j|, \quad (4.9)$$

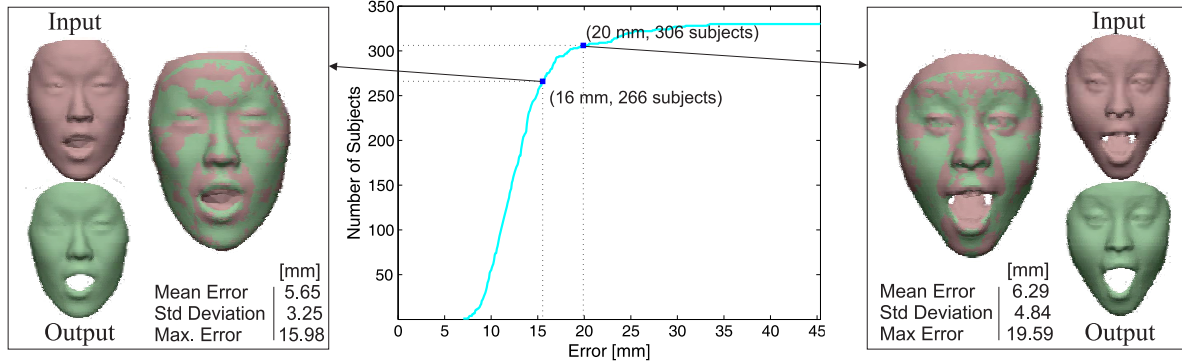


Figure 4.13: Cumulative distribution of the number of models where the error at all the landmark points not used for registration is below a threshold. Example of registration results (left and right). Error distribution (center).

where $|p_i, f_j|$ is the Euclidean distance between vertices of the template P and the vertices of the input model F , and N_p is the number of vertices of P . The *MHD* represents the average of the minimum Euclidean distance of the vertices of P , to which F is registered [112]. The values of the average, standard deviation and maximum of the *MHD* for the 332 tested models were 1.42mm, 0.56mm and 3.66mm, respectively. This shows that the proposed method has the ability of keeping the overall shape during the fitting.

In addition, the bottom row of Figure 4.17 shows the histograms and the false color visualization of the mean magnitude and standard deviation of the distance between the surfaces F and P computed over all 332 models. For every point p_r on P , its nearest neighbor $NN(p_r)$ on F is determined, the distance from p_r to the tangent plane of $NN(p_r)$ corresponds to distance between the surfaces. As most of the values of the distances are concentrated between 0 and 1mm, in order to improve the visualization, the color map was clamped to this range. Notice the variation in the lower lip and chin area, which are the regions where the surface is deformed most due to the facial expressions.

4.4.3.3 Visual Evaluation

Next, some examples that summarize the results of the expression and shape matching stages of the registration process are shown. The third column of Figure 4.14 shows examples of the expression fitting results for six different kinds of facial expression. In all cases, the expression of the mouth region of the input model is properly matched after linear blending. The fourth column of Figure 4.14 shows examples of the shape fitting results. The models are color-coded with respect to the signed distance from the input scan. Note that most points on the models are within 2mm of the scan. Furthermore, notice how the different expressions in the eyebrows are properly fitted. In order to visualize the quality of the correspondences,

a chess-board texture (with some facial features colored) was applied to the template model (see right of Figure 4.14). Results of the texture transferring show that in most of the face regions, the shape of the deformed template matches the shape of the input model.

The proposed method, which uses nearest neighbors to guide the deformation, the highest level of expression is the most difficult to register. All of the experiments outlined so far have considered this case. Figure 4.15 illustrates two examples of registering different levels of the same subject in the same expression. Note that the visual differences between the quality of the results are insignificant.

Finally, the running time of the here presented method is discussed. On a standard PC (2.4 GHz processor), the typical time to predict the set of landmarks for the initial alignment is about 5 seconds for rough alignment and about 176 seconds for the refinement of the position. The typical time for expression and shape fitting is about 6 seconds and 28 seconds, respectively.

4.4.4 Comparison to 3D Morphable Model

The registration results are compared to the results obtained using the commonly used 3D morphable model (3DMM) [117], which is a statistical model that encodes information about a set of training shapes. In order to use the morphable model for fitting, such a model has to be built. To this end, 50 subjects in highest expression levels (that were used for training for the landmark detection part) are used. Before computing the model, it is necessary to parameterize the training shapes. This is achieved by using manually placed marker positions that guide a non-rigid iterative closest point deformation. This step to parameterize a training set in a semi-automatic way is time-consuming. The morphable model was analyzed and it was found that retaining 50 principal components yields a compact, yet general model.

The morphable model was fitted to the data by first using the landmarks predicted by the proposed prediction method to rigidly align the scan to the model, and by subsequently minimizing the energy E_{data} defined in Equation 4.5 with respect to the model parameters.

Note that unlike 3DMM, the method introduced does not require a parameterized training set as a start. Furthermore, in the future, the proposed method could help building statistical models without the need to parameterize a training set in a semi-automatic way.

The comparison with the 3DMM is carried out in two ways. First, an evaluation of the obtained fitting results is provided. Since for the 3DMM, the amount of displacement during

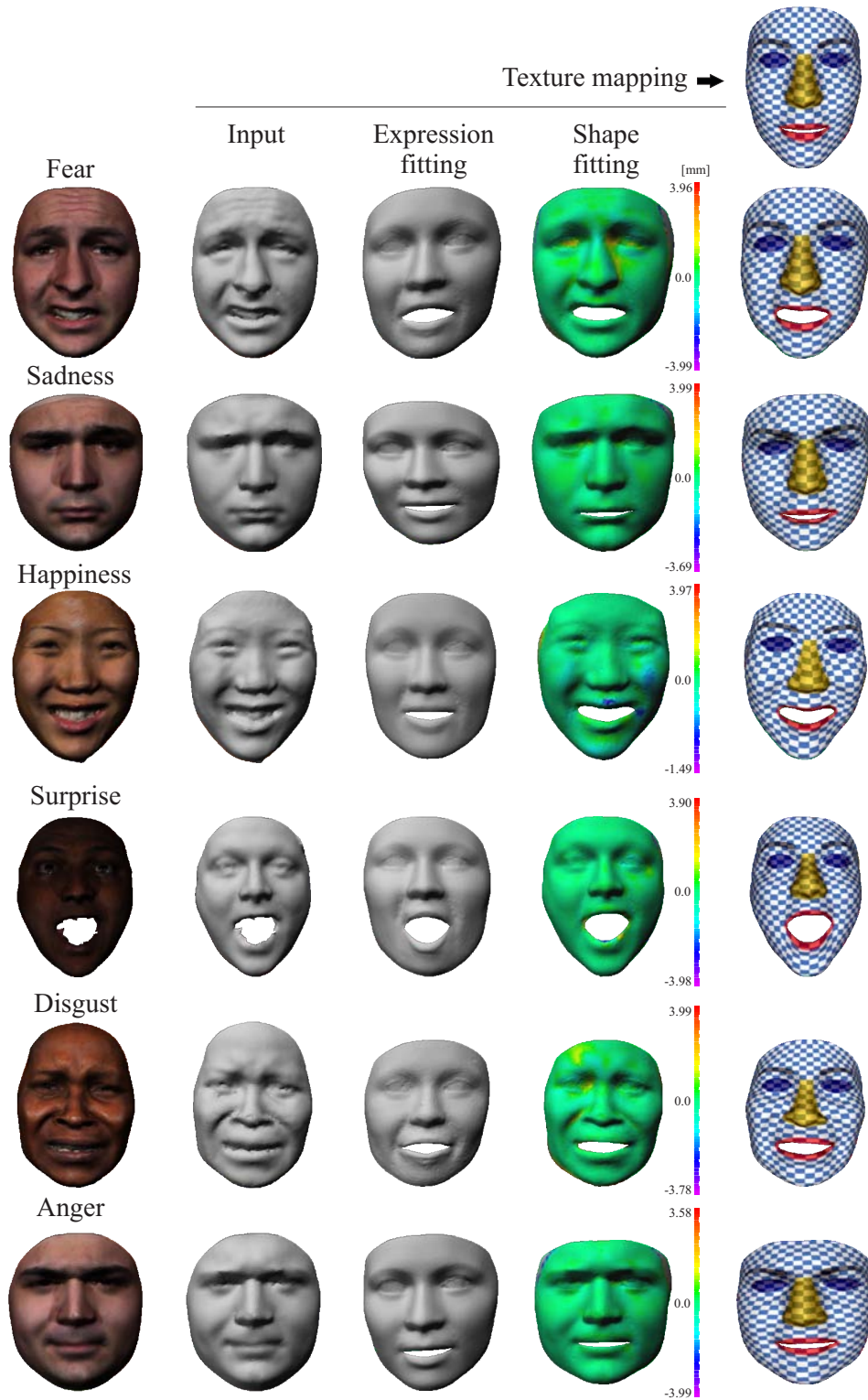


Figure 4.14: Examples of registration results. The input, fitted expression, error mapped, and texture mapped models are provided for each example.

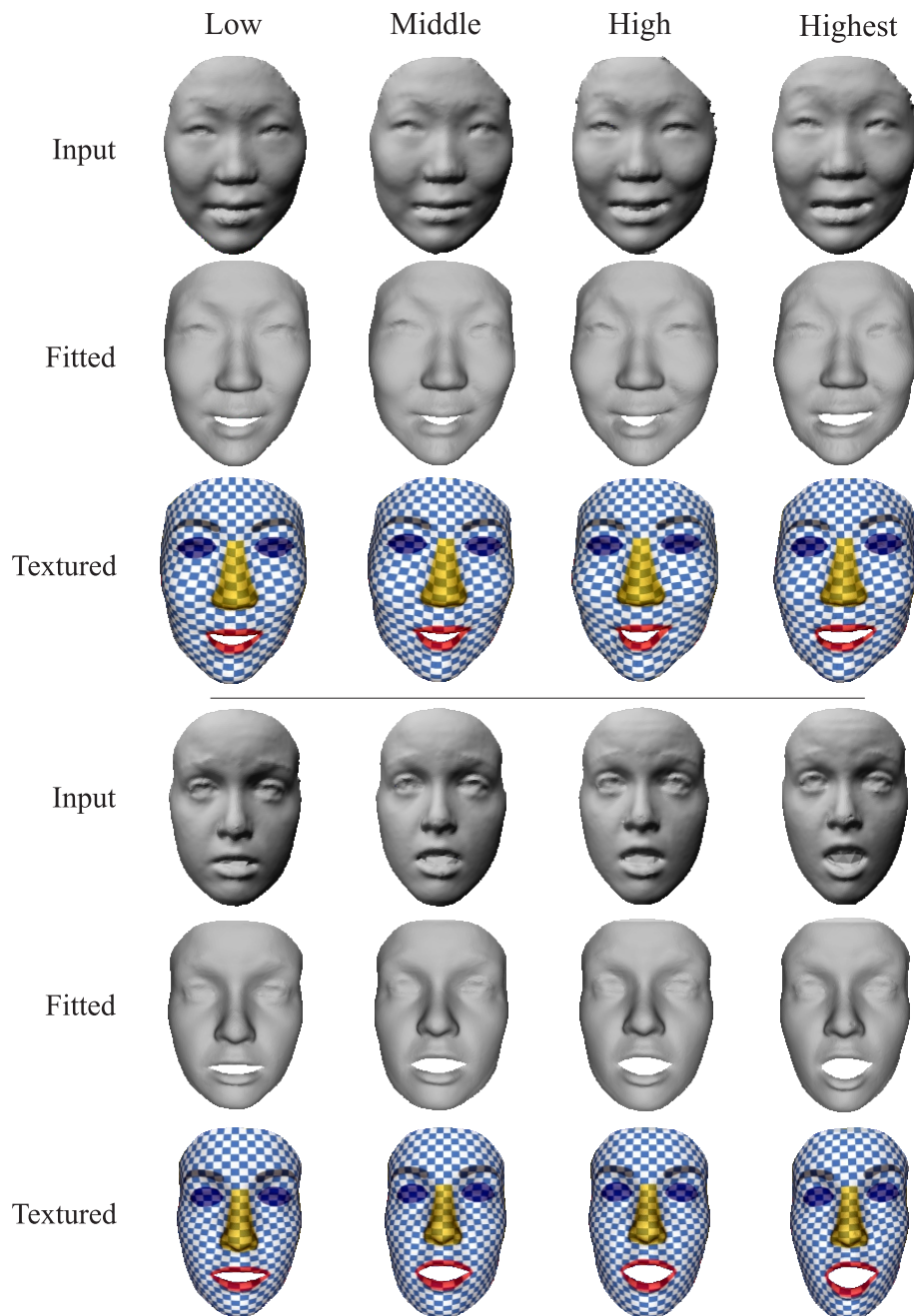


Figure 4.15: Examples of fitting to models of the same subject performing an expression in different levels. Fear (first three rows). Surprise (last three rows). For each example, first, second, and third rows are the input, output, and textured models.

the fitting is restricted to the one learned from the training data, the proposed method can fit local shape details more accurately than 3DMM, as can be seen in the four examples shown in Figure 4.16. As most of the values of the distances are concentrated between 0 and 1 mm, in order to improve the visualization, the color map was clamped to this range.

Figure 4.17 compares the histograms and the false color visualization of the mean magnitude and standard deviation of the distance between the surfaces F and P computed over all 332 models. Notice that while both methods lead to good fitting results overall, the proposed method has lower mean error in localized areas such as the tip of the nose or the eyebrows. The reason is that unlike the proposed method, 3DMM cannot fit to localized shape detail such as raised eyebrows, because 3DMM restricts the search space for the correspondence search to the variations observed in the training data.

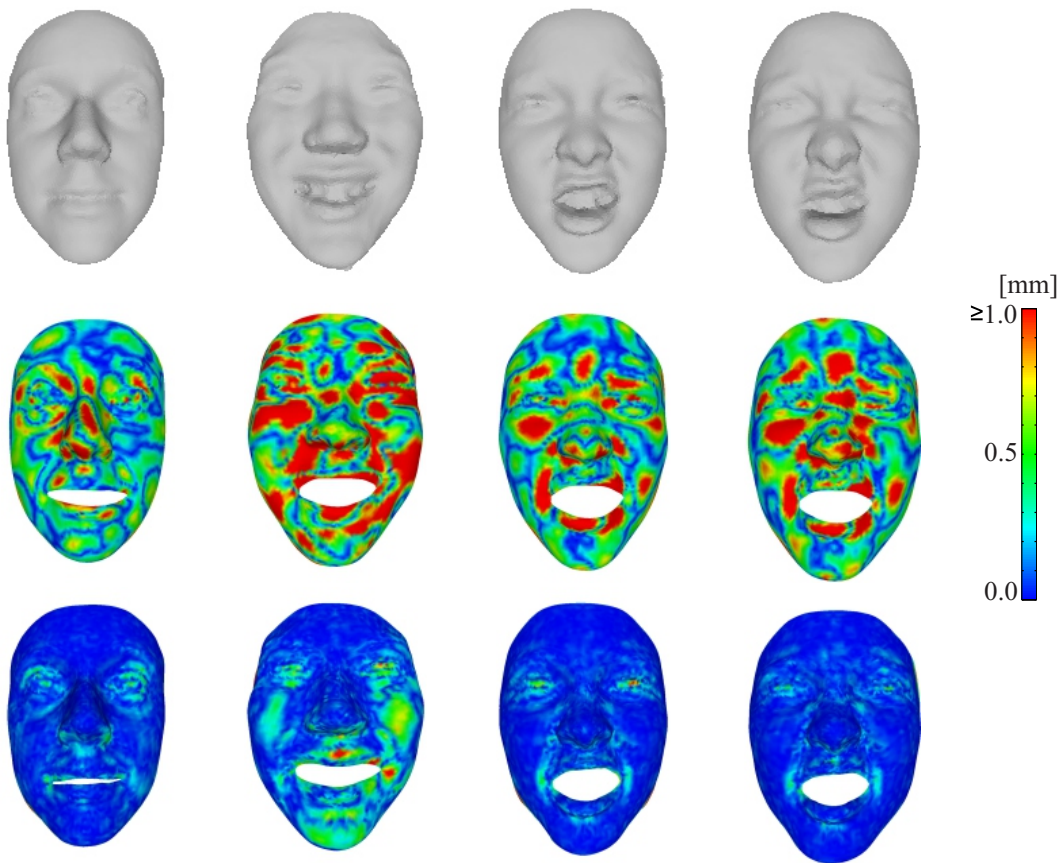


Figure 4.16: Comparison of shape distance of 3DMM fitting and the results of the proposed method. Top to bottom: input scan, 3DMM fitting, proposed method result.

