Reconstruction and Analysis of 3D Individualized Facial Expressions

Jing Wang

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements for the degree

Master of Applied Science

Ottawa-Carleton Institute for Electrical and Computer Engineering

School of Electrical Engineering and Computer Engineering

University of Ottawa

Ottawa, Ontario, Canada

July 2015

Abstract

This thesis proposes a new way to analyze facial expressions through 3D scanned faces of real-life people. The expression analysis is based on learning the facial motion vectors that are the differences between a neutral face and a face with an expression. There are several expression analysis based on real-life face database such as 2D image-based Cohn-Kanade AU-Coded Facial Expression Database and Binghamton University 3D Facial Expression Database. To handle large pose variations and increase the general understanding of facial behavior, 2D image-based expression database is not enough. The Binghamton University 3D Facial Expression Database is mainly used for facial expression recognition and it is difficult to compare, resolve, and extend the problems related detailed 3D facial expression analysis. Our work aims to find a new and an intuitively way of visualizing the detailed point by point movements of 3D face model for a facial expression.

In our work, we have created our own 3D facial expression database on a detailed level, which each expression model has been processed to have the same structure to compare differences between different people for a given expression.

The first step is to obtain same structured but individually shaped face models. All the head models are recreated by deforming a generic model to adapt a laser-scanned individualized face shape in both coarse level and fine level. We repeat this recreation method on different human subjects to establish a database. The second step is expression cloning. The motion vectors are obtained by subtracting two head models with/without expression. The extracted facial motion vectors are applied onto a different human subject's neutral face. Facial expression cloning is

proved to be robust and fast as well as easy to use. The last step is about analyzing the facial motion vectors obtained from the second step. First we transferred several human subjects' expressions on a single human neutral face. Then the analysis is done to compare different expression pairs in two main regions: the whole face surface analysis and facial muscle analysis.

Through our work where smiling has been chosen for the experiment, we find our approach to analysis through face scanning a good way to visualize how differently people move their facial muscles for the same expression. People smile in a similar manner moving their mouths and cheeks in similar orientations, but each person shows her/his own unique way of moving. The difference between individual smiles is the differences of movements they make.

Acknowledgments

First of all, I would like to thank to my parents who showed endless support through my Master studying life. I am also very grateful for my friend Chao Sun who can always put up with my joys and sorrows. My deepest thanks to my supervisor Prof. WonSook Lee from whom I received inspiration and encouragement, and others who associates with my thesis: Chao Sun, Shiyi Huang, Iman Eshraghi and Zhongrui Li. Also, I would like to give special thanks to Niloofar Aghayan who contributes parts of her system to initialize our system.

Dedication

To my

Mother and Father and My Friends

for their invaluable love and endless support

Table of Contents

List of Figures

List of Tables

Abbreviations

2D	Two Dimensional
3D	Three Dimensional
DFFD	Dirichlet Free Form Deformation
RBF	Radial Basis Function
PCA	Principal Component Analysis
TPS	Thin Plate Spline
kCCA	kernel Canonical Correlation Analysis
GA	Genetic Algorithms
FACS	Facial Action Coding System
AU	Action Unit
RVD	Regional Volumetric Difference

Chapter 1.Introduction

Facial expressions are very important in human life. The smile of Marilyn Monroe is very unique and popular in the world. If we can extract her smiling style, people can try to mimic her smile on their own face. The famous TV series *lie to me* is a kind of using micro expressions. People make micro expressions unconsciously. Police can make use of micro expressions to detect criminals. The script of *lie to me* is based on the knowledge of Facial Action Coding System (FACS). FACS is a common standard to describe the movements on the face. Also, for people who are incapable of performing any expression, they can transfer other people's expressions onto their own neutral faces. Electronic communication has been so popular, text messages have been followed immediately with facial expression represented by *Emoji*. For example, when people are texting messages or using any telecommunication software to chat, they often express their feelings by sending *emoji*. *Emoji* can be created by inserting extracted facial expressions on animated faces.

1.1 Motivations

The main motivation of this work is to analyze people's different expressions and find out a way to generate more natural animated expressions.

Facial expression often plays an important role in conveying non-verbal information when people are communicating. There are a lot of available facial expression databases. In [19], the database is in 2D image space. However, to handle large facial poses and increase the understanding of facial behavior, it is better that the database is organized in 3D. Also, there are

databases in 3D space like [51]. They also developed the 3D database to dynamic data. But their work is designed to the purpose of facial expression recognition. What we want is to find a new and an intuitively way of visualizing the movements of any facial expression in the database.

Producing facial expression is also prominent in a wide spectrum of facial animation. The core concept of producing facial expressions is how to extract facial expression motion data. Some researchers choose to reuse existing animation data to produce facial expression models. For human beings, different individual subject shows a strong shape resemblance, reusing existing animation data is a fast and accurate way of producing facial expressions. Once the motion vectors are extracted, the data can be duplicated on new faces. However, people have different expressions and duplicating the same facial expression (the same motion vectors) does not represent the true life smiling of a particular person. What we are trying to do is to establish an expression database. Once given a real human face, the new expressional face is created by the expressions in the database.

1.2 Objectives

In this thesis, our main goal is to provide a way to extract facial expressions and to analyze how differently people make facial expression. To analyze facial expressions, we have chosen to use 3D real-life scanning of people when they smile. Therefore, we could examine the true 3D movements of facial surface by extracting 3D motion vectors from neutral faces to smiling faces.

Among numerous facial expressions, we choose smile as our experimental expression for this thesis. Smile is one of the seven basic expressions defined by Facial Action Coding System (FACS). Smile is formed by a sequence of muscles expanding and contracting around the mouth.

Some smile is also accompanied with muscle contraction around eyes. In this work, the smile analysis starts with an overall analysis of the whole face surface. Then we examine four kinds of facial muscles which are most related to the movements of smiling.

1.3 Contributions

What we have contributed can be concluded in three main aspects:

a) We use real human 3D data for facial expression analysis.

b) We model and visualize facial expressions using 3D motion vectors.

c) We propose a method of facial expression analysis for the whole face as well as for facial muscles for expressions.

Even though people smile in similar style, the smile expressions are unique for each individual. This thesis provides a new methodology and a framework of processing a scanned face for expression analysis, but it is hard to make general conclusions for facial expressions. Our future research direction is to expand our database to perform further analysis including making classification of various smiling styles and exploring the correlation among ethnic groups, ages, genders, situations, facial shapes and characteristics.

1.4 Overview of the System

Our work contains three parts: (i) reconstruction of 3D head system; (ii) expression cloning system; (iii) analysis of 3D expressions.

The first part is to reconstruct the scanned real human faces and consistently parameterize the recreated faces to be used for our facial analysis system. Real human faces are captured by a 3D

laser scanner. The data cannot be used directly due to the inside holes and intersections. Thus, all the input faces are processed through Poisson Surface Construction that can smooth the surface and fill the holes. After that, 128 sparse feature pairs are collected from the generic model and input face model. Since this feature-collection work is finished manually, a 3D geometrical fitting stage is applied on the input face. The 3D geometrical fitting stage aims to bring the input face to the space of the generic model face. The generic model is first deformed by Radial Basis Functions that approach the feature points from the generic model to their correspondence points from the input model. To ensure a complete surface match, each vertex in the input face model is projected onto the generic model by finding the first intersections along the cylindrical projection center line. To make the recreated face surface smoothly and naturally, we employ Single Laplacian Smoothing on the deformed generic model. The preceding process is applied on different faces so that a database is established for our analyzing system.

The second part of our work is to transplant smile expressions onto different individual subjects. Since all the input faces are recreated though the same generic model, they all share the same geometry structure. In this part, the essential step is to extract expression motion vectors. The motion vectors are obtained by subtracting expressional faces to neutral faces. For expression exchange, the new faces are only the sum of motion vectors and their neutral faces.

The third part is smile analysis. For smile analysis, the motion vectors are parsed in length and orientation. Length is examined in three directions: absolute length, vertical length and horizontal length. We will detail the analysis in chapter 4. Figure 1.1 visualizes the overall flow diagrams of this work.

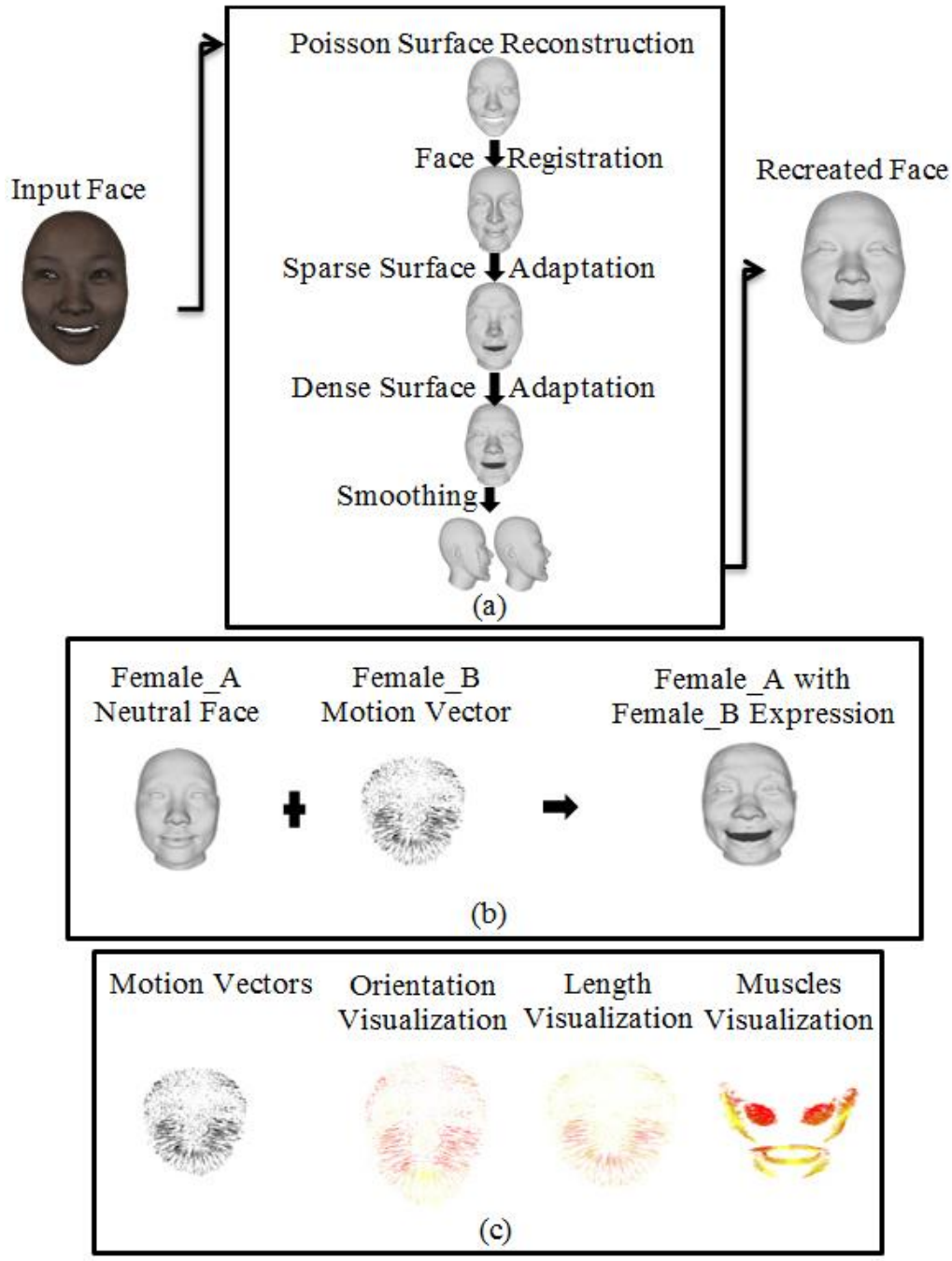


Figure 1.1: The framework of our system. (a) is the part of face reconstruction which we example Female_B's face as the input face; (b) is the part of facial expression exchange which we example Female_B's small smile expression onto Female_A's neutral face; (c) is the visualization for the facial motion vectors and muscle units.

1.5 Thesis Organization

The remainder of this thesis is organized as follows:

- Chapter 2 illustrates the related work of the three fields of face animation that are used in our work: head modeling, facial expression cloning and facial expression analysis.
- Chapter 3 shows the overview of our proposed methodology. It details how to generate a complete head from modifying a generic model. It also covers a fast and simple algorithm of transferring a person's expression onto another person's face.
- Chapter 4 discusses the experiments and results from our system and the finds of comparing different people's smiles. A brief introduction of smile analysis is provided at first. Then, data tables and figures are displayed to keep all the conclusions clear.
- Chapter 5 contains conclusions and contributions of what we have done. We also discuss the future research of our work.

