# Analysis of Human Face Shape Abnormalities Using Machine Learning

Jia Wu

A dissertationsubmitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Linda G. Shapiro, Chair

Jenq-Neng Hwang

James F. Brinkley

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

**Abstract**

Analysis of Human Face Shape Abnormalities Using Machine Learning

Jia Wu

Chair of the Supervisory Committee:
Dr. Linda G. Shapiro
Department of Computer Science and Engineering

Because the capture of 3D facial form for people with medical conditions has become a practical reality, much research has been done using computer-based automatic methods to study 3D facial characteristics. This study focuses on face abnormality, especially, children with cleft lip and/or palate.

Cleft lip is a birth defect that results in deformity of the upper lip and nose. Its severity is widely variable and the results of treatment are influenced by the initial deformity. Objective assessment of severity would help to guide prognosis and treatment. Given that facial asymmetry increases directly with increasing cleft severity, the main focus of this thesis is the quantification of asymmetry and nasal deformity.

In our study, we developed an automated computer-based system for facial analysis that can greatly facilitate medical researchers. Our system takes raw 3D data, automatically crops the region of interest (the face), and normalizes the 3D face mesh. Then 20 landmarks are automatically located using geometric information followed by a deformable registration method. These landmarks are used for locating the mid-facial reference plane, which sets up the reference to compare the left and right side of the face. The plane is used to provide a ranking order based on the severity of cleft lip using a learning-to-rank algorithm. Last but not least, asymmetry and nasal deformity descriptors are calculated to provide quantitative scores. The main contributions of this work are: an automated methodology for cleaning raw data and pose normalization; an automated landmark location method; a machine

learning technique that detect landmark-related regions; two original algorithms to find the mid-facial reference planes and the evaluation; features for ranking the severity of cleft lip; and six descriptors for asymmetry and nasal deformity, which are highly correlated to the ranking orders provided by medical expert.

## TABLE OF CONTENTS

Page

## LIST OF FIGURES

## LIST OF TABLES

## ACKNOWLEDGMENTS

First and foremost, I want to express my deepest gratitude to my advisor Dr. Linda Shapiro, without whose guidance and support I would not become the researcher I am today.

I am very grateful to all my committee members, Dr. Jenq-Neng Hwang, Dr. James Brinkley, and Dr. Mark Ganter, for their excellent feedback and comments.

I would like to express my sincere gratitude to Dr. Raymond Tse, for providing me with the medical 3D data for this study as well as his engaging discussions and suggestions.

I also thank my collaborators at Seattle Childrens Hospital Craniofacial Center: Dr. Carrie Heike, Dr. Clinton Morrison, Dr. Murat Maga, Dr. Craig Birgfeld, Dr. Kelly Evans and Dr. Babette Saltzman for providing their medical opinions.

I would like to thank Shu Liang for collaborating with me on automatic landmark detection, Ravensara Travillian for her advice on dissertation writing, and members of my research group for their countless suggestions which improved my work.

I owe an indescribable amount of gratitude to my parents, for their encouragement and support.

Finally, I reserve special thanks for my husband, Canxing Xu, for always being my best friend and supporter, and my daughter, Cynthia, who teaches me how great life is.

This research was supported by NIH/NIDCR under grant number 1U01DE020050-01 (PI: L. Shapiro)

## DEDICATION

to my family for their unconditional love and support

# Chapter 1

# INTRODUCTION

## *1.1 Motivation*

Recently, multi-camera commercial stereo systems that produce 3D models from images of patients have been used by medical researchers to analyze and quantify 3D face shape using computers and new techniques from machine learning. Doctors hope these new methods can aid them in diagnosis, surgical planning, establishing long-term outcome measurement, and studying correlations between symptoms and causes. Our collaborators in the Craniofacial Center at Seattle Children's Hospital are interested in quantifying facial abnormalities, especially for cleft lip and/or palate.

Cleft lip and/or palate is a type of clefting congenital deformity caused by abnormal facial development during gestation [1]. A cleft lip and/or palate is a non-fusion in the lip and/or palate that occurs before birth. Approximately 1 in 700 children is born with a cleft lip and/or a cleft palate [2] . A cleft lip and/or palate can be treated with surgery, and if it is performed soon after birth or in early childhood, the results will be more successful. Because we only look at the surface of the patient's face, we will use the term "cleft lip" or "cleft" for all the kinds of cleft lip and/or palate in the rest of this thesis.

Research on face abnormalities is further complicated by the fact that there are no "gold standards" for most cases. Our collaborating medical experts are willing to fill in surveys, ranking and giving a score on certain features, but they confirm that their scores are very subjective, and as we found in previous studies, the correlation between two different medical experts can be very low [3].

Last but not least, the facial abnormalities that we are dealing with are quite varied. Clefting can range from multiple deep severe clefts in the palate to a single incomplete or hardly noticeable cleft in the lip. After the surgery, the cleft lip is typically repaired sufficiently that it is hard for human eyes to pick up whether there is an abnormality or

not. Traditionally, the approaches that the medical researchers used to analysis the 3D face of cleft are relied on landmarks, which may provide some information about abnormalities. But the drawback is that no shape information is provided about regions between landmark points, which are scientifically interesting in some cases.

### *1.2 Research Objectives*

Given that facial asymmetry increases directly with increasing cleft severity, this study focuses on, but is not limited to, the quantification of facial asymmetry and deformations caused by clefts.

**Objective 1: quantification of facial asymmetry**

Given a 3D face, can we tell how asymmetric this face is? Given 3D data from a cleft patient, can we compute meaningful asymmetry scores before and after surgery? Where are the areas of least symmetry and how asymmetric are they?

**Objective 2: evaluation of mid-facial reference plane methods**

Motivated by the first objective, is there a best method to find a mid-facial reference plane? How do computer-based methods perform compared to human-based methods and what are their performances in terms of facial asymmetry scores?

**Objective 3: localization of landmarks**

Traditional morphometric analyses rely heavily on landmarks; how can computer-based methods reliably detect landmarks?

**Objective 4: preprocessing system**

Given raw data from a multi-camera commercial stereo system, how can the data automatically be processed so that the face is cropped out for further analysis for medical experts and computer-based algorithms?

### *1.3 Thesis Outline*

The rest of this paper is organized as follows. Chapter 2 discusses the existing 3D human face analysis methods. Chapter 3 introduces our data formats and data sets that are used in this study. Chapter 4 describes our work on locating landmarks using geometrical information and deformable registration. A machine-learning based method to find landmark-related

regions and its applications are discussed in Chapter 5. Chapter 6 discusses the comparison between the mid-facial reference planes provided by two human-based methods and three computer-based methods. These methods are further used and compared for learning to rank the severity of unicleft lip in Chapter 7. In Chapter 8, the quantification measurements for facial asymmetry are introduced and experimental results with cleft lip, both before and after surgery, are provided. A novel face extraction and normalization method is described in Chapter 9. Finally, a summary is provided in Chapter 10.

# Chapter 2

# RELATED LITERATURE

In this section the related literature on 3D facial feature assessment in computer vision and medicine will be described. Then the automated methods on quantification of 3D cleft lip data will be discussed. These will be followed by a general description of relevant work on symmetry detection and analysis in computer vision. Although some of the work mentioned here used a different data sources (2D photos, for example), the focus of our study will be on 3D surface meshes.

## *2.1 Related Work on 3D Facial Feature*

The blossoming of 2D face recognition techniques, such as elastic graph matching [4] and eigenface [5] in the last 20 years has brought a great interest in 3D face recognition as well. Several successful 2D face recognition methods have been adapted to 3D data [6, 7] and there are also 2D-3D hybrid methods, which use both gray tone images and depth information to help with face recognition [8, 9, 10, 11].

There have also been several works dealing with only 3D mesh face data without texture information for face recognition. Berritti, Bimbo and Pala [12] used iso-geodesic stripes and encoded the relevant information into a compact representation in the form of a graph. A multiple nose region matching method for face recognition was described by Chang, Bowyer and Flynn [13]. These methods could distinguish facial differences between distinct individuals from differences induced by non-neutral expressions. Nair and Cavallaro [14] presented a framework which was based on a 3-D point distribution model for detecting and segmenting faces, localizing landmarks, and achieving fine registration of face meshes.

While most 3D face analysis methods in computer vision have been developed with a focus on biometric authentication and recognition, very few of them attempt to address medical problems and facial abnormalities. Given the fact that the motivations are totally

different, the vision work in facial feature detection has not transferred well to medical applications, although some ideas can be adapted.

In medical craniofacial assessment, the traditional way to identify an individual with facial dysmorphism has been a thorough clinical examination combined with craniofacial anthropometric measurements [15]. Newer methods involving stereoscopic imaging data are often manually measured or manually labeled. The human effort in this work is still very labor-intensive.

Morphometrics is a branch of statistics combining tools from geometry, computer graphics and biometrics in techniques for the multivariate analysis of biological shape variation. Most morphometrics methods have emphasized the domain of landmarks. Procrustes analysis is one of the most powerful tools of landmark-based morphometrics [16, 17]. This method uses only the positions of the landmarks. It translates, scales and rotates landmarks from multiple objects and tries to align them in a single space.

Procrustes analysis also acted as a first step to compute mean landmarks in the dense surface model (DSM) technique that Hammond's group [18, 19] used. Then thin-plated-splines were used to warp face surfaces to the mean landmarks and to align all the faces. Points on each face can be mapped to produce a dense correspondence across the whole dataset. An average face surface of the set was then computed. Principal components analysis [20] was applied to the differences between the positions of the densely corresponded points on each face surface to the average face, and the Euclidean distance between DSM-based representations of face surface were analyzed. Hammond applied this model to several diseases, such as autism spectrum disorders [21], Fabry disease [22] and Cornelia De Lange syndrome [23].

### *2.2 Related Work on Cleft and Asymmetry Analysis*

With respect to the studies of 3D face analysis on cleft lip and palate, the use of 3D data are very limited, and the focus has been on the asymmetry of the nose after surgery. Nakamura *et al.* [24] described the surgical techniques for secondary correction of unilateral cleft lip-nose deformity with 3D observations of preoperative and postoperative nasal forms. Optical surface data were used to quantify the degree of asymmetry in individuals with craniofacial

birth defects before and after corrective surgery [25].

Benz *et al.* [26] introduced a common method for 3D facial symmetry analysis in which the original data was mirrored at an arbitrary plane, and then the original mesh and the mirrored mesh were registered using the iterated-closest-point algorithm [27]. Based on the registered data, the symmetry plane was determined from the centers of associated points. This method is reliable when the data is properly aligned [28], but the results rely heavily on the choice of the initial plane about which the data is mirrored, and, as a result, the method is not robust to noise [26].

More work has been done on face symmetry study in computer vision. Symmetry analyses have been used for studying facial attractiveness [29, 30] and analysis of facial expression for human identification [31]. Besides symmetry studies of human faces, there are many applications based on analysis of symmetry in general computer vision and graphics, such as image editing [32, 33], symmetrizing deformations for 3D data [34], and data repair [35, 36].

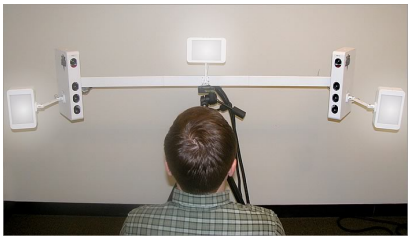
There have been several recent efforts in shape analysis to investigate symmetry in 2D images and 3D data. For two-dimensional images, Loy and Eklundh [37] proposed an algorithm to extract symmetric pairs of points based on local features and to detect bilateral or rotational symmetry by voting in a Hough-style space. Lee and Liu [38] developed a local feature-based method for point matching, then grouped the pairs to form a curved glide-reflection. Mitra, Guibas and Pauly [39] introduced a method for three-dimensional mesh data that is based on matching simple local shape signatures in pairs and using these matches to accumulate evidence for symmetry in an appropriate transformation space by clustering. Ovsjanikov, Sun and Guibas [40] proposed an algorithm for detecting and categorizing global intrinsic symmetries of shapes, in the signature space defined by the eigenfunctions of the Laplace-Beltrami operator. Xu *et al.* [36] also discussed a voting scheme to obtain a symmetry axis transform and find partial instrinsic reflectional symmetry curves. In general, all of these methods find local properties for matching points, and then compute symmetry by clustering, voting or grouping.

Chapter 3

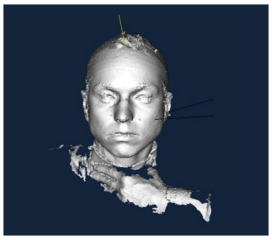
# DATA FORMATS AND STANDARD LANDMARK SELECTION

The 3D human head data used in our study were collected by craniofacial investigators using the 3dMDcranial system or the 3dMDface system [41] . The 3dMDcranial system is made up of four camera stands, each containing three cameras, as well as one stand from the top. Of the three cameras in each stand, one captures a direct photo, one captures from an overhead perspective and one captures from a perspective beneath the face to yield a three-dimensional view of the face through stereo analysis, as shown in Fig. 3.1 (a). The resulting range maps are stitched together using 3dMD proprietary methods to yield the final 3D head mesh and a texture map of the face. The 3dMDface system only uses two camera stands, on the left and right side of the face, as shown in Fig. 3.1 (b). Due to human subjects requirements (IRB), most of the data used in the research described in this work are only the 3D meshes without texture data. Fig. 3.1 (c) shows an example of the 3D raw mesh data that are used in this work. For the last part of the research, we were allowed to use texture data to crop out the faces from the raw data. That data format will be described in detail in Chapter 9.

(a) 3dMDcranial system



(b) 3dMDface system



(c) raw mesh data

Figure 3.1: 3dMD imaging system set up and raw mesh images, shaded for viewing.

Twenty-four standard cephalometric facial landmarks were selected by medical experts. These landmarks are widely used in craniofacial anthropometric measurements [42] and

also general 3D face analysis [14]. The names and definitions of the landmarks are shown in Table 3.1. In the following chapters, these landmarks, or a subset of them, will be used for different tasks. Several datasets were accumulated for the different studies. In the following sections, these datasets will be used.

**Cleft survey dataset:** Seattle Children's Hospital collected data for 45 individuals with various types of cleft lip, including 35 children with unrepaired unilateral cleft lip and 10 children who had unrepaired bilateral cleft lip. 5 normal control children who had no congenital craniofacial anomalies are also included for comparison. All of them are ages 3 to 6 months. All of the images in this dataset are taken by the 3dMDcranial system.

**Unicleft dataset:** This dataset is also collected by Seattle Children's Hospital and contains 35 infants with unilateral cleft lip, both before and after surgery. This dataset is also captured by the 3dMDcranial system.

**3D facial norms dataset:** The 3D norms dataset consists of 994 3D facial meshes from people who are of European Caucasian ancestry, ages 3 to 40 years old. All subjects are screened on the following exclusion criteria:

1. the presence of conspicuous facial hair
2. the presence of any facial piercings other than small studs
3. a personal history of facial trauma or disfigurement
4. a personal history of facial reconstructive or plastic surgery
5. a personal history of orthognathic/jaw surgery or jaw advancement
6. a personal history of any facial prosthetics or implants
7. a personal history of any palsy, stroke or neurological condition affecting the face
8. a personal or family history of any facial anomaly or birth defect
9. a personal or family history of any syndrome or congenital condition with a facial manifestation

An affirmative response to one or more of the above items would disqualify an individual from participation in the study. This dataset is collected as part of the FaceBase Consortium [43], mainly at the University of Pittsburgh, Seattle Children's Hospital and University of Texas Health Science Center at Houston. All the images in this dataset are captured by

the 3dMDface system.

Table 3.1: Names and descriptions of 24 standard landmarks

| Landmark Name | Landmark Label | Description |
|---|---|---|
| Nasion | n | Midline point where the frontal and nasal bones contact nasofrontal suture. Corresponds to the underlying bony landmark. |
| Pronasale | prn | Midline point marking the maximum protrusion of the nasal tip. |
| Subnasale | sn | Midline point marking the junction between the inferior border of the nasal septum and the cutaneous upper lip. It is the apex of the nasolabial angle. |
| Labiale Superius | ls | Midline point of the vermilion border of the upper lip, at the base of the philtrum. |
| Stomion | sto | Midpoint of the labial fissure. |
| Labiale Inferius | li | Midline point of the vermilion border of the lower lip. |
| Sublabiale | sl | Midpoint along the inferior margin of the cutaneous lower lip (labiomental sulcus). |
| Gnathion | gn | Midline point on the inferior border of the mandible. Corresponds to the underlying bony landmark. |
| Endocanthion (Right) | en_r | Apex of the angle formed at the inner corner of the palpebral fissure where the upper and lower eyelids meet. |
| Endocanthion (Left) | en_l | Same as above. |
| Exocanthion (Right) | ex_r | Apex of the angle formed at the outer corner of the palpebral fissure where the upper and lower eyelids meet. |
| Exocanthion (Left) | ex_l | Same as above. |
| Alare (Right) | al_r | Most lateral point on the nasal ala. |
| Alare (Left) | al_l | Same as above. |
| Alar Curvature Point (Right) | ac_r | Most posterolateral point on the alar cartilage, located within the crease formed by the union of the alar cartilage and the skin of the cheek. |
| Alar Curvature Point (Left) | ac_l | Same as above. |
| Subalare (Right) | sbal_r | Point located at the lower margin of the nasal ala, where the cartilage insterts in the cutaneous upper lip. |
| Subalare (Left) | sbal_l | Same as above. |
| Crista Philtri (Right) | cph_r | Point marking the lateral crest of the philtrum at the vermilion border of the upper lip. |
| Crista Philtri (Left) | cph_l | Same as above. |
| Chelion (Right) | ch_r | Point marking the lateral extent of the labial fissure. |
| Chelion (Left) | ch_l | Same as above. |
| Tragion (Right) | t_r | Point marking the notch at the superior margin of the tragus, where the cartilage meets the skin of the face. |
| Tragion (Left) | t_l | Same as above. |

# Chapter 4

# AUTOMATIC LANDMARK LOCATION

Craniofacial researchers make heavy use of established facial landmarks in their morphometric analyses. For studies on very large facial image datasets, the standard approach of manual landmarking is very labor intensive. With the goal of producing 20 established landmarks, we have developed a geometric methodology that can automatically locate 10 established landmark points and 7 other supporting points on human 3D facial scans. Then, to improve accuracy and produce all 20 landmarks, a deformable matching procedure establishes a dense correspondence from a template 3D mesh with a full set of 20 landmarks to each individual 3D mesh. The 17 geometrically-determined points on the individual 3D mesh are used for the initial correspondence required by the deformable matching. The method is evaluated on 994 3D facial meshes of normal adults (3D facial norms dataset), and results are compared to landmarks manually identified by medical experts. Our results show an improvement over prior results in the recent literature. This chapter is a collaboration with Shu Liang.