Second, the results for the application of expression recognition are compared. Note that

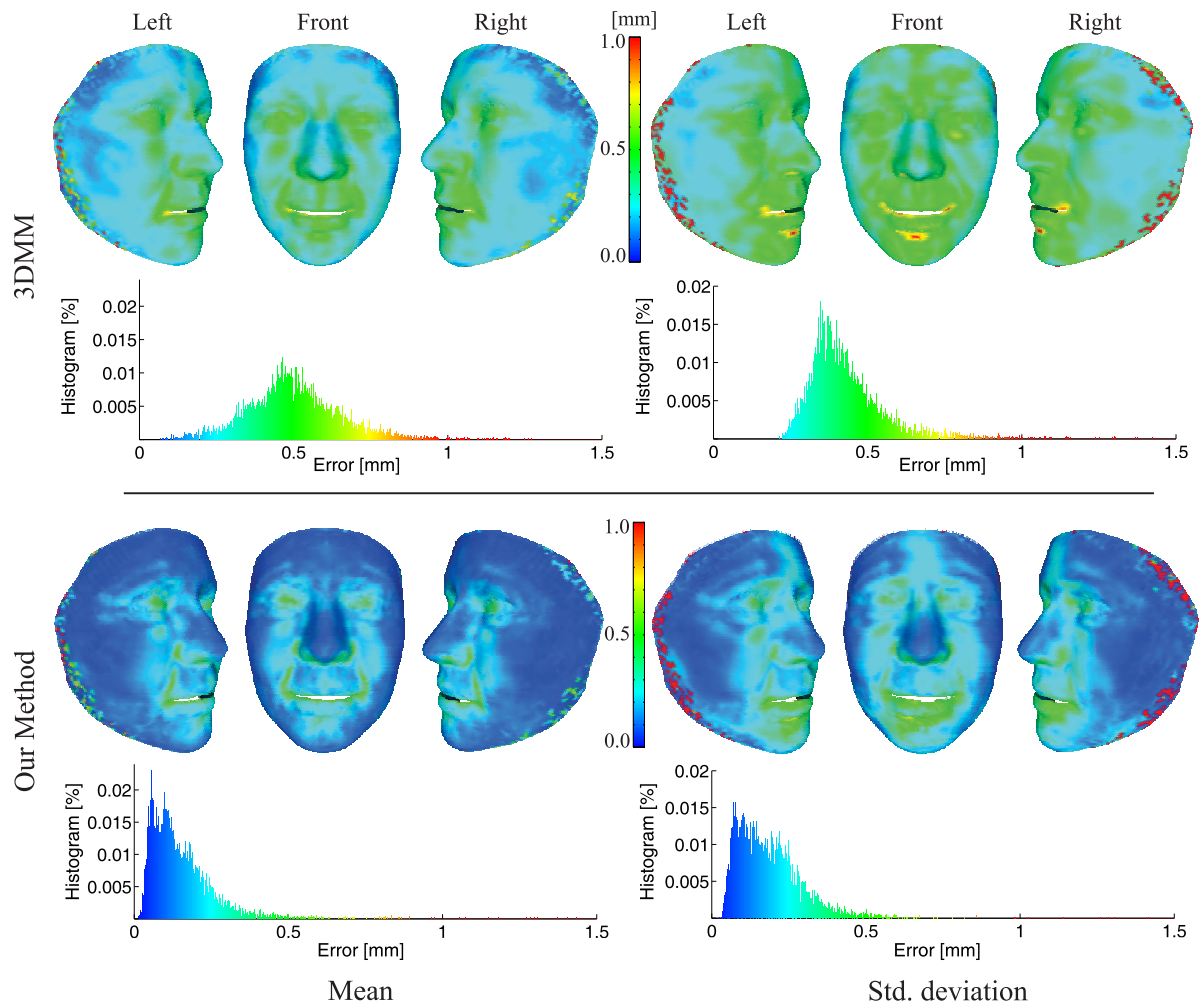


Figure 4.17: Distance between the surface of the template P and the surface of input model F. Histograms and the false color visualization (different views) of the magnitude of the mean and standard deviation of the distance.

this experiment is primarily intended to give a comparative evaluation between 3DMM and the proposed method, and not to introduce a new method for expression recognition.

In the following experiment, the aim was to recognize (the highest expression levels of) the expressions anger, happiness, and surprise. The features used for this experiment are based on anatomical facial landmarks and are computed following the methodology described in Rabiou et al. [133]. The feature selection, classification and evaluation is carried out using the pattern recognition tool developed by Duin et al. [134] with a support-vector classifier based on a 2nd order polynomial kernel. For training, features derived from the ground truth landmarks of the 50 subjects that were used for training for the landmark detection part are used. For testing, all fitting results (with expressions anger, happiness, or surprise) are used. The overall expression recognition rate using the models fitted with 3DMM is 61.1%, while the overall expression recognition rate using the models fitted using the proposed method is 77.7%. While neither of these results is competitive with human experts, who achieve a recognition rate of 98.1% [131], the experiment shows that the proposed method achieves significantly higher recognition rates than 3DMM. The reason is that the proposed method can fit better to local shape details, as discussed above.

4.4.5 Application

Finally, the fitting results are applied to build a statistical shape space that allows to explore the identity and expression variations of a database of faces separately. To this end, the registration results are used to compute a multilinear model [135]. The multilinear model expresses each face using one weight vector ω_i for identity and a second weight vector ω_e for expression. The expression of a subject can be modified by keeping ω_i fixed while modifying ω_e . Similarly, the identity can be modified while preserving the expression by keeping ω_e fixed while modifying ω_i . This is shown in Figure 4.18. Here, the faces shown in boxes are the registered faces of the database that were used to compute the multilinear model, and the remaining faces were generated by fixing ω_i to one identity and varying ω_e (top row) and by fixing ω_e to the weight of happiness and varying ω_i (bottom row). Note that in this way, realistic looking new expressions and identities can be generated, respectively.

4.4.6 Limitations

The method introduced in this chapter has some limitations. Sometimes, not all local areas of a face are fitted accurately. Most of the incorrect shape fitting occurs on the inner parts of the lips. As the input scans have information in the area of the teeth, which is not considered



Figure 4.18: Real models used to compute the multilinear model (shown in boxes) and synthetic models generated from the multilinear model.

in the template model, the algorithm converges to this region, thereby causing mismatches during the shape fitting. Figure 4.19 shows an example of the limitations in the shape fitting. Notice how the expression is matched correctly, but the corners of the mouth are not well located, which causes an incorrect fitting on the mouth and chin regions.

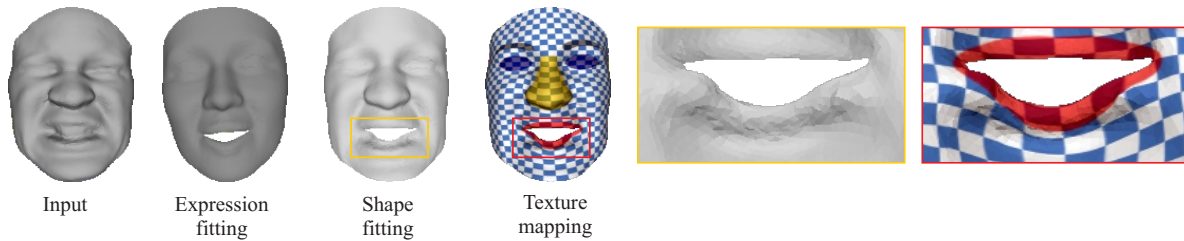


Figure 4.19: Incorrect shape fitting. The differences in topology of the input and template meshes cause incorrect expression and shape fitting.

Another limitation occurs for models with occluded parts. Figure 4.20 shows the result of the proposed point-to-point correspondence approach for a model of a subject where the mouth is occluded by a hand. In this case, the template is correctly fitted to areas not affected by the occlusion, but occluded regions cause unlikely face shapes.

4.5 Conclusions

This chapter presented a fully automatic method to compute dense point-to-point correspondences between a set of human face scans with varying expressions. The proposed approach proceeds by learning local shape descriptors and spatial relationships for a set of landmark

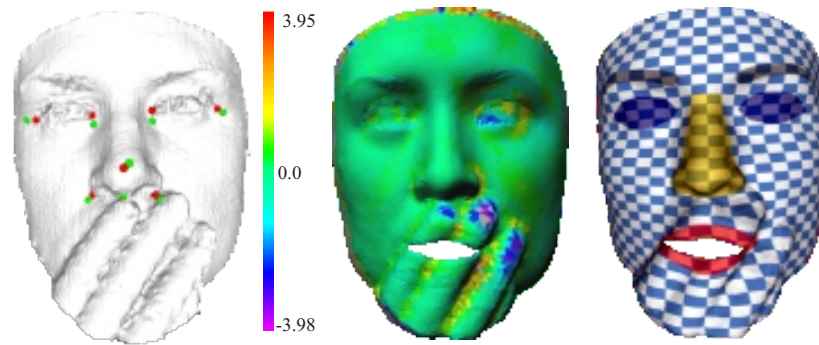


Figure 4.20: Challenging test scenario. Mapped error models correspond to the fitting result. Test was carried out over one model of the Bosphorus database.

points. For a new scan, the approach first predicts the landmark points by performing statistical inference on the learned model. The approach then fits a template to the scan in two stages. The first stage fits the expression of the template to the expression of the scan using the predicted landmark points. The second stage fits the shape of the template to the shape of the scan using a non-rigid iterative closest point technique. This approach was applied to 350 models of the BU-3DFE database, and evaluated the results both qualitatively and quantitatively. It was shown that for 94.9% of the models, the landmarks are predicted with an error below 30mm, and that for most of the models, a consistent correspondence is found. Furthermore, the algorithm was evaluated on a challenging case of a face with occlusion.

The failure cases of the algorithm are mostly caused by noisy data in the mouth area. For future work it would be necessary to design algorithms that can handle this challenging scenario. It is also of interest to test the algorithm on a large database of models with different types of occlusion, such as models wearing eyeglasses (e.g., models from Bosphorus database [136]) and on data acquired using different types of sensors. Finally, with the availability of inexpensive depth cameras, dynamic data is becoming increasingly important. Interesting future work includes to extend the proposed algorithm to compute correspondences of dynamic facial data in a fully automatic framework.

3D Anthropometry of the Face

Since this work is aiming to develop methods that help to overcome the disadvantages of the traditional anthropometry (which were stated in Chapter 1), this chapter presents the evaluation of the proposed non-rigid registration as a tool for the detection of several facial attributes corresponding to landmarks, anatomical regions, and the mouth contour.

So far, only the 3D information of the face geometry has been used as the input of the algorithms presented in this work. However, because the methods that combine different kinds of information achieve more accurate results [137], a landmark detection algorithm that combines 3D and texture information is used to improve the initialization of the non-rigid fitting procedure.

This chapter is organized as follows Section 5.1 reviews literature related to finding face attributes using multimodal approaches. Next, the pipeline of the algorithm for automatic face anthropometry is depicted in Section 5.2. The methodology used for the testing of the facial features detection algorithms is described in Section 5.3. Finally, the experimental results and conclusions are shown in Sections 5.4 and 5.5, respectively.

5.1 Related Work

Facial features extraction, whether realized on texture, depth images, point clouds, or 3D meshes, always will have to deal with some drawbacks, but each of the methods has its own advantages. Based on this, it has been demonstrated that combinations of different kinds of information produce better results than using each of them individually [137]. Some works which use more than one kind of information for facial feature detection are described below.

Xue and Ding [88] present a method to integrate range and intensity information. The algorithm is used to detect the face and three facial landmarks (tip of the nose and the two centers of the eyes). The detection of the 3D facial features is based on an extension of the *AdaBoost* method. The 2D analysis is performed using the Haar features described in [138]. H and K curvatures are used to describe the shape in 3D. ROC curves and Euclidean distance are the base to evaluate the performance of the face detector and facial landmarks localization methods, respectively.

Gong and Wang [79] introduce a face segmentation method from the texture of a 3D model. Twenty-eight regions containing ears, eyes, eyebrows, nose, and mouth are segmented by combining 2D texture and 3D geometry. Although, the segmented regions have a semantic meaning, the boundaries of the regions are rough. In addition, the accuracy of the method is determined by visual inspection.

Zou et al. [139] propose a model to extract features on range images based on a 3D deformable model. An extra optimization stage allows the extraction of features from the texture. The mapping of a 3D model to an input range image is made using a linear transformation, which simplifies and reduces the computational complexity. As the method described before, the evaluation is done by visual inspection.

Wang and Sung [140] introduce a method for facial feature detection on infrared images. Extraction of eyes and mouth corners is based on color analysis and edge detection using the *SUSAN* operator. The nose tip is detected on the disparity map. The head pose is estimated from the location of previously extracted features and based on the camera calibration. Using stereo matching, a 3D model of the face enables the location of more features. The experiments show that the method works well to correct the pose. However, no quantitative results of the accuracy of the detected facial features were presented because the ground truth was not available.

Guo et al. [115] propose a method where 17 facial landmarks are automatically annotated using a combination of a PCA-based feature recognition and a 3D-to-2D data transformation. The detected landmarks are the guidance to establish dense anatomical correspondence between facial images using a thin-plate spline protocol. Despite the methods are shown to be accurate for facial feature detection, the consistency of the registration is not evaluated quantitatively.

The method proposed in this chapter allows the location of a set of 39 landmarks, the se-

mantic segmentation of the face into 26 regions, and the extraction of the three dimensional contour of the mouth. The algorithm is divided in two stages. In the first part, a multimodal approach is used to detect a set of 12 landmarks. These landmarks are the base of a non-rigid fitting procedure that matches the shape of a target 3D model. Figure 5.1 shows an overview of the method. Unlike some of the works described in this section, all the methods are evaluated quantitatively. In addition, the potential of these methods for automatic face anthropometry is also studied.