Chapter 2.Related Work

This chapter sequentially introduces related research in face reconstruction, 3D facial expression generation and expression analysis. An animated face can be recreated from 2D image-based systems, stereo video sequence systems, performance-driven systems and other available approaches. To acquire a plausible and realistic human face, capturing high quality performances is a significant task. High-quality face data can be tackled from a 3D scanner or structured light systems. The following section depicts the literature of 3D facial expression generation. 3D Facial expressions can either be generated from facial retargeting or expression cloning. Other researchers produce 3D facial expressions by synthesizing existing expressions. The third part presents the related literature of facial expression analysis. The analysis of facial expressions is mostly based on learning Facial Action Coding System (FACS). We also give examples of available facial expression databases. After that, a brief description of facial muscles is provided which we will get to know it thoroughly in chapter 4.

2.1 3D Face Reconstruction

Reconstructing realistic face model has been a compelling problem since the pioneer work of Parke [21]. Human face data can be captured from image-based systems, facial motion capture systems or structured light based systems. Whether the data is captured by any of these three approaches, the key problem is to find dense correspondences between 2D points or 3D vectors to 3D facial points. For image-based approach, 2D points are extracted from single or multiple images. Key feature points are manually selected and then these 2D points are projected onto 3D

space. During this process, errors are gradually accumulated. Stereo system is efficient in finding 3D facial correspondences, but processing the data needs a lot of estimation, convergence and calibration. In a motion capture system, it is better performed by a professional actor than an average person so that the facial movements can be grabbed by the system.

2.1.1 Synthesizing a 3D Face with a Morphable Model

In 1999, Blanz and Vetter [1] introduced a parametric face modeling technique of generating textured 3D faces. Morphable model means that an virtual face can be linearly combined by a large number of 3D faces in a pre-set database. There are 200 young adult faces with different genders and races in their database. Each face has approximately 70,000 vertices and the same amount of color values. In the database, each model's geometry is represented by two vector sets that one is a shape-vector S and the other is a texture-vector T. By scaling these two vector sets, arbitrary new face can be computed by a linear equation (Equation 2.1),

$$S_n = \sum_{i=1}^{m}(a_i S_i), \quad T_n = \sum_{i=1}^{m}(b_i T_i), \quad \sum_{i=1}^{m}(a_i) = \sum_{i=1}^{m}(b_i) = 1 \qquad (2.1)$$

where a= (a_1, a2,…,a_m) and b= (b_1, b_2,…,b_m) are scaling factors. The system starts with a face analyzer after 2D input images are given. Key techniques are perspective projection and Phong illumination model. In order to reconstruct a realistic face, they first project the 3D model into 2D image plane. An analysis-by-synthesis loop is assigned to shrink the gap between the reconstructed image and the input image. This process can be described as: initialize the morphable model, render an image and update parameters according to the residual difference until rendered images become quite close to the input images. After face analyzer, a new neutral textured face is produced whose only input is a single image. Other face manipulations are also processed by using the modeler in the system such as changing face attributes, making new

expressions and loosing or gaining weight of this person. This work's framework is described in Figure 2.1(reprinted from [1]).

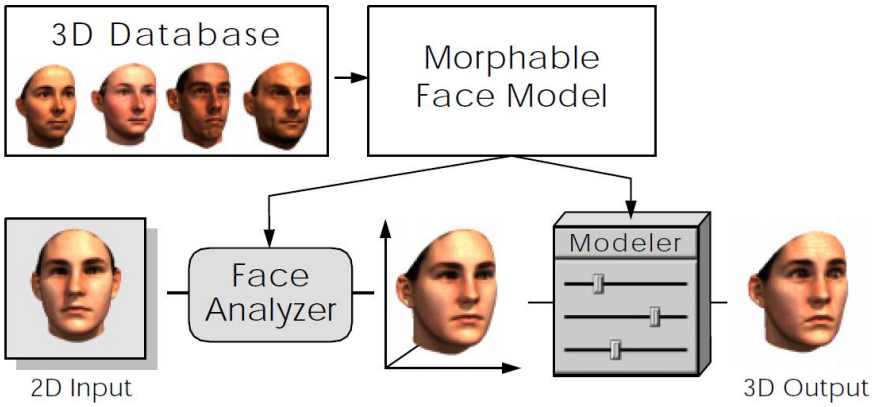


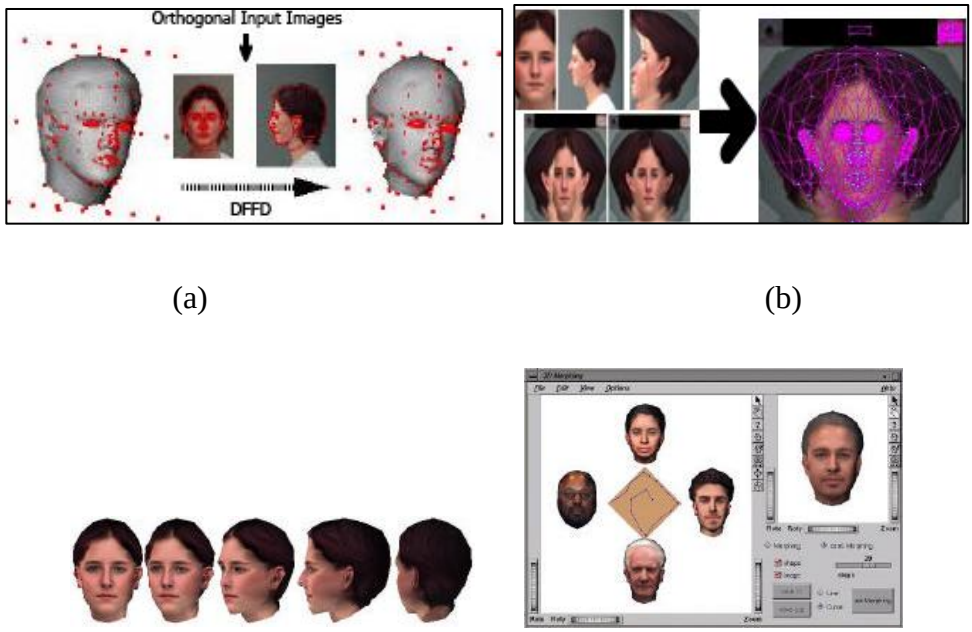
Figure 2.1: Input a 2D image to the system, face analyzer is used to adjust the coefficients of the faces in the database and then face attributes, expressions or other face manipulations depending on the needs of users are processed in the Modeler.

The morphable model referred to in this paper is a milestone in 3D face reconstruction. However, computation of derivatives in each iteration step is time consuming. Also, new arbitrary faces may not be generated only using this 200-young-adults database.

2.1.2 Images-based Animated Faces Generation

Lee and Magnenat-Thalmann [2] proposed another method of 3D face reconstruction. A pre-defined generic model is deformed to adjust the geometry of two orthogonal input images according to the feature points. These two orthogonal images are then normalized in the same space such as the same height of the head and the same level of jaw part. Salient feature points are extracted like nose, eyes and forehead. Given feature points, the generic model is deformed by Dirichlet-Free-Form-Deformation (DFFD). The deformed generic model is only feature-registered to the input face. According to the feature lines which are defined manually, the two

input images are imbedded into one texture image. The texture image contains triangles in highly curved regions and large triangles in elsewhere plain so that it matches the triangular generic face model. Finally, resultant faces are created by mapping texturing points onto deformed generic model. Four basic face models are obtained from the above technique. Furthermore, for user friendly, a dynamic 3D system is constructed for continuous morphing. The workflow of this images-based animated face is shown in Figure 2.2 (reprinted from [2]). Besides texture mapping in this framework, generic model is only sparse registered with a few feature points. However, dense point-to-point surface matching is very important in reconstructing a high-quality face model.



(a) (b)

(c) (d)

Figure 2.2: (a) Generic model deformation based on feature points. (b) Texture generation and texture coordinate fitting. (c) Reconstructed face in different views. (d) Dynamic system for 3D morphing.

In 2010, Choi et al. [3] deployed another images-based face reconstruction approach. The model is generated from a single image or multiple images. Whether the input image is single or multiple, it is first fed to a pose classifier, which is the 'core module' in their approach. Core module is a two-step classifier system. Firstly, it locks the location of the face in input images using generic classifiers. Secondly, three specific classifiers are employed on the locked face. Through Principal Component Analysis (PCA), the system decides the input image belongs to which pose (frontal, half-frontal and profile). In their view-based method, facial landmarks are detected by a range of different view images. The detector consists of a shape model and a texture model. Shape model is a linear combination of basis shapes while texture model is linked

to highlights of a human face such as lip and eye corners. If the input is a single image, 3D face reconstruction is created after texture warping. Otherwise, before texture warping, a face mask is constructed according to the detected facial landmarks. A generic model is then deformed by Thin Plate Splines (TPS) to produce a dense face structure. Their workflow is explained in Figure 2.3 (reprinted from [3]).

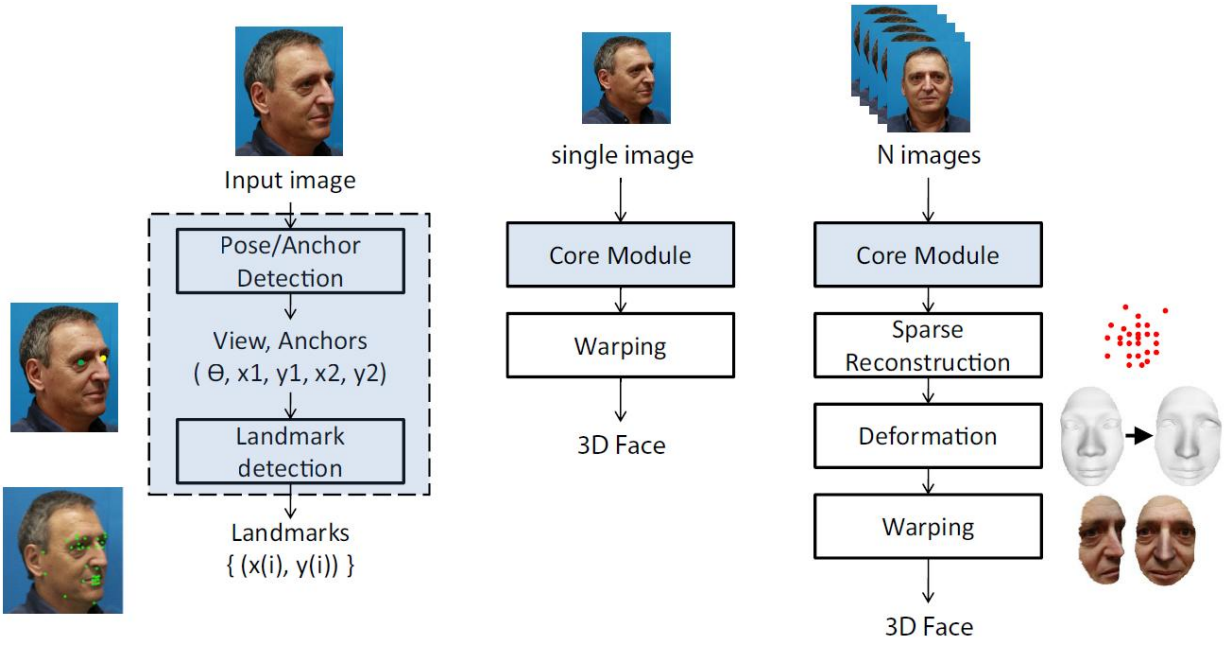


Figure 2.3: The left part is the ‘core module’. Single image processing is in the middle. The right part is Multi-image processing.

2.1.3 3D Face Reconstruction from Stereo Video

A stereo camera is different from an ordinary camera. A stereo camera has the ability of simulating binocular vision. It is capable of capturing three-dimensional images. In 2006, Park and Jain [4] proposed a stereo video method of creating a 3D face. In their approach, two images in specific views are taken by two stereo cameras. They are essential in extracting texture

coordinates and facial landmarks. Afterwards, it is sparse feature point reconstruction. In brief, this step is to build a silhouette of input faces with only 72 sparse feature points. Given camera projection matrices and facial landmarks, it transfers 2D feature points into 3D space. A generic model is modified according to 3D feature points. In the modification, they first do coarse alignment between 3D feature points of the input face and the corresponding vertices of the generic model. After then, non-linear deformation Thin Plate Spline (TPS) is applied for the rest vertices. Figure 2.4 (reprinted from [4]) illustrates how this method works.