## *4.1 Background*

Using 3D human face data to measure facial features is of great practical importance in craniofacial research and practice. Traditionally, direct anthropometry using calipers has been the standard technique for quantifying craniofacial dysmorphology, as well as for surgical planning and outcome assessment [44]. Some of the major downsides to direct anthropometry include the excessive time and invasiveness of the method, the amount of training required, the extent of measurement error, and limitations in the kinds of data that can be collected. Following the introduction of cost-effective 3D surface imaging solutions, computerized anthropometry has largely replaced more traditional direct methods for collecting quantitative information on human faces [45]. These systems are capable of capturing the

full 3D geometry of the human face and head in just a fraction of a second.

While computerized 3D anthropometry represents a major advance, to obtain measurements, points on the face and head corresponding to traditional anthropometric landmarks must still be captured manually through the use of software. This can be a time-consuming process, requiring a fair amount of training to master. Efficiency is particularly important when dealing with very large 3D facial datasets, such as the 3D norms dataset. Recognizing the need to move beyond manual data collection, our goal is to provide an automatic method to detect landmarks from 3D facial surfaces. Such a method requires the resulting automatically-generated landmarks to be located in the correct anatomical positions and the process to be extendable to as many landmarks as needed.

In the computer vision community, facial landmark detection methods can be classified as those that are solely dependent on geometric information and those that are supported by trained statistical feature models [46]. Of the methods that are dependent on geometric information, surface curvatures, shape index, and geometric relationships are heavily used. For example, Lu and Jain [47] and Colbry *et al.* [48] find eye/mouth corners and nose/chin tips based on a fusion scheme of shape index on range maps and the "cornerness response," making use of distance relationships to find the points. Lin *et al.* [49] used curvature analysis to determine the eye sockets and detected the nose tip as the extreme vertex along the normal direction of the eye sockets.

Using not only geometric information but also trained models, Yu and Moon [50] located the nose tip and inner eye corners in 3D range maps with a genetic-algorithm-trained detector. Xu *et al.* [51] used the concept of "effective energy" to describe the relationships between neighboring points and an SVM classifier to select the correct nose tip points. Nair and Cavallaro [52] used a point distribution model to locate landmarks and register faces. Romero-Huertas and Pears [53] developed a graph-matching approach to locate the positions of the nose tip and inner eye corners. All of the above methods are limited to a fixed number of predefined points. If any new landmarks are needed, a new type of model must be created.

In this chapter, we describe an improved method to automatically locate an arbitrary number of landmarks on a 3D facial mesh. First, a partial set of landmarks is located

on each individual mesh by geometric techniques. Then, in order to improve positional accuracy and add additional landmarks, the geometrically-detected landmarks are used to initialize a deformable transformation that is used to create a dense correspondence between a template mesh and the individual (target) mesh, so the set of established landmarks on the template mesh can be transferred to the target. The distance between manually annotated landmarks labeled by experts in anthropometry and the automatically-detected final landmarks is computed to evaluate the accuracy of the method.

The main contribution of this work is the fully working system that can produce all 20 established landmarks and outperforms all competing methods in the literature. A second contribution is the geometric methodology for finding the initial set of 17 landmark points that are used to intialize the deformable matching procedure and are critical to its success. The use of deformable matching for this purpose is not a new idea. However, the specific matching algorithm used [54], which has not been used for this purpose before, is particularly good for this task, because it executes an order of magnitude faster than other algorithms we tested (e.g. about 5 minutes for the process in [54], compared to more than 4 hours for the method in [55]).

### *4.2 Method*

There are 27 landmarks mentioned in this chapter, including the nasion (n), sellion (se), pronasale (prn), subnasale (sn), right and left alare (al_r & al_l), right and left alar curvature point (ac_r & ac_l), labiale superius (ls), stomion (sto), labiale inferius (li), sublabiale (sl), right and left subalare (sbal_r & sbal_l), right and left crista philtri (cph_r & cph_l), right and left chelion (ch_r & ch_l), gnathion (gn), right and left endocanthion (en_r & en_l), right and left exocanthion (ex_r & ex_l), right and left superaurale (sa), right and left postaurale (pa), as shown in Fig. 4.1. The 20 established facial landmarks that we sought are listed in Table 1 and also marked with a triangle in Fig. 4.1. For ground truth, these 20 landmarks were located manually on each surface by a single trained personnel. All landmarking was performed on 3D models with color and texture mapping active, which made them more accurate than if they were placed on the mesh alone. However, our programs only had access to 3D meshes with no color/texture.

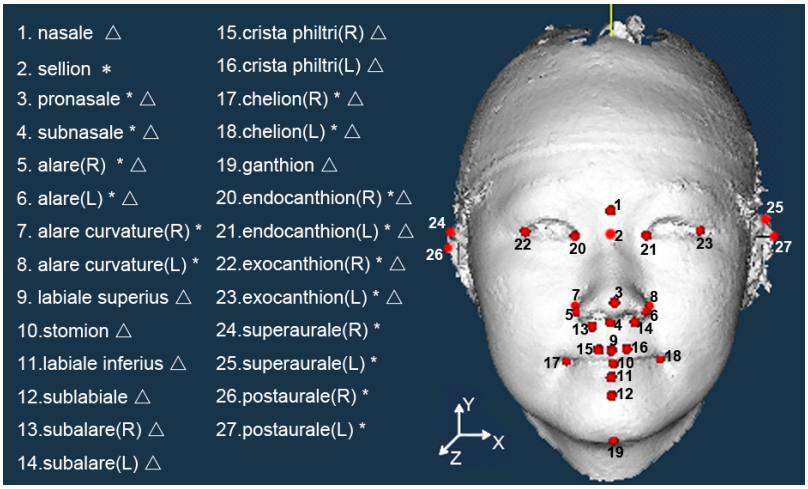



Figure 4.1: Normalized head with landmarks. The geometric method finds the points marked with ⋆. The points marked with △ are found using the deformation method and compared with the ground truth.

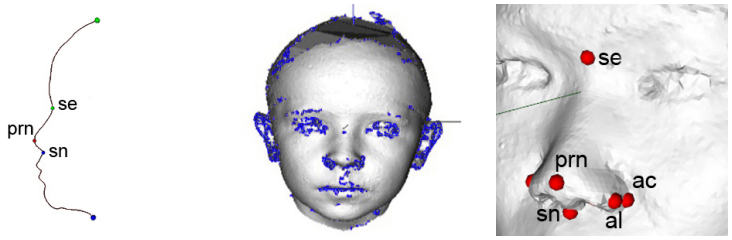
Surface meshes were pose-normalized for our automatic detection system using a method described in [56].

Our method to find landmarks using deformable registration can be described in the following steps:

1. Compute 17 initial landmarks on the nose, mouth, eyes, and ears of both the template mesh T and the target mesh D, using fully automatic geometric methods.
2. Use deformable registration with the geometric landmarks from step 1 for initialization to find a dense correspondence from T to D.
3. Transfer the 20 manually-marked established landmarks from T to the corresponding points of D to obtain final landmarks.

### *4.2.1 Generating Initial Landmarks*

The geometric method for automatic detection of initial landmarks requires that the head is normalized to face forward, the origin point lies in the center of the head, and the data are in $(x, y, z)$ coordinates, as shown in Fig. 4.1. The face width is defined along the $x$-axis, the height along the $y$-axis, and the depth along the $z$-axis. Our geometric methodology can




(a) Nose curve (b) Sharp edges (c) Nose points

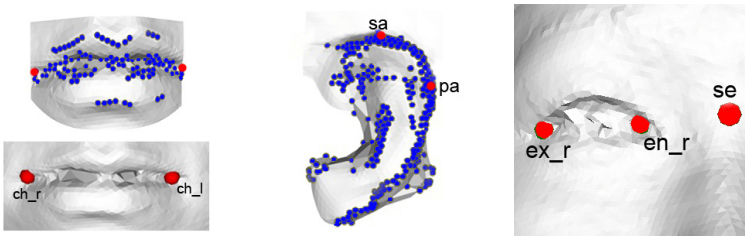
Figure 4.2: Initial landmarks on nose.

find 17 landmarks automatically, including seven nose points, four eye points, two mouth points and four ear points, marked with an asterisk in Fig. 4.1. Of the 17, ten are established landmark points and seven more (the sellion, the right and left alar curvature, right and left superaurale, right and left postaurale) are support points that help the deformable matching procedure to find a better correspondence from the template to the target.

*Nose Landmarks*

For each 3D head, there is a set of points $Z_{max}$ at the maximum $z$-value. The geometric center of these points ($x_{prn}$,$y_{prn}$) is the pronasale (prn). The sellion (se) and subnasale (sn) can be found as the local minima on either side of the pronasale on the line with the same $x$-value as the pronasale, as shown in Fig. 4.2(a). (Note that the nasion, an established landmark, is easy to find on skull CT data, but impossible to localize on 3D mesh data, so we find the sellion instead and leave it to the deformable matching to estimate the nasion.) To find the left and right alare and alar curvature, the region is restricted to $y_{sn} < y < y_{prn}$. The left and right alare (al_l & al_r) are located by calculating the surface normal angles of the points. Given the surface normal vector $n(n_x, n_y, n_z)$ of a point in the region, the point with the largest $n_x$ value has a surface normal pointing toward the left side of the face. This point is selected as al_l, and the point with the smallest $n_x$ value is selected as al_r (Fig. 4.2(c)).

Given the surface normal vectors of two neighboring triangles, the angle between these two vectors is a dihedral angle. All the points on the edges with dihedral angles larger than




(a) Mouth points (b) Ear points (c) Eye points

Figure 4.3: Initial landmarks on mouth, ears, and eyes.

30° make up the set of sharp edge points S shown in Fig. 4.2(b).

Two more nose points, the left and right alar curvature (ac_l & ac_r), are defined by equations (4.1) and (4.2). The ac_l point is selected as the leftmost point (with the largest x-value) of the region, and the ac_r is the rightmost point (with the smallest x-value) as shown in Fig. 4.2(c).

$$ac\_l = \{(x,y,z) | \arg\max_{x,y,z} x, y_{sn} < y < y_{prn}, (x,y,z) \in S\} \tag{4.1}$$

$$ac\_r = \{(x,y,z) | \arg\min_{x,y,z} x, y_{sn} < y < y_{prn}, (x,y,z) \in S\} \tag{4.2}$$

Both points are selected from sharp edge point set $S$, lower than the pronasale and higher than the subnasale.

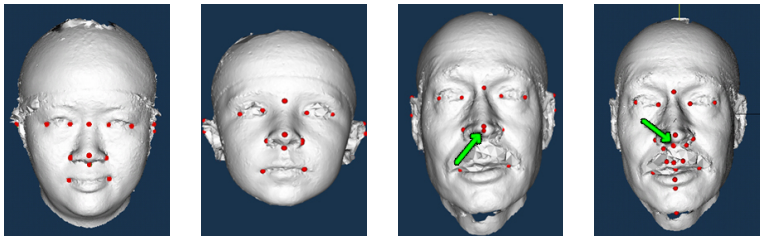
*Mouth, Ear and Eye Points*

As shown in Fig. 4.3(a), to locate the left and right chelion (ch_l & ch_r) points, the mouth is extracted by first restricting analysis to the region where $y < y_{sn}$. Among the points on sharp edges in this region, we select the point with the largest x-value as ch_l and the point with the smallest x-value as ch_r, as defined in (4.3) and (4.4).

$$ch\_l = \{(x,y,z) | \arg\max_{x,y,z} x, y < y_{sn}, (x,y,z) \in S\} \tag{4.3}$$

$$ch\_r = \{(x,y,z) | \arg\min_{x,y,z} x, y < y_{sn}, (x,y,z) \in S\} \tag{4.4}$$

Two points on each ear are used to help define the eye position and registration: the point



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

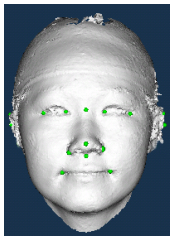
Figure 4.4: Initial landmarks generated by the geometric method. (a)(b) show good results while (c) is a poor result which is improved by deformable registration. (d) is the result after deformable registration for (c), the subnasale (on nose tip in (c)) is largely improved (below nose in (d)).

with the largest y value as superaurale (sa) and the point with the smallest z value as the postaurale (pa) on the outline of the ear (Fig. 4.3(b)).

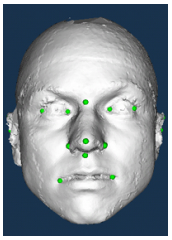
To locate the endocanthion (en_l) and exocanthion (ex_l) points on the left eye, the region $E_l$ is restricted to $y_{pa} < y < y_{se}$ along the $y$-axis and $x_{se} < x < 0.8x_{pa}$ along the $x$-axis. Then ex_l and en_l can be found as those points with the least Euclidean distance to the corner with largest $x$-value and smallest $y$-value (the right-back-most point) and smallest $x$- and $z$-values (the left-back-most point), respectively, among all the candidates on the sharp edges. The ex_r and en_r points can be detected similarly as shown in Fig. 4.3(c). All the geometric restrictions are based on the relationship of points to already-detected landmarks. Thus the pose normalization step [56] is critical to the success of the geometric methods. For most individuals, including adults (Fig. 4.4(a)) and children (Fig. 4.4(b)), the geometric method produces good initial results. In some cases, such as in Fig. 4.4(c), where the subnasale is mistakenly located because of the mustache in the upper lip area, initial results can be improved by the following deformation step.

### *4.2.2 Deformable Registration*

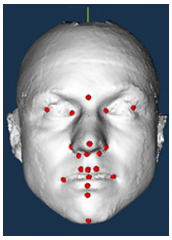
The purpose of this step is to improve the accuracy of the initial landmarks and to add additional landmarks that cannot be easily detected on meshes. Given a template 3D mesh T and a target 3D mesh D with sets of geometrically-generated initial landmark points $I_T$



(a) T with initial landmarks



(b) D with initial landmarks



(c) D with full landmarks

Figure 4.5: The deformable registration process: (a) and (b) show T (target) and D (destination) with initial landmarks $I_T$ and $I_D$ respectively. (c) shows D with transferred landmarks, which is the final result of our method.

and $I_D$, respectively (Fig 4.5(a) and (b)), deformable registration is a process that can be applied to find a dense correspondence from mesh T to mesh D. Due to its speed, we used the deformable registration method of Allen *et al.* [54], initialized by the correspondences between the points of $I_T$ and the corresponding points of $I_D$. After completion, all landmarks marked by an expert on the template mesh T are transferred to the corresponding points on the target mesh D, and these become our final landmarks. Note that the number of final landmarks is determined by the number of the landmarks on the template mesh, and the final landmarks need not have any relation to the initial landmarks.

The deformable registration algorithm uses an optimization framework to find a set of transformations that move all the points in T to a deformed surface that matches well with D, minimizing an energy function

$$E = \alpha E_d + \beta E_s + \gamma E_m \tag{4.5}$$

whose terms include the data error $E_d$ (how well the deformed template matches the target), the smoothness error $E_s$ (of the deformation) and the marker error $E_m$ (how well the initial landmarks match). The process is iterative. In early iterations, the marker error contributes more to the global optimization. As the process moves towards the end of the registration, the data error dominates the transformation. See [54] for details. Once the point to point correspondence is established, the landmarks can be transferred from the template to the

target mesh, as shown in Fig. 4.5 (c).

Only one template mesh was used in the experiments reported here. We did try several different template meshes with similar results.

Table 4.1: Average distances and the standard deviations of automatically generated landmarks compared with the ground truth given by medical experts

| Point Name | Geometric Method Average Distance(mm) | Deformable Method Average Distance(mm) |
|---|---|---|
| n | —— | 3.13±2.31 |
| prn | 1.54±1.68 | 1.72±1.11 |
| sn | 3.25±2.16 | 2.38±1.61 |
| al_r | 3.24±2.61 | 2.07±2.24 |
| al_l | 3.14±2.41 | 2.69±3.15 |
| ls | —— | 2.62±2.15 |
| sto | —— | 2.87±3.90 |
| li | —— | 2.27±1.41 |
| sl | —— | 3.29±2.87 |
| sbal_r | —— | 2.09±3.06 |
| sbal_l | —— | 2.46±2.93 |
| cph_r | —— | 2.30±3.27 |
| cph_l | —— | 3.67±1.03 |
| ch_r | 3.14±2.41 | 3.08±2.14 |
| ch_l | 2.80±2.38 | 3.06±1.64 |
| gn | —— | 5.21±3.54 |
| en_r | 4.78±4.45 | 3.37±4.09 |
| en_l | 4.58±4.70 | 3.78±4.50 |
| ex_r | 3.15±4.21 | 4.40±5.63 |
| ex_l | 3.73±4.86 | 4.96±5.91 |

### *4.3 Results*

In order to validate our automatic landmark results, we compared the landmark positions with the landmarks manually placed by the experts. The average distances from our landmarks to the ground truth given by medical experts are shown in Table 4.1 with their means and standard deviations. Recall that the automatic landmarks are computed based on the shape information of the heads without textures, while the experts used the color/texture data.

The average distance of the 10 points generated by the geometric method alone to the expert points is 3.30 mm. After the deformable registration with one template to all targets, the average distance of these 10 landmarks reduces to 3.14mm, with a smaller standard deviation for seven out of the 10 points (though a few points increase in average error). The

final results with all 20 points have an average distance of 3.07 mm to the ground truth. Results in the literature on similar data are much worse with average distances ranging from 5 to 13 mm, as shown in Table 4.2. Note that none of the other methods attempted to find the full set of 20 landmarks.

Table 4.2: Average distances (mm) and the standard deviations of our method and methods in the literature.

| Point Name | Our method | Yu [50] | Nair [52] | Lu [47] | Colbry [48] | Perakis [46] |
|---|---|---|---|---|---|---|
| prn | 1.7±1.1 | 2.2±6.8 | 8.8 | 8.3±19.4 | 4.1±5.1 | 4.9±2.4 |
| ch_r | 3.1±2.1 | —— | —— | 6.0±16.9 | 6.9±8.6 | 5.6±4.3 |
| ch_l | 3.1±1.6 | —— | —— | 6.2±17.9 | 6.7±9.3 | 6.4±4.2 |
| gn | 5.2±3.5 | —— | —— | —— | 11.0±7.6 | 6.0±4.3 |
| en_r | 3.4±4.1 | 4.7±9.8 | 12.1 | 8.3±17.2 | 5.5±4.9 | 5.1±2.5 |
| en_l | 3.8±4.5 | 5.6±16.1 | 11.9 | 8.2±17.2 | 6.3±5.0 | 5.5±2.6 |
| ex_r | 3.1±5.6 | —— | 20.5 | 9.5±17.1 | —— | 5.8±3.4 |
| ex_l | 5.0±5.9 | —— | 19.4 | 10.3±18.1 | —— | 5.7±3.5 |

### *4.4 Summary*

In this chapter, an improved method was introduced to automatically detect landmarks from 3D human facial data. First, geometric information was used to locate 17 prominent points. Then a deformable transformation between target mesh and data mesh determined 20 established landmarks and located them more accurately than with the geometric method alone. Our method has an average error of 3.07 mm over a sample of 994 heads, and is superior to prior published methods. The follow-up application of this work is to determine the mid-facial reference plane using the automatically detected landmarks and some landmark-based descriptors, which will be discussed in detail in chapter 6 and chapter 7.