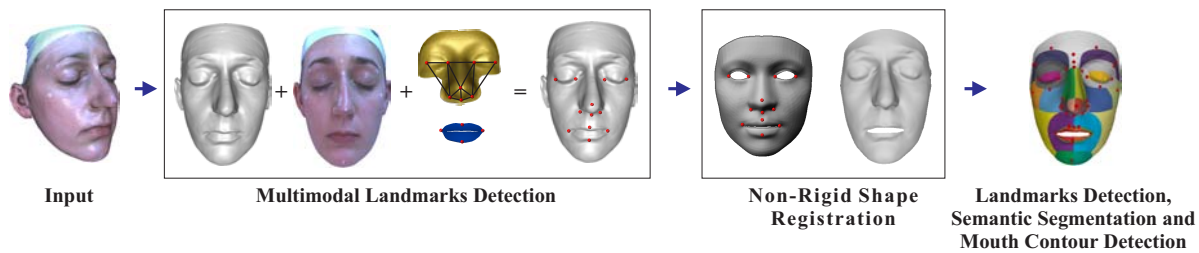


Figure 5.1: Overview of the fully automatic face anthropometry approach

5.2 Non-contact Face Anthropometry

This section describes the algorithm used to extract different facial features that are the base of an automatic face anthropometry procedure. This algorithm is a complement of the method presented in the previous chapter. Unlike the landmark detection procedure described in Section 4.2, the texture is used to increase the number (from eight to twelve) and accuracy of the landmark's location used for the initial alignment of the face template. Aiming in the improvement of the initial estimation of the face height, a set of four new landmarks are located in the mouth area. Before locating these landmarks, a combination of rigid and non-rigid ICP procedures is used to refine the location of the predicted landmarks in the nose and eyes areas.

5.2.1 Nose and Eyes Landmarks Detection

Approaches that extract facial features from 2D images rely on the proper frontal orientation of the subject with respect to the camera position. In the case of textured 3D models, one of the first steps is to orientate the model in a way that the appearance can be properly projected to 2D, maximizing the visibility of the facial features. As the eyes, nose and mouth are the features of interest, detecting some points in these areas, is enough to find a proper point of view. Here, the pose of the 3D model is corrected based on the locations of eight landmarks

(four in the nose and four in the eyes). The approach to find these landmarks proceeds as follows:

1. An initial estimation v_e of the landmarks location is found using the method depicted in Figure 4.6.
2. Using Procrustes analysis, a transformation T_e that best aligns the point set v_{ep} from a template P_e with the point set v_e is found. Using T_e , the template P_e is rigidly aligned to the input scan F .
3. In order to improve the initial alignment, a rigid ICP procedure is used to minimize the distance between P_e and F . This step aims for correction of the rotation and translation of P_e .
4. The eyes and nose landmarks are located on the final positions of v_{ep} after the non-rigid alignment of P_e to F .

The second row of Figure 5.2 illustrates the procedure used for the detection of the landmarks in the nose and eyes region.

5.2.2 Mouth Landmarks Detection

Once the nose and eyes landmarks have been detected, their locations are used to correct the pose of F and to restrict the area where the mouth landmarks will be located. The mouth landmarks detection is performed as follows:

1. Based on the location of six of the located landmarks, namely the inner and outer corners of the eyes and both subalare points, the pose of F is corrected. The reference plane P_R is set to be the best-fit plane to the six landmarks by least squares. Afterwards, the texture of F is projected to P_R and image I_t is obtained.
2. I_t is transformed to the YCbCr color space.
3. In order to filter the red zones, the RGB and YCbCr color information is combined using

$$I_{enh} = (I_R + I_{Cr}) - (I_G + I_B + I_{Cb}), \quad (5.1)$$

where I_R , I_G and I_B are the RGB components of I_t ; I_{Cr} and I_{Cb} are the red and blue chrominance components of the YCbCr image, respectively.

4. A binary image I_b is obtained selecting the pixels of I_{enh} that have an intensity value larger than $\tau \max(I_{enh})$. Where $\max(I_{enh})$ corresponds to the maximum intensity value of I_{enh} and τ is a weight value. In the experiments performed here, τ was set to 0.1.
5. The white pixels of I_b that are located above the subnasal point are discarded. Using morphological analysis the biggest region of I_b is selected as the mouth region. The right and left mouth corners are selected as the points with the minimum and maximum x coordinates within the mouth region, respectively. The y coordinates of the upper and lower points of the mouth are extracted from the points with the minimum and maximum y coordinates within the mouth region. The x coordinates correspond to the x coordinate of the mass center of the region of the mouth's bounding box.
6. An initial estimation v_m of the mouth landmarks is founded by re-projecting the above-located points to 3D.
7. Using Procrustes analysis, a transformation T_m that best aligns the point set v_{mp} from a template P_m with the point set v_m is found. Using T_m , the template P_m is rigidly aligned to the input scan F .
8. The mouth landmarks are located on the final positions of v_{mp} after the non-rigid alignment of P_m to F .

The mouth region and mouth landmarks detection procedures are depicted in the third and fourth row of Figure 5.2.

The twelve detected landmarks are used for the initial alignment and registration of the input scan F to a face template P following the procedures described in Section 4.3.

5.3 Experimental Setup

The above-described methods were tested as a tool for automatic face anthropometry. Four different kinds of analysis were performed First, the ability of the method to detect facial landmarks and its accuracy; second, the reliability of a set of anthropometric dimensions measured from the detected landmarks; third, the quality of a semantic region segmentation of the face; last, a quantitative evaluation of the extracted external contour of the lips. The next sections describes the database, the set of landmarks, the set of dimensions, the semantic segmentation, a template of the external lips contour, and the methods used to carry out a quantitative evaluation for each of the analysis that were described above.

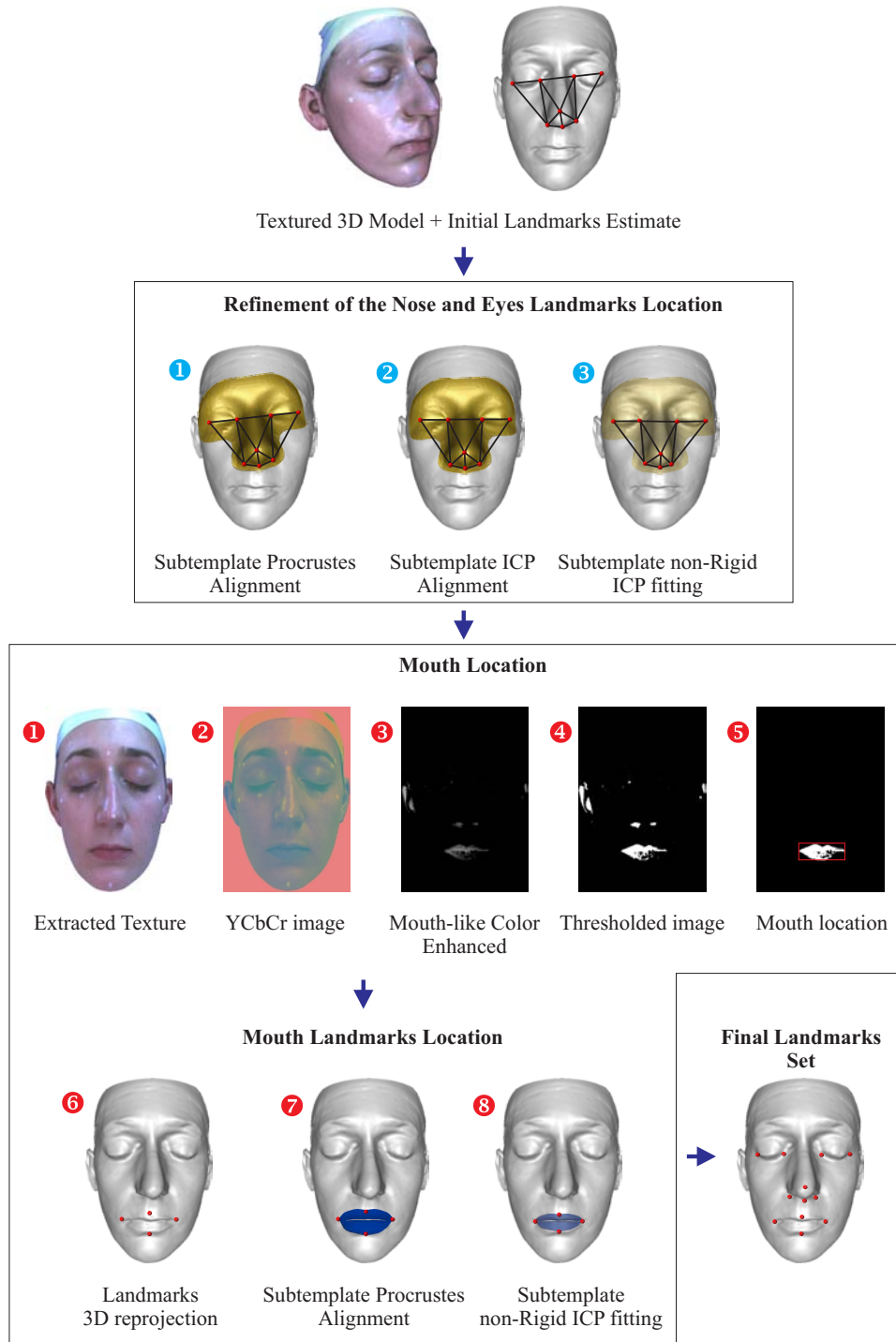


Figure 5.2: Framework of the proposed multimodal landmark detection method.

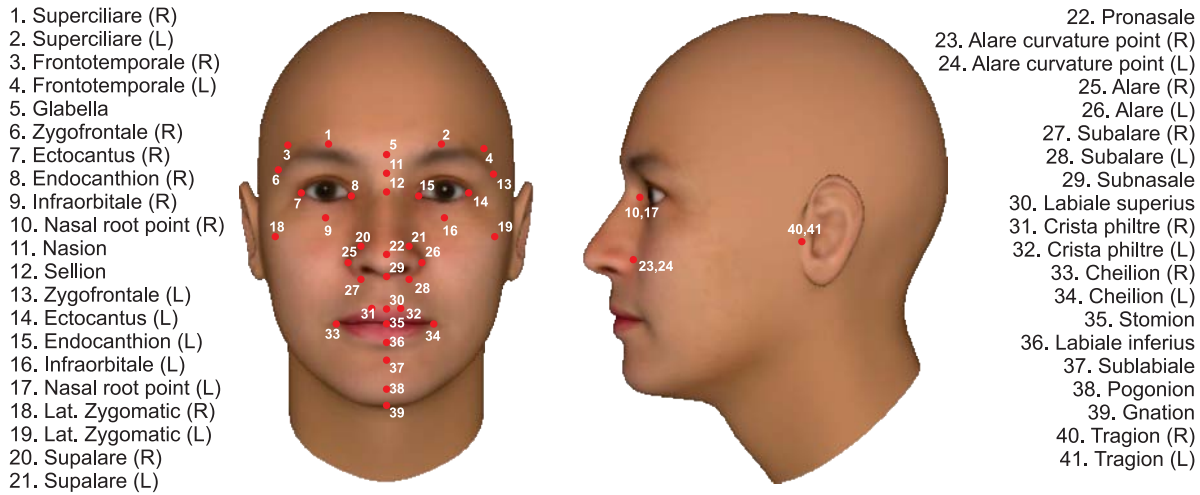


Figure 5.3: Set of landmarks

5.3.1 Data Collection

To test the system an in-house database (DB_H) was used. The database consists of 20 subjects (10 female and 10 male) between 18 and 35 years old. The subjects were asked to wear a silicone cap to prevent that the hair interferes during the further 3D data acquisition process. An expert physically located fourteen landmarks, which coincide with the critical bone structures, on the face surface of each subject. The 3D geometry and texture data were captured with a Minolta Vivid 9i. Four scans of the same subject from different angles were necessary to capture the whole face surface properly. The final 3D face model of each subject was obtained after the registration of the four scans using a commercial software. Once the 3D face model was obtained, another twenty-seven landmarks were manually located. Figure 5.3 shows the full set of landmarks. The landmarks labeled as 3, 4, 5, 6, 9, 12, 13, 16, 18, 19, 22, 38, 40 and 41 are the anatomical landmarks located before the 3D scanning. The set of landmarks was selected based on the methodology described in the work by Luximon *et al.* [141].

Figures 5.4a and 5.4b show an example of one model from the DB_H database. The white dots correspond to the locations of the anatomical landmarks. An expert manually located the anatomical and physical landmarks onto the surface of the 3D models. As the ratio of a white dot is about 5mm, the location of a landmark corresponds to the mass center of four points that were located on the edge of the each white dot (see Figure 5.4c). This procedure helps to reduce the variability of the final location of the point. These points are available for all the models of the DB_H database and correspond to the ground truth for the anthropometric measurements analyzed in this chapter. Figures 5.4d and 5.4e show an example of a model with the ground truth landmarks.

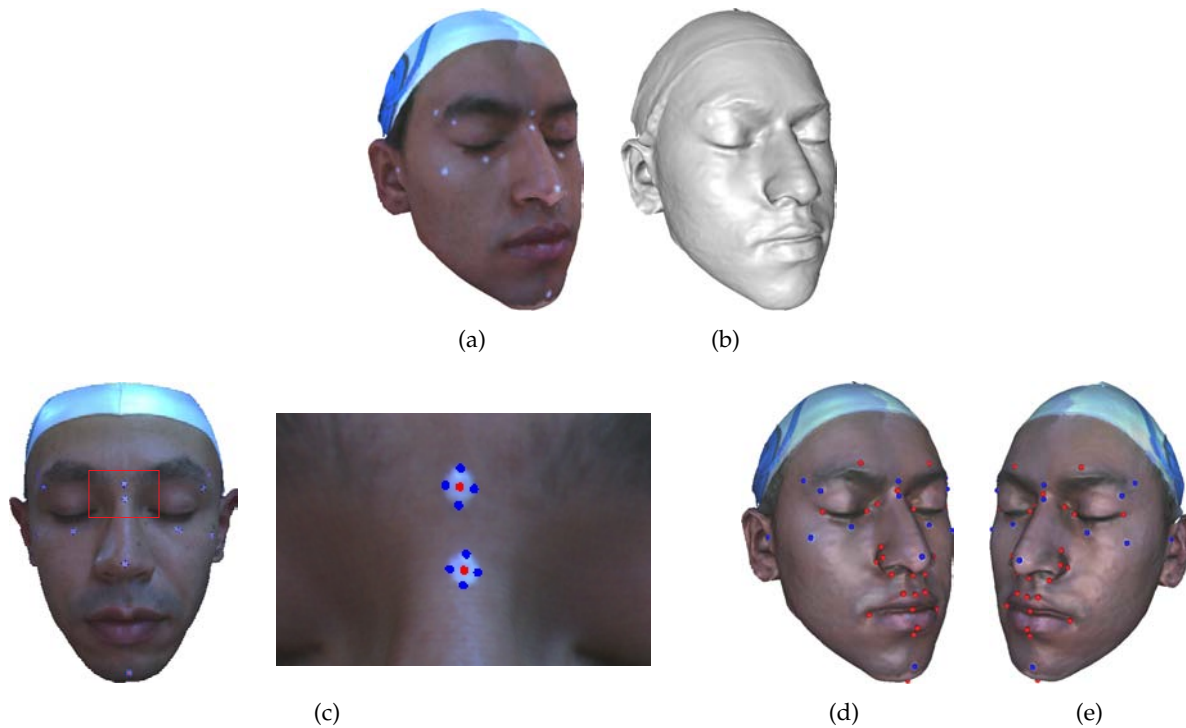


Figure 5.4: Example of model from the DB_H database. (a) Textured and (b) Geometry. (c) Points (blue) used to compute the absolute (red) position of the landmarks. (d) and (e) show the location of the anatomical (blue) and physical (red) landmarks computed from the manually annotated points.