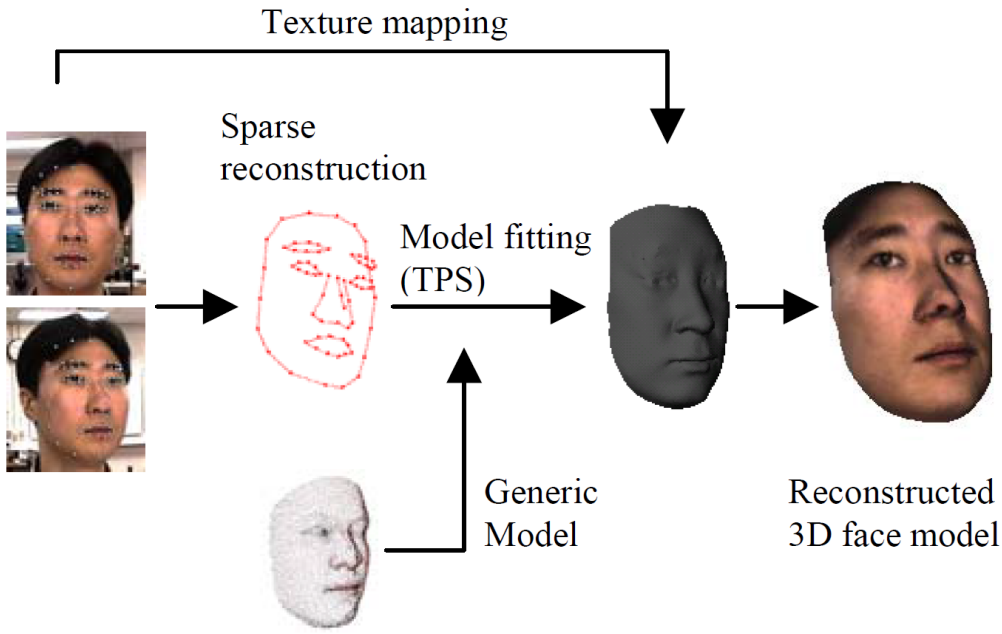


Figure 2.4: System diagram of 3D face reconstruction from stereo video.

2.1.4 High-quality Face Reconstruction Using Monocular Video

In recent research, Garrido et al. [6] embark on reconstructing detailed face mesh from monocular video while Beeler et al. [8] use a binocular-stereo camera to generate a high-quality model. Their system starts from capturing the data from a binocular stereo camera. 29 facial

landmarks are chosen for model registration. Each blend shape model shares the same structure of 200K vertices and the same amount of triangles. After registration, face tracking system detects 66 facial points in 2D plane. To avoid local errors, they perform optical flow correction. Optical flow correction begins with selecting some key frames compared with a reference frame which has a clear appearance. Calculate optical flow between feature locations in each frame. Some local errors are thereby corrected. In their approach, finding dense correspondences is finished by closing the bridge between a synthetic frame and its corresponding true frame. Once a synthetic frame is found, dense 3D points are obtained through projecting 2D points. For a more detailed dynamic face, they estimate capture system's lighting, albedo and the geometry of input face so as to a single environment map can be used for other time stamps. Figure 2.5 (reprinted from [6]) is their system workflow.

Figure 2.5: Left pane is input image. Middle pane is optical flow correction.

Right pane is output.

2.1.5 High-fidelity Facial Performance Acquisition

Another high-quality face reconstruction is introduced by Huang et al. In [5], a sequence of facial expressions with pre-set 100 markers is captured by a twelve-camera *VICON*^{TM} [22] motion capture system. The recording rate is 240 frames per second. A few key frames are chosen for time saving. The actor redisplays the selected facial expressions afterwards. Motion

markers are registered both in two capture system and two expressions (one original expression shown in motion capture system and one redisplayed by the actor). After motion marker registration, it is the dense correspondence registration. They deform a generic model so that it fits salient facial features of the input face. Local refinement is applied on the four segmented face regions. Finally, 3D facial performance is the combination of different weight topology-consistent blend shape models. The weights are calculated from solving a non-negative least square problem. There are a few resultant faces coming out from their system in Figure 2.6 (reprinted from [5]).

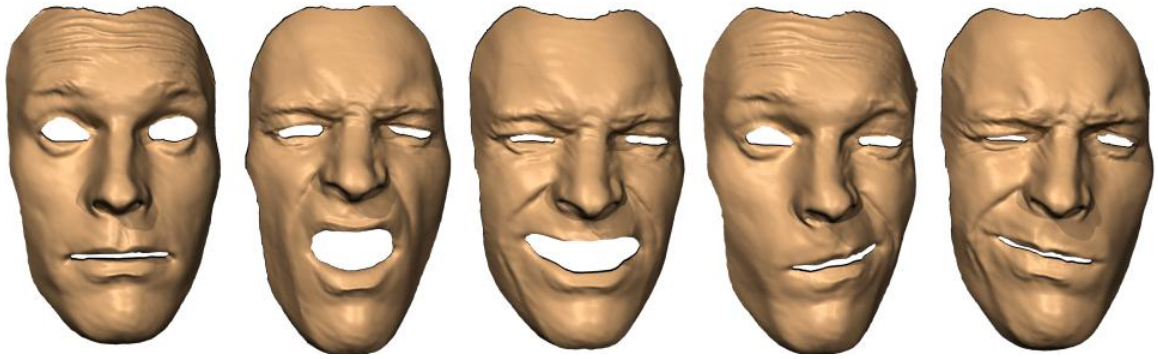


Figure 2.6: Resultant faces of high-fidelity facial performance acquisition.

2.1.6 Other Approaches

In 2011, J. Rafael Tena et al. [9] proposed a method by using linear models based on PCA. In their work, a region-based linear face modeling is introduced. The face is automatically segmented into different parts to increase a model's expressiveness. They do detailed enhancement in sub-models. Users can manipulate and constrain vertices on the face mesh.

Lee et al. [7] propose a way of exaggerating a 3D face model for small or large sized model manufacturing. They construct a plaster cast of a human face. First, a low resolution generic

mesh is deformed based on input face features. Multi-resolution models are achieved by subdivision loops. Figure 2.7 (reprinted from [7]) shows how they produce a plaster face model using a silicone-based substance.

Figure 2.7: The process of constructing the negative and positive moulds of a human face (the results are in the right pane).

Bermano et al. [10] use dynamic shape space to enhance facial performance. They add subtle face attributes such as wrinkles and pores on a low-resolution face. They perform spatial and temporal enhancement to increase the expressiveness of a low-resolution face.

2.2 3D Facial Expression Generation

Turning to 3D expression generation, there have been extensive works on the acquisition of 3D facial expressions. Both using expression cloning [11] [12] and facial retargeting [13] [14] [15] can receive good achievements. Meanwhile, motion capture system (e.g. *VICON*^{TM} system) can also solve this problem. In this section, expression cloning is first introduced while facial retargeting follows. Lastly, other expression approaches are illustrated.

2.2.1 Expression Cloning

In [11], Noh and Neumann pose a scheme on expression cloning. Combining generic model deformation and motion vector manipulation together, new facial characters are obtained. Their scheme consists of two steps. Firstly, they try to find the relationship between the target model and the source model. By deciding 15 to 35 feature vertices from the source model, the correspondences are automatically found in the target model. Secondly, transfer the motion vectors onto the target model. At first, a generic model is sparsely deformed by RBF. Only a few vertices from the source model are projected onto the target model. For an overall complete surface match, cylindrical projection is followed which aims to find a correspondence intersection for each vertex. In expression cloning, their main thought is to transfer motion vectors onto another face. The main attributes of a motion vector are the direction and the magnitude. Direction is adjusted to fit target model according to the transmission of local coordinate and world coordinate. If the proportions of input model and generic model are similar, a simple scale can be fit to all motion vectors using overall bounding box. Otherwise, local scale is computed by calculating the ratio of the source bounding box and the deformed bounding box. The whole processing can be explained in Figure 2.8 (reprinted from [11]).

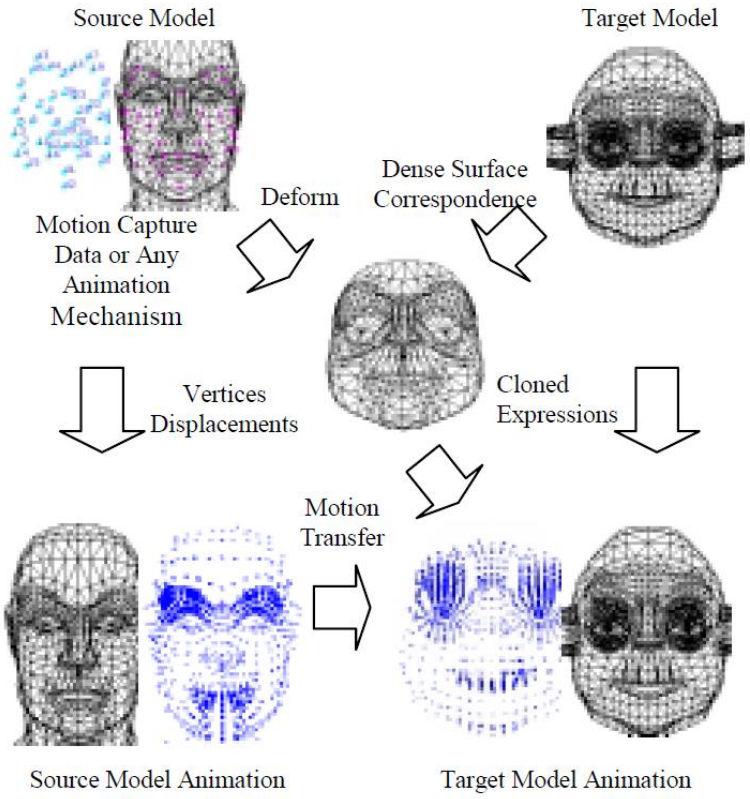


Figure 2.8 : The expression cloning system.

After Noh and Neumann's work, Pyun et al. [12] propose an improved method of expression cloning. Compared to Noh's work, they use an example-based expression cloning technique, and the results are more stable and accurate. Their three-step cloning approach is: key-model construction, parameterization and expression blending. Key models comprise six basic expressional models with 13 verbal models for each expression. They craft the whole 84 key models (6 basic expressions+ 6*13 verbal) by combining the pure expressional models and 13 verbal models by adding weights on each vertex. Key models are composited based on the contributions to face expressions. The expressions around eyes are mainly credit to pure expressional models and the mouth region contributes to both expressional and verbal models. They define the purely neutral expression model as the base model. On the base model, 20

feature points are chosen manually to represent a human face. After finding correspondences of feature points on each key source model, the displacement is obtained by subtracting each key model to the base model using Equation 2.2 [12].

$$\boldsymbol{V}_i = \boldsymbol{S}_i - \boldsymbol{S}_{\boldsymbol{B}}, 1 \leq i \leq M \tag{2.2}$$

All the key models are sorted by a specific order. M represents the number of source key models. Their next step is to evaluate the weights of the target key models. The output model is the blending of those target key models with respect to the computed weights. A weight function is predefined based on the cardinal basis function. The cardinal basis function consists of two parts: linear basis function and radial basis function. The linear basis function is used to approximate the global shape of the weight function. Radial basis function is used to locally interpolate the corresponding key models. Figure 2.9 (reprinted from [12]) is their example-based facial expression cloning.

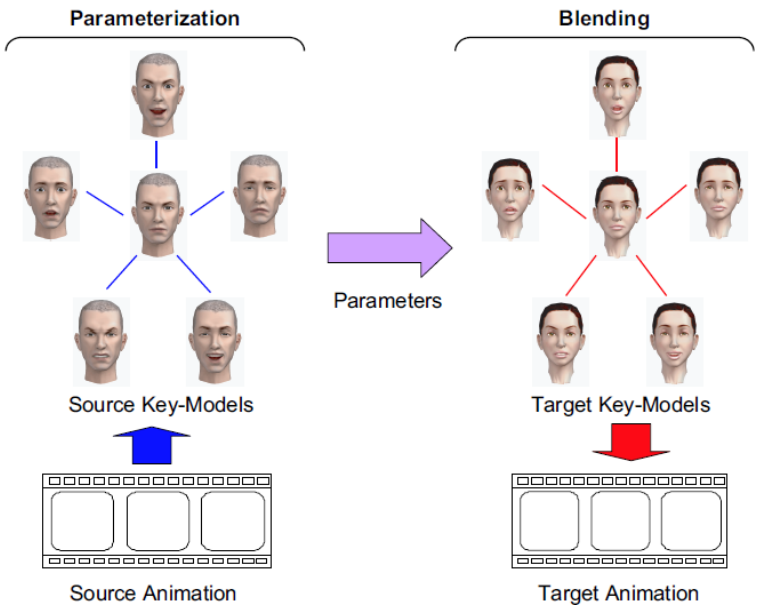


Figure 2.9: Example-based facial expression cloning.

2.2.2 Facial Retargeting

Na and Jung [13] present an example-based approach for mesh retargeting. Their face mesh data is recovered from stripe camera images. Figure 2.10 (reprinted from [13]) shows an input face recovered from a stripe image.