# Chapter 5

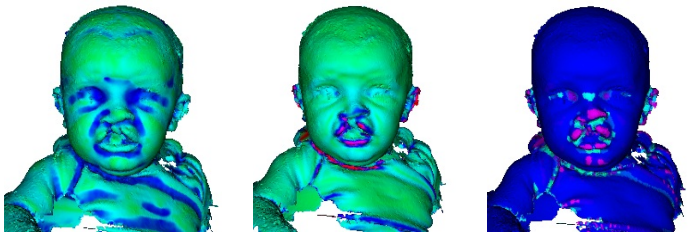
# LEARNING LANDMARK-RELATED REGIONS

Detection of "interesting" regions has become an important aspect of computer vision systems for use in object recognition and image matching [57, 58]. The interesting regions in this thesis are related to facial landmarks. Chapter 4 has already introduced some existing techniques that use a trained model to detect landmarks. In this chapter, instead of detecting landmarks, the purpose of learning from a trained model is to find regions that are related to landmarks, and use these regions to perform further analysis. Two tasks, finding the mid-face reference plane and rotating the data to get a frontal position, are introduced in the latter part of this chapter.

A surface mesh, which represents a 3D face, consists of points $p_i$ on the objects surface and information regarding the connectivity of the points. Our framework starts by executing two phases: low-level feature extraction (Section 5.1) and multi-scale histogram (Section 5.2). The low-level feature extraction starts by applying a low-level operator to every point on the surface mesh. After the first phase, every point $p_i$ on the surface mesh will have several single low-level feature values. The second phase computes a vector of values for a given neighborhood of every point $p_i$ on the surface mesh. Then the features are used for training and detecting landmark related regions.

## *5.1 Low-level Operators*

Our methodology starts by applying a low-level operator to every point on the surface mesh. The low-level operators extract local properties of the surface points by computing a single feature value $v_i$ for every point $p_i$ on the mesh surface. In this work, shape index, curvedness and Besl-Jain curvature value are used for our experiments. Figure. 5.1 (a), (b) and (c) show examples of the shape index, curvedness and Besl-Jain curvature values of a 3D model, respectively.



(a) Shape index (b) Curvedness (c)Besl-Jain

Figure 5.1: Local properties of the surface points. (a) The shape index of surface. Green means low and blue means high in shape index. (b) The curvedness of each point on one mesh. High curvedness is shown in red and low in green. (c) The Besl-Jain surface value on the surface. Blue means plane surface, dark purple means peak, light purple means pit, and cyan means saddle ridge.

Shape index is a quantitative measure of the shape of a surface at a point [59]. It gives a numerical value to a shape, thus making it possible to mathematically compare shapes and categorize them. Shape index is defined as follows:

$$s = \frac{1}{2} - \frac{1}{\pi} \times arctan(\frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2}) \tag{5.1}$$

where $\kappa_1$ and $\kappa_2$ are the principal curvature values at point $p_i$ and $\kappa_1 \geq \kappa_2$. Figure. 5.1 (a) shows the shape index of each point on one mesh.

The curvedness low-level feature value is derived from the shape index properties of the surface points [59]. The curvedness properties specify the 'intensity' of a surface curvature. The curvedness is defined as the distance from the origin in the ($\kappa_1$,$\kappa_2$) plane. The curvedness $c$ for a point $p$ is calculated as:

$$c = \sqrt{\frac{\kappa_1^2 + \kappa_2^2}{2}} \tag{5.2}$$

Figure. 5.1 (b) shows the curvedness of each point on one mesh. High curvedness is shown in red and low in green.

Besl and Jain [27] suggested surface characterization of a point $p_i$ using only the sign of the mean curvature $H$ and the Gaussian curvature $K$. These surface characterizations

Table 5.1: Besl-Jain Surface Characterization

| Label | Category | H | K |
|---|---|---|---|
| 1 | saddle valley | $H > 0$ | $K < 0$ |
| 2 | minimal | $H = 0$ | $K < 0$ |
| 3 | saddle ridge | $H < 0$ | $K < 0$ |
| 4 | valley | $H > 0$ | $K = 0$ |
| 5 | plane | $H = 0$ | $K = 0$ |
| 6 | ridge | $H < 0$ | $K = 0$ |
| 7 | pit | $H > 0$ | $K > 0$ |
| 8 | peak | $H < 0$ | $K > 0$ |

result in a scalar surface feature for each point that is invariant to rotation, translation and changes in parametrization. The eight different categories are: (1) saddle valley, (2) minimal surface, (3) saddle ridge surface, (4) valley surface, (5) plane surface, (6) ridge surface, (7) pit surface, and (8) peak surface. Table 5.1 lists the different surface categories with their respective curvature signs. Figure. 5.1 (c) shows the Besl-Jain surface value of each point on one mesh. The eye inner conners are a pit surface while the nose tip is a peak.

### *5.2 Multi-scale Histogram*

Next, instead of just using one value for a point $p_i$, a local histogram is used to aggregate the low-level feature values of each point. The histograms are computed by taking a neighborhood around the point and accumulating the single-valued features in that neighborhood. The size of the neighborhood is $r$, which is a key parameter of the histogram. With different sizes of the neighborhood, the histograms could appear very different. In order to capture the histogram shape properties when $r$ differs, all the histograms with multiple sizes of neighborhoods are combined to form a large feature vector. The histograms with multiple sizes of neighborhoods are then combined together to form another large feature vector. Note that the three single-valued low-level features and the multiple scale histogram are all rotation invariant. This ensures that the eye and nose candidates can be selected despite the rotation of the original data.

After that, we taught a classifier the characteristics of points that are near landmarks. We tried several landmarks models, including eight midline points (n, prn, sn, ls, sto, li,

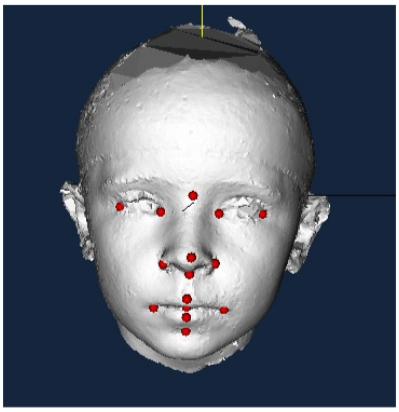
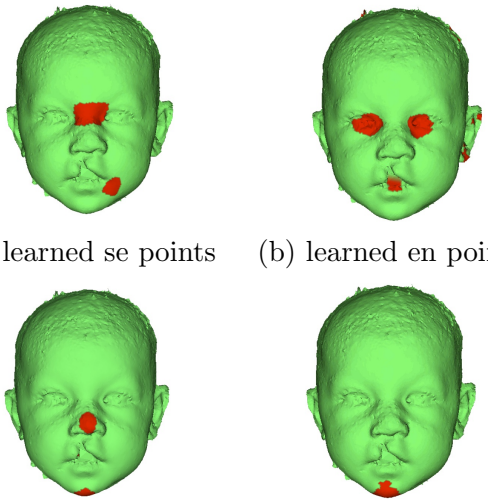


Figure 5.2: The 3D head mesh data with landmarks labeled by medical experts.

sl and gn) and four paired landmarks (en_l, en_r), (ex_l, ex_r), (ac_l, ac_r), (ch_l, ch_r), as shown in Fig. 5.2. Histograms of low-level features were used to train a Support Vector Machine (SVM) classifier [60] to learn these 12 points on the 3D surface mesh. We used the SVM implemented in WEKA for our experiments [61]. The training data for supervised learning for the classifiers was obtained by manually marking the landmark points on the surface of each training object. The histogram of low-level features of each of the marked points was saved and used for the training.

A small training set, consisting of 40 head meshes from the 3D facial norms dataset was used to train the classifier to learn the characteristics of the landmark points in terms of the histograms of their low-level features. After training was complete, the classifier was able to label each of the points of any face mesh as either a specific landmark or not, and provided a confidence score for its decision. A threshold $T_{score}$ was applied to the confidence scores for the landmarks. In our experiments, we used $T_{score} = 0.98$ to keep only the points with high confidence scores from the classifiers.

From our experiments, the inner eye corners (en), the nose tip (prn), the nose bridge (n) and the chin (gn) are the most reliable landmark models. Figure 5.3 shows the points with scores larger than $T_{score}$ from the se, en, prn and gn classifiers. The points are colored according to the classifier confidence score assigned to the point. Non-landmark points are colored in green, while points selected for specific landmarks are colored in red.



(a) learned se points (b) learned en points

(c) learned prn points (d) learned gn points

Figure 5.3: Positive results from four trained landmark models.

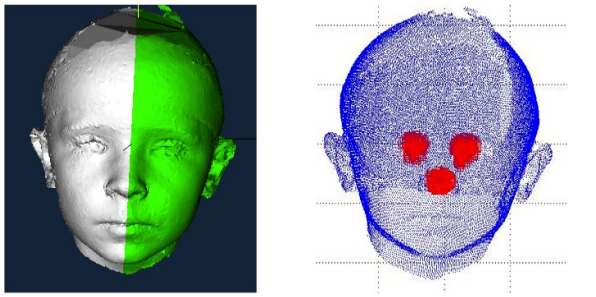
### *5.3 Learning the Mid-facial Reference Plane*

Here the learned landmark-related regions are used and a symmetry model is trained to find the mid-facial reference plane. The main purpose of this step is to find a most suitable reference plane for measuring facial asymmetry of children with cleft lip. So we ruled out the part near the mouth and nose tip, where the cleft lip deformation mainly occurs, and used only the landmark regions learned from models related to inner eye corners (en), nose bridge (se) and chin (gn).

After the learning step that uses multi-scale histograms to find regions, the second training step, which is used to build the symmetry model, is based on the interesting regions. The points obtained from the last step are extracted and used to form connected components. Components of less than 10 points are removed. For each component $m$, the center coordinates $C_m$ are extracted, as well as the eigenvalues $\lambda_{m1}$, $\lambda_{m2}$ and $\lambda_{m3}$ of the matrix formed by the 3D coordinates of all the points in component $m$. Eigenvalues are used because they reflect the size of the interesting regions and are invariant to rotation.

In the training set, the standard symmetry plane $SSP$ is determined from the 24 ground-truth landmarks. This plane is selected by minimizing the distance of the plane to each

of the single landmarks and to the average of each of the paired landmarks as shown in Figure 5.4(a).



(a) $SSP$ (b) positive components

Figure 5.4: Training symmetry model. (a) Standard symmetry plane ($SSP$) ground truth. (b) A positive single component and a positive pair of components.

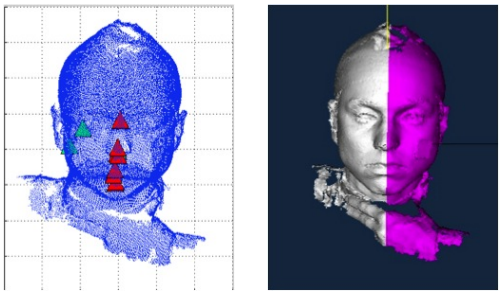
A single component which lies on the mid-facial reference plane and pairs of components that are approximately symmetric to the mid-facial reference plane are marked as positive samples, which means they are suitable for use in symmetry plane construction. The rest of the components are negative samples. Figure 5.4(b) shows a positive single component and a positive pair of components.

The feature vector, $Csingle$, to describe single component $m$ includes the number of points in component $m$, as well as the 3 eigenvalues. The feature vector, $Cpair$, for paired components $m$ and $n$ is designed to capture the difference between these two components. It contains the distance between their areas, eigenvalues, and centers. Using the feature $Csingle$ along with positive and negative samples for useful single components, a REPTree classifier was trained to learn if a component is suitable for construction of the mid-facial reference plane. Similarly, using the features $Cpair$ and positive and negative samples for useful pairs, another REPTree classifier was trained to learn if a pair of components is suitable.

Then for new data classification, the steps to identify useful components are:

1. Compute the multi-scale histogram composed of low-level operator values for original data.

2. Use the en, se and gn landmark models to get a landmark-related confidence score for every point on the mesh.
3. Pick all points with score larger than $T = 0.98$ and form components of 10 or more points. Compute $Csingle$ and $Cpair$ for all components.
4. Run the model for useful single components and the model for useful pair components to label components as useful singles, useful pairs, or neither.
5. Accumulate the useful singles and useful pairs for all the landmark models.



(a) Centers of useful components (b) Computed symmetry plane

Figure 5.5: Centers of useful singles and useful pairs and the mid-facial reference plane. (a) RANSAC prunes out false positives (green). (b) the mid-facial reference plane is shown on an original 3D mesh.

After all the useful components have been classified, a symmetry plane is determined that satisfies two criteria: 1) the plane should be perpendicular to the line connecting each useful pair and lie along the centers of useful pairs. 2), the plane should be close to the centers of useful single components. The plane $P$ was chosen to minimize the function $f$

$$f = \sum D(P, C_s) + \sum D(P, C_p) + \gamma Angle(P, C_p) \tag{5.3}$$

where $D(P, C_s)$ is the distance of plane $P$ to the centers of useful single components, $D(P, C_p)$ is the distance of plane $P$ to the centers of useful pairs of components, $Angle(P, C_p)$ stands for the angle between the normal of plane $P$ and the vector connecting a useful pair of components, and $\gamma$ is a weight. In order to obtain the optimal plane, the RANSAC [62]

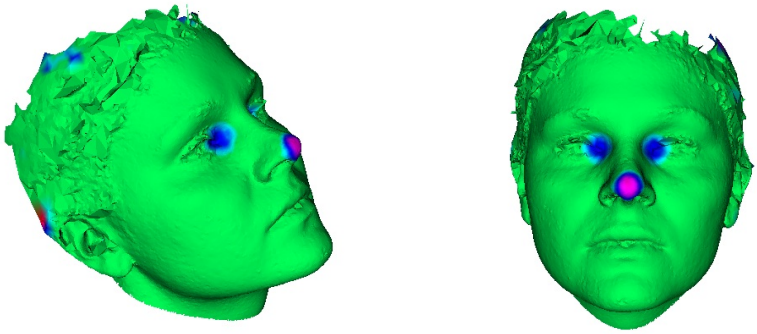
algorithm was applied to get rid of false positive useful components, which are not consistent with the others. (See Figure 5.5.)

### *5.4 Detecting the Eye-nose Triangle Area*

Another application for using landmark related regions is to find the eye and nose triangle area, and use that area to rotate the data so that the face is in a frontal position. The inner eye corner and nose tip are chosen for this task because they can be reliably detected.

Although the classifiers for eye and nose are very powerful, there are still false positive regions in the predicted results. We developed a systematic way to find the eye-nose-eye triangle area. First, the candidate points from both classifiers are grouped to form regions. Second, small regions ($< 50$ points) are removed. Third, a pair of eye regions that have a similar number of points, along with a nose region that lies almost equidistant from those two eye regions are picked. Last, some geometric thresholds are used to rule out regions that are too small or too big.

After the eye-nose triangle area is determined, the head is rotated so that the eye regions are leveled and symmetric and the nose appears to be directly underneath the eye center, as shown in Fig. 5.6.



(a) Initial position with detected inner eye corner and nose tip

(b) Rotated frontal position with inner eye corner and nose tip

Figure 5.6: Eye-nose triangle area. The red color means nose tip and blue means inner eye corner detected by the classifiers.

### *5.5 Summary*

This chapter introduced the method for detecting landmark-related regions by a machine learning approach and its applications. Our method does not require pose-alignment processing and is invariant to rotation [63]. In chapter 6, we will compare the methods for finding the mid-facial reference plane in detail. In chapter 9, the method for detecting the eye-nose triangle area and rotating the face to a frontal position is used to produced a frontal face screenshot for the subsequent steps to crop out the face from the raw data.

# Chapter 6

# DETERMINING THE BEST MID-FACIAL REFERENCE PLANE

## *6.1 Background*

In order to measure the most important face feature for children with cleft lip—symmetry—we needed to define a plane across which asymmetry will be calculated. In a perfectly symmetric face this plane would be the plane of symmetry and would equally divide the two halves, but in a face with cleft abnormalities, the asymmetry of the nasolabial region will alter the plane of symmetry relative to the other part of the face and consequently, it is hard to define a mid-facial reference plane.

Multiple automatic computer-based approaches have been developed. One method was introduced by Benz *et al.* [26], in which the original data is mirrored at an arbitrary plane. Then the original mesh and the mirrored mesh are registered using the iterated-closest-point algorithm [27]. In this dissertation, this method is referred as *the mirror method*. In the second method, referred to as *the a-lmk method* and descried in chapter 4, 20 landmarks are automatically located by a deformable registration algorithm from a template mesh to a target mesh, which is initialized by a geometric point detector [64]. After these landmarks are found, the mid-facial reference plane is calculated using only the landmarks on the eyes and chin area. The last method takes two steps of processing: landmark-related region detection and mid-facial reference plane computation using these regions, as described in chapter 5. It is referred as *the learning method*.

In additional to the three automatic methods, two sets of mid-facial reference planes were provided manually by two craniofacial specialists to be used as ground truth for the automatic methods and further comparison. One (*the m-lmk method*) is based on landmarks, in which the medical experts provide the landmarks on cleft patient data, and the mid-facial reference plane is calculated using the eyes and chin landmark positions. In the second one (*the direct method*), the mid-facial reference plane was placed directly on the 3D mesh data

Table 6.1: Brief summary of the 5 approaches for computing mid-facial reference plane

| Method | Approach | Details of the approach |
|---|---|---|
| mirror | Computer | mirror the data and register the mirrored data and original data |
| a-lmk | Computer | calculated using landmarks from deformable registration |
| learning | Computer | mid-facial reference plane models learned from training data |
| m-lmk | Manual | calculated using landmarks from medical expert |
| direct | Manual | directly put on 3D mesh data |

by an expert. A brief summary of these five methods is given in Table 6.1.

There are two parts of the evaluation: comparing the computer-based methods' results to the ground truth, and a survey to evaluate which method is the best for finding the mid-facial reference plane. This study is applied to the cleft survey dataset, including 50 individuals from 3 to 6 months old. 35 of them have unilateral cleft lip, 10 have bilateral cleft lip and 5 comprise a control group.