In addition, a plaster cast of the face of a subject was made and scanned (see Figure 5.5). As the plaster casting procedure is highly invasive, only one subject was willing to collaborate. A set of anthropometric dimensions were obtained from both the real model and the 3D scan model. The aim of this experiment is to evaluate the discrepancy between the magnitude of the real and digital dimensions. The set of anthropometric dimensions and the results of the comparison are shown in the next section.

5.3.2 Anthropometric Dimensions

There are several anthropometric dimensions that can be measured by using the above-described facial landmarks as a reference. The set of anthropometric dimensions used in this chapter was selected because they are the base for the computation of anthropometric ratios, which are useful in different contexts such as design of respirators [142], cráneo-facial surgery planning [17], anthropometric 3D face recognition [143], among others. Figure 5.6 shows the set of dimensions that were selected. Appendix A contains the descriptions of the

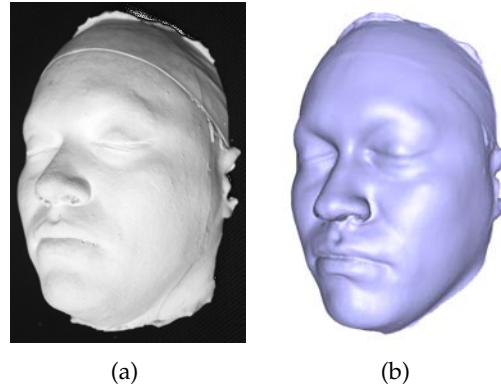


Figure 5.5: Plaster model. (a) Picture and (b) 3D Geometry.

anthropometric dimensions.

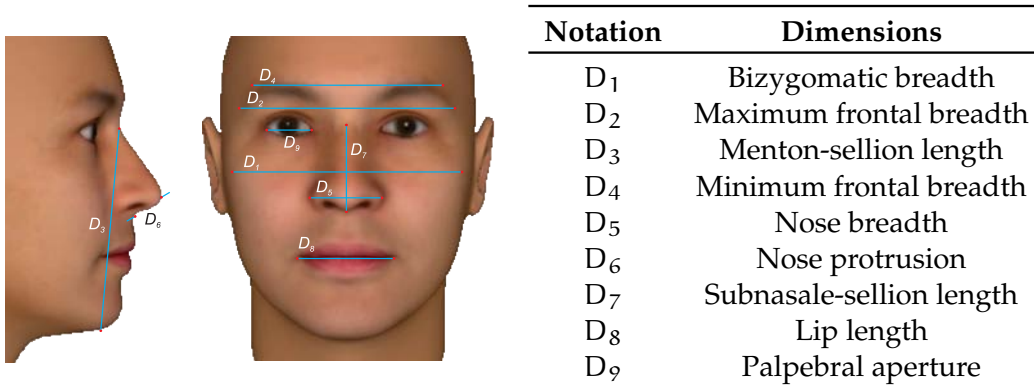


Figure 5.6: Set of dimensions

In order to verify the reliability of the acquisition device, a comparison between the magnitude of the dimensions measured on the plaster cast (described in the previous section) and the dimensions obtained from its 3D model, was carried out. An expert made the physical measurements following the procedures described in Appendix A. The measurements obtained from the 3D model were computed based on the 3D coordinates of the landmarks that were located by the expert. The measurement of the dimensions D_1 , D_2 , and D_4 , was carried out with a spreading caliper, which has a resolution of 0.5mm. The rest of the dimensions were measured with a sliding caliper, which has a resolution of 0.05mm.

The results of the comparison are shown in Table 5.1. The discrepancy of each measure corresponds to the absolute difference between the physical measure and its digital counterpart. The overall value of discrepancy obtained was $0.45 \pm 0.1637\text{mm}$, which is within the

Dimension	Plaster Cast [mm]	3D model [mm]	Discrepancy [mm]/[%]
Bizygomatic breadth	143.50	143.89	0.39/0.27
Maximum frontal breadth	122.0	122.68	0.68/0.55
Menton-sellion length	125.25	125.48	0.23/0.18
Minimum frontal breadth	112.50	113.17	0.67/0.59
Nose breadth	42.95	42.49	0.46/1.07
Nose protrusion	21.05	20.56	0.49/2.32
Subnasale-sellion length	50.80	51.41	0.61/1.2
Lip length	61.95	61.61	0.34/0.54
Palpebral aperture (right)	38.55	38.86	0.31/0.8
Palpebral aperture (left)	38.80	39.08	0.28/0.72

Table 5.1: Discrepancy between real and 3D measurements.

resolution of the measurement devices ($\approx 0.5\text{mm}$). However, a smaller value of discrepancy is not enough to conclude that a measure is reliable. Notice that the discrepancy values (in percentage) for the dimensions measured on the nose region were the highest. As the nose is one of the face regions where the geometry varies drastically in a small area, this leads to difficulties in the acquisition due to self-occlusions. Therefore, the nose area is one of the regions where more geometry corrections (hole filling, smoothing, etc.) have to be made during the registration, which increases the error of the surface reconstruction.

5.3.3 Semantic Segmentation

The convergence of the classical 3D segmentation algorithms is mostly guided by rules, which are inspired in the observations of the values of a certain feature or a set of certain features computed from the surface. Therefore, the results of the segmentation do not necessarily have a semantic meaning. In the case of faces, this situation becomes evident since the face can be divided into many anatomical regions, each one with a well-defined description. Despite the proper anatomical definitions, the variations of the geometry on the boundaries of the region are very small, which complicates the definition of rules that can be incorporated into a system for the automatic segmentation of faces into semantic regions. The non-rigid registration of surfaces offers an alternative to this problem since the semantic information can be attached to the template used for fitting. Thus, the semantic segmentation is intrinsic to the registration. In this chapter, a face template with 26 anatomical regions is used to ob-

tain a semantic segmentation of the face. Figure 5.7 shows the template with the anatomical regions highlighted in different colors.

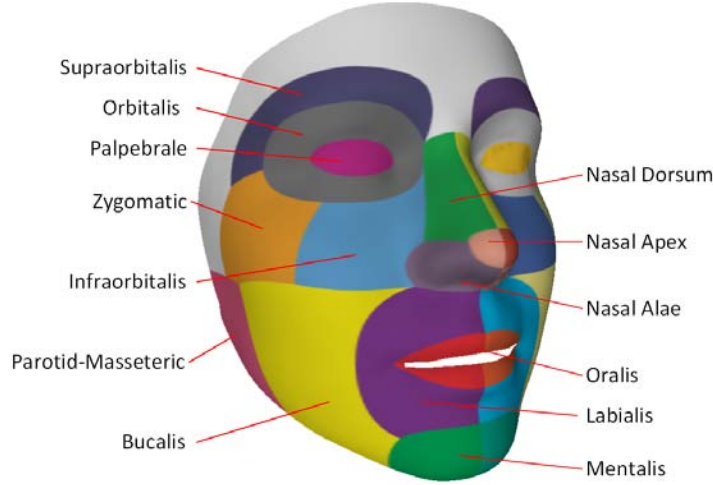


Figure 5.7: Face template used for the semantic segmentation of the face.

For the quantitative analysis of the quality of the segmentation, the four metrics described below were used. The description of the metrics was taken from [144]:

Cut Discrepancy. It summates the distances from points along the cuts (segment boundaries) in the computed segmentation to the closest cuts in the ground truth segmentation, and vice-versa. This method is boundary-based thus, it measures the distances between cuts. Assuming C_1 and C_2 are sets of all points on the segment boundaries of segmentations S_1 and S_2 , respectively, and $d_G(p_1, p_2)$ measures the geodesic distance between two points on a mesh, then the geodesic distance from a point $p_1 \in C_1$ to a set of cuts C_2 is defined as:

$$d_G(p_1, C_2) = \min\{d_G(p_1, p_2), \forall p_2 \in C_2\},$$

and the Directional Cut Discrepancy, $DCD(S_1 \Rightarrow S_2)$, of S_1 with respect to S_2 is defined as the mean of the distribution of $d_G(p_1, C_2)$ for all points $p_1 \in C_1$:

$$DCD(S_1 \Rightarrow S_2) = \text{mean}\{d_G(p_1, C_2), \forall p_1 \in C_1\}.$$

The Cut Discrepancy, $CD(S_1, S_2)$, is the mean of the directional functions in both directions, divided by the average Euclidean distance from a point on the surface to the

centroid of the mesh (avgRadius), in order to ensure symmetry of the metric and to avoid effects due to scale:

$$CD(S_1, S_2) = \frac{DCD(S_1 \Rightarrow S_2) + DCD(S_2 \Rightarrow S_1)}{\text{avgRadius}}.$$

Hamming Distance. Given two mesh segmentation $S_1 = \{S_1^1, S_1^2, \dots, S_1^m\}$ and $S_2 = \{S_2^1, S_2^2, \dots, S_2^n\}$ with m and n segments, respectively, the Directional Hamming Distance is defined as

$$D_H(S_1 \Rightarrow S_2) = \sum_i \left\| S_2^i \setminus S_1^{i_t} \right\|,$$

where \setminus is the set difference operator, $\|x\|$ is a measure for set x (e.g., the size of set x , or the total area of all faces in a face set), and $i_t = \max_k \|S_2^i \cap S_1^k\|$. The general idea is to find a best corresponding segment in S_1 for each segment in S_2 , and sum up the set difference. If S_2 is regarded as the ground truth, then Directional Hamming Distance can be used to define the missing rate R_m and false alarm rate R_f as follows:

$$R_m(S_1, S_2) = \frac{D_H(S_1 \Rightarrow S_2)}{\|S\|},$$

$$R_f(S_1, S_2) = \frac{D_H(S_2 \Rightarrow S_1)}{\|S\|},$$

where $\|S\|$ is the total surface area of the polygonal model. The Hamming Distance is simply defined as the average of missing rate and false alarm rate:

$$HD(S_1, S_2) = \frac{1}{2} (R_m(S_1, S_2) + R_f(S_1, S_2)).$$

Since $R_m(S_1, S_2) = R_f(S_2, S_1)$, the Hamming Distance is symmetric.

Rand Index. Measures the likelihood that a pair of faces are either in the same segment in two segmentations, or in different segments in both segmentations. Let S_1 and S_2 two segmentations, S_i^1 and S_i^2 as the segment identities of face i in S_1 and S_2 , and N as the number of faces in the polygonal mesh. $C_{ij} = 1$ iff $S_i^1 = S_j^1$, and $P_{ij} = 1$ iff $S_i^2 = S_j^2$, then the Rand Index is defined as:

$$RI(S_1, S_2) = \binom{2}{N}^{-1} \sum_{i,j,i < j} [C_{ij}P_{ij} + (1 - C_{ij})(1 - P_{ij})].$$

$C_{ij}P_{ij} = 1$ indicates that face i and j have the same identities in both S_1 and S_2 . Also, $(1 - C_{ij})(1 - P_{ij}) = 1$ indicates that face i and j have different identities in both S_1 and S_2 . Thus, $RI(S_1, S_2)$ tells the proportion of face pairs that agree or disagree jointly on their segment group identities in segmentations S_1 and S_2 . In this case, the reported value corresponds to $1 - RI(S_1, S_2)$, in order to be consistent with the other metrics that report dissimilarities rather than similarities (the lower the number, the better the segmentation result).

Consistency Error Denoting S_1 and S_2 as two segmentation results for a model, t_i as a mesh face, \setminus as the set difference operator, and $\|x\|$ as a measure for set x (as in the case of the Hamming Distance), $R(S, f_i)$ as the segment (a set of connected faces) in segmentation S that contains face f_i , and n as the number of faces in the polygonal model, the local refinement error is defined as:

$$E(S_1, S_2, f_i) = \frac{\|R(S_1, f_i) \setminus R(S_2, f_i)\|}{\|R(S_1, f_i)\|}.$$

Given the refinement error for each face, two metrics are defined for the entire 3D mesh, Global Consistency Error (GCE) and Local Consistency Error (LCE), as follows:

$$GCE(S_1, S_2) = \frac{1}{n} \min \left\{ \sum_i E(S_1, S_2, f_i), \sum_i E(S_2, S_1, f_i) \right\},$$

$$LCE(S_1, S_2) = \frac{1}{n} \sum_i \min \{E(S_1, S_2, f_i), E(S_2, S_1, f_i)\}.$$

Both GCE and LCE are symmetric. The difference between them is that GCE forces all local refinements to be in the same direction, while LCE allows refinement in different directions in different parts of the 3D model. As a result, $GCE(S_1, S_2) \leq LCE(S_1, S_2)$.

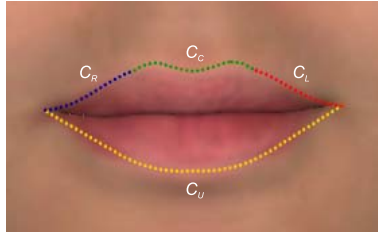
For more details about the definition of the metrics see [144] and the references within. The computation of the metrics was carried out using a publicly available software¹. The evaluation of the semantic segmentation proposed in this chapter will be presented in Section 5.4.2.

5.3.4 Mouth Contour Extraction

A template of the external contour of the lips was designed. The template is composed by four parametric functions. The order of these functions was chosen in a way that the resulting

¹<http://segeval.cs.princeton.edu/>

curves match the shape of the contour of the lips as close as possible. Figure 5.8 shows the four curves that were used to define the external contour of the mouth.



Notation	Contour
C_C	Cupid's arc
C_R	Upper lip external contour (right)
C_L	Upper lip external contour (left)
C_U	Lower lip external contour

Figure 5.8: Curves used to define the external contour of the lips.

Lower Lip. To characterize the lower lip one curve was used. It was observed that a third order polynomial was not able to fit the whole contour of the lower lip (see Figure 5.9a). Therefore, a fourth order polynomial was chosen as the curve to characterize the lower lip shape (see Figure 5.9b). This also helps to catch shape variations due to asymmetry.

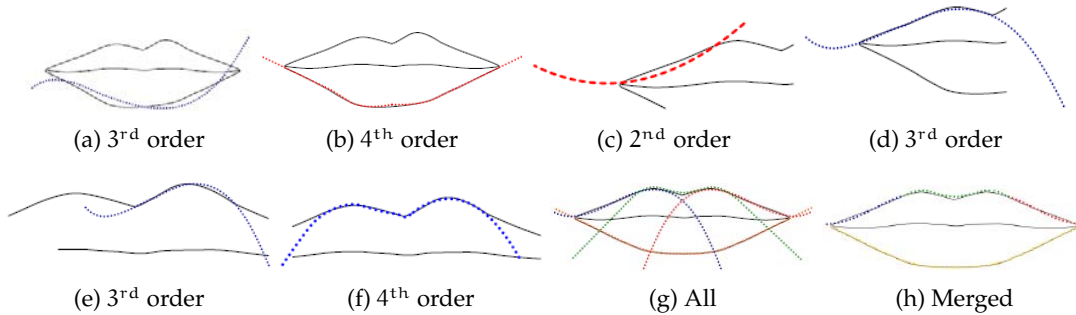


Figure 5.9: Parametric functions used to define a external lips contour template.

Upper Lip. Three functions were used to characterize the shape of the upper lip. First, as a second order polynomial did not offer a proper match of the shape of the C_R and C_L upper lip contour sides (see Figure 5.9c), a third order polynomial was used (see Figure 5.9d). Last, a third and fourth order polynomial were tested to characterize C_C . It was observed that a fourth order polynomial allows a proper matching of the shape of C_C (see Figure 5.9e and Figure 5.9f).