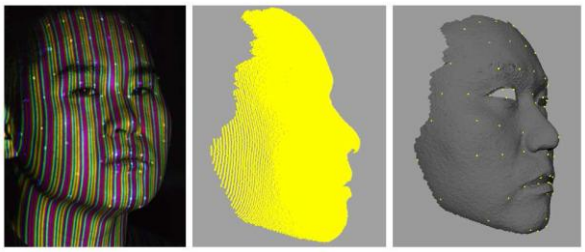


Figure 2.10 : Right pane is a recovered dense 3D face mesh from a stripe image.

They use hierarchical retargeting method for fine facial motions. They do facial retargeting level by level. It stars with a low-resolution base model by connecting only feature points. In temporal, base models in each frame share the same topology. In spatial, users determine feature points both on source and target meshes. Hierarchy retargeting begins with the base model. They use a motion map transformation to retarget each point of the base source model on the target base model. The motion vector of a new expression of the target model is the sum of corresponding points multiplied by varied weights of the example faces. Other level normal models are firstly derived from subdividing base models. After each subdivision scheme, detail information reduces. To make a mimic expression just like the input face, detailed mesh retargeting is applied. Normal difference offsets, which are subtracted from new expression to neutral expression face, are added on target models' normal offsets. Rather than cumbersome work of creating high level mesh, they render the target motion at a low level. In the ensuing methods, Song and Choi [14] extended human facial expression to cartoon characters. They cope

with the problem by a hybrid-retargeting model. The hybrid-retargeting model is the combination of a linear interpolation RBF and a non-linear interpolation kernel Canonical Correlation Analysis (kCCA). The output can be represented in equation 2.3.

$$\boldsymbol{t}_{final} = \alpha(\boldsymbol{t}_{RBF}) + (1 - \alpha)(\boldsymbol{t}_{\boldsymbol{kCCA}}) \tag{2.3}$$

where $\boldsymbol{t}$ is the retargeted vector and α is a user-controlled blending parameter. In [15], Ductreve tackles facial retargeting by a skeleton-based deformation. For the models with different topologies, they use RBF to do space transformation. In the section of target mesh modification, linear blend skinning deformation is conducted. Given a vertex, they compute the distance from this vertex to each feature point. Among those distances, only the one with the most weight is considered as the influence value. All the influence values are normalized so that the sum of the values is 1. Thus, a target motion vector is the product of an influence value multiplies the displacement of feature-point pairs (from the feature point of the expressional face to the feature point of the neutral face).

2.2.3 Other Approaches

Except expression cloning and facial retargeting, Zhu and Lee [17] [18] propose other approaches of 3D expression generation. In [17], they use Genetic Algorithms (GA) to optimize. GA is a regressive technique.

In the first step, 30-expression-faces data bank is constructed where the faces are from the database produced by L. Zhang et al [23]. The data can be found on the website [24] at which two models are chosen in this thesis as well. Figure 2.11 (reprinted from [17]) are the six basic expression faces they use.

(a) (b)

Figure 2.11: (a) Six basic expression faces. (b) Left face is the input face and left is the output of GA system.

Once an expressional face is passed through the system, this face expression is the blending of three expression faces in the databank with weights calculated from genetic algorithms. To find optimal weights, three expression faces are selected manually at the initialization step from 30-expressions databank. The gap is narrowed from judging angles and lengths between the three initial faces' feature motion vectors and the corresponding feature motion vectors of the input face. A new intermediate face is produced in each generation. The new intermediate face replaces the least related face in current round and is counted in to the next round. They also decompose the whole face into left and right parts with middle region overlaid to enhance facial performance. In this work, they only find weights from feature points. It saves time in regression process and makes convergence more quickly, but it makes little use of the other vertices.

After this work, Zhu and Lee propose another method of producing high quality facial expression with motion capture data [18]. Eleven expressions and fourteen visemes are captured from a laser scanner. To make all the models in consistent parameters, they first deform a pre-defined generic model depends on the feature points from scanned faces. Afterwards, Loop's subdivision is performed on the deformed generic model. They then map all the surface points onto subdivided face by cylindrical projection. Generic model adaptation is also deployed in this

thesis, which will be detailed in chapter 3. Then, 32 facial motion markers are captured by an optical tracking system *VICON*^{TM}. For each frame, space normalization transformation is obtained by computing nose tip from the current frame to the neutral frame. This transformation is applied on each motion marker. Figure 2.12 (reprinted from [18]) is their resultant faces from facial motion reconstruction system.

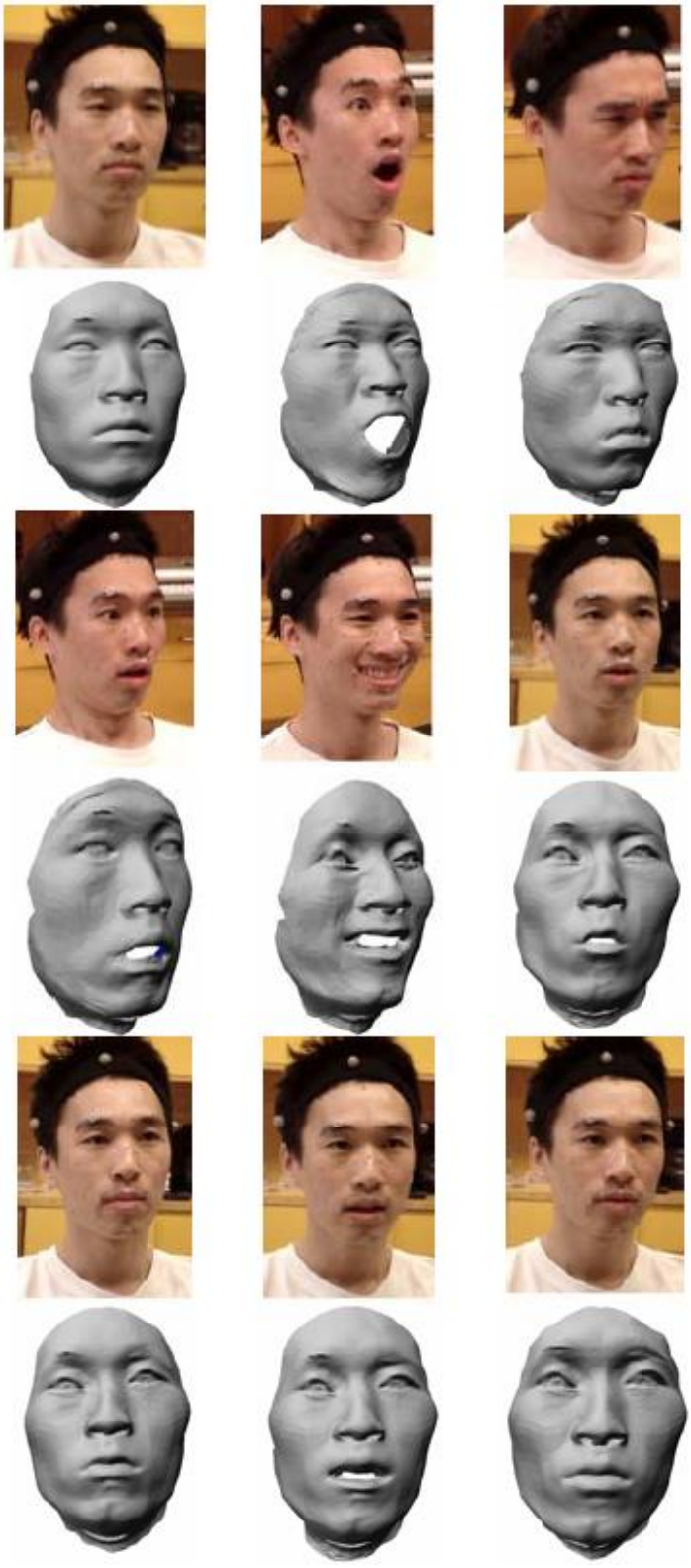


Figure 2.12: Results of reconstructed facial motions.

Motion markers are also converted to the data bank space by using surface blending technique. Again, by applying the aforementioned GA, the final motion vectors are a blending of several example faces from the data bank. Blended motion vectors are retargeted on the neutral face from input face.

2.3 The Analysis of Human Facial Expression

Facial expression analysis helps a lot in exploring new insights of psychological problems, especially in the study of emotions, communication and social interactions. Facial expression provides a behavioral measurement for emotion studying. In [50], Dacher Keltner and Paul Ekman review different scholars' views of expression of emotion. In the review, they give three main fields in expression of emotions: how emotions are expressed, how emotions are perceived and how emotions vary among individuals beyond different backgrounds. In their review, they point out that: emotions are strong related with central nervous system activity; some scholars make inferences that fleeting facial expression indicates intentions, personality, social status and living environment; some authors inquire whether emotional expressions are culturally constructed or universal. A lot of factors can make a difference on the human expression of emotions such as personalities, noisy environment, background music and human voices. Especially, people who have emotional disorders may deliver emotion signals different from average people. Even though great improvement has been made in the study of facial expression, ascertain statements and new discoveries are still being awaited.

2.3.1 Facial Action Coding System (FACS) Action Units

FACS [26] Action Units clearly define various human expressions with related facial muscles. FACS presents precise measurements for facial movement and internal muscle state. There are forty-four action units in FACS. Thirty of them are related to contraction specific muscles and the other fourteen are unspecified. Table 2.1 [27] shows thirty action units, which are related to the contraction of specific facial muscles.

Table 2.1: Thirty FACS Action Units [27].

AU	Facial muscle	Description of muscle movement
1	Frontalis, pars medialis	Inner corner of eyebrow raised
2	Frontalis, pars lateralis	Outer corner of eyebrow raised
4	Corrugator supercilii, Depressor supercilii	Eyebrows drawn medially and down
5	Levator palpebrae superioris	Eyes widened
6	Orbicularis oculi, pars orbitalis	Cheeks raised; eyes narrowed
7	Orbicularis oculi, pars palpebralis	Lower eyelid raised and drawn medially
9	Levator labii superioris alaeque nasi	Upper lip raised and inverted; superior part of the nasolabial furrow deepened; nostril dilated by the medial slip of the muscle
10	Levator labii superioris	Upper lip raised; nasolabial furrow deepened producing square-like furrows around nostril
11	Levator anguli oris (a.k.a. Caninus)	Lower to medial part of the nasolabial furrow deepened
12	Zygomaticus major	Lip corners pulled up and laterally
13	Zygomaticus minor	Angle of the mouth elevated; only muscle in the deep layer of muscles that opens the lips
14	Buccinators	Lip corners tightened. Cheeks compressed against teeth

15	Depressor anguli oris (a.k.a Triangularis)	Corner of the mouth pulled downward and inward
16	Depressor labii inferioris	Lower lip pulled down and laterally
17	Mentalis	Skin of chin elevated
18	Incisivii labii superioris and incisivii labii inferioris	Lips pursed
20	Risorius w/ platysma	Lip corners pulled laterally
22	Orbicularis oris	Lips everted (funneled)
23	Orbicularis oris	Lips tightened
24	Orbicularis oris	Lips pressed together
25	Depressor labii inferioris, or relaxation of mentalis, or orbicularis oris	Lips parted
26	Masseter; relaxed temporal and internal pterygoid	Jaw dropped
27	Pterygoids and digastric	Mouth stretched open
28	Orbicularis oris	Lips sucked
41	Relaxation of levator palpebrae superioris	Lips sucked
42	Orbicularis oculi	Eyelid slit
43	Relaxation of levator palpebrae superioris; orbicularis oculi, pars palpebralis	Eyes closed
44	Orbicularis oculi, pars palpebralis	Eyes squinted
45	Relaxation of levator palpebrae superioris; orbicularis oculi, pars palpebralis	Blink
46	Relaxation of levator palpebrae superioris; orbicularis oculi, pars palpebralis	Wink

In [19], they provide the problem space for facial expression analyze. Their definition of the problem space includes level of description, transitions among expressions, deliberate versus spontaneous expressions, reliability of expression data, individual differences among subjects, head orientation and scene complexity, image acquisition and resolution and relation to non-facial behavior.

They also evaluate their database based on the aforementioned problem. Their AU-coded expression database is either presented with a single AU or the combination of different AUs. A brief description of their database can be seen in Table 2.2.

CMU-Pittsburgh AU-Coded Facial Expression Database	
Subjects	
Number of subjects	210
Age	18-50 years
Women	69%
Men	31%
Euro-American	81%
Afro-American	13%
Other	6%
Digitized sequences	
Number of subjects	182
Resolution	640x490 for grayscale
Frontal view	2105
30-degree view	Videotape only
Sequence duration	9-60 frames/sequence
Action units	
AUs with specific muscles	All but AU 13
AUs with unspecific muscles	AU 8, 38 and 39

Table 2.2: CMU-Pittsburgh AU-Coded Facial Expression Database [19].