### *6.2 Comparing to the Ground Truth*

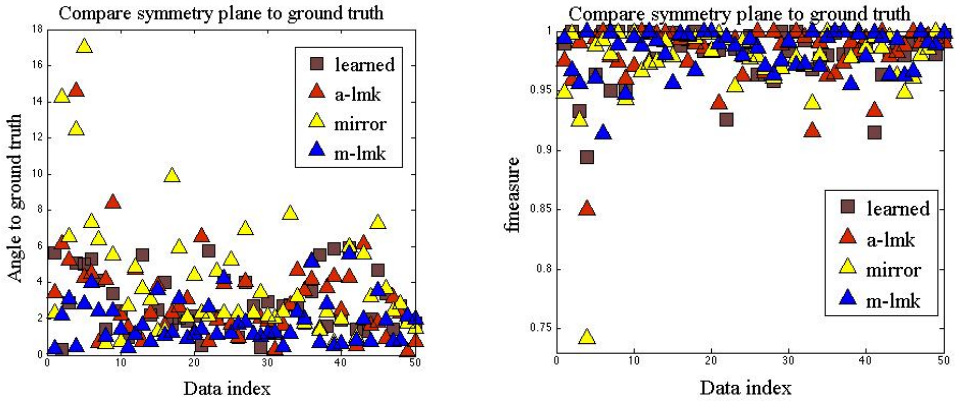
In this section we use the plane provided by the direct method as the ground truth and compare the planes calculated using the computer-based methods, as well as estimated from the landmarks given by the experts, to this ground truth plane. Two metrics are used to evaluate the results of these four methods (the m-lmk method, the mirror method, the a-lmk method and the learning method). The first metric is the angle $Dif_a$ between the ground truth plane and the other plane, where a score of 0 means the best match. The second metric F-measure reflects how the planes from the ground truth and the other method agree with each other on a point-by-point basis. The closer the F-measure value is to 1, the closer the result of that method is to the ground truth. Figure 6.1 gives the comparisons of both the angle and F-measure for the m-lmk method, the mirror method, the a-lmk method and the learning method. Most of the planes are within $10^\circ$ of angle difference to the ground truth. On the point-to-point comparision base, most planes agree more than 95% with the ground truth. The other human-based method, the m-lmk method, gives the closest result to the ground truth. The mirror method has a few data that are significantly different from the

Table 6.2: The average of angle differences and F-measure

| Method | m-lmk | mirror | a-lmk | learning |
|---|---|---|---|---|
| $Dif_a$ | 1.793 | 3.088 | 3.067 | 2.793 |
| F-measure | 0.983 | 0.976 | 0.982 | 0.980 |

ground truth.

Table 6.2 shows the average angle differences and F-measures for all the data in the cleft survey dataset. Out of the other four methods, the m-lmk method is the closest one to the direct plane, as it is another human-based method. It also shows that the experts are close to each other when determining the mid-facial reference plane. The a-lmk method performs the best out of all the three automated methods on F-measure at the average of 0.982. The learning method is the closest one out of three computer-based methods to the ground truth in terms of the angle difference, with an average of 2.79 °.




(a) $Dif_a$
(lower is better)

(b) F-measure
(higher is better)

Figure 6.1: Comparison of the m-lmk method, the learning method, the a-lmk method and the mirror method to the direct plane method (a) Angle differences (b) F-measure.

### *6.3 A Survey to Evaluate the Human and the Computer Based Methods*

Although the planes given the medical experts are treated as ground truth in the computer vision field, in our study, the medical experts were still curious about how good the human-based methods are, compared to the computer-based methods in a blind test. So we set up a survey to evaluate all five methods of defining a mid-facial reference plane for quantification of facial asymmetry on 3D images of children with and without cleft lip.

The survey was set up on a computer with a 27-inch display to ensure that the 3D images were big enough for the raters to examine all the details. Six raters (3 cleft surgeons; 2 craniofacial pediatricians; 1 craniofacial morphology researcher) examined six 3D images of each subject to determine relative success in approximating the mid-facial reference plane. Five 3D meshes represented the five planes from those five methods in a random order. One mesh with no results on it was placed on the top left for the raters to control the rotation of the head. The six 3D facial meshes of each individual rotated synchronously in order to allow for placement evaluation of the mid-facial reference plane.

The raters were given very specific instructions when taking the survey. They were asked to use the forehead, the eyes, the ears, the commissures, and the chin to give their thoughts. They were told not to evaluate this plane using the entire head or the nose and lips. For each subject, the raters were asked to rank them in the order that most closely approximates what they thought was is the mid-facial reference plane for measurement of facial asymmetry. They ranked all the results from 1 to 5, where 1 is the best approximation and 5 is the worst. The raters were also asked to rate the precision of planes according to a 1-7 numeric scale. These ratings evaluated the deviation (in degrees) from each rater's imagined best mid-facial reference plane, as well as the lateral displacement in millimeters on the level of the tip of the nose. A carefully designed rating scale and examples were provided when they were rating. The scale is in table 6.3.

#### *6.3.1 Ranking and Rating Scores*

Figure 6.2 shows one data mesh with five methods that are ranked and rated by two experts. When they took the survey, they could rotate and look at the 3D models and write down

Table 6.3: Rating scale for the survey

| Scale | | Define |
|---|---|---|
| 1 | Absolute match | |
| 2 | Probably | within 1° of rotation and 1/128 facial width of translation |
| 3 | Very close | within 2° of rotation and 1/64 facial width of translation |
| 4 | Slightly off | within 3° of rotation and 1/32 facial width of translation |
| 5 | Moderately off | within 5° of rotation and 1/16 facial width of translation |
| 6 | Severely off | within 10° of rotation and 1/8 facial width of translation |
| 7 | Unacceptable | greater than 10° of rotation or 1/8 facial width of translation |

their rank for each method. Then they compared them to the examples we provided and checked the rating score.

After gathering all the ranking and rating scores from the six raters, the average scores for each method and for each sub group were calculated. Table 6.4 shows the average ranking score for each method. Two human-based methods, the direct method and the m-lmk method, are ranked at the first and second place on average. Out of three computer-based methods, the a-lmk method was ranked as the best method for all the data and also for each sub group. Comparing the ranking score for the a-lmk method and two human-based methods, the direct method and the a-lmk method, the Friedman test scores are 0.1 and 0.4, respectively. These scores are larger than 0.05, which means there is no statistically significant difference between the a-lmk method and the two human-based methods. Comparing to the other two computer-based methods, the Friedman test scores are 0.00 and 0.00, which means the a-lmk method is statistically significantly better than those two methods in terms of ranking. Table 6.5 gives the average rating score for each method. Again two human-based methods are rated as the best methods for all the data. The a-lmk method is rated as the best computer-based method for unilateral and bilateral cleft lip, but the mirror method gets the lowest rating score for the very small control group.

Figure 6.3 shows the rating and ranking distribution for five methods, in which the a-lmk gets more rank 1 than all the other methods and the mirror method gets rank 5 for the most. The m-lmk and direct methods get rating 1 more than the other three methods and very rarely get ratings lower than 6. Comparing the rating scores for the a-lmk method and

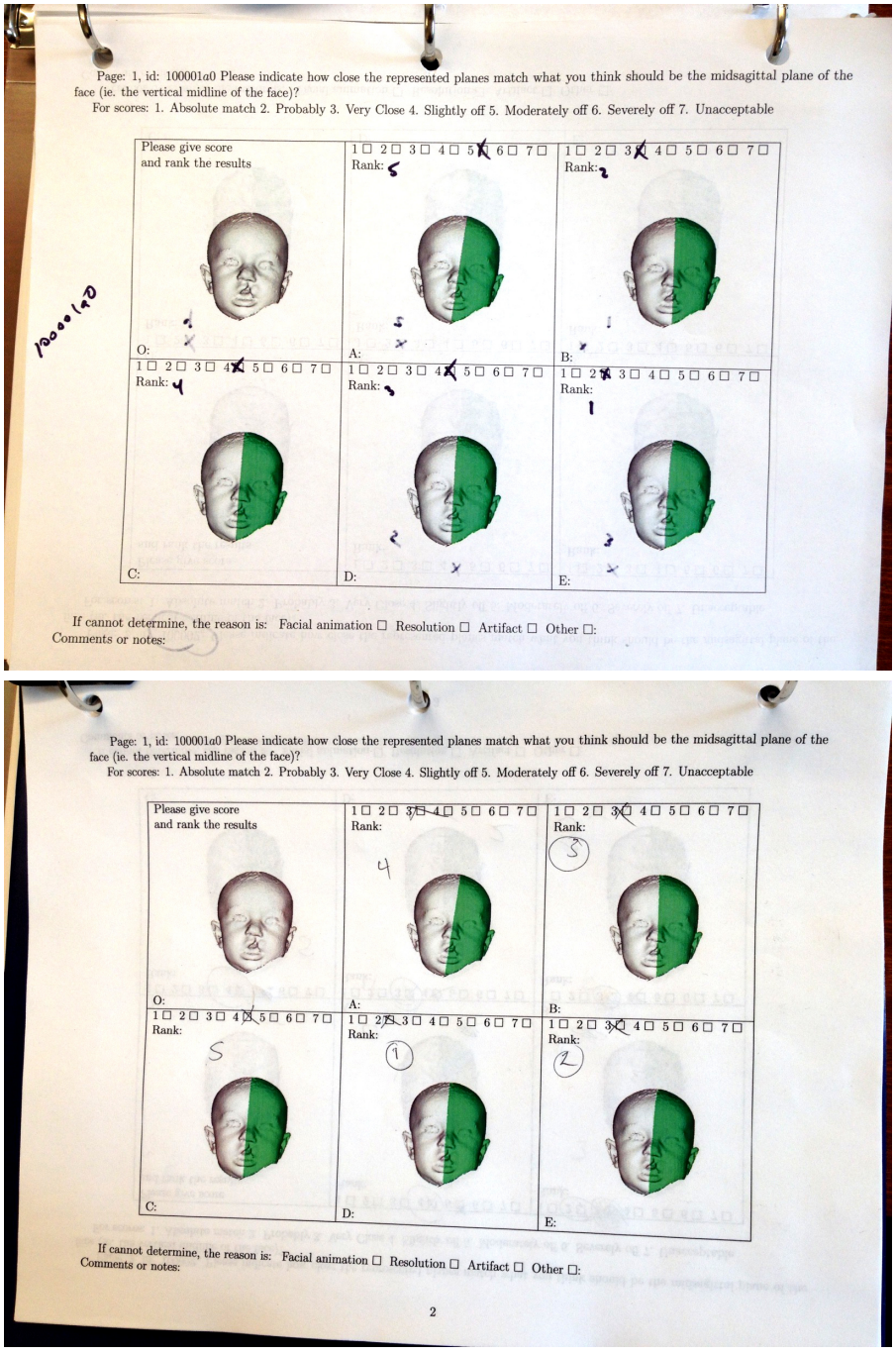



Figure 6.2: Ranking and rating score sheets from two experts for the same individual.

the human-based method, the distributions are very similar.

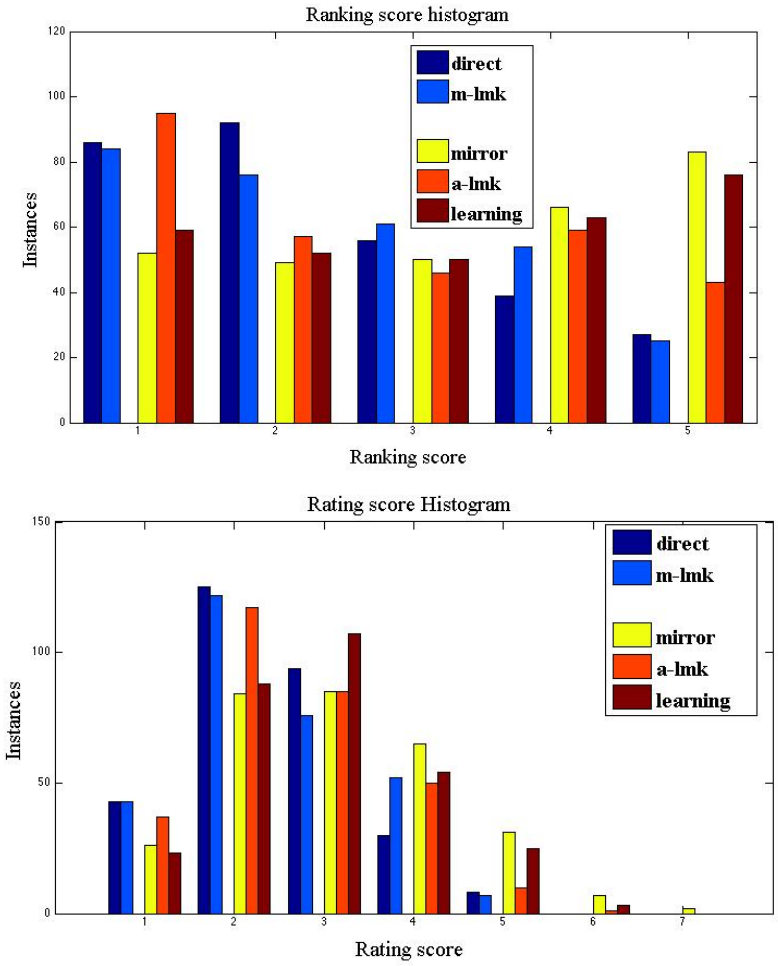



Figure 6.3: Ranking and rating distribution for all five methods.

Table 6.4: The average ranking score for all methods

| Method | direct | m-lmk | mirror | a-lmk | learning |
|---|---|---|---|---|---|
| All (50) | 2.43 | 2.54 | 3.27 | 2.66 | 3.15 |
| Unilateral (35) | 2.62 | 2.52 | 3.32 | 2.70 | 3.17 |
| Bilateral (10) | 2.03 | 2.79 | 3.44 | 2.66 | 3.36 |
| Control (5) | 1.90 | 2.13 | 2.47 | 2.36 | 2.63 |

Table 6.5: The average rating score for all methods

| Method | direct | m-lmk | mirror | a-lmk | learning |
|---|---|---|---|---|---|
| All (50) | 2.45 | 2.53 | 3.07 | 2.61 | 2.93 |
| Unilateral (35) | 2.60 | 2.57 | 3.15 | 2.63 | 2.99 |
| Bilateral (10) | 2.16 | 2.56 | 3.23 | 2.64 | 3.04 |
| Control (5) | 2.00 | 2.20 | 2.17 | 2.37 | 2.30 |

### *6.3.2 Inter-rater Reliability*

Inter-rater reliability is determined using Pearson correlation coefficients for all raters. Table 6.6 shows the Pearson correlation coefficients for rating scores provided by the raters. The number at column two and row one means the correlation coefficient between rater one and rater two. If the correlation coefficient is larger than 0.5, it means the scores given by the two raters are correlated. Otherwise it means no correlation between them. At an average of 0.43, the reliability between the raters is poor even though they were given very specific instructions. The reason behind this is that all the planes are vey similar to each other. Recall the angle differences in Fig. 6.2, the $Dif_a$ are very small. Most of them get an angle difference less than $10\,^{\circ}$. When the differences are so subtle and the planes are so similar to each other, it is very hard for human eyes to differentiate among them. Figure 6.4 illustrates the average Pearson correlation coefficient rises up when the angle difference are enlarged. When only looking at the planes that are different from the direct method for more than $6\,^{\circ}$, the average of Pearson correlation coefficients goes up to 0.66 as shown in table 6.7.

Table 6.6: Pearson correlation coefficients for inter-rater reliability in terms of rating score

| Rater | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 1 | 0.48 | 0.41 | 0.46 | 0.50 | 0.48 |
| 2 | | 0.53 | 0.35 | 0.50 | 0.51 |
| 3 | | | 0.18 | 0.37 | 0.46 |
| 4 | | | | 0.38 | 0.35 |
| 5 | | | | | 0.50 |

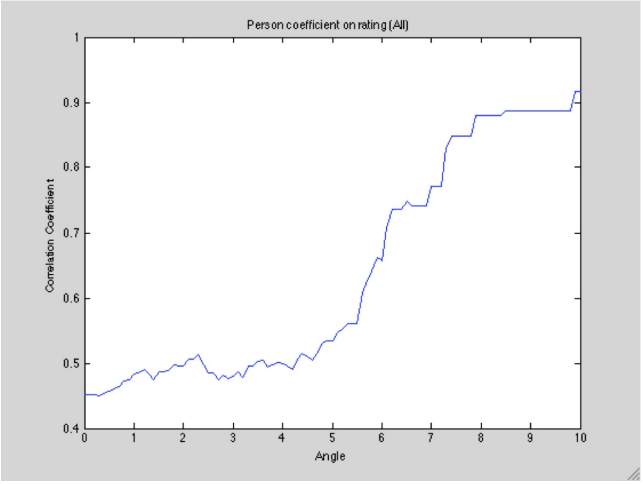



Figure 6.4: The average Pearson correlation coefficient for rating score increases when the angles between the planes increase.

Table 6.7: Pearson correlation coefficients for rating score when angle difference larger than 6° to the ground truth

| Rater | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 1 | 0.50 | 0.53 | 0.84 | 0.50 | 0.82 |
| 2 | | 0.76 | 0.70 | 0.69 | 0.56 |
| 3 | | | 0.64 | 0.69 | 0.64 |
| 4 | | | | 0.68 | 0.73 |
| 5 | | | | | 0.57 |

## *6.4 Summary*

This chapter compared three computer-based methods and two human-based methods to determine the best mid-facial reference plane of children with cleft lip. The two human methods are very reliable and correspond closely to each other. The three computer-based methods are very close to the ground truth, at the average angle difference around 3° and the average F-measure around 0.98. Out of three computer-based methods, the a-lmk method is the best in terms of F-measure comparing to the direct method, as well as medical experts' ranking and rating scores.

Chapter 7

# LEARNING TO RANK THE SEVERITY OF CLEFT LIP

The severity of cleft lip is widely variable and the results of treatment are influenced by the initial deformity. Objective assessment of severity would help to guide prognosis and treatment. However, most assessments are subjective. The purpose of chapter 7 and chapter 8 is to develop and test quantitative computer-based methods of measuring cleft lip severity. In this chapter, a grid-patch based measurement of symmetry is introduced, with which a computer program learns to rank the severity of cleft lip on 3D meshes of human infant faces. Three computer-based methods to define the mid-facial reference plane were compared to two human-based methods. Four different symmetry features were calculated based upon these reference planes, and evaluated. The result shows that the rankings predicted by the proposed features are highly correlated with the ranking orders provided by experts that were used as the ground truth.

## *7.1 Background*

For children born with cleft lip, surgical treatment can produce a dramatic change in appearance of the lip, However, stable correction of the nose remains a challenge, and treatment strategies continue to be debated. Given that the potential results of treatment are limited by the cleft severity, objective assessment of the deformity is important for prognosis and treatment outcomes.