The coefficients of the polynomials are obtained by Singular Value Decomposition of the over-constrained system of equations using the contour points, obtained from the face template after the shape fitting stage. To preserve the continuity among the functions, common points from the adjacent curves are used (see Figure 5.9g and Figure 5.9h).

5.4 Experimental Results

This section is divided into four parts. First, the results of the automatic landmark detection algorithm are introduced in this chapter. The second part is the comparison between the ground truth anthropometric dimensions and the ones obtained automatically. Then, a third section shows the results of the quantitative evaluation of the semantic segmentation of the face. Last, but not least follows the evaluation of the quality of the external lips contour extraction algorithm.

5.4.1 Automatic Landmarks Location Accuracy

The accuracy of the automatic landmark location was computed following the same procedures described in Section 4.4.2. The only difference is that the detection rate was computed by counting the percentage of test models where the landmark \hat{l}_i was predicted with an error below 5mm ($T < 5$). Table 5.2 shows the results of the evaluation of the automatic location for the twelve detected landmarks.

Notation	Landmark	Mean \pm Std [mm]	Max. [mm]	T < 5 [%]
l_1	Right Endocanthion	2,72 \pm 1,87	5,55	90
l_2	Right Ectocantus	4,34 \pm 1,92	6,76	75
l_3	Left Endocanthion	2,47 \pm 1,08	5,77	90
l_4	Left Ectocantus	4,16 \pm 1,58	6,81	75
l_5	Right Subalare	2,80 \pm 1,84	5,72	90
l_6	Left Subalare	2,96 \pm 1,78	5,58	90
l_7	Pronasale	1,65 \pm 1,73	4,35	100
l_8	Subnasale	1,94 \pm 1,36	5,26	95
l_9	Right Cheilion	2,53 \pm 1,45	5,64	90
l_{10}	Left Cheilion	2,14 \pm 1,86	5,44	90
l_{11}	Labiale Superius	2,35 \pm 1,77	5,65	95
l_{12}	Labiale Inferius	2,57 \pm 1,83	6,04	95

Table 5.2: Error of landmark detection used the multimodal approach. $T < 5$ corresponds to the detection rates with a tolerance of 5mm.

The Pronasale and Subnasale landmarks are detected with the lowest errors and both Ectocantus are predicted with the highest error. As the scaling factors for the initial alignment are derived from the non-rigid fitting of the nose template, where the horizontal scale

is less influenced, this could cause that the initial estimation of the Ectocantus is less accurate than for the other points. For the rest of the points the mean error is within the same range ($< 3\text{mm}$). Notice that for most of the landmarks, the detection rate $T < 5$ is 90% or higher, especially for the Pronasale and both Labiale landmarks, which are the points with the highest detection rate. Figure 5.10 shows some examples of the landmark detection results.

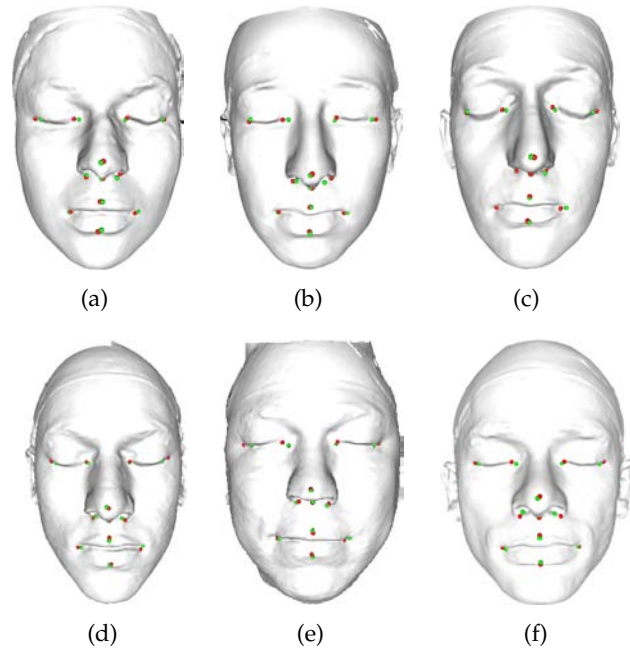


Figure 5.10: Examples of the landmark detection results. Red and green spheres correspond to the manually placed and predicted landmarks, respectively. (a)-(c): female subjects; (d)-(f): male subjects.

5.4.2 Segmentation

In order to verify the influence of the set of points (excluding and including the landmarks on the mouth area) used for the initial alignment in the quality of the segmentation and also to see if the expression fitting step offers advantages for the automatic segmentation, four tests were performed:

1. Test T_1 uses only eight landmarks (landmarks l_1 to l_8 in Table 5.2) for the affine alignment and the non-rigid registration without expression fitting.
2. Test T_2 uses eight landmarks (same set used for the test T_1) for the affine alignment and the non-rigid registration with expression fitting.
3. Test T_3 uses twelve landmarks (landmarks l_1 to l_{12} in Table 5.2) for the affine alignment and the non-rigid registration without expression fitting.

4. Test T_4 using twelve landmarks (same set used for the test T_3) for the affine alignment and the non-rigid registration with expression fitting.

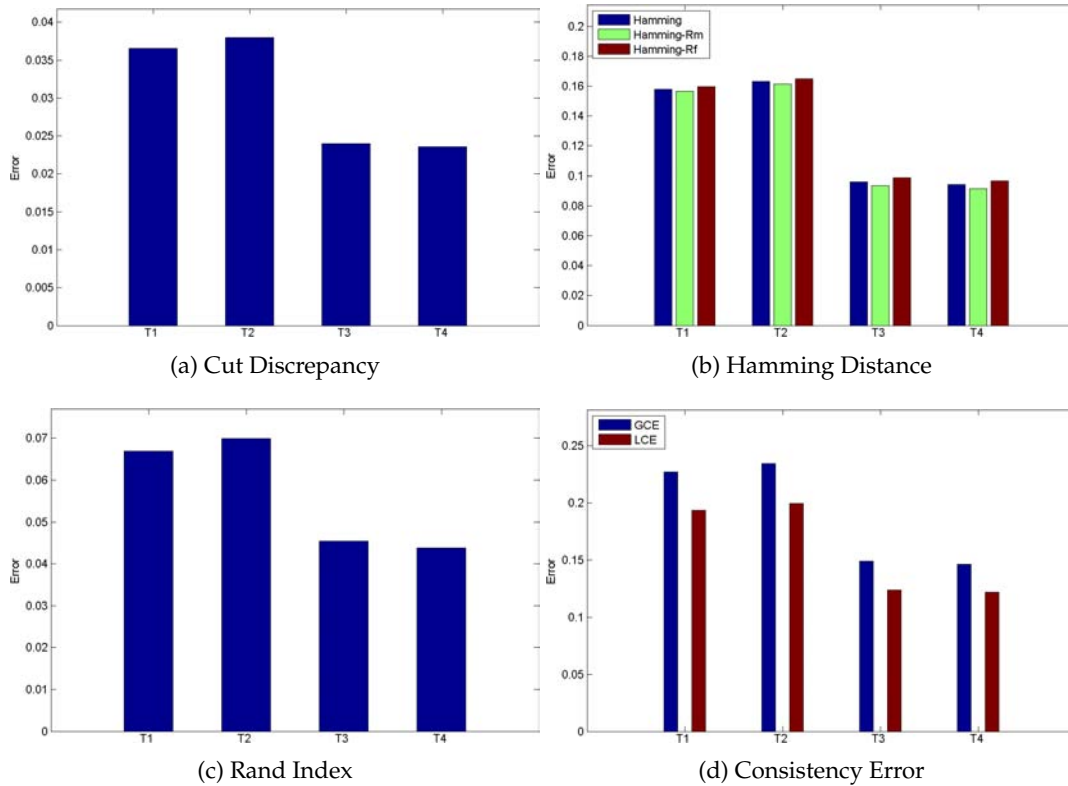


Figure 5.11: Evaluation of the segmentation for the tests T_1 to T_4 .

The evaluation of the segmentation quality was carried out based on the principles described in [144]. Figure 5.11 shows the results of the evaluation of the four previously described tests. It is not easy to draw a conclusion about the magnitude of the metrics. As was mentioned before, the lower the value of the metric the best. Based on this, the best results were obtained using the T_4 configuration, which shows that the mouth landmarks are important for the affine alignment, resulting on a better initial estimation of the vertical scaling of the template. In addition, the results show that during the expression fitting step, the subtle variations from the rest pose are caught, which contributes with a better initial estimate of the face shape, resulting in a more consistent segmentation of the regions.

Figure 5.12 shows examples of the segmentation results for each of the four configurations. Notice the inconsistency of the results when the landmarks of the mouth are not considered, causing an inappropriate fitting mostly in the mouth and chin regions. Also, notice that it is quite difficult to perceive the difference between the T_3 and T_4 segmentation results.

This situation highlights the importance and efficacy of the quantitative evaluation.

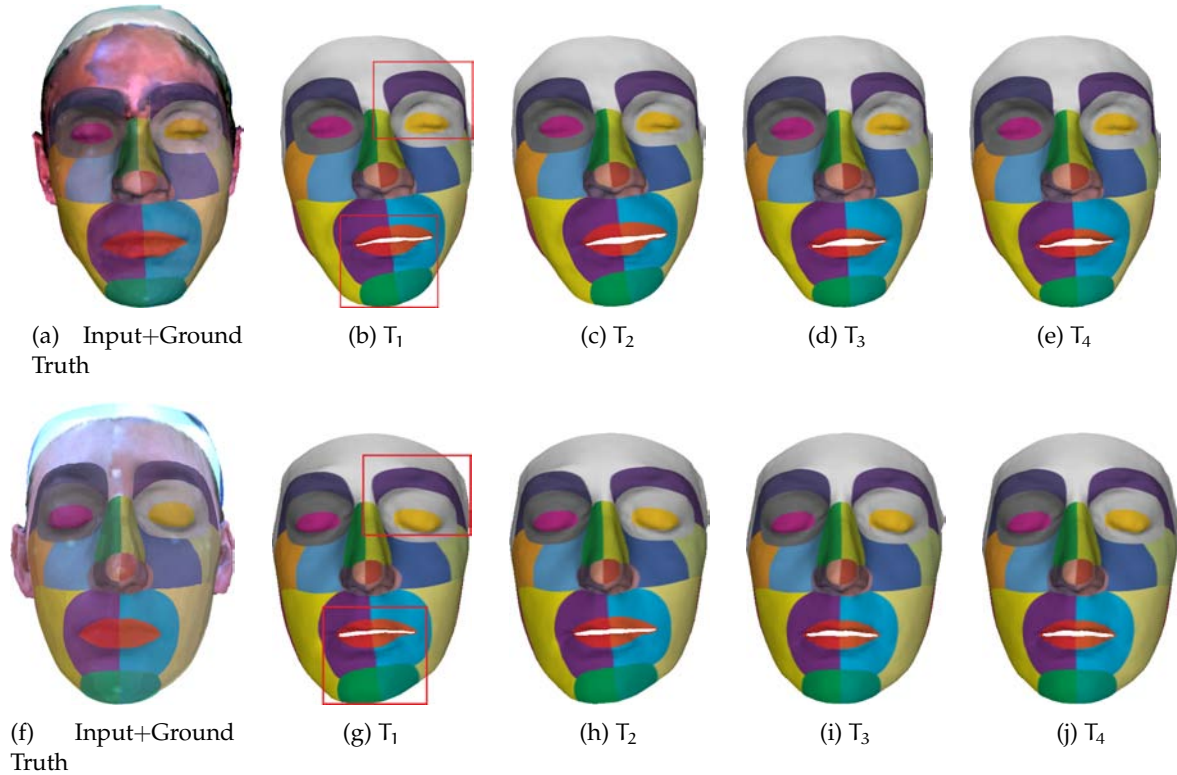
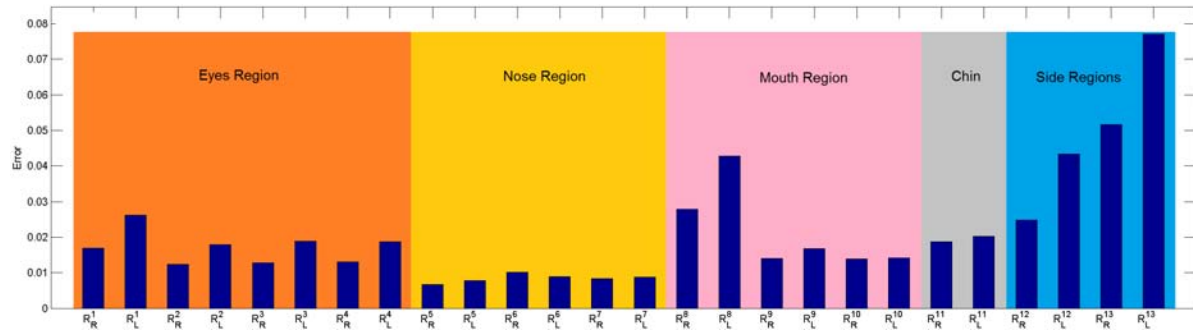


Figure 5.12: Results of the semantic segmentation.

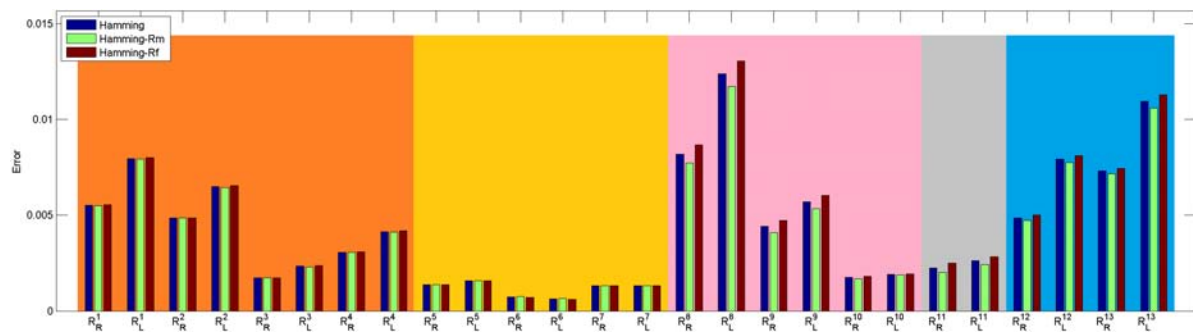
Next, the segmentation results obtained using the T_4 configuration were used for a more detailed analysis. In this case, the segmentation metrics were computed individually for each region of each side of the face. Table 5.3 lists the set of regions used for the region evaluation.

Figure 5.13 shows the results of the evaluation of the segmentation for each region. Notice that the magnitude of the metrics is slightly different for the right and left side of the face. Regarding to the regions, the best results were obtained in the nose region. Similar results were obtained for the regions: Palpebrale, Mentalis, and Oralis. On the other hand, the worst results were obtained for the regions: Bucalis, Zygomatic, and Parotid-Masseteric; which are the regions where the anatomical boundaries are not well defined.