This paper covers quite diverse topics, but their analysis is based on 2D images rather than 3D face models. Our analysis is based on 3D real human faces. Also, instead of tracking facial motions between frames, we analyze facial expression directly with motion vectors. In addition, the analysis is deepened into facial muscle units.

2.3.2 Facial Muscles

In our work, we not only establish a database, but also analysis the obtained data. The analysis includes two parts: the first part is conducted on the whole face region and the second part is about the muscles on the face.

Apparently, there are various kinds of muscles on the face. Facial expressions are made of different combinations of a set of muscles. Table 2.3 [28] is the list of muscles on the face. Figure 2.13[1] is the anatomical photo of a human face.

Table 2.3: The list of facial muscles [28].

Number	Muscle
2	Temporoparietalis muscle
3	Procerus muscle
4	Nasalis muscle
5	Depressor septi nasi muscle
6	Orbicularis oculi muscle
7	Corrugator supercilii muscle
8	Depressor supercilii muscle
9	Auricular muscles (anterior, superior and posterior)
10	Orbicularis oris muscle
11	Depressor anguli oris muscle

[1] http://hair-and-makeup-artist.com/facial-anatomy-proportions/

12	Risorius
13	Zygomaticus major muscle
14	Zygomaticus minor muscle
15	Levator labii superioris
16	Levator labii superioris alaeque nasi muscle
17	Depressor labii inferioris muscle
18	Levator anguli oris
19	Buccinator muscle
20	Mentalis

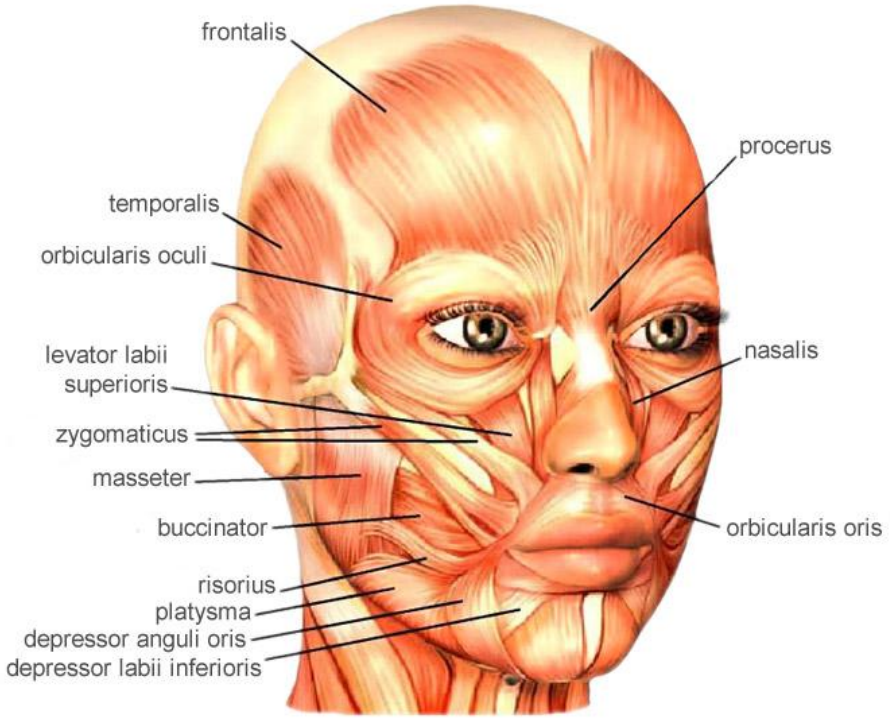


Figure 2.13 : An example of facial muscles of a human head.

2.3.3 MPEG-4 Face Animation

In 1999, MPEG (Moving Picture Experts Group) released a new ISO/IEC standard called MPEG-4 [52]. In MPEG-4 International Standard, one part is the MPEG-4 Face and Body Animation (FBA). It deals with animation of human or human-like characters. For the standard of the face part, MPEG-4 specifies a set of Face Animation Parameters (FAPs) including visemes, primary expressions, 66 low-level and two high-level parameters to allow the display of most facial expressions. FAPs manipulate the key moving points on the face when people are trying to make expressions. The aim of MPEG-4 of defining those FAPs is to allow the interpretation of the facial movements on any facial model in a consistent way. On a neutral human face, it defines 84 feature points. Figure 2.14[2] displays at which the feature points locate on the face.

[2] http://what-when-how.com/face-recognition/facial-action-tracking-face-recognition-techniques-part-1/

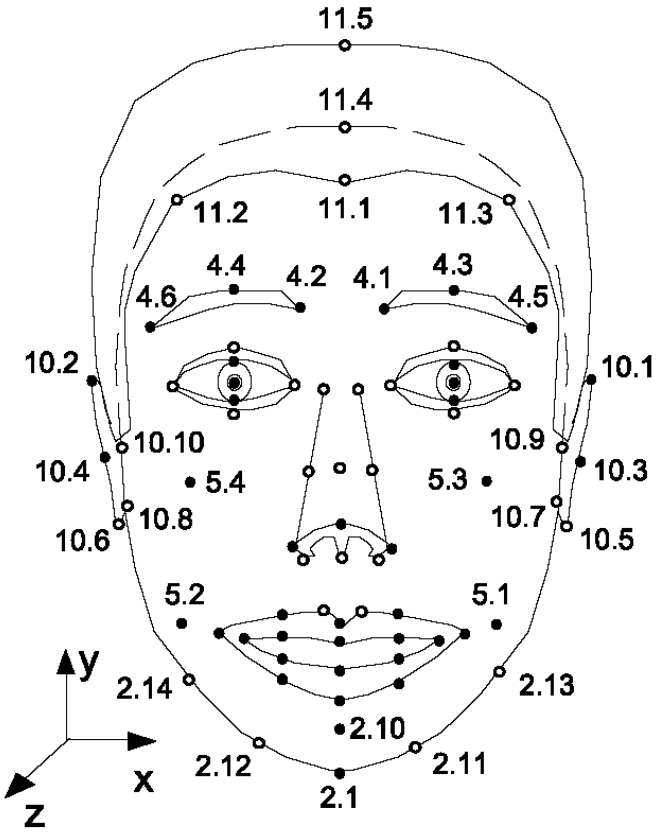


Figure 2.14 : MPEG-4 feature points set.

2.3.4 3D Facial Expression Database

Even though, human face has a relatively flat pattern, 2D image-based facial expression database limits the variations on real human facial expressions like [19]. Figure 2.15 shows frontal and 30-degree views from the CMU-PITTSBURGH AU-Coded Face Expression Image Database[3]. Each sequence begins with a neutral expression and proceeds to a target expression. In Figure 2.15, the target expression is surprise.

[3] http://www.pitt.edu/~emotion/ck-spread.htm

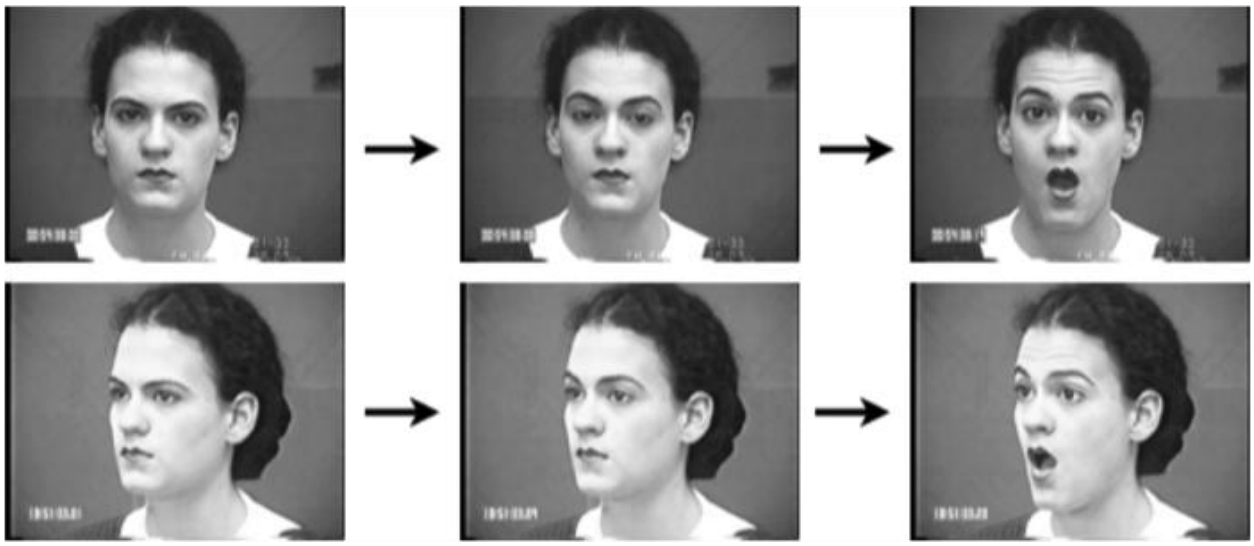


Figure 2.15 : Frontal and 30-degree views from the CMU-PITTSBURGH AU-Coded Face Expression Image Database.

In 2006, Lijun Yin et al [51] create 3D facial expression database. Their database contains 2500 face models of 100 human subjects. The set-up of their 3D face imaging system includes six high-quality cameras and two light pattern projectors. Each individual subject is instructed to perform seven universal expressions (neutral, happiness, disgust, surprise, fear, sadness, and angry) in front of the capture system for a short period of time. They also ask all the subjects to perform four levels (low, middle, high, and highest) of intensity for each expression. The outcome face models are recreated using a stereo photogrammetry technique. Figure 2.16 (reprinted from [51]) shows sample models in their database.

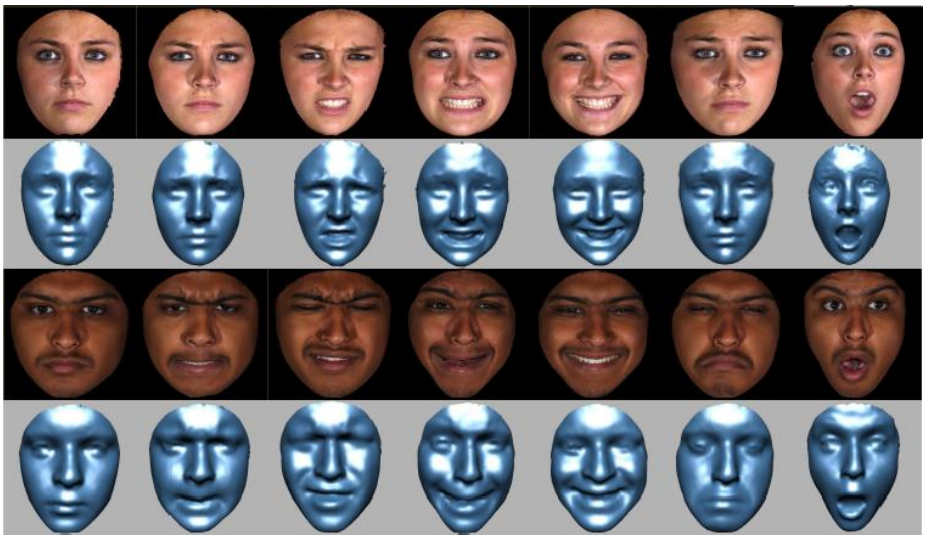


Figure 2.16 : sample subjects showing seven expressions (neutral, angry, disgust, fear, happiness, sadness, and surprise).

Facial expression analysis can be conducted in a complete 3D space. Human face is a mobile, bump surface [51]. In 3D facial expression analysis, researchers can analyze plain expressions as well as extreme expressions. Not only the 3D face database can be used for facial expression analysis, it can also be used in a wide spectrum in the facial animation field such as face recognition, face expression recognition, and facial behavior research.

Chapter 3.Expression Modeling and Expression Cloning

Chapter 3 details the system we build including how to produce face models and how to exchange smile expressions between different people. This chapter has two sections: head reconstruction and expression cloning. In section 1, we introduce complete head model recreation by modifying a generic face model. We enumerate the techniques used in the system along with exemplar pictures. As a few head models have been obtained from section 1, section 2 gives a very stable and fast way of transferring expressions on different faces. We use these two parts to establish our database. Our research analysis of smile expressions in the database is discussed in chapter 4.

3.1 Scanning and Reconstruction of Complete Head Models with Consistent Parameterization

The first part of our system is to produce human heads for expression cloning and expression analysis. We use a part of Niloofar Aghayan's system [29] to initialize our head reconstruction system.

All the recreated head models are generated from the generic model in Figure 3.1. It is a triangle-based mesh with 13074 faces and 6587 vertices. The generic model is deformed and modified to fit input faces. Since most expressions in our input faces are mouth-opened, the mouth of the generic model is separated.