Traditional evaluation of cleft deformities was relied on clinical description and landmark-based measurements that are taken directly with calipers. Neither of these is ideal, given that clinical descriptors are somewhat subjective, and anthropometric measurements on young infants are difficult and burdensome.

Our goal is to develop novel tools for analysis of shape in children with cleft lip. Specifically, we want these tools to be automated, computer-based, and quantitative. Given

that facial asymmetry increases directly with increasing cleft severity, this study focuses on quantifying nasolabial symmetry.

In this chapter, we present a system that learns to rank the severity of the abnormalities of the 3D infant faces with cleft. Our system uses a mid-facial plane as a reference to compute the difference between the left and the right side of the face according to four different features. After the differences are extracted, a machine learning algorithm takes the ranking orders provided by an expert as the ground truth to train a classifier to order the data according to the severity of the clefts.

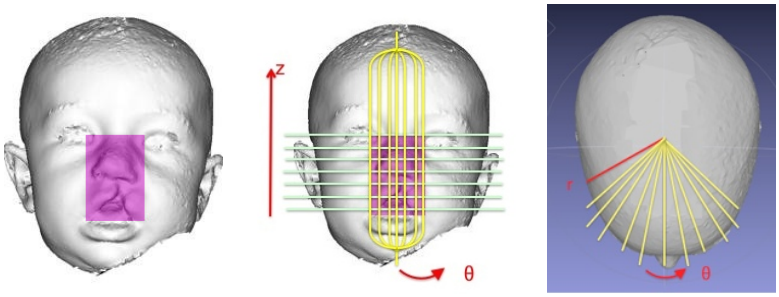
## *7.2 Ground Truth*

In this study we used 35 meshes from infants with unrepaired unilateral cleft lip and 5 normal infant controls. In terms of acquiring the ground truth, although rating specific facial features can produce variable results [65], ranking a group of subjects in a side-by-side comparison can be performed reliably [66]. In order to facilitate comparison and ranking of digital 3D images, we developed an interface that allows the user to freely shuffle 3D images and examine them in a side-by-side manner. Subjects were ranked in order of severity of the cleft lip nasal deformity by an expert cleft surgeon to serve as the ground truth.

## *7.3 Quantifying the Asymmetry of the Face*

The mid-facial reference plane we use in this chapter are the same as we described in chapter 6. The accuracy of this mid-facial reference plane is critical for all of the next steps, as we will show that the performance of different plane detection methods varies using the same features and the same ranking algorithms.

Using the mid-facial reference plane, several different symmetry measures are defined based on grid patches. A grid is placed over the face and it measures the difference between the left side and right side in terms of features of the corresponding grid patches. The features used are radius difference (from a central point), angle difference, curvature difference, and edge feature difference.




(a) Area selection (b) Front view (c) Top view

Figure 7.1: Selected area, grid patches and $r$, $\theta$ and $z$ directions

### *7.3.1 Area Selection*

Because the deformity mainly occurs in the nose and mouth area, the grid will be placed only in the center part of the face. Using mouth corners and a nose bridge point generated automatically by the method described in chapter 4, a rectangle area is cropped from the face as shown in Fig. 7.1(a).

### *7.3.2 Grid-Patch-Based Quantification*

Grid-patch-based quantifications divide the area selected from the face into several patches (as shown in Fig. 7.1). Each patch is represented by the average value of the points inside it. Half of the rectangle area is divided into $M$ by $M$ squares, equally divided in the $z$ and $\theta$ directions, as shown in Fig. 7.1(b). In our experiments, $M = 10$.

Four differences are compared for each corresponding reflected patch pair: the radius difference ($RD$), the angle difference ($AD$), the curvature difference ($CD$), and the sharp edge difference ($ED$). The radius difference defined for a grid patch at position $(\theta, z)$ is

$$RD(\theta, z) = |r(\theta, z) - r(-\theta, z)| \tag{7.1}$$

where $r$ takes the average radius value in that grid patch, and $(-\theta, z)$ is the reflected grid patch of $(\theta, z)$ with respect to the mid-facial reference plane. This gives the actual surface

distance. The angle difference defined for grid patch $(\theta, z)$ is

$$AD(\theta, z) = cos(\beta_{v_{\theta,z}, v_{-\theta,z}}) \tag{7.2}$$

where $\beta_{v_{\theta,z}, v_{-\theta,z}}$ is the angle between the surface normal vector of the face mesh at grid patch $(\theta, z)$ and its reflected grid patch. This shows how differently the two patches are oriented. For curvature difference, the average Gaussian curvature of each grid patch is calculated, represented as $K(\theta, z)$. The Gaussian curvature difference is

$$CD(\theta, z) = |K(\theta, z) - K(-\theta, z)| \tag{7.3}$$

for grid patch $(\theta, z)$, where $K(-\theta, z)$ is the average Gaussian curvature value for the reflected grid patch. Last but not least, the sharp edge for one grid patch is defined as the ratio of points with a dihedral angle larger than a certain threshold angle $Th$ to the total number of points in the grid patch. The shape angle difference is defined as:

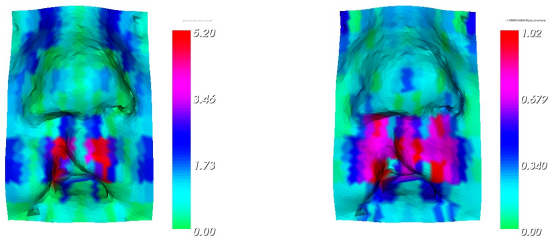
$$ED(\theta, z) = |\frac{\#points(\theta, z) > Th}{\#points(\theta, z)} - \frac{\#points(-\theta, z) > Th}{\#points(-\theta, z)}| \tag{7.4}$$

representing, in two corresponding reflected grid patches, the difference between the ratio of points with sharp edges.

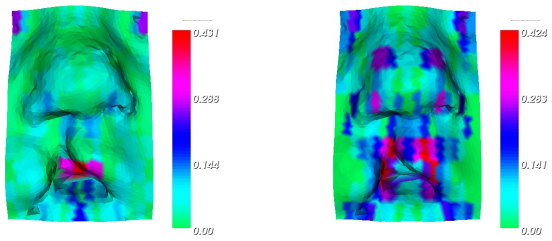
Given the local differences of $RD, AD, CD$, and $ED$, a vector of length $M \times M$ is formed to represent the asymmetry of the center part of the face. Figure 7.2 illustrates the four local differences with a $10 \times 10$ grid.

### *7.4 Learning to Rank*

The task of learning from the expert's ranking to compare the severity of face abnormalities falls into a learning-to-rank problem, which has been studied heavily with applications in many information retrieval problems, such as document retrieval, collaborative filtering, and computational advertising. Liu [67] categorized the algorithms to train a rank model into three groups: pointwise, pairwise, and listwise. In our study, we will compare two pointwise and two pairwise algorithms. The listwise algorithms are not used in this study because the



(a) Radius difference (b) Angle difference



(c) Curvature difference (d) Edge difference

Figure 7.2: Grid-patch-based asymmetry measurements. Red means a big difference between that grid patch and its reflecting patch. Green means a small difference.

algorithms are not suitable for such a small data set.

In this chapter, linear regression [68] and SVM regression [69] are used as the pointwise algorithms, and RankBoost [70] and RankNet [71] are used as the pairwise algorithms to learn how to rank the data. The pointwise methods approximate the problem as a regression problem: given a single instance, predict its score. The pairwise algorithms take the features of every pair instance and transfer the problem into a classification problem: learning a binary classifier that can tell which instance is better (higher rank) in a given pair of instances.

RankBoost trains the model in rounds. It starts with all pairs being assigned an equal weight. At each round, the learner selects the weak ranker that achieves the smallest pairwise loss on the training data with respect to the current weight distribution. Pairs that are correctly ranked have their weights decreased and those that are incorrectly ranked have their weights increased so that the learner will focus more on the hard samples in the

next round. The final model is essentially a linear combination of weak rankers. Weak rankers theoretically can be of any type but they are most commonly chosen as binary functions with a single feature and a threshold [72].

RankNet is a probabilistic pairwise ranking framework based on neural networks. For every pair that is correctly ranked, each instance is propagated through the net separately. The difference between the two outputs is mapped to a probability by the logistic function. The cross entropy loss is then computed from that probability and the true label for that pair. Next, all weights in the network are updated using the error back propagation and the gradient descent method [72].

## *7.5 Experiments and Results*

The actual data with unrepaired unilateral cleft lip includes left and right clefts. For better comparison, all the individuals with left cleft lip were mirrored by the plane given by the expert, so they appear to be right cleft in the pictures and 3D meshes. Thus, 35 right cleft meshes and 5 control meshes were shown to the expert. A user interface was created to allow the user to arrange 3D meshes. The expert can click and drag the pictures in any order desired, and also open up a window with 3 neighboring meshes to rotate, enlarge and compare the details in the 3 meshes to carefully determine the order. After the expert is finished ranking, the images have an assigned ground truth rank from 1 to 40.

### *7.5.1 Measurements*

Because the ground truth and the scores predicted are all ranks instead of actual quantified linear scores, Spearman rank correlation coefficient $\rho$ and the Kendall rank correlation coefficient $\tau$ are used to evaluate the experimental results.

The Spearman correlation coefficient is defined as the Pearson correlation coefficient between the ranked variables. For a sample of size $n$, the $n$ raw scores $X_i$, $Y_i$ are converted to ranks $x_i$, $y_i$ , and $\rho$ is computed from these. The closer the $\rho$ value is to 1, the better the two ranks are correlated:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}} \tag{7.5}$$

The Kendall $\tau$ test is a non-parametric hypothesis test for statistical dependence based on the $\tau$ coefficient. It is a pairwise error that reflects how many pairs are ranked discordant. The best matching ranks get a $\tau$ value of 1.

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{\frac{1}{2}n(n-1)} \tag{7.6}$$

### *7.5.2 Experiments*

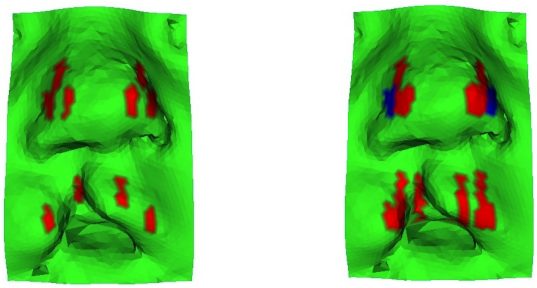
For our experiments, we used the WEKA [61] implementation for linear and SVM regression. RankLib [73] was used for RankNet and RankBoost training and testing. The feature length is 400, with $10 \times 10$ grids and 4 scores in each grid. 4-fold cross validation was performed, and the results are in Table 7.1, with the first number as the Spearman correlation coefficient $\rho$ and the second number as the Kendall correlation coefficient $\tau$. A small dataset (40 instances) with large feature vectors (400 features) does not perform very well with the regression methods and RankNet. The pairwise method RankBoost performs significantly better than all the other regression and ranking methods. The reason is that RankBoost used weak rankers in every round to pick up a feature that is most distinguishable. In this experiment, it used 10 features in these weak rankers, instead of all 400. In terms of the mid-facial reference plane finding methods, the plane produced by the a-lmk method has the highest performance of the three automatic methods, with a score similar to that obtained by the derict method in which the experts actually drew the plane on the data.

### *7.5.3 Feature Selection and Results*

Based on the good performance accomplished by RankBoost and the reasons behind it, a feature selection was done by best-first search to come up with the most distinguishing features in all 400 features. Out of the top five features selected by the best-first search approach, three of them are angle differences, one is an edge difference, and one is a curvature difference. The grid positions of the top five grid patches are shown in Fig. 7.3(a). The

Table 7.1: Ranking correlations for all features (feature length 400, CV4). Each box contains Spearman correlation coefficient $\rho$ followed by Kendall correlation coefficient $\tau$.

| Method | Linear R | SVM R | RankNet | RankBoost |
|---|---|---|---|---|
| mirror | 0.661 0.522 | 0.636 0.511 | 0.512 0.389 | 0.683 0.515 |
| a-lmk | 0.597 0.489 | 0.599 0.489 | 0.513 0.389 | **0.773 0.615** |
| learning | 0.574 0.482 | 0.589 0.515 | 0.669 0.541 | 0.746 0.582 |
| m-lmk | 0.560 0.478 | 0.549 0.452 | 0.632 0.485 | 0.635 0.493 |
| direct | 0.524 0.422 | 0.521 0.400 | 0.630 0.533 | **0.771 0.615** |



(a) Top 5 selected grids (b) Top 10 selected grids

Figure 7.3: Top discriminative features. (a) The red colored areas are the positions for top 5 selected grid patches. (b) The red and blue areas are the top 10 selected grid patches. The red grid patches are selected once, and the blue patch is selected twice with two features.

grid patches are located on the side of the nose area and upper mouth area, which are exactly the areas the experts are looking at when ranking. The top 10 features are shown in Fig. 7.3(b). The blue-colored grid patch near the nose side is selected twice, with one edge difference and one angle difference. The others contain four angle features, two edge features, one radius difference, and one curvature difference.

Using only the top five features for 4-fold cross validation on the dataset and repeating the same experiments as in 7.5.2, the results are boosted by 0.06 for $\rho$ (from around 0.77 to more than 0.83), and 0.1 for $\tau$ (from around 0.61 to 0.71), as shown in Table 7.2. Out of all four ranking algorithms, RankNet now obtains the best performance. The reason is that RankNet is based on neural networks which are known to be hard to train. When it is dealing with a large number of features, it is less effective than RankBoost [72]. However, when using less features, because the task is much easier, RankNet is trained more efficiently.

Table 7.2: Ranking correlations for selected features (feature length 5 CV4)

| Method | Linear R | SVM R | RankNet | RankBoost |
|---|---|---|---|---|
| mirror | 0.729 0.589 | 0.730 0.574 | 0.719 0.570 | 0.687 0.528 |
| a-lmk | 0.787 0.641 | 0.780 0.633 | 0.809 0.663 | 0.707 0.559 |
| learning | 0.792 0.637 | 0.812 0.644 | **0.843 0.700** | 0.750 0.612 |
| m-lmk | 0.800 0.648 | 0.813 0.652 | 0.831 0.696 | 0.772 0.611 |
| direct | 0.795 0.659 | 0.813 0.670 | 0.827 0.711 | 0.752 0.626 |

| expert's order | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| images |  | | | | | | | | | |
| learning | 1 | 3 | 2 | 4 | 5 | 6 | 8 | 9 | 7 | 10 |
| a-lmk | 1 | 2 | 3 | 5 | 6 | 4 | 8 | 7 | 9 | 10 |
| mirror | 1 | 2 | 4 | 8 | 5 | 6 | 9 | 3 | 7 | 10 |
| m-lmk | 1 | 2 | 3 | 4 | 5 | 6 | 9 | 7 | 10 | 8 |
| direct | 1 | 2 | 3 | 5 | 4 | 6 | 7 | 9 | 10 | 8 |

Figure 7.4: Ranking results: ten sample images with nose and cleft areas are shown, nine with unilateral cleft and one control. They are ordered by the expert's rank. Under the images are their ranks by our system (ranked 1-10, with 1 being most severe).

Out of the three automatic mid-facial reference plane finding methods, the learning method is now the winner and is able to beat both human-based methods by a small margin. Both the learning method and the a-lmk method beat the mirror method by a substantial margin. A subset of the results is shown in Fig. 7.4. Ten images with nose and cleft areas are shown from nine unilateral cleft infants and one control. They are ordered by the expert's rank. The other ranks for each mid-facial plane method are predicted by RankNet.

### *7.6 Summary*

This chapter introduced a system to learn to rank the severity of abnormalities on 3D faces with cleft lip. The system takes a mid-facial reference plane, extracts symmetry measurements based on grid patches determined by that mid-facial reference plane, and uses a machine learning algorithm to train a model to predict the ranks of how difficult it is to repair the cleft lip nasal deformity. For the mid-facial reference plane, three automated

methods along with two human-based approaches were compared. Four machine learning algorithms were tested to learn from the experts' ranking orders and to build a model to predict the severity of clefts based on the mid-facial plane and four patch-based features. The learning method for plane construction along with the algorithm RankNet performs the best. The results show that the rankings predicted by the proposed features are highly correlated with the clinicians' ranking order.

# Chapter 8

# QUANTIFYING THE SEVERITY OF CLEFT LIP

In this chapter, the descriptors to present asymmetry and nasal deformity are introduced and the scores for children with cleft before and after reconstruction surgery are compared. Because the automated landmark location method described in Chapter 4 is confirmed to be very reliable for not only finding the landmarks but also determining the mid-facial reference plane, the descriptors discussed in this chapter are all based on this method.

## *8.1 Quantifying the Asymmetry of the Nose and Upper Lip*

Instead of applying the asymmetry quantification to all the points for the whole face, we restrict the area to be only the nose and upper lip area in order to reduce the noise. Given the mid-facial reference plane, the face is translated and rotated so that the mid-facial reference plane is the $X = 0$ plane.

### *8.1.1 Point-based Quantification Descriptors*

For every point $p$ with coordinate $(x, y, z)$ , the point that is symmetric to $p$ is labeled as $p_s$ with the coordinate $(-x, y, z)$. The point on the other side of the face that is closest to $p_s$ is $q$. Then the overall difference is calculated as

$$PD_a = \frac{1}{n} \sum_{p} distance(p_s, q) \tag{8.1}$$

where $n$ is the number of points in the nose and upper-lip area, and *distance* is the Euclidean distance. The area selected and the scores for each point are shown in Fig. 8.1. The red color means a bigger distance and the green color means 0 distance. For this individual, the side of the nose and the cleft on the upper lip contributes to the $PD_a$ score. The bigger $PD_a$ is, the more asymmetry the nose and upper lip area has.

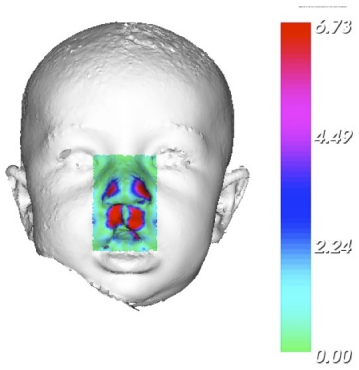



Figure 8.1: The point-based distance.

### *8.1.2 Grid-based Quantification Descriptors*

The radius distance ($RD$) and angle distance ($AD$) are introduced and used for learning the severity of cleft lip in Chapter 7. In this section, these two descriptors are summed once each grid in the nose and upper lip area form two scores for each individual, $RD_a$ and $AD_a$. A larger number means a more severe asymmetry in this area.

$$RD_a = \frac{1}{m \times m} \sum_{\theta,z} RD(\theta, z) \tag{8.2}$$

$$AD_a = \frac{1}{m \times m} \sum_{\theta,z} AD(\theta, z) \tag{8.3}$$

where $m \times m$ is the number of grids. In our experiments, $m = 10$.