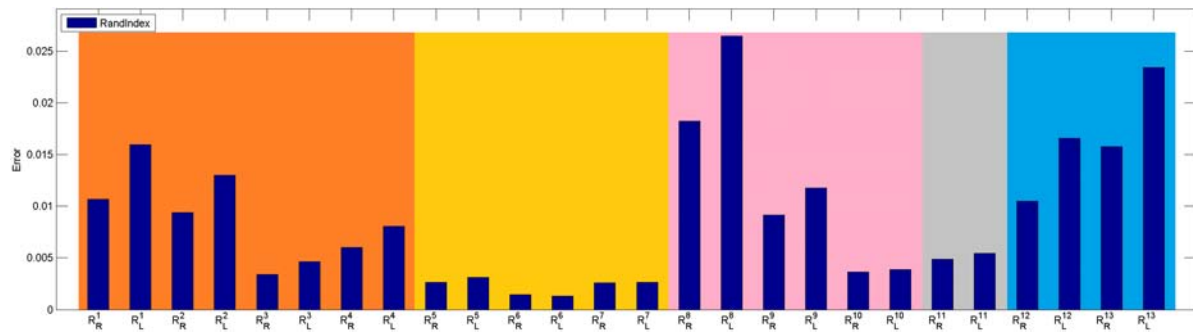
Last, the error in the location of landmark points that are not considered for the alignment is computed. The error corresponds to the Euclidean distance between a manually placed point and its corresponding location after registration. The set of points considered for the evaluation corresponds to the remainder points from the landmarks set shown in Figure 5.3.



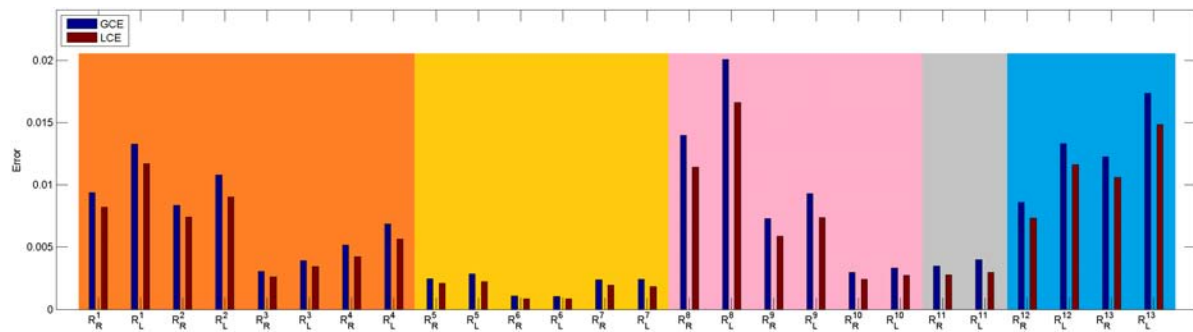
(a) Cut Discrepancy



(b) Hamming Distance



(c) Rand Index



(d) Consistency Error

Figure 5.13: Evaluation of the segmentation for each region individually.

Region	Notation (Right/Left)
Supraorbitalis	R_R^1/R_L^1
Orbitalis	R_R^2/R_L^2
Palpebrale	R_R^3/R_L^3
Infraorbitalis	R_R^4/R_L^4
Nasal Dorsum	R_R^5/R_L^5
Nasal Apex	R_R^6/R_L^6
Nasal Alae	R_R^7/R_L^7
Bucalis	R_R^8/R_L^8
Labialis	R_R^9/R_L^9
Oralis	R_R^{10}/R_L^{10}
Mentalis	R_R^{11}/R_L^{11}
Zygomatic	R_R^{12}/R_L^{12}
Parotid-Masseteric	R_R^{13}/R_L^{13}

Table 5.3: Set of evaluated regions.

Table 5.4 shows the results of the evaluation. Notice that the points located on the nose base and mouth were located with a low error. The points with the highest location error were the points located on the eyebrows and cheeks. Similar to the segmentation results, this situation is caused by the small geometry variations around the areas where the above-mentioned points are located. Regarding to the side of the face, no considerable differences were observed between the landmarks location errors of the left and right side.

5.4.3 Dimensions Magnitude

Once the automatic location of landmarks has been evaluated, the next step is to analyze how reliable the magnitude of the dimensions computed using the detected landmarks, is. Table 5.5 shows the values of the mean, standard deviation, and maximum of the absolute difference of automatic magnitudes with respect to the ground truth ones. As the dimensions are derived from the above-analyzed landmarks, the best result were obtained for the dimensions of the nose and mouth region. The results for the dimensions derived from the landmarks of the eyebrows and cheeks regions were the worst.

5.4.4 Automatic Mouth Contour Location

In this section, the results of the automatic detection of the external contour of the lips are shown. First, in order to evaluate the consistency of the shape of the detected contour, a visual inspection of the results was made. Figure 5.14 shows examples of the detected lips

Landmark	Mean \pm Std [mm]	Max. [mm]	Mean \pm Std [mm]	Max. [mm]
Superciliare	5,88 \pm 2,47	7,28	5,29 \pm 2,72	7,65
Frontotemporale	5,54 \pm 3,21	7,75	5,14 \pm 3,18	7,96
Glabella	3,29 \pm 1,62	7,07	N/A	N/A
Zygofrontale	3,65 \pm 2,58	7,32	3,65 \pm 2,16	7,69
Infraorbitale	2,69 \pm 2,41	6,02	3,87 \pm 2,70	6,70
Nasal root point	3,41 \pm 1,75	6,05	3,29 \pm 1,59	6,67
Nasion	2,15 \pm 0,95	4,15	N.A.	N.A.
Sellion	2,02 \pm 0,80	3,52	N.A.	N.A.
Zygomatic	4,37 \pm 1,13	6,77	4,14 \pm 1,39	6,39
Supalare	2,46 \pm 1,43	4,95	3,12 \pm 1,88	5,16
Alare curvature point	2,30 \pm 1,14	4,75	2,10 \pm 1,37	5,24
Alare	2,38 \pm 1,03	4,30	2,44 \pm 1,18	5,31
Christa philtre	1,89 \pm 1,27	5,56	1,93 \pm 1,53	6,53
Sublabiale	2,06 \pm 1,12	4,60	N.A.	N.A.
Pogonion	4,24 \pm 1,79	6,89	N.A.	N.A.
Gnation	3,97 \pm 1,70	7,02	N.A.	N.A.

Table 5.4: Error at landmarks points not used for the initial alignment. Gray column corresponds to the landmarks located at the left side of the face and the other column to the landmarks located at the right. N.A. means that the point is located on the center line of the face.

Dimension	Mean \pm Std [mm]	Max [mm]
Bizygomatic breadth	4,76 \pm 2,32	8,95
Maximum frontal breadth	4,52 \pm 2,26	7,56
Menton-sellion length	3,13 \pm 2,12	7,30
Minimum frontal breadth	5,96 \pm 3,63	9,66
Nose breadth	1,26 \pm 0,98	3,27
Nose protrusion	1,04 \pm 0,81	3,01
Subnasale-sellion length	2,99 \pm 1,39	6,31
Lip length	2,05 \pm 1,35	6,26
Palpebrale aperture (right)	2,37 \pm 1,98	5,84
Palpebrale aperture (left)	2,06 \pm 1,65	5,97

Table 5.5: Error of the magnitude of the dimensions.

contours, red and green points correspond to the ground truth and the automatically detected contours, respectively. Notice that the shape of the contours is consistent and non-aberrant (local and global) shapes were obtained.

Second, a quantitative analysis of the automatic lips contours location was performed. The error corresponds to the Euclidean distance between the points of the ground truth contour and their corresponding contour points of the fitted template. Figure 5.15 shows the mean and standard error for each of the contour points. For a better visualization, the error was clamped to 3mm. Most of the points were located with an error between 0.8mm and 2.25mm, just a few points around the right corner of the lips were located out of this range.

Last, the surface measurements that can be derived from the detected contour were evaluated. Table 5.6 shows the values of the mean, standard deviation, and maximum of the absolute difference of automatic measurements with respect to the ground truth ones. The measurements with the biggest error were the contour of the lower lip and the cupid's arc, which are derived from points lying on a surface where the geometry variations are the highest.

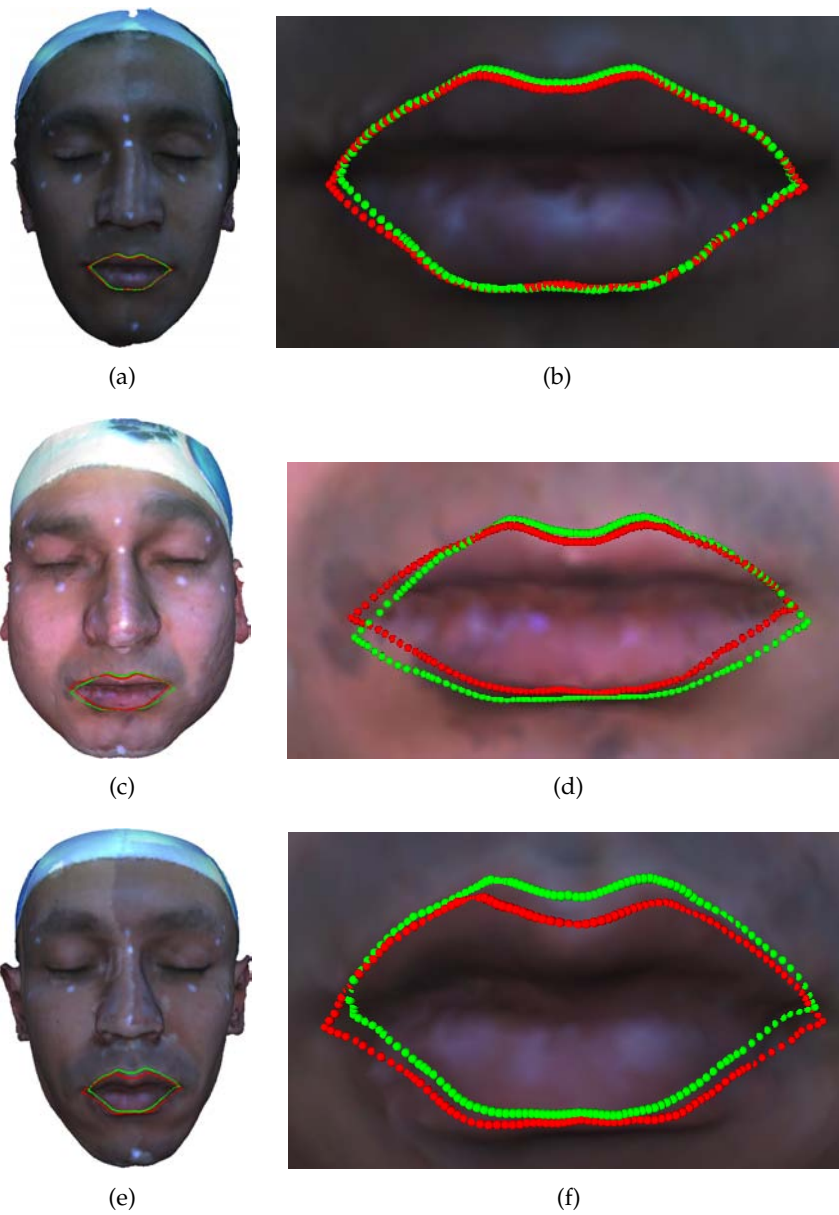


Figure 5.14: Mouth contour detection results. Red: Ground truth. Green: Detected.

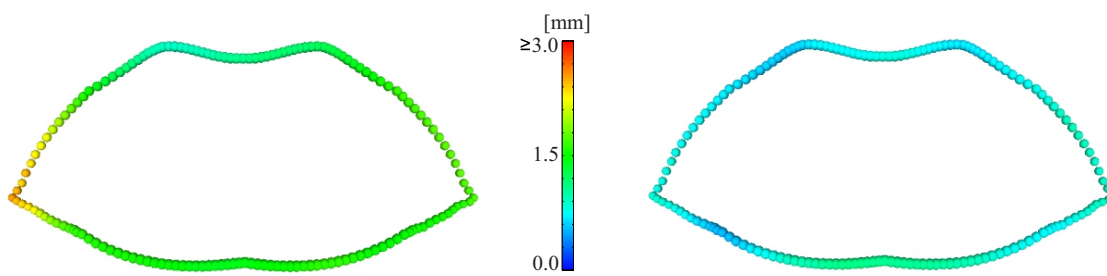


Figure 5.15: Lips contour location error. Mean (left) and Standard deviation (right).

Dimension	Mean \pm Std [mm]	Max [mm]
Cupid's arc	2,48 \pm 1,48	5,99
Upper lip external contour (right)	1,40 \pm 0,95	3,68
Upper lip external contour (left)	1,30 \pm 0,83	3,53
Lower lip external contour	2,82 \pm 1,61	5,85

Table 5.6: Error of the surface measurements of the mouth contour.

5.5 Conclusions

In this chapter, the potential of a dense point-to-point fully automatic correspondence method as a tool for face anthropometry was studied. The shape fitting method is initialized with a set of 12 landmarks located on the eyes, nose, and mouth region. These landmarks are located using a multimodal approach. After the shape fitting procedure the location of 39 landmarks, 26 anatomical regions, and the lips contour are extracted. In addition, 10 anthropometric dimensions were derived from the detected landmarks; and four surface measurements were carried out over the lips contour.

Regarding to the landmark location, the set of points considered in this chapter included osseous and soft landmarks. In this case, the soft landmarks on the nose and mouth regions were more accurately detected than the osseous landmarks. The detection of osseous landmarks is a challenging task because the geometry variation around their locations is very subtle, thus, there are no convergence cues for the shape-matching algorithm. Consequently, the anthropometric dimensions derived from the detected osseous landmarks were the ones with the highest error.

The consistency of the registration was established through a quantitative analysis of the face region segmentation. The nose and lips were the regions with the highest consistency. On the other hand, the structures located around the cheeks and the eyebrows were the regions where the consistency was the lowest. The initialization that includes mouth landmarks in combination with the expression and shape fitting procedures was shown to be the best configuration for the registration.

The shape of the extracted lips contours was correct (no degenerate shapes were obtained) and close to the target ground truth contours. This leads to a proper measurement of the surface length of the lips contours. This result shows the potential of the method to be incor-

porated to the planning and monitoring of lips reconstruction surgery in the event of malformation or trauma.

Conclusion and Further Work

In this thesis, several methods for automatic detection of facial features were presented. With these methods it is possible to locate a set of 39 landmarks, to segment a face in 26 anatomical regions, and to extract the three dimensional contour of the mouth. This is achieved by means of the non-rigid registration of a raw scan to a face template. Thus, a dense point-to-point correspondence of surfaces across a database also can be obtained.

The non-rigid registration method relies on the proper location of a small set of landmarks, which serve as the initial position of the face template. Two different methods for the detection of the initial landmarks were proposed. The first method encodes 3D information of the location and spatial relationships and predicts the location of the landmarks using statistical inference. The second method performs better than the first one, but requires the information of the shape and the appearance of the surface. It was shown that the quality of the final registration highly depends on the proper location of the initial set of landmarks. Likewise, the inclusion of some landmarks in the mouth area improves the consistency of the final registration.

To cope with the high variation in the cheeks, chin and mouth areas, an expression-fitting step was added to the process. It was demonstrated that with a small set of blendshapes, it is possible to obtain a proper point-to point registration in presence of expressions such as surprise, happiness, disgust, sadness, anger and fear. In addition, through a quantitative evaluation of the consistency of the segmentation, it was proven that for the segmentation of neutral pose models, the facial expression fitting step positively affects the performance of the algorithm.

The proposed methods were tested for dense point-to-point registration, facial expression recognition, expression synthesis, and face anthropometry; reliable results were obtained in

all the tests carried out. This shows the versatility of the methods and their potential to be extended in order to be used in a large field of disciplines.

The automatic detection of facial features is a task that always will offer many ways for future research. One of the main drawbacks of the template registration methods is the requirement of a proper initialization. Hence, the development of more accurate landmarks detection methods is of great importance.