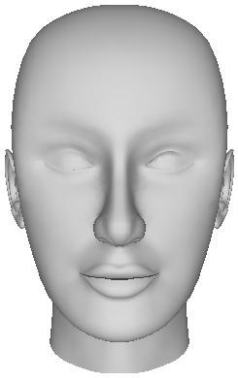


Figure 3.1: The generic head model.

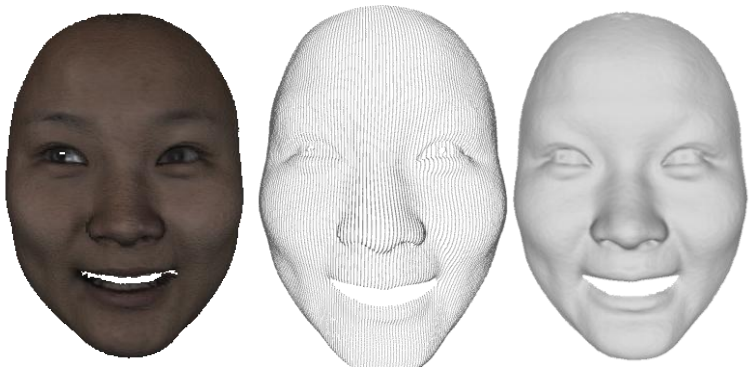
3.1.1 Mesh Construction from Point Clouds Scanned from the Laser Scanner

Before performing any technical process, our input models are either captured from a 3D laser scanner or from online sources. For models from a 3D laser scanner [30], subjects are seated in front of the machine for a few seconds and each output from the scanner is a high-quality triangular mesh (350k vertices and 700k faces). Figure 3.2[4] shows the real person sits in front of the laser scanner. The online data we searched out is from [25] which are produced by L. Zhang et al [24]. The online data is well refined that we can use directly. However, the raw data from the laser scanner is quite noisy because of the missing holes on the face and other useless parts. Plus, the surface is unevenly distributed. The raw data, therefore, is first manually preprocessed. The preprocess includes cutting unnecessary parts, flipping and rotating the model to coarsely fit the space of generic model.

[4] http://www.dgp.toronto.edu/~jacky/masters.html

Figure 3.2 : the Cyberware laser scanner captures the real 3D human face data.

In order to convert all the input data into the same file format (for example, PLY files to OBJ files), Poisson Surface Reconstruction [31] technique is employed. It can fill scattered missing holes of the input face. Poisson method reconstructs the surface with oriented points. It considers all the points at once, assuming that all the given points are inside one surface. Since the surface of online source model is still not smooth enough, Poisson Surface Reconstruction is also applied on. After reconstruction, the surface is well connected. Figure 3.3 is an example using Poisson Surface Reconstruction.



(a)

(b)

Figure 3.3: (a) is the scanned face of Female_B and its Poisson reconstructed surface is on right; the middle is the original point clouds of Female_B; (b) is the online source model with its reconstructed face beside.

3.1.2 Face Registration

The main idea of creating a complete head is to adapt a generic model according to the salient features from the source model. Before applying any deformation on the generic model, the first thing is to transform the space of input model to the space of the generic model. This process is called *Face Registration*.

Two screenshots are undertaken on the generic face and the input face. Both the screenshots are in the same size of width 800 by height 800. Based on the MPEG-4 Face and Body Animation Standard, we exclude some feature points on tongue, ear and hair areas. Our thirteen feature points are the subset of the total 84 feature points defined in the standard. Like the points around eyes or mouth, they are salient features of a human face which makes very little visual errors when undertaken manually. The points obtained from this step is only the 2D coordinates of the features on the image plane, then we do back-projecting method which projects 2D points into 3D coordinates. After seizing 3D coordinates of the generic model and the input model, Procrustes Analysis [32][33][34] is used to register these two faces. Figure 3.4 displays the place at which 13 feature points locate.

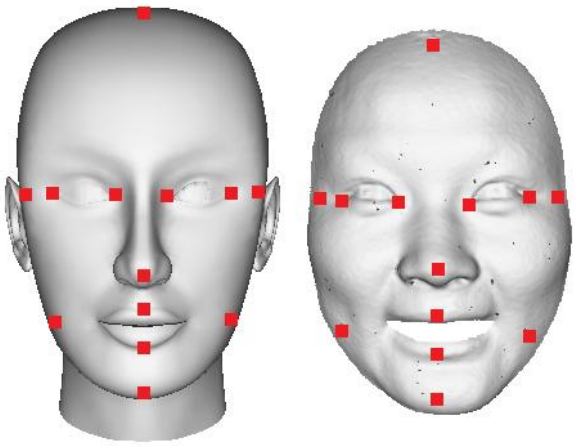


Figure 3.4: The figure shows thirteen feature points on the face (the left is the generic face models and the right is the input face).

Given those two point-sets with each set has the same number of parameters, Procrustes Analysis returns the best matching transform matrix mapping the points from the input to the points of the generic model. Procrustes Analysis also returns the linear transformation (includes translation, scale and rotation or reflection) which can adjust the point of the input mesh so that it

best conforms the corresponding point of the generic mesh. Afterwards, each vertex in the whole input mesh is reset by the Equation 3.1:

$$v_r = p_b * (v_e * p_T) + p_c \tag{3.1}$$

p_c , p_T , p_t respectively represent scale component, orthogonal rotation and reflection component, and translation component. v_r is the point v_e after face registration. Figure 3.5 illustrates the results after face registration.

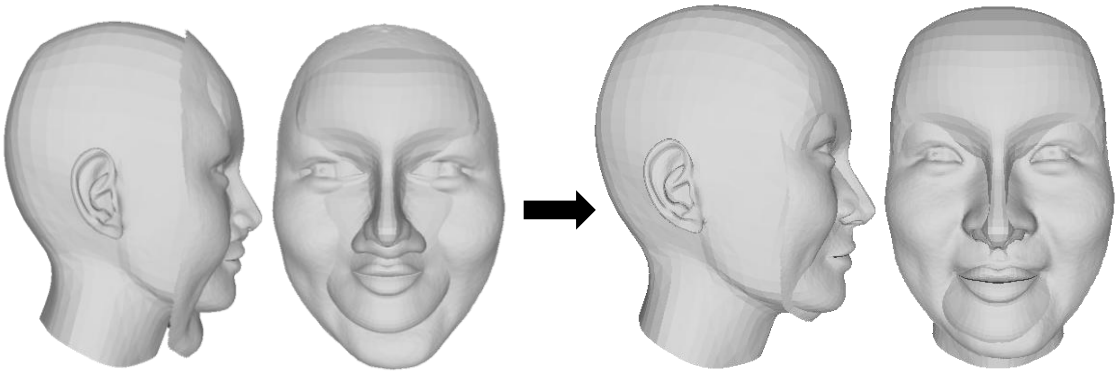


Figure 3.5: The left two are the input face and the generic face; the left two are the front and profile views of the registered face superimposed on the generic model.

3.1.3 Generic Model Adaptation for Consistent Mesh Structure

Feature Extraction

In generic model adaptation, RBF is the main technique we use. The generic model is first deformed according to sparse feature points. In this step, we extend the 86 feature points, which have already been defined in MPEG-4 FBA standard, to 128 feature points. Those 128 feature points on the face clearly show personal facial characters. There are 39 marks around the face: 26 marks on the face contour, 8 marks on ears and 5 marks on the neck; 89 marks inside the

surface of the face: 10 marks on eyebrows, 24 marks around eyes, 18 marks on upper and lower lips, 26 marks on the nose area and another 11 marks scattered in cheeks, the nose root, chin, and the forehead.

For people with exaggerated expressions, especially movements around mouth, regional care should be taken into considered. Two separate regions are predefined when extracting feature points: seven points are set just over the upper lip (upper-lip region); nine points trails the curve of the jaw (lower-lip region). These sixteen points enclose the whole movements of the mouth. Figure 3.6 displays the 128 feature points on the face.

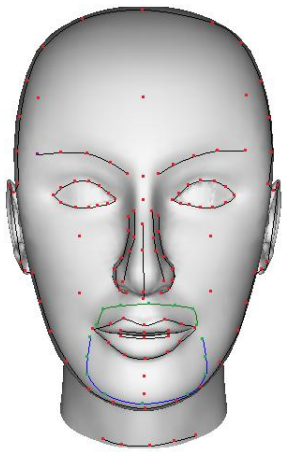


Figure 3.6: 128 feature points (red dots) on the surface of the generic model; 16 feature points encloses the mouth (7 green dots in upper-lip region and 9 blue dots in lower-lip region).

Sparse Feature Points Adaptation

In sparse feature-points adaptation, we employ RBFs [35] [37] [49] to achieve this process. The basic idea of using radial basis functions in the head modeling system is to interpolate a vertex of the input model with the RBF network function. The new position of this point is the sum of its original value and the displacement obtained from solving the interpolation function.

Only the positions of the input vertices are needed without considering the orientation normal of each point. There are a lot of forms of radial basis functions, the function which is used in our system to calculate the distance between a vertex and a feature point is the shifted log function (Equation 3.2):

$$\Phi(\mathrm{r}) = \sqrt{log_{10}(r^2 + 1)} \tag{3.2}$$

Two sets of 128 feature points are obtained from feature-extraction process. The other points are approximated by RBF interpolation. Given 128 feature points on the surface M, the RBF interpolation function F for all the other vertices inside the same surface is:

$$F(X) = \sum_{i=1}^{N} c_i \Phi(||P_i - X||) + a_0 + a_1 X_x + a_2 X_y + a_3 X_z$$

$$X = (X_x, X_y, X_z) \tag{3.3}$$

The function Φ is the radial basis function in Equation 3.2. Equation 3.3 returns the displacement of a vertex. The term N is equal to 128, which represents the maximum amount of feature points. P_i is the position of a feature point. X is the vertex position we want to estimate. The coordinates of P_i and X are both in 3D space. c_i is the scalar coefficient.

In order to determine the values of c_i and a_i, define $C = (c_1, c_2, \dots, c_N, a_0, a_1, a_2, a_3)$ and $F = (F_1, F_2, \dots, F_N, 0,0,0,0)$. Equation 3.3 is shortened to:

$$F = GC \rightarrow G = FC^{-1}$$

$$F = GC \rightarrow C = G^{-1}F \tag{3.4}$$

Therefore,

$$G = \begin{pmatrix} \Phi_{11} & \Phi_{12} & \bullet & \Phi_{1N} & 1 & x_1 & y_1 & z_1 \\ \Phi_{21} & \Phi_{22} & \bullet & \Phi_{2N} & 1 & x_2 & y_2 & z_2 \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \Phi_{N1} & \Phi_{N2} & \bullet & \Phi_{NN} & 1 & x_N & y_N & z_N \\ 1 & 1 & \bullet & 1 & 0 & 0 & 0 & 0 \\ x_1 & x_2 & \bullet & x_N & 0 & 0 & 0 & 0 \\ y_1 & y_2 & \bullet & y_N & 0 & 0 & 0 & 0 \\ z_1 & z_2 & \bullet & z_N & 0 & 0 & 0 & 0 \end{pmatrix}$$

Where, $\Phi_{ij} = \Phi(||P_i - P_j||)$

Since we have 128 feature point values, *(c_1, c_2,..., c_N)* and *(a_0, a_1, a_2, a_3)* are separately solved from the equation 3.4. New position of a vertex on the head model is computed with the resulting coefficients of c_i and a_i. For each dimension, there is an interpolator which has the same *G* but different *F*. Hence, there are different values of *C* in each dimension. For a point of the input face, the new position is determined by adding its original value and the displacement in each dimension. Figure 3.7 contains the generic face, input face and the face after RBF deformation.

Figure 3.7: The left is the generic face; the middle is the input face; the right is the deformed generic face.

Loop's Surface Subdivision

After feature based RBF deformation, the surface of the deformed model is not well connected. This disconnectivity is like something promptly pops up between the two ends of a

bridge. Plus, the deformed generic model after RBF process is not in a high-quality state. In order to make up with these flaws, Loop's surface subdivision is deployed.

Loop's Surface Subdivision [38] [39] [40] is able to increase the resolution as well as to smooth the mesh surface. There are three passes in Loop's surface subdivision. The first pass is to add mid-point vertices in the three edges of a triangle. The second pass is to use new edges to connect the generated mid-point vertices and old edges are updated. A triangle is now changed into four new triangles. The process of these two passes can be seen in Figure 3.8. The last pass is to smooth the whole mesh surface.

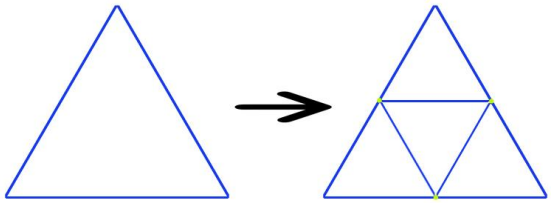


Figure 3.8: the left is the first pass of Loop's Surface Subdivision and the right is the second pass of Loop's Surface Subdivision.