### *8.1.3 Experiment Results for Asymmetry Descriptors*

The dataset we used in this chapter is the unilateral cleft dataset, which include 35 unilateral cleft patients for both before and after constructive surgery. The point distance ($PD_a$), the radius distance ($RD_a$) and angle distance ($AD_a$) are applied to the nose and upper lip area. The average scores for all 35 patiens before surgery are 2.04 mm for the radius distance,

Table 8.1: Comparing the three asymmetry scores before and after construction surgery

| Score | $RD_a$ | $AD_a$ | $PD_a$ |
|---|---|---|---|
| Before surgery | 2.04 | 0.39 | 4.33 |
| After surgery | 1.07 | 0.26 | 1.67 |

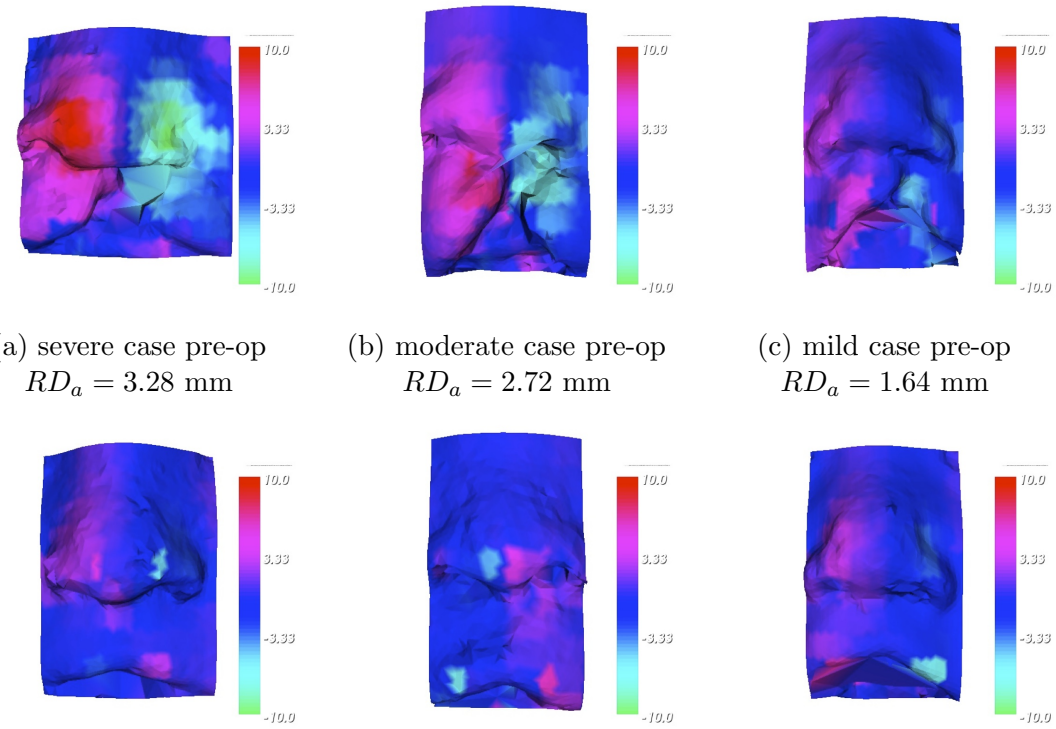
Table 8.2: Correlation coefficient of asymmetry descriptors with ranks given by medical expert based on the severity of cleft before surgery

| Correlation Coefficient | $RD_a$ | $AD_a$ | $PD_a$ |
|---|---|---|---|
| Before surgery | 0.71 | 0.70 | 0.72 |
| After surgery | 0.27 | 0.02 | 0.19 |
| Improvement | 0.70 | 0.61 | 0.70 |

0.39 for the angle distance and 4.33 mm for point distance. After surgery, the average radius distance was reduced to 1.07 mm, the average angle difference was reduced to 0.26 and the average point distance was reduced to 1.67 mm. For these three measurements, the asymmetry scores decreased by 48%, 33% and 61% after the surgery, respectively.

Figure 8.2 shows three examples with the radius difference before (upper row) and after (lower row) surgery. Red color means that one side has a larger radius than the other side, and green color means smaller radius. Blue color means no big difference between the two sides. Figure 8.2 (a) is a severe case, with $RD_a$ score equal to 3.28 mm. The main asymmetry occurs in the side of the nose. After the surgery, as shown in Fig. 8.2 (d), the asymmetry of the nose is largely reduced, and the $RD_a$ score decreased to 1.03 mm. The same trend of change happens for the moderate case in Fig 8.2 (b) and the mild case (c); their scores after surgery are reduced as well because the asymmetries in the nose area are largely corrected by the surgery. Although in the mild case, the reduction in $RD_a$ score is not as high as for the severe cases, the improvement is still noticeable.

The asymmetry scores for all the 35 data are plotted in Fig. 8.3. The horizontal axis is the rank given by the expert in terms of the severity of the cleft lip nasal deformity before surgery. 1 means the most severe case and 35 is the least severe case. The vertical axis is the asymmetry score calculated from the data. Blue squares stand for the scores before




(a) severe case pre-op $RD_a = 3.28$ mm

(b) moderate case pre-op $RD_a = 2.72$ mm

(c) mild case pre-op $RD_a = 1.64$ mm

a) severe case post-op $RD_a = 1.03$ mm

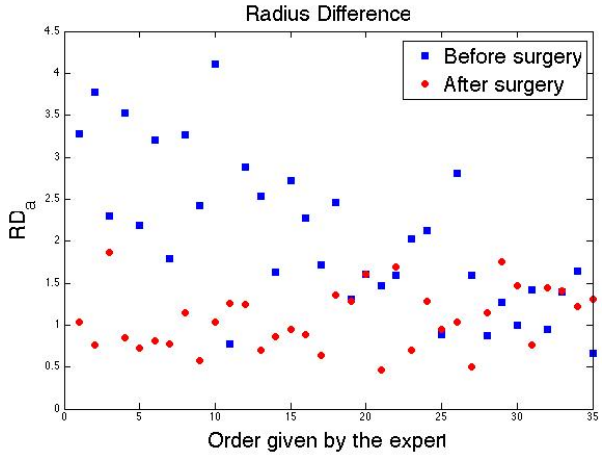
(b) moderate case post-op $RD_a = 0.95$ mm

(c) mild case post-op $RD_a = 1.22$ mm

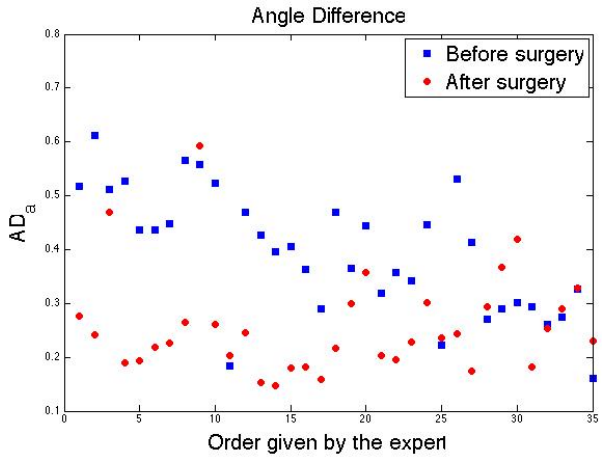
Figure 8.2: $RD_a$ reduction after the surgery for three cases. Red and green shows big difference between the left and right side. Red means higher and green means lower. Blue means small difference between the two sides.

surgery and red dots for the scores after surgery. In Fig. 8.3 (a), the $RD_a$ scores before the surgery are larger, especially for the severe cases, while the after surgery scores are mostly lower than 2 mm. Figure 8.3 (b) and (c) show the same trend as in (a), for the cases before surgery, the scores are higher and more scattered versus lower and more concentrated in the data after surgery.

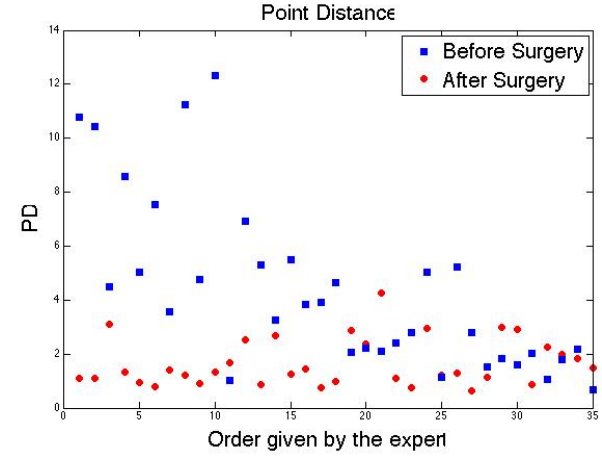
In addition to comparing with before and after surgery scores for each individual, the asymmetry scores are also compared to the ranking orders given by the expert. The Pearson correlation coefficients are calculated to see the correlation between the ground truth (the ranking order given by the expert) and the asymmetry scores. With an average of correlation




(a) $RD_a$ before and after surgery




(b) $AD_a$ before and after surgery




(b) $PD_a$ before and after surgery

Figure 8.3: The comparison for $RD_a$, $AD_a$ and $PD_a$ before and after surgery.

coefficient at 0.71, the before-surgery scores are highly correlated with the ground truth, as show in Table 8.2. There is no such correlation between the ground truth and the after-surgery scores. This reflects the fact that after surgery, the asymmetry of the nose and lip is largely corrected. The score improvements are also correlated with the severity of the initial deformity, which means the more severe the initial deformity is, the larger the improvement.

## *8.2 Descriptors to Quantify the Deformation of Noses*

Three descriptors on the nose of patients with cleft lip have been developed, including the angle of columella, the distance from the nasal tip to the mid-facial reference plane, and the angle between the plane of the nose and the mid-facial reference plane.

### *8.2.1 The Angle of Columella*

The columella is the strip of skin running from the tip of the nose to the upper lip, which separates the nostrils. In cleft patients, the columella is not straight as in normal subjects. Here the angle of columella is a feature of how abnormally the nose is deformed. The angles $\alpha_R$ and $\alpha_L$ in figure 8.4 are extracted from three automatically detected landmarks (nose tip, subnasale and alar curvature points (left and right). The prn, sn and ac_l are used to calculate $\alpha_L$, and the prn, sn and ac_r are used for $\alpha_R$. The difference between angle $\alpha_R$ and $\alpha_L$ is calculated. The closer $\alpha$ is to 0, the more normal the nose is. Patients with right cleft tend to have negative $\alpha$ scores while patients with left cleft are more likely to have positive $\alpha$ scores. For the experiments to compare this score before and after surgery, the absolute value of $\alpha$ is used, with the label $|\alpha|$.

$$\alpha = \alpha_R - \alpha_L \tag{8.4}$$

### *8.2.2 The Distance from the Nose Tip to the Mid-facial Reference Plane*

For normal faces, the nose tip (prn) will position very close to the mid-facial reference plane. But due to the impact of cleft lip, the nose tip is pushed to the opposite direction of the cleft. This distance (labeled as $d_p$) from the prn to the mid-facial reference plane captures how much the nose tip is away from the normal position. The further the distance is, the

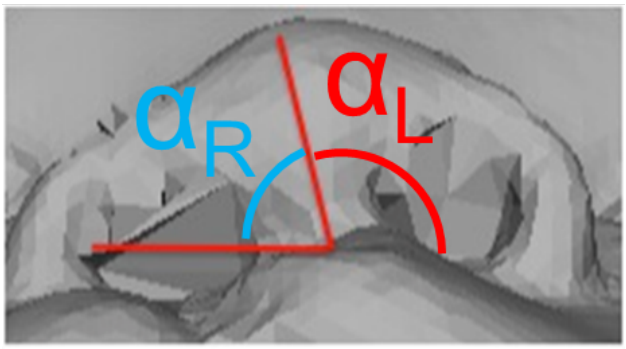
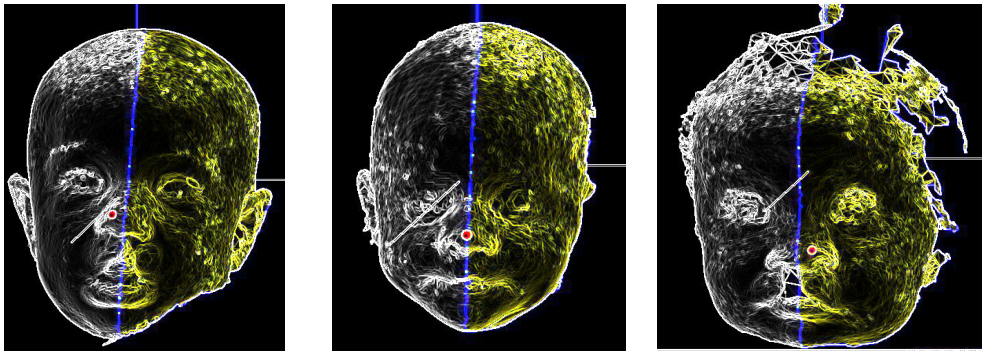



Figure 8.4: The angle of columella.



(a) Left cleft with $d_p$ as a large negative value

(b) Cleft with $d_p$ close to 0

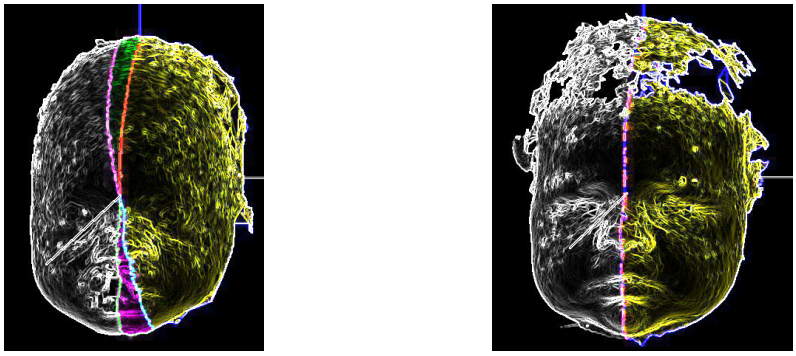
(c) Right cleft with $d_p$ as a large positive value

Figure 8.5: The distance from nose tip to the mid-facial reference plane $d_p$.

more severe the cleft is. Patients with left clefts tend to have negative $d_p$ scores while patients with right clefts are on the positive side, as shown in figure 8.5. For before and after surgery comparison, the absolute value is used as $|d_p|$.

### *8.2.3 The Angle Between the Plane of the Nose and the Mid-facial Reference Plane*

The plane of the nose is calculated using three points, the nose tip (prn), the nasion (n) and the subnasale (sn). Normally, the plane of the nose should be the same as the mid-facial reference plane. But for a unilateral cleft patient, the nose is deformed into one side. So the angle difference (labeled as $\beta$) between the plane of the nose and the mid-facial reference



(a) Cleft patient with a large angle $\beta$ (b) Cleft patient with a small angle $\beta$

Figure 8.6: The angle between the plane of the nose and the mid-facial reference plane.

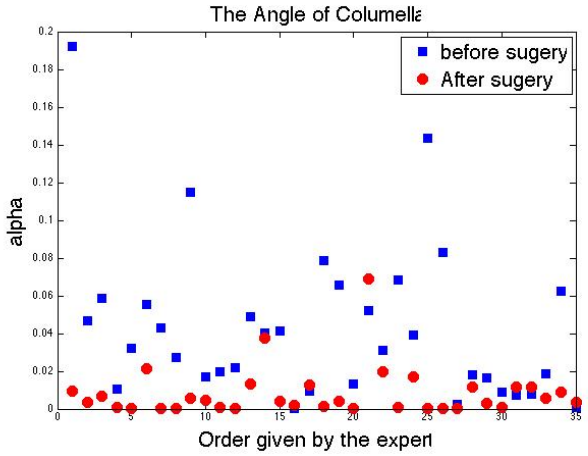
Table 8.3: Comparing the three nose deformity scores before and after construction surgery

| Score | $\lvert\alpha\rvert$ | $\lvert d_p\rvert$ | $\beta$ |
|---|---|---|---|
| Before surgery | 0.0427 | 3.2916 | 0.1919 |
| After surgery | 0.0084 | 1.3823 | 0.1080 |

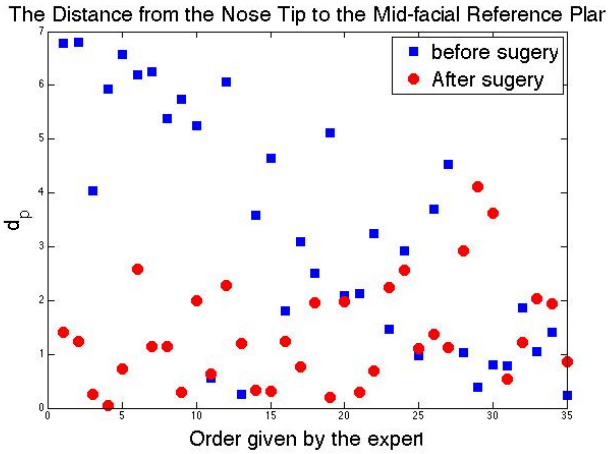
plane can give a quantified value of how severe the nose is deformed. Fig. 8.6(a) shows an example of a cleft patient with a big difference between the plane of the nose and the mid-facial reference plane, while in figure 8.6(b) these two planes are almost the same. The situation in figure 8.6(a) is more severe than in figure 8.6(b).

### *8.2.4 Experiment Results for Nose deformity Descriptors*

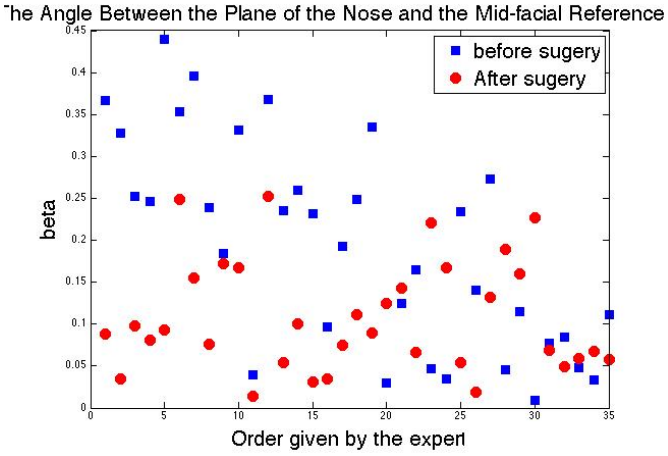
The angle of columella ($\alpha$), the distance from nose tip to the mid-facial reference plane ($d_p$), and the angle between the plane of the nose and the mid-facial reference plane ($\beta$) are calculated for the unicleft dataset. The average scores for all these 35 patients before surgery are 0.04 for $|\alpha|$, 3.29 mm for $|d_p|$ and 0.19 for $\beta$. After surgery, the average absolute value for the angle of columella is reduced to 0.01, the average absolute value for the distance from nose tip to the mid-facial reference plane is reduced to 1.38 mm, and the average angle $\beta$ reduced to 0.11. For these three measurements, the nose deformity scores decreased by 80%, 58% and 44%, respectively.