Another advancement would be the extension of the methods to work in presence of noise and occlusions. In this sense, the statistical models are an alternative to face this challenge. Building a statistical model requires an appropriate registration of the training data, thus, the methods proposed in this thesis could help building statistical models without the need to parameterize a training set. For instance, in the comparison of statistical models presented in [16], the registration proposed in this work was used to generate the parameterized database in which the statistical models were based. These models shows robustness to different types of severe occlusion. The models are publicly available at [145].

Particularly in terms of the set of detected features, these could be increased in order to cover the whole surface of the head (e.g., to analyze the ear morphology) and the whole body. Additionally, as there are databases available that include the dynamics of facial expression [146], it should be considered to extend the registration method to enable the inclusion of this kind of data.



Anthropometric Dimensions Description

The descriptions of the set of linear dimensions considered in this work are listed below.

Bizygomatic breadth The maximum horizontal breadth of the face between the zygomatic arches is measured with a spreading caliper. The subject is sitting, looking straight ahead and with the teeth together (lightly occluded). Only enough pressure to ensure that the caliper tips are on the zygomatic arches is exerted.

Maximum frontal breadth The straight-line distance between the right and left Zygofrontale landmarks at the upper margin of each bony eye socket is measured with a spreading caliper. The subject sits looking straight ahead. Only enough pressure to ensure that the caliper tips are on the landmarks is exerted.

Menton-sellion length The distance in the midsagittal plane between the Menton landmark at the bottom of the chin and the Sellion landmark at the deepest point of the nasal root depression is measured with a sliding caliper. The subject is sitting, looking straight ahead and with the teeth together (lightly occluded). The fixed blade of the caliper is placed on Sellion. Only enough pressure to attain contact between the caliper and the skin is exerted.

Minimum frontal breadth The straight-line distance between the right and left Frontotemporale landmarks on the temporal crest on each side of the forehead is measured with a spreading caliper. The subject sits looking straight ahead. Only enough pressure to ensure that the caliper tips are on the landmarks is exerted.

Nose breadth The straight-line distance between the right and left Alare landmarks on the sides of the nostrils is measured with a sliding caliper. The subject sits looking straight

ahead. Only enough pressure to attain contact between the caliper and the skin is exerted.

Nose protrusion The straight-line distance between the Pronasale landmark at the tip of the nose and the Subnasale landmark under the nose is measured with a sliding caliper. The subject sits looking straight ahead.

Subnasale-sellion length The straight-line distance between the Subnasale landmark under the nose and the Sellion landmark at the deepest point of the nasal root is measured with a sliding caliper. The subject sits looking straight ahead. Only enough pressure to attain contact between the caliper and the skin is exerted.

Lip length The straight-line distance between the right and left Chelion landmarks at the corners of the closed mouth is measured with a sliding caliper. The subject is sitting, looking straight ahead with the teeth together (lightly occluded). The facial muscles are relaxed, and the mouth is closed.

References

- [1] L. Li and W. Zhang, "Using 3D body scans for shaping effects testing developed by foundation garment," in *International Conference on Electronic Measurement and Instruments*, vol. 4, 2007, pp. 951–954. 1
- [2] V. Ferrario, C. Sforza, C. Dellavia, G. Tartaglia, D. Sozzi, and A. Caru, "A quantitative three-dimensional assessment of abnormal variations in facial soft tissues of adult patients with cleft lip and palate," *Cleft Palate–Craniofacial Journal*, vol. 40(5), pp. 544–549, 2003. 1
- [3] A. Heliovaara, J. Hukki, R. Ranta, and A. Rintala, "Changes in soft tissue thickness after Le Fort I osteotomy in different cleft types," *International Journal of Adult Orthodontics and Orthognathic Surgery*, vol. 16, no. 3, pp. 207–213, 2001. 1
- [4] G. Singh, D. Levy-Bercowski, M. Yañez, and P. Santiago, "Three-dimensional facial morphology following surgical repair of unilateral cleft lip and palate in patients after nasoalveolar molding," *Orthodontics and Craniofacial Research*, vol. 10, pp. 161–166, 2007. 1
- [5] L. Farkas, T. Hreczko, J. Kolar, and I. Munro, "Vertical and horizontal proportions of the face in young adult north american caucasians: revision of neoclassical canons," *Plastic and Reconstructive Surgery*, vol. 75, no. 3, pp. 328–337, 1985. 1
- [6] M. Krimmel, S. Kluba, M. Bacher, K. Dietz, and S. Reinert, "Digital surface photogrammetry for anthropometric analysis of the cleft infant face," *Cleft Palate–Craniofacial Journal*, vol. 43(3), no. 3, pp. 350–355, 2006. 1
- [7] L. Farkas, K. Hajnis, and J. C. Posnick, "Anthropometric and anthroposcopic findings of the nasal and facial region in cleft patients before and after primary lip and palate repair," *Cleft Palate–Craniofacial Journal*, vol. 40, no. 5, pp. 544–549, 1993. 1
- [8] T. Yamada, Y. Mori, K. Minami, K. Mishima, and Y. Tsukamoto, "Surgical results of primary lip repair using the triangular flap method for the treatment of complete uni-

- lateral cleft lip and palate: A three-dimensional study in infants to four-year-old children," *Cleft Palate–Craniofacial Journal*, vol. 39(5), pp. 497–502, 2002. 1
- [9] S. C. Aung, R. C. Ngim, and S. T. Lee, "Evaluation of the laser scanner as a surface measuring tool and its accuracy compared with direct facial anthropometric measurements," *British Journal of Plastic Surgery*, vol. 48-8, pp. 551–558, 1995. 1
- [10] K. M. Robinette, H. Daanen, and E. Paquet, "The CAESAR project: a 3D surface anthropometry survey," in *International Conference on 3D Digital Imaging and Modeling*, October 1999, pp. 380–386. 1
- [11] C. Boehnen and P. J. Flynn, "Accuracy of 3D scanning technologies in a face scanning scenario," in *IEEE International Conference on 3D Digital Imaging and Modeling*, 2005. 1
- [12] H. Lee, S. Kil, Y. Han, and S. Hong, "Automatic face and facial features detection," in *IEEE International Symposium on Industrial Electronics*, 2001, pp. 254–259. 2
- [13] A. Salazar, A. Cerón, and F. Prieto, "3D curvature-based shape descriptors for face segmentation: An anatomical-based analysis," in *Advances in Visual Computing*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, vol. 6455, pp. 349–358. 2
- [14] A. Cerón, A. Salazar, and F. Prieto, "Relevance analysis of 3D shape descriptors on interest points and regions of the face," *International Journal of Signal and Imaging Systems Engineering*, vol. 5, no. 2, pp. 110–122, 2012. 2
- [15] A. Salazar, S. Wuhrer, C. Shu, and F. Prieto, "Fully automatic expression-invariant face correspondence," *Machine Vision and Applications*, vol. 25, no. 4, pp. 859–879, 2014. 3
- [16] A. Brunton, A. Salazar, T. Bolkart, and S. Wuhrer, "Review of statistical shape spaces for 3D data with comparative analysis for human faces," (*Computer Vision and Image Understanding*, in press). 3, 90
- [17] L. Farkas, *Anthropometric Facial Proportions in Medicine*. Thomas Books, 1987. 5, 70
- [18] G. G. Yen and N. Nithianandan, "Facial feature extraction using genetic algorithm," in *Congress on Evolutionary Computation*, vol. 2, 2002, pp. 1895–1900. 6, 7
- [19] T. Kanade, "Picture processing system by computer complex recognition," Ph.D. dissertation, Department of Information Science, Kyoto University, 1973. 6, 7
- [20] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," *Pattern Recognition*, vol. 40, pp. 1106–1122, 2007. 6

- [21] J. Chen and B. Tiddeman, "Robust facial feature tracking under various illuminations," in *IEEE International Conference on Image Processing*, 2006, pp. 2829–2832. 6, 7
- [22] B. Zhang, G. Gao, J. Lu, and Y. Zhu, "Human face location based on gradient distributions," in *IEEE Computer Society International Workshop on Knowledge Discovery and Data Mining*, 2010, pp. 320–322. 6, 7
- [23] P.-J. Lai and J.-H. Wang, "Facial image database for law enforcement application: an implementation," in *IEEE Annual International Carnahan Conference on Security Technology*, vol. 2, 2003, pp. 285–289. 6, 7
- [24] Y. Zhao, X. Shen, and N. Geroganas, "Combining integral projection and gabor transformation for automatic facial feature detection and extraction," *IEEE International Workshop on Haptic Audio Visual Environments and their Application*, pp. 103–107, 2008. 6
- [25] B. Amarapur and N. Patil, "The facial features extraction for face recognition based on geometrical approach," in *Canadian Conference on Electrical and Computer Engineering*, 2006, pp. 1936–1939. 6, 7
- [26] N. Tokuda, T. Hoshino, T. Watanabe, T. Funahashi, T. Fujiwara, and H. Koshimizu, "Caricature generation system PICASSO-2 exhibited at Expo2005 and its performance improvement," in *International Conference on Control, Automation and Systems*, 2007, pp. 1354–1358. 6, 7
- [27] Z.-F. Liu, Z.-S. You, K. Anil, and Y.-Q. Wang, "Face detection and facial feature extraction in color image," in *International Conference on Computational Intelligence and Multimedia Applications*, 2003, pp. 126–130. 6, 8
- [28] J. Chen and B. Tiddeman, "A real-time stereo head pose tracking system," in *IEEE International Symposium on Signal Processing and Information Technology*, 2005, pp. 258–263. 7
- [29] M. Kirby and L. Sirovich, "Application of the karhunen-loeve procedure for the characterization of human faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 103–108, 1990. 7
- [30] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991. 7
- [31] B. Moghaddam and A. Pentland, "Maximum likelihood detection of faces and hands," in *International Workshop on Automatic Face- and Gesture-Recognition*, 1995, pp. 122–128. 7

- [32] D. Shah and S. Marshall, "Statistical coding method for facial features," in *IEEE International Conference on Computational Cybernetics and Simulation*, vol. 145, no. 3, 1998, pp. 187–192. 7
- [33] L. Qing, S. Shan, and W. Gao, "Eigen-harmonics faces: Face recognition under generic lighting," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2004. 7
- [34] K. Fukui and O. Yamaguchi, "Facial feature point extraction method based on combination of shape extraction and pattern matching," *Systems and Computers in Japan*, vol. 29, no. 6, pp. 49–58, 1998. 7, 8
- [35] Y.-S. Ryu and S.-Y. Oh, "Automatic extraction of eye and mouth fields from a face images using eigenfeatures and ensemble networks," *Applied Intelligence*, vol. 17, no. 2, pp. 171–185, 2002. 7, 8
- [36] K.-A. Kim, S.-Y. Oh, and H.-C. Choi, "Facial feature extraction using PCA and wavelet multi-resolution images," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2004. 7, 8
- [37] Y. Zhao, X. Shen, N. Georganas, and E. Petriu, "Part-based PCA for facial feature extraction and classification," in *IEEE International Workshop on Haptic Audio visual Environments and Games*, 2009, pp. 99–104. 7
- [38] J. Tu, Y. Fu, and T. Huang, "Locating nose-tips and estimating head poses in images by tensorposes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 1, pp. 90–102, 2009. 8
- [39] M. Covell, "Eigen-points: Control-point location using principal component analyses," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 1996, pp. 122–127. 8
- [40] K. Yoshiki, H. Saito, and M. Mochimaru, "Reconstruction of 3D face model from single shading image based on anatomical database," in *International Conference on Pattern Recognition*, 2006, pp. 350–353. 8
- [41] S. Menet, P. Saint-Marc, and G. Medioni, "Active contour models: Overview, implementation and applications," in *IEEE International Conference on Systems, Man and Cybernetics*, 1990, pp. 194–199. 8
- [42] M. Kass, A. Witkin, and D. Terzopoulos, "Active contour models," in *International Conference on Computer Vision*, 1987, pp. 259–268. 8, 9

- [43] —, “Snakes: Active contour model,” *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988. 8
- [44] B. Widrow, “The “rubber mask” technique - I. and II.” *Pattern Recognition*, vol. 5, pp. 175–211, 1973. 8
- [45] D. B. Cooper, “Maximum likelihood estimation of markov-process blob boundaries in noisy images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 1, no. 4, pp. 372–384, October 1979. 8
- [46] R. B. Schudy, “Harmonic surfaces and parametric image operators: Their use in locating the moving endocardial surface from three-dimensional cardiac ultrasound data,” Computer Science Technical Report , University of Rochester, Tech. Rep., Rochester, New York, March 1981. 8
- [47] C. W. K. Gritton and J. E. A. Parrish, “Boundary location from an initial plan: The bead chain algorithm,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 1, pp. 8–13, January 1983. 8
- [48] L. H. Staib and J. S. Duncant, “Parametrically deformable contour models,” *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 93–103, 1989. 8
- [49] Y. Yagi, “Facial feature extraction from frontal face image,” in *International Conference on SignalProcessing Proceedings*, 2000, pp. 1225–1232. 9
- [50] R.-L. Hsu and A. K. Jain, “Generating discriminating cartoon faces using interacting snakes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 1, pp. 1388–1398, 2003. 9
- [51] P. Yan and K. W. Bowyer, “Biometric recognition using 3D ear shape,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 8, pp. 1297–1308, 2007. 9, 11
- [52] B. Tiddeman, N. Duffy, and G. Rabey, “Construction and visualisation of three-dimensional facial statistics,” *Computer Methods and Programs in Biomedicine*, vol. 63, no. 1, pp. 9–20, 2000. 9
- [53] Y. Yokogawa, N. Funabiki, T. Higashino, M. Oda, and Y. Mori, “A proposal of improved lip contour extraction method using deformable template matching and its application to dental treatment,” *Systems and Computers in Japan*, vol. 38, no. 5, pp. 80–89, 2007. 9
- [54] Z. Baizhen and R. Qiuqi, “Facial feature extraction using improved deformable templates,” in *International Conference on Signal Processing*, 2006. 9

- [55] A. Nikolaidis and I. Pitas, "Facial feature extraction and pose determination," *Pattern Recognition*, vol. 33, pp. 1783–1791, 2000. 9
- [56] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995. 10
- [57] A. Lanitis, C. Taylor, and T. F. Cootes, "Automatic interpretation and coding of face images using flexible models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 743–756, 1997. 10
- [58] M. H. Mahoor and M. Abdel-Mottaleb, "Facial features extraction in color images using enhanced active shape model," in *International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 144–148. 10, 11
- [59] F. Zuo and P. H. N. de With, "Fast facial feature extraction using a deformable shape model with haar-wavelet based local texture attributes," in *International Conference on Image Processing*, 2004, pp. 1425–1428. 10
- [60] K.-W. Wan, K.-M. Lam, and K.-C. Ng, "An accurate active shape model for facial feature extraction," *Pattern Recognition Letters archive*, vol. 26, no. 15, pp. 2409–2423, 2005. 10
- [61] Z.-L. Zheng and F. Yang, "Enhanced active shape model for facial feature localization," in *International Conference on Machine Learning and Cybernetics*, vol. 5, 2008, pp. 2841–2845. 10, 11
- [62] C. Sun and M. Xie, "Enhanced active shape model for facial features extraction," in *IEEE International Conference on Communication Technology*, 2008, pp. 661–664. 10, 11, 12
- [63] H. S. Lee and D. J. Kim, "Tensor-based AAM with continuous variation estimation: Application to variation-robust face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 1102–1116, 2009. 10
- [64] B. Jiang, J. Bu, and C. Chen, "Improving visual awareness by real-time 2D facial animation for ubiquitous collaboration," in *International Conference on Computer Supported Cooperative Work in Design*, April 2007, pp. 151–156. 10, 11
- [65] G. G. Gordon, "Face recognition based on depth maps and surface curvature," in *SPIE Geometric methods in Computer Vision*, 1991, pp. 234–247. 11
- [66] P. J. Flynn and A. K. Jain, "Surface classification: Hypothesis testing and parameter estimation," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 1988, pp. 261–267. 11