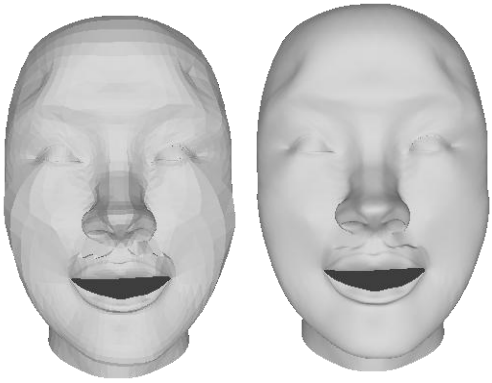


Figure 3.9: The left is the deformed generic face with 6587 vertices; the right is the subdivided deformed generic face with 52296 vertices.

In our experiment, if the generic model is a mesh with 6587 vertices and 26248 faces, after subdivision it increases to a mesh with 13074 vertices and 52296 faces (Figure 3.9). The triangles of the subdivided mesh are four times of the ones of the original mesh.

K-D Tree

After feature-based deformation and surface subdivision, the system turns to surface-points adaptation. In our database, each face surface has at least 20K vertices. Before cylindrical projection, the vertices are sorted in a specific way. We choose tree data structure. All the data are stacked in a balanced binary tree.

K-D Tree [41] [42] is used to partition the data and organize all the vertices inside a "tree". For a 3D vertex, it has three dimensions. Hence, we are building a 3-D tree. The tree starts with the root setting. The root is the median point in the first dimension X. The components in this dimension are less than the root's value are piled up in the left branch and the others are stacked into the right branch. The setting of next level's nodes follows the same strategy in the second dimension Y. In the Y dimension, the points bigger than the node are partitioned to the right branch and the others are put into the left branch. This procedure repeats on both the left and right branches of each node until the last level's nodes are the only components to be partitioned.

K-D tree data structure provides a fast way of searching nearest neighbor if a target point is given. This advantage increases the speed of finding dense correspondences between two point-sets. The nearest distance is calculated in Euclidean Distance. Three-dimension Euclidean Distance equation (Equation 3.2) is:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z - z_2)^2} \tag{3.5}$$

$(x_1, y_{1,} z_1)$ and $(x_2, y_{2,} z_2)$ are two 3D example coordinates.

Given a target point, the search starts from the root. If the x coordinate value is smaller than the x coordinate value of the root, the search goes to the left branch. Otherwise, the search goes to the right branch. The next level search is along the Y coordinate and this procedure continues until finding the leaf (the terminal level) which can be regarded as the 'nearest point'. After that, Euclidean distance is computed between this leaf point and the target point. This distance is considered as the nearest distance. Then, the search goes back to the parent node of this leaf point. If the Euclidean distance of this parent node and the target point is smaller than the nearest distance. The nearest distance updates and the nearest neighbor also updates.

In out experiment, we both find the nearest point of a target point and the points inside a range circle. All the detected points are sorted along Euclidean distance. The one with the smallest value is chosen as the corresponding point. Figure 3.10 shows an example of a 3D tree.

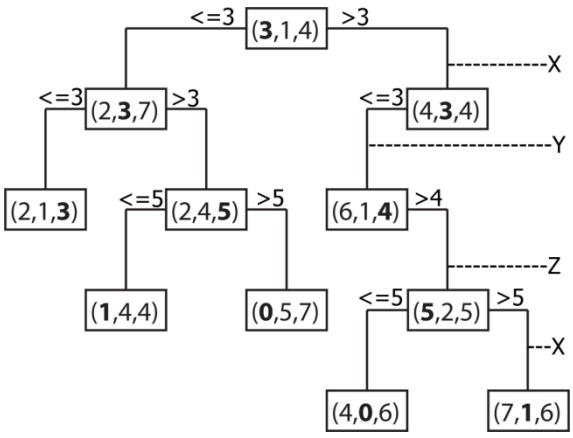


Figure 3.10: An example of k-d Tree. At each level, the points are split along a dimension (X, Y, or Z).

Dense Surface Adaptation

The generic face model is modified according to the 128 feature points. This, however, is not enough for a plausible face model. To find a complete perfect match from the input surface to the

generic model, we use cylindrical projection. A cylindrical projection centerline is established as a vertical line through the centroid of the head. A ray perpendicular to the projection centerline is passed through each vertex in the source model and intersected with triangles in the target model. The first intersection is used in cases of multiple valid intersections.

Every 3D point of the two models are transformed into 2D image plane using the Equation 3.6.

$$u = \arctan\left(\frac{x}{z}\right), \mathrm{v} = \mathrm{y} \tag{3.6}$$

Each 3D vertex (for example, (x, y, z) is marked as coordinates of the modified generic model and (x', y', z') is marked as coordinates of the input surface) is now represented as (u, v) and (u', v'). To decide if a point lies inside a triangle of the modified generic model, barycentric coordinates (α, β) are used to evaluate. If a triangle is represented with ABC (A, B, C are three vertices of this triangle), (α, β) are calculated using Equation 3.7:

$$\alpha = \frac{(V_{AB} \bullet V_{AB}) * (V_{AC} \bullet V_{AX}) - (V_{AC} \bullet V_{AB}) * (V_{AB} \bullet V_{AX})}{(V_{AC} \bullet V_{AC}) * (V_{AB} \bullet V_{AB}) - (V_{AC} \bullet V_{AB}) * (V_{AC} \bullet V_{AB})},$$

$$\beta = \frac{(V_{AC} \bullet V_{AC}) * (V_{AB} \bullet V_{AX}) - (V_{AC} \bullet V_{AB}) * (V_{AC} \bullet V_{AX})}{(V_{AC} \bullet V_{AC}) * (V_{AB} \bullet V_{AB}) - (V_{AC} \bullet V_{AB}) * (V_{AC} \bullet V_{AB})} \tag{3.7}$$

X is the given point which is going to be evaluated. If (α, β >=0) and (α + β <=1), X is considered to be the point lies inside the triangle ABC. In Figure 3.11, there are the results coming out of dense surface adaptation process.

Figure 3.11: the left is input registered face with the deformed subdivided generic model in the middle, and the right face is the one after dense surface adaptation

3.1.4 Smoothing of the Model Surface

Since the input faces do not have back part of the head. After surface point adaptation, the border which connects the input face and the back head of the generic model is not evenly combined. To settle down this problem, a smoothing method is applied. The first step of smoothing is to find the “not well-connected” points. The “not well-connected” points are attained by comparing each vertex of the generic model and each vertex from the unevenly surface model. The attained points can be regarded as board points.

After finding the border points, two-iterative-step smoothing is utilized. During the first iteration, attained border points are replaced with the average values of their neighbor points. Then, the new values are considered as the inputs for the second iteration, the first iteration is repeated on the new values. After two-step iterations, the uneven model is smoothed. Figure 3.12 examples the model before and after applying smooth method.

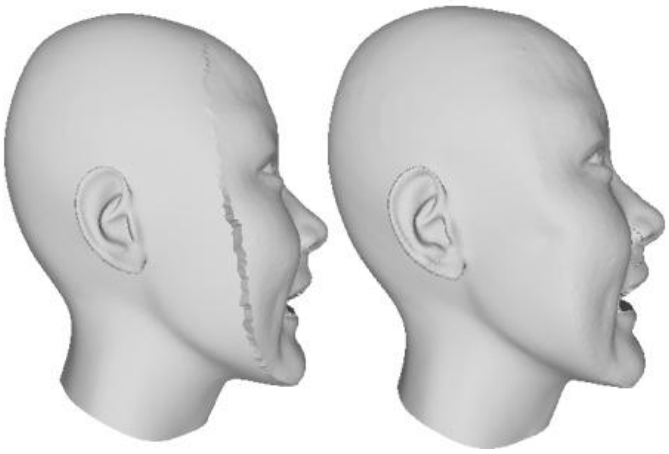


Figure 3.12: Surface smoothing using Average Points Calculation. The left is the side view of the model before smoothing and the right is the side view after smoothing.

3.2 Facial Expression Cloning

3.2.1 Motion Vector

After head reconstruction system, all the models are in the same geometry structure: 89557 vertices and 178800 triangle faces. Since they all come from the same generic face model, it is easy to perform our facial expression cloning.

For each reconstructed model, the back part of the head and ears are merely changed. The main part of the face where the vertices move the most is the face surface. Hence, the motion vector of a facial expression is the difference subtracting from a non-neutral face to a neutral face. Equation 3.8 is the subtraction formula:

$$\text{MV} = v_e - v_n \tag{3.8}$$

In this equation, neutral face vertices is presented by v_n while small smile expression or big smile expression is v_e.

Expression cloning means transferring the motion vector onto another people's neutral face. After attaining the motion vector from the source model, the vertices of the target model is calculated using Equation 3.9:

$$\boldsymbol{v}_{\boldsymbol{target}} = \boldsymbol{v}_{\boldsymbol{target}} + \boldsymbol{mv}_{\boldsymbol{source}} \tag{3.9}$$

3.2.2 Adjust the Magnitude of the Motion Vector

As the models in our system, they all share the similar face shape which means the directions of the motion vectors do not need to be adjusted. However, if the motion vectors from a long face are directly transplanted onto a rectangular face, the result may not look pleased. To adjust the motion vectors from the source model, the motion vectors are simply scaled in proportion to the model size. A global bounding box (BB), therefore, is utilized to scale the motion.

The face model is assumed to be enclosed by a rectangle box. By deciding the edges of the box, we calculate the maximum and minimum values in each dimension (x, y, and z). The calculation process is both performed on the source model (non-neutral face) and the target model (neutral face). Figure 3.13 shows the bounding box of a 3D face model.

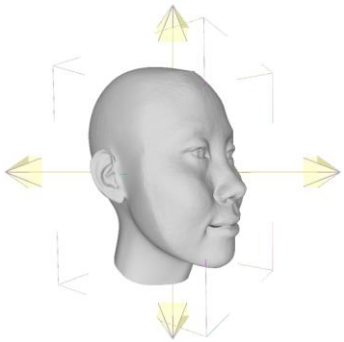


Figure 3.13: A three-dimensional face object sits inside its bounding box (colored lines).

The global scale factor is computed in respective dimension (Equation 3.10):

$$\boldsymbol{f_{x-scale}} = \left|\frac{\boldsymbol{Maximum_{x-target} - Minimum_{x-target}}}{\boldsymbol{Maximum_{x-source} - Minimum_{x-source}}}\right|$$

$$\boldsymbol{f_{y-scale}} = \left|\frac{\boldsymbol{Maximum_{y-target} - Minimum_{y-target}}}{\boldsymbol{Maximum_{y-source} - Minimum_{y-source}}}\right|$$

$$\boldsymbol{f_{z-scale}} = \left|\frac{\boldsymbol{Maximum_{z-target} - Minimum_{z-target}}}{\boldsymbol{Maximum_{z-source} - Minimum_{z-source}}}\right| \quad (3.10)$$

Each motion vector is multiplied by the above three scale factors in respective dimension. Finally, a vertex in the target model $v_{target}(v_{x-target}, v_{y-target}, v_{z-target})$ is displaced by the Equation 3.11:

$$v'_{x-target} = v_{x-target} + f_{x-scale} * mv_{x-source}$$

$$v'_{y-target} = v_{y-target} + f_{y-scale} * mv_{y-source}$$

$$v'_{z-target} = v_{z-target} + f_{z-scale} * mv_{z-source} \quad (3.11)$$

3.2.3 Smoothing of the Model after Expression Cloning

The main function of applying mesh smoothing is to remove the noise with the least destruction to the whole surface. Not only remove the unexpected points, but also maintain the overall shape and salient features of the original models. Various mesh smoothing algorithms are available such as Smoothing spline [47] [48], Laplacian Smoothing [45] [46], Taubin Smoothing [43] [44], etc. In our method, we choose Single Laplacian Smooth. Single Laplacian Smooth is a fast and simple way to smooth a surface. Figure 3.14 is an example of defining a vertex's 1-ring neighbors.

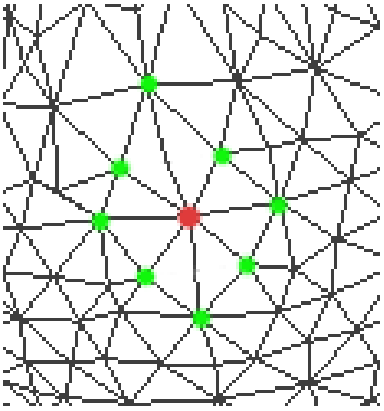


Figure 3.14: The red dot is a vertex from a face model; the green dots are its 1-ring neighbors.