(a) $\alpha$ before and after surgery




(b) $D_p$ before and after surgery




(b) $\beta$ before and after surgery

Figure 8.7: The comparison for $RD_a$, $AD_a$ and $PD_a$ before and after surgery.

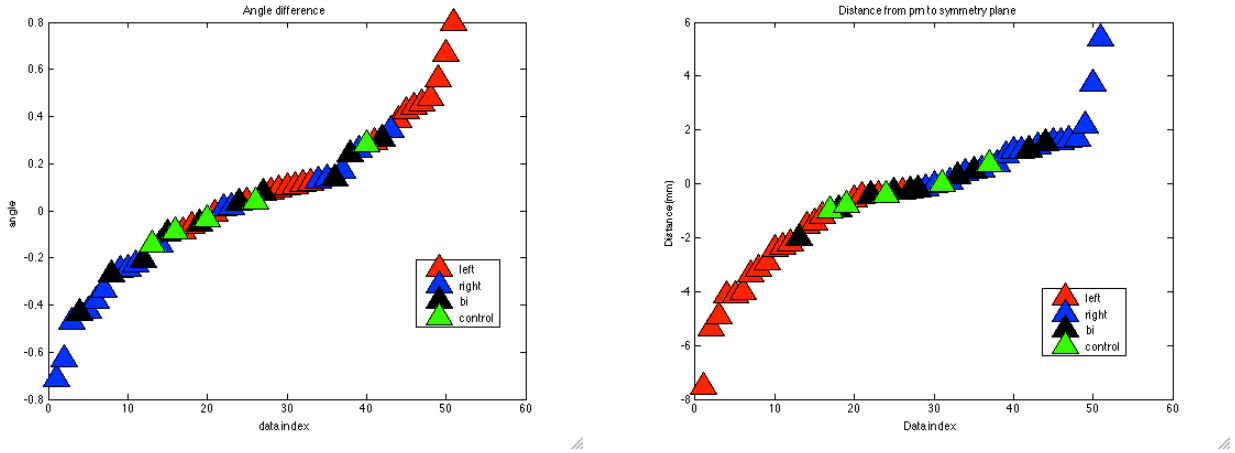
Table 8.4: Correlation coefficient of nose deformity descriptors with ranks given by medical expert based on the severity of cleft before surgery

| Correlation Coefficient | $\|\alpha\|$ | $\|d_p\|$ | $\beta$ |
|---|---|---|---|
| Pre-op | 0.29 | 0.76 | 0.72 |
| Post-op | 0.05 | 0.35 | 0.04 |
| Improvement | 0.30 | 0.76 | 0.64 |

The changes of the scores for all 35 data are plotted in Fig. 8.7. The horizontal axis is the rank given by the expert in terms of the severity of the cleft lip nasal deformity before surgery. The vertical axis is the nose score calculated from the data. Blue squares stand for the scores before surgery and red dots for the scores after surgery. Just like the cases with the asymmetry score, the scores before surgery tend to be higher and more scattered, and the scores after surgery are significantly lower.

The Pearson correlation coefficients are also calculated to see the correlation between the ground truth (the ranking order given by the expert) and the nasal descriptor scores. The latter two descriptors, distance from nose tip to the mid-facial reference plane ($d_p$) and the angle between the plane of the nose and the mid-facial reference plane ($\beta$) are highly correlated with the ground truth, with correlation coefficient as high as 0.76 and 0.72 respectively, as shown in table 8.4. Again, there is no such correlation between the ground truth and the after-surgery scores. The score improvement are also correlated with the severity of the initial deformity for $d_p$ and $\beta$, which means the more severe the initial deformity is, the larger the improvement.

There is another important characteristic that is captured by the angle of columella $\alpha$ and the distance from the nasal tip to the mid-facial reference plane $d_p$: on which side has the cleft occurred. Because there are positive value and negative value for $\alpha$ and $d_p$, it is very easy to tell from the sign which side the nose is pushed to. In Fig. 8.8, the nose descriptor values are calculated for the cleft survey dataset, which contains unilateral cleft, bilateral cleft and control. The green triangles mean left unilateral cleft, the blue triangles mean right unilateral cleft, the black triangles stand for bilateral cleft, and the green triangles are for control data. For the angle of columella, most left cleft instances score a positive




(a) $\alpha$ (b) $d_p$

Figure 8.8: $\alpha$ and $d_p$ values for unilateral cleft, bilateral cleft and control data.

value, which means the prn-sn-ac angle on the left side is larger than the angle on the right side. The bilateral cleft and control have values that lie in between, not as extreme as the most severe unilateral cleft. For the distance from prn to the mid-facial reference plane, 0 is actually the threshold to differentiate the left cleft and right cleft. All the left cleft data have $d_p$ value less than 0 and all the right cleft have $d_p$ value larger than 0. The bilateral and control data have a value that is close to 0.

### *8.2.5 Summary*

In this chapter, three asymmetry descriptors and three nasal deformity descriptors are introduced. For before and after surgery comparison, all of these six descriptors capture the fact that the severity scores are decreased after the surgery. Five out of six descriptors are highly correlated to the severity order provided by the experts. It is also proven that the improvement of surgery is correlated to the severity of initial deformity.

# Chapter 9

# AUTOMATED FACE EXTRACTION AND NORMALIZATION

3D stereo photography is rapidly being adopted by medical researchers for analysis of facial forms and features. An essential step for many applications using 3D face data is to first crop the head and face from the raw images. The goal of this chapter is to develop a reliable automatic methodology for extracting the face from raw data with texture acquired from a stereo imaging system, based on the medical researchers' specific requirements. We present an automated process, including eye and nose estimation, face detection, Procrustes analysis, and final noise removal to crop out the faces and normalize them. The proposed method shows very reliable results on several datasets, including a normal adult dataset and a very challenging dataset consisting of infants with cleft lip.

## *9.1 Background*

In all of the previous methods described in chapter 4 through chapter 7, and in other works that using computer-based automatic methods to study 3D face characteristics with diseases such as autism [21], plagiocephaly [74] and 22q11.2 deletion syndrome [56], the data are required to be preprocessed, which means the face and head data are cropped from the background, so that the shape descriptors can be applied only on the region of interest. However, this step is usually done manually.

Our collaborators from Seattle Children's Hospital also analyze 3D characteristics of the face, including manually landmarking the data, completing some craniofacial anthropometric measurements, and evaluating the data after acquisition of the data. For all of those tasks, they want the data to be cleaned in a specific way, so that clothing, body under the neck, and other noise are removed. The face, forehead and front part of skull are integral, and the ears are kept. These requirements are very specific and different from most 3D face recognition applications, in which only the face (from eyebrow to chin) is extracted, so that

(a) Texture image from 5 cameras (b) Mesh data

Figure 9.1: Raw data. (a) The texture image from 5 different cameras on left, right, front, back and top. (b) The 3D mesh composed of vertices and triangular connections.

the existing methods for 3D face extraction are not suitable for this purpose [75].

In this chapter, we present a system that takes the raw data captured by the 5-pod 3dMDcranial system as an input, automatically detects the face, and extracts it to meet the clinicians' requirements. Four steps, including curvature classification, face detection, Procrustes analysis, and surface normal and color thresholding, are investigated to ensure that the face, including the forehead and ears, is extracted from the whole noisy, raw image.

## *9.2 Data Format and Datasets*

Our dataset consists of 3D craniofacial surface meshes obtained from the 5-pod 3dMDcranial imaging system. In addition to the mesh data as used in previous chapters, the texture data are used only for the purpose of extracting the faces. Each 3D image is composed of two parts: the 3D mesh part and the texture part. The mesh part is a point cloud with connections between the points, as shown in Fig. 9.1(b). The texture part is composed of RGB images, viewed from 4 - 5 perspectives, as in Fig. 9.1(a). The two parts are connected by the texture coordinates, which are associated with every vertex in the mesh data and XY pixel positions in the RGB image.

There are three datasets in our study. One contains normal adults, with 21 scans from 10 individuals. The second one contains 64 meshes from 52 infants with unrepaired cleft lip.

The last one consists of 35 meshes from 35 infants with repaired cleft lip. The latter two are extremely challenging, because these infants are too young to sit unsupported and need to be held by an adult or a positioning device. Although it is suggested that the subjects should face one of the cameras and have a relaxed expression when the scan is taken, the orientation and expression of infants are hard to control and the data can be very noisy.

## *9.3 Method*

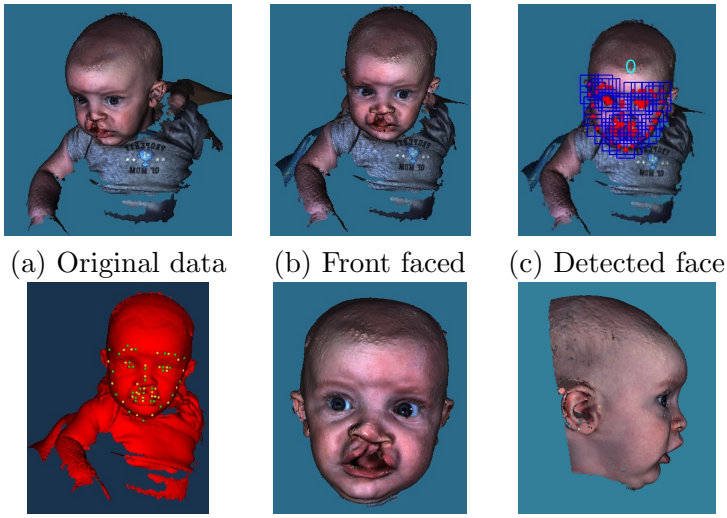
### *9.3.1 System Design*

Our system is built for automatically cleaning the raw data and normalizing the 3D faces. With an input of 3D textured mesh data, our system first performs eye and nose region detection based on a machine learning technique and rotates the mesh so that the face is forward. Then, a face detection algorithm is used on the screenshot of the 3D mesh to detect the human face and a set of landmarks. After that, Procrustes alignment is applied to normalize the data so that a standard box can be used to cut the data uniformly. Last, some final cleaning methods are employed to further improve the data. Figure 9.2 illustrates the steps of our system.

### *9.3.2 Eyes and Nose Tip Detection on Mesh*

There are already some successful face recognition algorithms for 2D photos. In our dataset, the input contains both the reconstructed 3D mesh and 2D photos from several cameras. However, when capturing these photos, it is hard for the photographer to control the postures of the infants, so it is not guaranteed that one of the cameras will capture a good presentation of the face. Therefore, instead of applying a face recognition algorithm directly onto the photos taken by cameras, our first step is to ensure that the data is rotated to a suitable angle, so that the advantage of already existing and evaluated face recognition algorithms for 2D photos can be enhanced.

Our system uses three steps to rotate the 3D mesh and construct a frontal face screenshot for subsequent processes. The first step is to form a feature vector composed of a multiple-scale histogram of single-valued, low-level features. Next, two classifiers, one for inner eye



(a) Original data (b) Front faced (c) Detected face

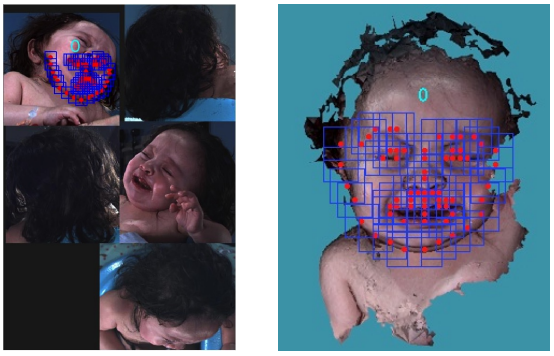
(d) Procrustes (e) Cleaned data (f) Side view

Figure 9.2: System design for automatic face extraction. (a) Original 3D mesh with texture. (b) Face rotated to a frontal position. (c) Face detected with 0 degree of rotation. (d) Procrustes aligned face with landmarks. (e) Front view of the cleaned data, the forehead and the front part of the skull are kept. (f) Side view of the cleaned data with the ears kept.

corners and one for nose tip, are trained and the candidate points for eye and nose are selected. The third step is to select possible eye-nose triangle regions and rotate the data. These steps are detailed in chapter 5.

### *9.3.3 2D Face Detection*

The problem of finding and analyzing faces from 2D images is a foundational task in computer vision and there are multiple existing techniques. In Zhu's work [76], the tasks of face detection, pose estimation and landmark estimation are jointly addressed by a model based on a mixture of trees with a shared pool of parts. Every facial landmark is modeled as a part, and global mixtures are used to capture topological changes. This method is applied to our data after the face is rotated to a frontal position. The results after this step include: face location, head pose estimation, and landmarks on the 2D screenshot. Figure 9.3 illustrates why the previous rotation step is essential. Because a baby's pose and expression are



(a) Face detection on the original data

(b) Face detection on the screenshot

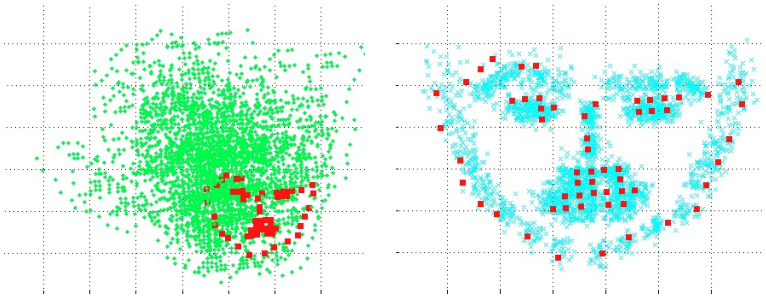
Figure 9.3: Same face detection algorithm applied to original data and screenshot after face rotation. The number 0 on the image means frontal face position. (a) shows the pose estimation is not true and the landmarks are not properly located if the face detection algorithm is applied directly to the original photo. (b) returns a true positive result, with a detection of frontal face.

very hard to control, if the face detection method is applied directly to the original photo without pose normalization, it sometimes leads to a failure in face detection or even a false positive result. After pose normalization, when there is a clear frontal face in the image, the task is much easier for both face detection and landmark localization algorithm.

### *9.3.4 Face Normalization*

Once the frontal face is detected, the 3D landmarks are calculated by projecting the extracted 2D landmarks on the screenshot from the previous step. The 3D mesh faces are then normalized by Procrustes analysis (PA) [77]. In this step, a random head mesh $D_m$ with landmarks $L_m$ is selected as the approximate mean shape. Then every head mesh $D_j$ with landmarks $L_j$ is aligned to the approximate mean shape by translation, scaling and rotation so that the sum of squared errors of the landmarks $L_j$ and $L_m$ is minimized. Then a new mean shape $D_m$ with landmarks $L_m$ is calculated based on the average of the aligned data. The steps are iterative until the mean shape is stable and each head mesh is aligned to the mean.

Figure 9.4 shows the experimental data of landmarks before and after Procrustes anal-



(a) Landmarks before PA (b) Landmarks after PA

Figure 9.4: Landmarks before and after procrustes analysis.

ysis. The red dots represent the landmarks $L_m$. Before Procrustes analysis, the green dots in Fig 9.4 (a) are all scattered, while the blue dots in Fig 9.4 (b) shows the landmarks after the Procrustes analysis gathered around the mean landmarks $L_m$.

### *9.3.5 Final Cleanup*

Because all the data are normalized to the same scale and rotation, a standard bounding box is used to cut the data and meet the requirements for keeping the ears and forehead area, as shown in Fig. 9.5(a). For an adult, this standard bounding box is adequate to separate the face from other parts because adults tend to have a long neck. However, for children especially young babies, there is still clothing or other noise under the chin area due to the shortness of their necks. Some additional cleanup steps, including surface normal thresholding and color thresholding, are used to remove this noise. For every point in the bounding box $p_i$, the surface normal vector is calculated, with $Nx_i$, $Ny_i$ and $Nz_i$ as the normal vector projection in $x$, $y$ and $z$ axis respectively. The larger $Ny_i$ is, the more the surface around point $p_i$ is facing up. Figure 9.5 (c) shows the normal value in the $y$ direction. Since the points around the lower jaw and chin are all facing downward and the points in the clothing and shoulders are facing up, a threshold $T_N y$ is used to differentiate the shoulder and chest from the face. Another threshold is used in color space. The colors of the points are converted from RGB space to HSV space; the hue values for every point in the bounding box are illustrated in Fig. 9.5 (d). The average hue value for the cheek area is calculated.

(a) Bounding box (b) After cleanup (c) $Ny$ values

(d) The hue value (e) Bounding box (f) After cleanup

Figure 9.5: Final cleanup. (a) Cut by a standard boundary box. (b) Cleaned by using normal and color thresholds. (c) The surface normal value in the $y$ direction, red means the surface normal is pointing up, and green means facing down. (d) The hue value for the face. A thresholds was chosen for cleaning the noise underneath the chin (e) (f)Another example before and after final cleanup.

Any point with a large difference from the average hue value in the cheek area is removed. Figure 9.5(b) shows that after these two cleanup steps, the clothing and noise under the chin are all removed.

### *9.4 Experiments and Results*

For our experiments, we use the WEKA [61] implementation of the SVM classifier in the eye and nose detection step. This step is considered accurate when the eye-nose triangle is detected correctly and the face is rotated to a frontal position. This step achieves 100% accuracy in the normal adult dataset, 94% in the unrepaired cleft dataset and 97% in the repaired cleft dataset. For the images in which the classifier failed in detecting eye-nose regions, the face is rotated manually to face forward for latter processes.

The software provided by Zhu [76] was used for face detection on the screenshot produced by the previous step. This algorithm is extremely successful when the input is a frontal

Table 9.1: Accuracy for Each Step in the Process

| Dataset | normal | unrepaired cleft | repaired cleft |
|---|---|---|---|
| number of instances | 21 | 64 | 35 |
| eyes and nose detection | 21 (100%) | 60 (94%) | 34 (97%) |
| face detection | 21 (100%) | 64 (100%) | 35 (100%) |
| ear and forehead | 21 (100%) | 64 (100%) | 35 (100%) |
| no clothes left | 21 (100%) | 60 (94%) | 32 (91%) |

face, achieving 100% in the face detection task with all three datasets. After Procrustes analysis and normalizing all the detected faces, the standard box successfully keeps the face, forehead, front part of the skull and ears for all the data from normal, unrepaired and repaired cleft lip data.

In order to avoid over cropping, in the last step of removing the clothes, chest and shoulder areas under the chin, we chose the threshold to be very relaxed to ensure that the chin is well maintained. We found there are 6% with very small amounts of clothes or skin in the chest remaining underneath the chin for babies with cleft lip in the datasets before surgery, and 9% after sugery, as show in Table 9.1.