- [67] J. Kittler, A. Hilton, M. Hamouz, and J. Illingworth, "3D assisted face recognition: A survey of 3D imaging, modelling and recognition approaches," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005. 11
- [68] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, J. Hoffman, K. and Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," *Computer Vision and Pattern Recognition*, pp. 947–954, 2005. 11
- [69] D. Deo and D. Sen, "Automatic recognition of facial features and land-marking of digital human head," in *Conference on Computer Aided Industrial Design and Conceptual Design*, 2005, pp. 506–602. 11, 12, 16
- [70] Y. Sun and L. Yin, "Automatic pose estimation of 3D facial models," in *International Conference on Pattern Recognition*, 2008, pp. 1–4. 11, 16
- [71] M. Segundo, L. Silva, O. Pereira, and C. Queirolo, "Automatic face segmentation and facial landmark detection in range images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 40, no. 5, pp. 1319–1330, 2010. 11, 12, 32, 33, 49, 51
- [72] X. Lu, D. Colbry, and A. K. Jain, "Three-dimensional model based face recognition," in *International Conference on Pattern Recognition*, vol. 1, 2004, pp. 362–366. 11, 12, 17
- [73] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Three-dimensional face recognition," *International Journal on Computer Vision*, vol. 64, no. 1, pp. 5–30, 2005. 11
- [74] C. Conde, L. J. Rodríguez-Aragón, and E. Cabello, "Automatic 3D face feature points extraction with spin images," in *Image Analysis and Recognition*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2006, vol. 4142, pp. 317–328. 11, 12, 17
- [75] H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama, "3D face recognition under expressions, occlusions, and pose variations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 9, pp. 2270–2283, 2013. 11
- [76] D. Colbry, G. Stockman, and A. K. Jain, "Detection of anchor points for 3D face verification," in *IEEE Workshop on Advanced 3D Imaging for Safety and Security*, 2005. 11, 12, 17
- [77] Z. Ben Azouz, C. Shu, and A. Mantel, "Automatic locating of anthropometric landmarks on 3D human models," in *International Symposium on 3D Data Processing, Visualization, and Transmission*, 2006, pp. 750–757. 11, 12, 17, 31, 32, 36
- [78] A. Colombo, C. Cusano, and R. Schettini, "3D face detection using curvature analysis," *Pattern Recognition*, vol. 39, pp. 444–455, 2006. 11, 12, 16

- [79] X. Gong and G. Wang, "Automatic 3D face segmentation based on facial feature extraction," in *IEEE International Conference on Industrial Technology*, 2006, pp. 1154–1159. 12, 64
- [80] A. Adán, M. Adán, S. Salamanca, and P. Merchán, "Using non local features for 3D shape grouping," in *Structural, Syntactic, and Statistical Pattern Recognition*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2008, vol. 5342, pp. 644–653. 15
- [81] T. Gatzke and C. Grimm, "Feature detection using curvature maps and the min-cut/max-flow algorithm," in *Geometric Modeling and Processing*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2006, vol. 4077, pp. 578–584. 15, 19
- [82] J. J. Koenderink and A. J. Van Doorn, "Surface shape and curvature scales," *Image and Vision Computing*, vol. 8, no. 10, pp. 557–564, 1992. 16, 17
- [83] A. Johnson, "Spin-images: A representation for 3D surface matching," Ph.D. dissertation, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, August 1997. 16, 17
- [84] Y. Sun and M. Abidi, "Surface matching by 3D point's fingerprint," in *IEEE International Conference on Computer Vision*, vol. 2, 2001, pp. 263–269. 16, 18, 32
- [85] C. Bishop, *Pattern Recognition and Machine Learning*. Springer Science Business + Media, LLC, 2006. 16, 19, 21
- [86] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression-invariant 3D face recognition," in *Audio- and Video-Based Biometric Person Authentication*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2003, vol. 2688, pp. 62–70. 16, 17
- [87] P. Hallinan, G. Gaile, A. Yuille, G. Peter, and M. David, *Two-and Threedimensional patterns of the face*. A. K. Peters, 1999. 16
- [88] F. Xue and X. Ding, "3D+2D face localization using boosting in multi-modal feature space," in *International Conference on Pattern Recognition*, 2006. 16, 64
- [89] G. Zhang and Y. Wang, "A 3D facial feature point localization method based on statistical shape model," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, 2007, pp. 249–252. 17
- [90] A. Jagannathan and E. L. Miller, "Three-dimensional surface mesh segmentation using curvedness-based region growing approach," *IEEE Transactions Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2195–2204, 2007. 17

- [91] T. Cox and M. Cox, *Multidimensional Scaling, Second Edition*. Chapman & Hall CRC, 2001. 18
- [92] S. Wuhrer, Z. Ben-Azouz, and C. Shu, "Semi-automatic prediction of landmarks on human models in varying poses," in *Canadian Conference on Computer and Robot Vision*, 2010, pp. 136–142. 18
- [93] J. Wang, L. Yin, X. Wei, and Y. Sun, "3D facial expression recognition based on primitive surface feature distribution," in *Conference of Computer Vision and Pattern Recognition*, 2006, pp. 1399–1406. 19
- [94] P. Xi and C. Shu, "Consistent parameterization and statistical analysis of human head scans," *The Visual Computer*, vol. 25(9), pp. 863–871, 2009. 29, 34
- [95] H. Li, T. Weise, and M. Pauly, "Example-based facial rigging," *ACM Transactions on Graphics (SIGGRAPH)*, vol. 29, no. 4, pp. 32:1–32:6, 2010. 30, 35, 43
- [96] E. Learned-Miller, "Data driven image models through continuous joint alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 2, pp. 236–250, 2006. 31
- [97] M. Cox, S. Sridharan, S. Lucey, and J. Cohn, "Least squares congealing for unsupervised alignment of images," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8. 31
- [98] Y. Tong, X. Liu, and P. Wheeler, F. and Tu, "Semi-supervised facial landmark annotation," *Computer Vision and Image Understanding*, vol. 116, pp. 922–935, 2012. 31
- [99] T. Cootes, G. Edwards, and C. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 116, no. 6, pp. 681–685, 2001. 31
- [100] S. Mehryar, K. Martin, K. Plataniotis, and S. Stergiopoulos, "Automatic landmark detection for 3D face image processing," in *IEEE Congress on Evolutionary Computation*, 2010, pp. 1–7. 31
- [101] E. Vezzetti and F. Marcolin, "3D human face description: landmarks measures and geometrical features," *Image and Vision Computing*, vol. 30, no. 10, pp. 750–761, 2012. 31
- [102] S. Berretti, B. Ben Amor, M. Daoudi, and A. del Bimbo, "3D facial expression recognition using SIFT descriptors of automatically detected keypoints," *The Visual Computer*, vol. 27, no. 11, pp. 1021–1036, 2011. 31, 33
- [103] C. Creusot, N. Pears, and J. Austin, "3D face landmark labelling," in *Proceedings ACM workshop on 3D object retrieval*, 2010, pp. 27–32. 31

- [104] P. Perakis, T. Theoharis, G. Passalis, and I. Kakadiaris, "Automatic 3D facial region retrieval from multi-pose facial datasets," in *Eurographics Workshop on 3D Object Retrieval*, 2009, pp. 37–44. 32
- [105] P. Perakis, G. Passalis, T. Theoharis, and I. Kakadiaris, "3D facial landmark detection & face registration: A 3D facial landmark model & 3D local shape descriptors approach," Computer Graphics Laboratory, University of Athens, 15784 Ilisia, Greece, Tech. Rep., 2010. 32
- [106] P. Nair and A. Cavallaro, "3D face detection, landmark localization, and registration using a point distribution model," *IEEE Transactions on Multimedia*, vol. 11(4), pp. 611–623, 2009. 32, 51
- [107] X. Lu and A. Jain, "Automatic feature extraction for multiview 3D face recognition," in *International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 585–590. 32
- [108] X. Lu, D. Colbry, and A. Jain, "Matching 2.5D scans to 3D models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 31–43, 2006. 32
- [109] A. Elad and R. Kimmel, "On bending invariant signatures for surfaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25(10), pp. 1285–1295, 2003. 32, 36
- [110] K. Chang, K. Bowyer, and P. Flynn, "Multiple nose region matching for 3D face recognition under varying facial expression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1695–1700, 2006. 32
- [111] O. van Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or, "A survey on shape correspondence," *Computer Graphics Forum*, vol. 3, no. 6, pp. 1681–1707, 2011. 33
- [112] G. Passalis, P. Perakis, T. Theoharis, and I. Kakadiaris, "Using facial symmetry to handle pose variations in real-world 3D face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1938–1951, 2011. 33, 53, 54
- [113] I. Kakadiaris, G. Passalis, G. Toderici, M. Murtuza, L. Yunliang, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 640–649, 2007. 33
- [114] I. Mpipiperis, S. Malassiotis, and M. Strintzis, "Bilinear models for 3D face and facial expression recognition," *IEEE Transactions on Information Forensics and Security*, pp. 498–511, 2008. 33

- [115] J. Guo, X. Mei, and K. Tang, "Automatic landmark annotation and dense correspondence registration for 3D human facial images," *Journal of Anthropological Sciences*, vol. 14, no. 232, pp. 1–12, 2013. 33, 64
- [116] Y. Huang, X. Zhang, Y. Fan, L. Yin, L. Seversky, J. Allen, T. Lei, and W. Dong, "Reshaping 3D facial scans for facial appearance modeling and 3D facial expression analysis," *Image and Vision Computing*, vol. 30, no. 10, pp. 681–796, 2012. 33
- [117] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Conference on Computer Graphics and Interactive Techniques*, 1999, pp. 187–194. 34, 55
- [118] L. Xiaoguang and A. Jain, "Deformation modeling for robust 3D face matching," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 1377–1383. 34
- [119] C. Basso, P. Paysan, and T. Vetter, "Registration of expressions data using a 3D morphable model," in *International Conference on Automatic Face and Gesture Recognition*, 2006, pp. 205–210. 34
- [120] B. Amberg, R. Knothe, and T. Vetter, "Expression invariant 3D face recognition with a morphable model," in *IEEE International Conference on Automatic Face Gesture Recognition*, 2008, pp. 1–6. 34
- [121] B. Allen, B. Curless, and Z. Popović, "The space of human body shapes: Reconstruction and parametrisation from range scans," *ACM Transactions on Graphics (SIGGRAPH)*, vol. 22, no. 3, pp. 587–594, 2003. 34, 47
- [122] S. Wuhrer, C. Shu, and P. Xi, "Landmark-free posture invariant human shape correspondence," *The Visual Computer*, vol. 27, no. 9, pp. 843–852, 2011. 34, 37
- [123] A. Bronstein, M. Bronstein, and R. Kimmel, "Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching," *Proceedings of the National Academy of Sciences*, vol. 103, no. 5, pp. 1168–1172, 2006. 34
- [124] —, "Expression-invariant representations of faces," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 188–197, 2007. 35
- [125] T. Weise, S. Bouaziz, H. Li, and M. Pauly, "Realtime performance-based facial animation," *ACM Transactions on Graphics (SIGGRAPH)*, vol. 30, no. 4, pp. 77:1–77:10, 2011. 35
- [126] J. Yedidia, W. Freeman, and Y. Weiss, *Understanding Belief Propagation and Its Generalizations*. Science & Technology Books, 2003. 37

- [127] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*. Morgan Kaufmann Publishers, 2006. 39
- [128] F. Cazals and M. Pouget, "Smooth surfaces, umbilics, lines of curvatures, foliations, ridges and the medial axis: a concise overview," INRIA, route des Lucioles, BP 93, 06902 Sophia Antipolis Cedex (France), Tech. Rep. RR-5138, 2004. 40
- [129] H. Li, B. Adams, L. Guibas, and M. Pauly, "Robust single-view geometry and motion reconstruction," *ACM Transactions on Graphics (SIGGRAPH Asia)*, vol. 28, no. 5, pp. 175:1–175:10, 2009. 47, 48
- [130] D. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical Programming*, vol. 45, pp. 503–528, 1989. 48
- [131] L. Yin, X. Wei, J. Wang, Y. Sun, and M. Rosato, "A 3D facial expression database for facial behavior research," in *IEEE International Conference on Automatic*, 2006. 48, 60
- [132] Y. Gao, "Efficiently comparing face images using a modified hausdorff distance," in *IEEE Conference on Vision, Image and Signal Processing*, 2003, pp. 346–350. 53
- [133] H. Rabiou, M. Saripan, S. Mashohor, and M. Marhaban, "3D facial expression recognition using maximum relevance minimum redundancy geometrical features," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–8, 2012. 60
- [134] R. Duin, P. Juszczak, P. Paclik, E. Pekalska, D. de Ridder, D. Tax, and S. Verzakov, *PRTools4.1, A Matlab Toolbox for Pattern Recognition*. Delft University of Technology, 2007. 60
- [135] D. Vlastic, M. Brand, H. Pfister, and J. Popovic, "Face transfer with multilinear models," *ACM Transactions on Graphics (SIGGRAPH)*, vol. 24, no. 3, 2015. 60
- [136] A. Savran, N. Alyüz, H. Dibekliouğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in *European Workshop on Biometrics and Identity Management*, 2008, pp. 47–56. 62
- [137] K. I. Chang, K. W. Bowyer, and P. J. Flynn, "An evaluation of multimodal 2D + 3D face biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 4, pp. 619–624, April 2005. 63
- [138] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, pp. 511–518. 64

- [139] L. Zou, S. Cheng, Z. Xiong, M. Lu, and K. Castleman, Castleman, "Facial feature extraction from range image using a 3D morphable model," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, 2007, pp. 241–244. 64
- [140] J.-G. Wang and E. Sung, "Facial feature extraction in an infrared image by proxy with a visible face image," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, no. 5, pp. 2057–2066, October 2007. 64
- [141] Y. Luximon, R. Ball, and L. Justice, "The 3D chinese head and face modeling," *Computer-Aided Design*, vol. 44, pp. 40–47, 2012. 69
- [142] Z. Zhuang, "A head-and-face anthropometric survey of u.s. respirators users," NIOSH/NPPTL, 626 Cochran Mill Road, Pennsylvania, U.S.A., Tech. Rep., 2004. 70
- [143] S. Gupta, M. K. Markey, and A. C. Bovik, "Anthropometric 3D face recognition," *International Journal of Computer Vision*, vol. 90, pp. 331–349, 2010. 70
- [144] X. Chen, A. Golovinskiy, and T. Funkhouser, "A benchmark for 3D mesh segmentation," *ACM Transactions on Graphics (SIGGRAPH)*, vol. 28, no. 3, Aug. 2009. 73, 75, 79
- [145] T. Bolkart, A. Brunton, A. Salazar, and S. Wuhrer, "Statistical 3D shape models of human faces," 2014. [Online]. Available: <http://statistical-face-models.mhci.uni-saarland.de/> 90
- [146] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3D dynamic facial expression database," in *International Conference on Automatic Face and Gesture Recognition*, 2008, pp. 17–19. 90