### *9.5 Summary*

This chapter introduces a system to crop out the face from 3D textured mesh data in infants as young as 3 months old and in adults to meet the requirements of medical researchers. The system takes a 3D mesh, detects the inner eye corners and the nose tip, rotates the data and saves a screenshot. Then the screenshot is analyzed by a 2D face detection and landmark estimation algorithm. After the landmarks are obtained, the data are normalized, allowing for a standard boundary crop. Finally, the surface normal and color are used to threshold some noise underneath the chin area. The results on normal and challenging patient datasets show that the whole process is very reliable. It detects the face and maintains the face area, the forehead, the ears and the front part of the skull with a 100% rate of success.

# Chapter 10

# CONCLUSIONS

With the method described in the last chapter, automatic face extraction and normalization, the whole system is complete. This computer-based system can take in raw data, clean the data and normalize the face, detect the landmarks, find the mid-facial reference plane, extract the features to rank based on the cleft lip severity, and provide scores for asymmetry and nasal deformity. Figure 10.1 shows the flowchart of how our system works.

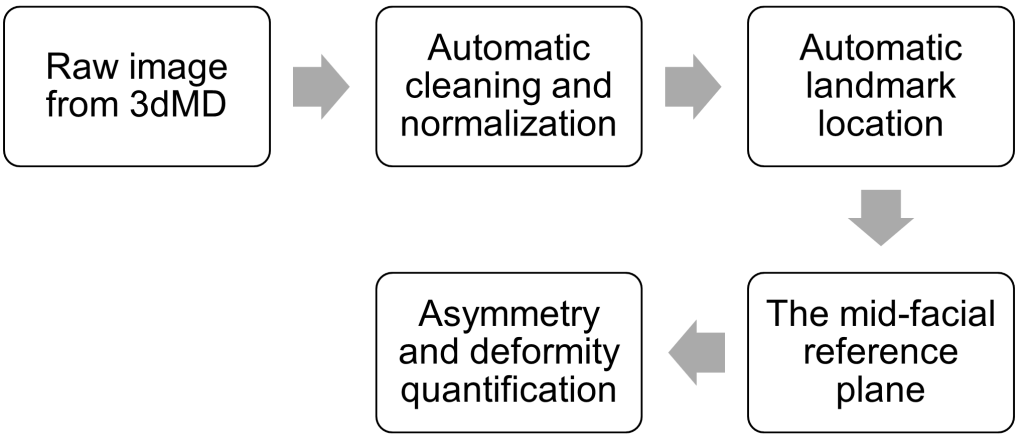



Figure 10.1: The whole system.

## *10.1 Contribution*

The main contributions of this work are:

1. Preprocessing system: given raw data from a multi-camera commercial stereo system, the face is cropped out and normalized automatically. This method detects the face

and maintains the face area, the forehead, the ears and the front part of the skull with a 100% rate of success.

2. Localization of landmarks: our fully automatic landmark locating algorithm has an average error of 3.07 mm on 20 landmarks over a sample of 994 heads, and is superior to prior published methods. These landmarks are also used to calculate the mid-facial reference plane and nose deformity descriptors.

3. Evaluation of mid-facial reference plane methods: the plane calculated used our automatic landmark locating algorithm is determined as the best computer-based method to find the mid-facial reference plane. This is evaluated by comparing to the ground truth provided by the experts, also by the rating and ranking scores provided by six medical experts.

4. Ranking based on the severity of cleft: a system to learn and rank the severity of cleft on 3D faces with cleft lip is introduced. This system extracts symmetry measurements based on grid patches determined by the mid-facial reference plane, and uses a machine learning algorithm to train a model to predict the ranks of how difficult it is to repair the cleft lip nasal deformity. The results show that the rankings predicted by the proposed features are highly correlated with the clinician's ranking order.

5. Quantification of the severity of cleft: three asymmetry descriptors and three nose deformity descriptors are developed. All six descriptors shows a decrease in abnormality for kids with cleft after constructive surgery comparing to their before surgery 3D images. Five out of six descriptors show high correlation with the clinician's ranking order.

### *10.2 Future Work*

The use of texture images as well as the 3D mesh would definitely opens up several research avenues. For example, the landmarks could be more precisely detected given the color information. Another example is that the nostril shape, which is very interesting to the clinicians, could be captured more easily with the texture information. To quantify the improvement of constructive surgery, other methods, such as deformable registration could be investigated. Lastly, a long term future goal is be to be able to translate the obtained 3D quantification into plain English language that could then be integrated into clinical practices for an aid in diagnosis.

# BIBLIOGRAPHY

[1] H. Kalter and J. Warkany. Experimental production of congenital malformations in strains of inbred mice by maternal treatment with hypervitaminosis a. *The American Journal of Pathology*, 38(1):1, 1961.

[2] H.H. Ardinger, K.H. Buetow, G.I. Bell, J. Bardach, D.R. VanDemark, and J.C. Murray. Association of genetic variation of the transforming growth factor-alpha gene with cleft lip and palate. *American journal of human genetics*, 45(3):348, 1989.

[3] K. Wilamowska. *Shape-based Quantification and Classification of 3D Face Data for Craniofacial Research*. PhD thesis, University of Washington, 2009.

[4] L. Wiskott, J.M. Fellous, N. Kuiger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):775–779, 1997.

[5] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[6] T. Heseltine, N. Pears, and J. Austin. Three-dimensional face recognition: A fisher-surface approach. *Image Analysis and Recognition*, pages 684–691, 2004.

[7] T. Heseltine, N. Pears, and J. Austin. Three-dimensional face recognition: An eigen-surface approach. In *Image Processing, 2004. ICIP'04. 2004 International Conference on*, volume 2, pages 1421–1424. IEEE, 2004.

[8] A.S. Mian, M. Bennamoun, and R. Owens. An efficient multimodal 2d-3d hybrid approach to automatic face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(11):1927–1943, 2007.

[9] C. Boehnen and T. Russ. A fast multi-modal approach to facial feature detection. In *Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on*, volume 1, pages 135–142. IEEE, 2005.

[10] K.I. Chang, K.W. Bowyer, and P.J. Flynn. An evaluation of multimodal 2d+ 3d face biometrics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(4):619–624, 2005.

[11] K.W. Bowyer, K. Chang, and P. Flynn. A survey of approaches and challenges in 3d and multi-modal 3d+ 2d face recognition. *Computer Vision and Image Understanding*, 101(1):1–15, 2006.

[12] S. Berretti, A. Del Bimbo, and P. Pala. 3d face recognition using isogeodesic stripes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(12):2162–2177, 2010.

[13] K.I. Chang, W. Bowyer, and P.J. Flynn. Multiple nose region matching for 3D face recognition under varying facial expression. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(10):1695–1700, 2006.

[14] P. Nair and A. Cavallaro. 3-d face detection, landmark localization, and registration using a point distribution model. *Multimedia, IEEE Transactions on*, 11(4):611–623, 2009.

[15] J.T. Richtsmeier, V. Burke Deleon, and S.R. Lele. The promise of geometric morphometrics. *American journal of physical anthropology*, 119(S35):63–91, 2002.

[16] F.L. Bookstein. Landmark methods for forms without landmarks: localizing group differences in outline shape. In *Mathematical Methods in Biomedical Image Analysis, 1996., Proceedings of the Workshop on*, pages 279–289. IEEE, 1996.

[17] P. Mitteroecker and P. Gunz. Advances in geometric morphometrics. *Evolutionary Biology*, 36(2):235–247, 2009.

[18] P. Hammond. The use of 3d face shape modelling in dysmorphology. *Archives of Disease in Childhood*, 92(12):1120, 2007.

[19] P. Hammond, T.J. Hutton, J.E. Allanson, L.E. Campbell, R. Hennekam, S. Holden, M.A. Patton, A. Shaw, I.K. Temple, M. Trotter, et al. 3d analysis of facial morphology. *American Journal of Medical Genetics Part A*, 126(4):339–348, 2004.

[20] I.T. Jolliffe and MyiLibrary. *Principal component analysis*, volume 2. Wiley Online Library, 2002.

[21] P. Hammond, C. Forster-Gibson, AE Chudley, JE Allanson, TJ Hutton, SA Farrell, J. McKenzie, JJA Holden, and MES Lewis. Face–brain asymmetry in autism spectrum disorders. *Molecular psychiatry*, 13(6):614–623, 2008.

[22] J. Cox-Brinkman, A. Vedder, C. Hollak, L. Richfield, A. Mehta, K. Orteu, F. Wijburg, and P. Hammond. Three-dimensional face shape in fabry disease. *European Journal of Human Genetics*, 15(5):535–542, 2007.

[23] Z.A. Bhuiyan, M. Klein, P. Hammond, A. Van Haeringen, M.M.A.M. Mannens, I. Van Berckelaer-Onnes, and R.C.M. Hennekam. Genotype-phenotype correlations of 39 patients with cornelia de lange syndrome: the dutch experience. *Journal of medical genetics*, 43(7):568–575, 2006.

[24] N. Nakamura, T. Okawachi, K. Nishihara, N. Hirahara, and E. Nozoe. Surgical Technique for Secondary Correction of Unilateral Cleft Lip-Nose Deformity: Clinical and 3-Dimensional Observations of Preoperative and Postoperative Nasal Forms. *Journal of Oral and Maxillofacial Surgery*, 68(9):2248–2257, September 2010.

[25] I. Stauber, E. Vairaktaris, A. Holst, M. Schuster, U. Hirschfelder, F.W. Neukam, and E. Nkenke. Three-dimensional analysis of facial symmetry in cleft lip and palate patients using optical surface data. *Journal of Orofacial Orthopedics/Fortschritte der Kieferorthopädie*, 69(4):268–282, 2008.

[26] M. Benz, X. Laboureux, T. Maier, E. Nkenke, S. Seeger, F.W. Neukam, and G. Häusler. The symmetry of faces. In *Proceedings of Vision, Modeling, and Visualization, G. Girod, H. Niemann, T. Ertl, B. Girod, and H.-P. Seidel, eds.(Akademische Verlagsgesellschaft, 2002)*, pages 43–50, 2002.

[27] P.J. Besl and N.D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on pattern analysis and machine intelligence*, pages 239–256, 1992.

[28] J. Hartmann, P. Meyer-Marcotty, M. Benz, G. Häusler, and A. Stellzig-Eisenhauer. Reliability of a method for computing facial symmetry plane and degree of asymmetry based on 3D-data. *Journal of Orofacial Orthopedics/Fortschritte der Kieferorthopädie*, 68(6):477–490, 2007.

[29] G. Rhodes, F. Proffitt, J.M. Grady, and A. Sumich. Facial symmetry and the perception of beauty. *Psychonomic Bulletin and Review*, 5:659–669, 1998.

[30] M. Komori, S. Kawamura, and S. Ishihara. Averageness or symmetry: Which is more important for facial attractiveness? *Acta psychologica*, 131(2):136–142, 2009.

[31] Y. Liu, K.L. Schmidt, J.F. Cohn, and S. Mitra. Facial asymmetry quantification for expression invariant human identification. *Computer Vision and Image Understanding*, 91(1-2):138–159, 2003.

[32] K. Huang, W. Hong, and Y. Ma. Symmetry-based photo-editing. *Pattern Recognition*, 38(6):825–834, 2005.

[33] Y. Liu, T. Belkina, J.H. Hays, and R. Lublinerman. Image de-fencing. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[34] N.J. Mitra, L.J. Guibas, and M. Pauly. Symmetrization. In *ACM SIGGRAPH 2007 papers*, page 63. ACM, 2007.

[35] M. Pauly, N.J. Mitra, J. Wallner, H. Pottmann, and L.J. Guibas. Discovering structural regularity in 3D geometry. *ACM Transactions on Graphics-TOG*, 27(3):43–43, 2008.

[36] K. Xu, H. Zhang, A. Tagliasacchi, L. Liu, G. Li, M. Meng, and Y. Xiong. Partial intrinsic reflectional symmetry of 3d shapes. *ACM Trans. Graph*, 28(5):1–10, 2009.

[37] G. Loy and J.O. Eklundh. Detecting symmetry and symmetric constellations of features. *Computer Vision–ECCV 2006*, pages 508–521, 2006.

[38] S. Lee and Y. Liu. Curved glide-reflection symmetry detection. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1046–1053. IEEE, 2009.

[39] N.J. Mitra, L.J. Guibas, and M. Pauly. Partial and approximate symmetry detection for 3D geometry. *ACM Transactions on Graphics (TOG)*, 25(3):560–568, 2006.

[40] M. Ovsjanikov, J. Sun, and L. Guibas. Global intrinsic symmetries of shapes. In *Computer graphics forum*, volume 27, pages 1341–1348. Wiley Online Library, 2008.

[41] 3dMD. http://www.3dmd.com.

[42] J.C. Kolar and E.M. Salter. *Craniofacial anthropometry: Practical measurement of the head and face for clinical, surgical, and research use*. CC Thomas, 1997.

[43] H. Hochheiser, B.J. Aronow, K. Artinger, T.H. Beaty, J.F. Brinkley, Y. Chai, D. Clouthier, M.L. Cunningham, et al. The facebase consortium: a comprehensive program to facilitate craniofacial research. *Developmental biology*, 355(2):175–182, 2011.

[44] J.Y. Wong, A.K. Oh, E. Ohta, A.T. Hunt, G.F. Rogers, J.B. Mulliken, and C.K. Deutsch. Validity and reliability of craniofacial anthropometric measurement of 3d digital photogrammetric images. *The Cleft Palate-Craniofacial Journal*, 45(3):232–239, 2008.

[45] C.L. Heike, M.L. Cunningham, A.V. Hing, E. Stuhaug, and J.R. Starr. Picture perfect? reliability of craniofacial anthropometry using three-dimensional digital stereophotogrammetry. *Plastic and reconstructive surgery*, 124(4):1261, 2009.

[46] P. Perakis, G. Passalis, T. Theoharis, and I.A. Kakadiaris. 3d facial landark detection and face registration. Technical report, University of Athens, Greece, 2010.

[47] X. Lu and A.K. Jain. Automatic feature extraction for multiview 3d face recognition. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 585–590. IEEE, 2006.

[48] D. Colbry, G. Stockman, and A. Jain. Detection of anchor points for 3d face verification. In *IEEE Workshop on Advanced 3D Imaging for Safety and Security*, volume 3, page 118, 2005.

[49] TH Lin, WP Shih, WC Chen, and WY Ho. 3d face authentication by mutual coupled 3d and 2d feature extraction. In *Proceedings of the 44th annual Southeast regional conference*, pages 423–427. ACM, 2006.

[50] T.H. Yu and Y.S. Moon. A novel genetic algorithm for 3d facial landmark localization. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6. IEEE, 2008.

[51] C. Xu, T. Tan, Y. Wang, and L. Quan. Combining local features for robust nose location in 3d facial data. *Pattern Recognition Letters*, 27(13):1487–1494, 2006.

[52] P. Nair and A. Cavallaro. 3-d face detection, landmark localization, and registration using a point distribution model. *IEEE Transactions on Multimedia*, 11(4):611–623, 2009.

[53] M. Romero-Huertas and N. Pears. 3d facial landmark localisation by matching simple descriptors. In *2nd IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–6. IEEE, 2008.

[54] B. Allen, B. Curless, and Z. Popović. The space of human body shapes: reconstruction and parameterization from range scans. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 587–594. ACM, 2003.

[55] C.C. Teng, M.M. Austin-Seymour, J. Barker, I.J. Kalet, L.G. Shapiro, and M. Whipple. Head and neck lymph node region delineation with 3-d ct image registration. In *Proceedings of the AMIA Symposium*, page 767. American Medical Informatics Association, 2002.

[56] K. Wilamowska, L. Shapiro, and C.L. Heike. Classification of 3D face shape in 22q11. 2 deletion syndrome. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, pages 534–537. IEEE, 2009.

[57] T. Kadir and M. Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

[58] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[59] J.J. Koenderink and A.J. van Doorn. Surface shape and curvature scales. *Image and vision computing*, 10(8):557–564, 1992.

[60] J. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

[61] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and L.H Witten. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*, 11(1):10–18, 2009.

[62] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[63] J. Wu, R. Tse, C.L. Heike, and L.G. Shapiro. Learning to compute the symmetry plane for human faces. In *ACM Conference on Bioinformatics, Computational Biology and Biomedical Informatics*. ACM, 2011.

[64] S. Liang, J. Wu, S. M. Weingberg, and L. G. Shapiro. Improved detection of landmarks on 3d human face data. In *IEEE Engineering in Medicine and Biology Society Annual Conference*. IEEE, 2013.

[65] C. Asher-McDade, C. Roberts, W.C. Shaw, and C. Gallager. Development of a method for rating nasolabial appearance in patients with clefts of the lip and palate. *The Cleft Palate-Craniofacial Journal*, 28(4):385–391, 1991.

[66] D. M. Fisher, R. Tse, and J.R. Marcus. Objective measurements for grading the primary unilateral cleft lip nasal deformity. *Plastic and reconstructive surgery*, 122(3):874–880, 2008.

[67] Tie-Yan Liu. Learning to rank for information retrieval. *Foundations and Trends in Information Retrieval*, 3(3):225–331, 2009.

[68] J. Neter, W. Wasserman, M.H. Kutner, et al. *Applied linear statistical models*, volume 4. Irwin Chicago, 1996.

[69] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K.R. Murthy. Improvements to the smo algorithm for svm regression. *Neural Networks, IEEE Transactions on*, 11(5):1188–1193, 2000.

[70] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer. An efficient boosting algorithm for combining preferences. *The Journal of machine learning research*, 4:933–969, 2003.

[71] C. Burges, T. Shaked, E. Renshaw, A. Lazier, M. Deeds, N. Hamilton, and G. Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd international conference on Machine learning*, pages 89–96. ACM, 2005.

[72] V. Dang and B.W. Croft. Feature selection for document ranking using best first search and coordinate ascent. In *Proc. of SIGIR 2010 Workshop on Feature Generation and Selection for Information Retrieval*, 2010.

[73] V. Dang. http://sourceforge.net/p/lemur/wiki/ranklib/, 2013.

[74] I. Atmosukarto, L.G. Shapiro, J.R. Starr, C.L. Heike, B. Collett, M.L. Cunningham, and M.L. Speltz. Three-dimensional head shape quantification for infants with and without deformational plagiocephaly. *The Cleft Palate-Craniofacial Journal*, 47(4):368–377, 2010.

[75] A. S. Mian, M. Bennamoun, and R. A. Owens. Automatic 3d face detection, normalization and recognition. In *3DPVT*, volume 6, pages 735–742, 2006.

[76] X. Zhu and D. Ramanan. Face detection, pose estimation, and landmark localization in the wild. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2879–2886. IEEE, 2012.

[77] J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.