Single Laplacian Smooth is utilized on the whole face model after expression cloning. For each vertex, if there is any triangle related to this vertex, the other two vertices of this triangle are counted as its 1-ring neighbors. After finding all the 1-ring neighbors, the difference is computed between each neighbor and the given vertex. Sum up all the differences and divided this summation by the number of the nearest triangles. A smoothing factor is set to balance the unexpected errors. The final vertex is calculated in the Equation 3.12,

$$v_{final} = v_{original} + factor * \frac{\sum_{i=1}^{N} v_i - v_{original}}{N} \quad (3.12)$$

where N is the number of 1-ring neighbors, the factor is set to be 0.5 in our system.

Figure 3.15 is an example of applying Single Laplacian Smoothing, it does not affect the surface structure but keeping the surface smoothly.

Figure 3.15: the left face is the original and the right is the face after smoothing.

Chapter 4.Facial Expression Analysis

In this chapter, we first present the recreated 3D faces from the head reconstruction system. The detail information of face models in the database is also provided. Then, we show the resultant faces after we apply our method of expression cloning. The third part is dedicated to the smile analysis. The main methods of data analysis are finding and comparing the facial motion vectors of the models in the database. In our experiment, we choose small smile and big smile expressions. It should be mentioned that the whole process could be applied on any facial expression. At last, we examine how facial muscles move when people express different smiles.

4.1 Head Model Database

We use our head reconstruction system to produce fourteen face models. The models in our database are all complete heads with consistent parameters. The detailed information of these models is shown in Table 4.1.

Table 4.1: The information of the five people in the database.

People	Data acquisition	Ethnic	Age
Female_A	Cyberware laser scanner	Asian	20~30
Female_B	Cyberware laser scanner	Asian	20~30
Male_A	Cyberware laser scanner	Asian	20~30
Male_B	Cyberware laser scanner	Caucasian	20~30
Male_C	Cyberware 3030 MS scanner	Asian	20~30

All the input faces are either captured either from the 3D laser scanner and two face models from online resources. The raw input faces could not be used directly. Therefore, we cut useless parts and only use the front surface. After applying the face reconstruction system, all the recreated face models are reconstructed in the same geometry structure: 89557 vertices and 178800 triangles. Also, their back heads are the same since they all inherit the generic model. Figure 4.1displays all the face models in the database.

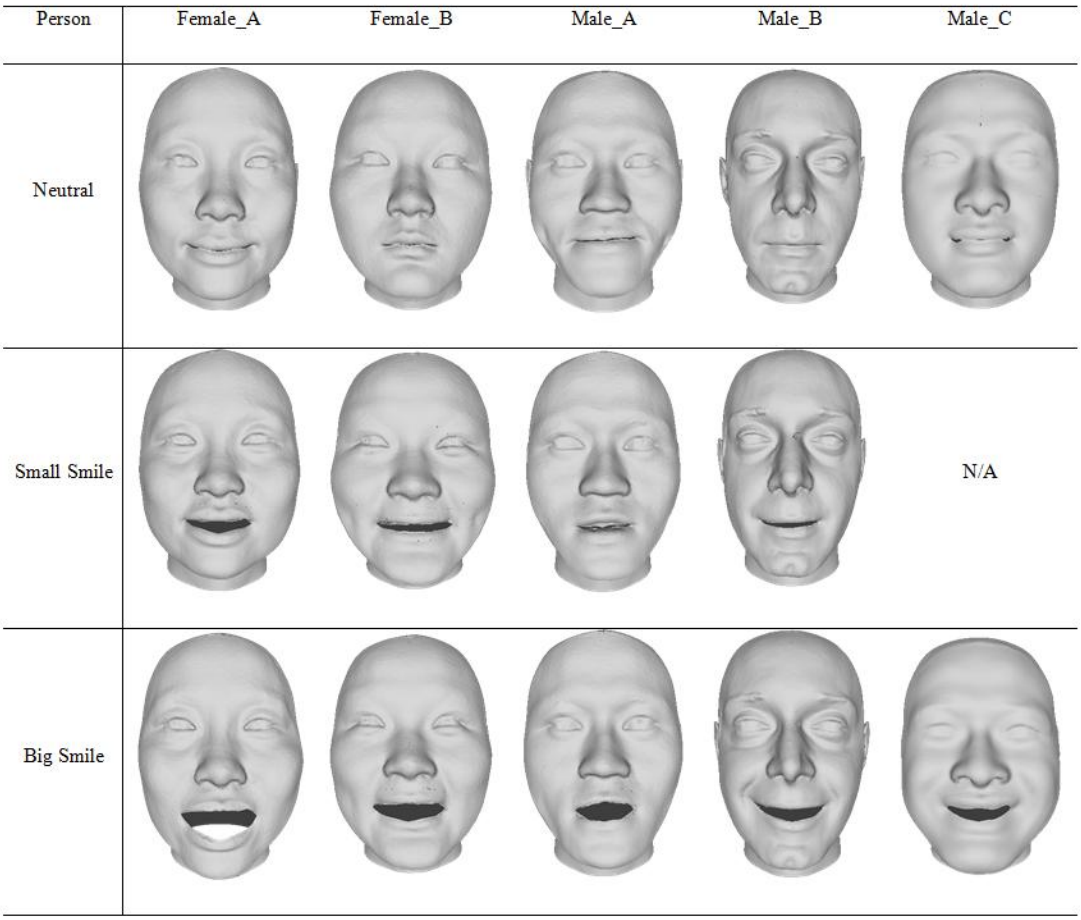


Figure 4.1: The face models in our database include four people's big and small smiles, their neutral expressions, and Male_C's big and neutral faces. In total, there are fourteen face models.

4.2 Expression Exchange

In this section, we will take a look at experimental results. We extract all the people's expression motion vectors and place them on Female_A's Neutral face. Figure 4.2 and Figure 4.3 shows the results. Our way of expression cloning is very stable and simple. The results are acceptable enough compared to the results obtained by other techniques.

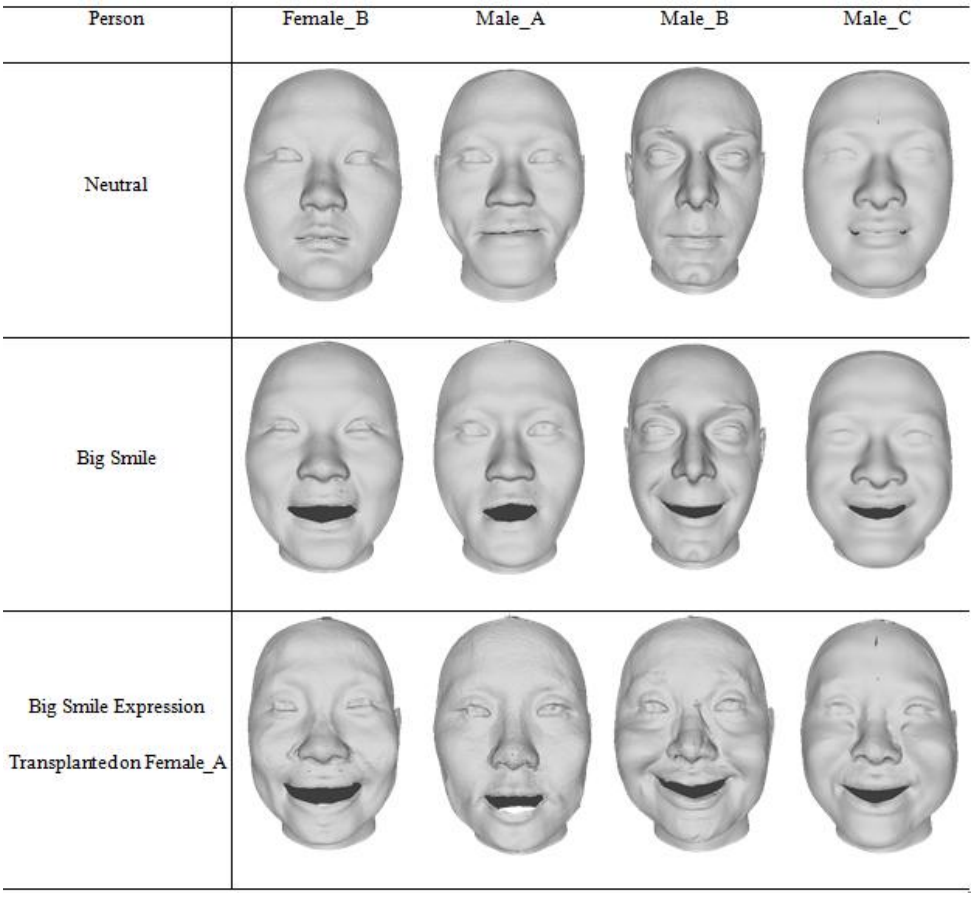


Figure 4.2: Big smile expression exchange. The first row is the neutral faces of Female_B, Male_A, Male_B, and Male_C. The second row is the big smile expressions of Female_B, Male_A, Male_B, and Male_C. The last row is the results of transplanting the four people's big smile expressions onto Female_A's neutral face.

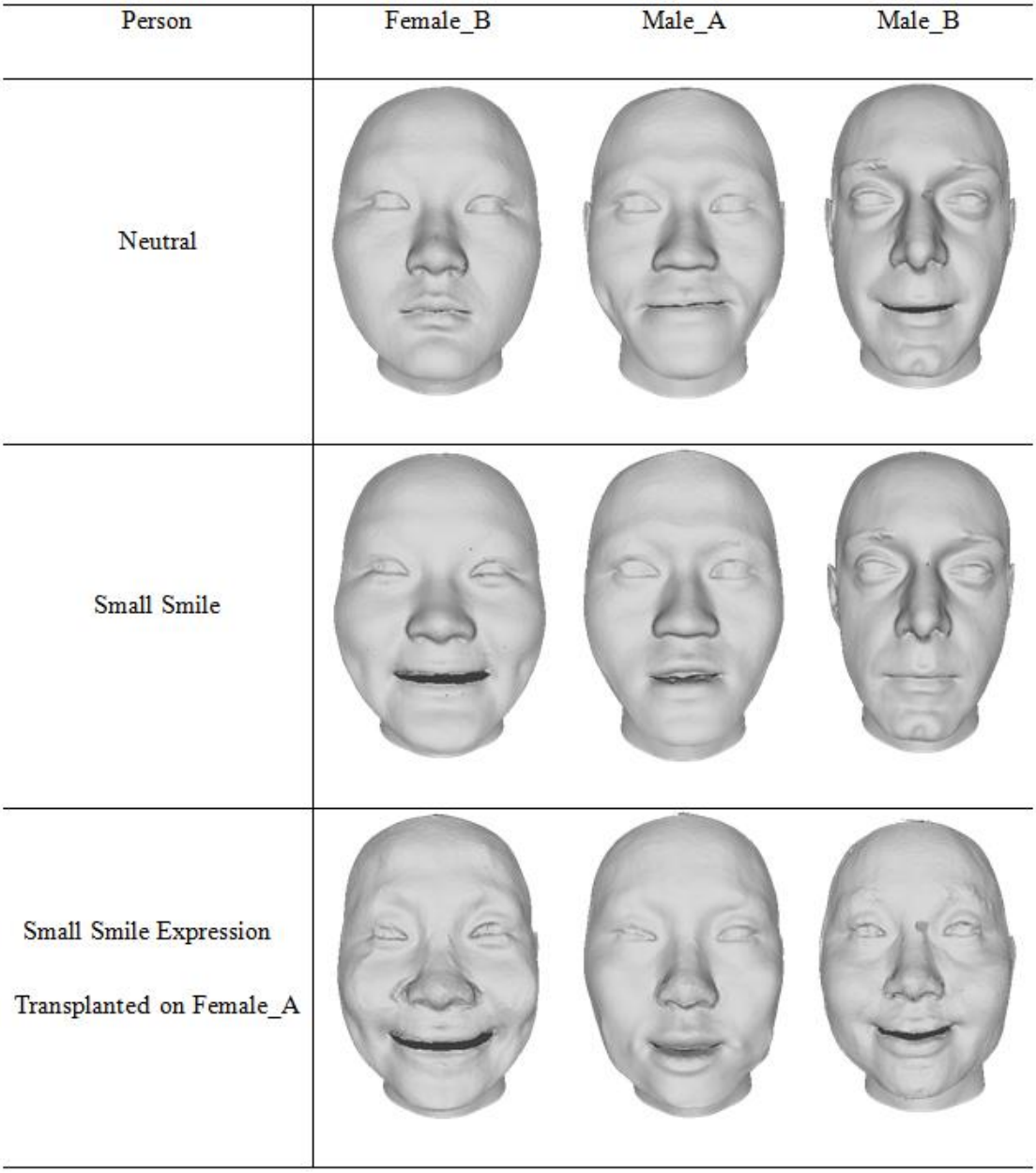


Figure 4.3: Small smile expression exchange. The first row is the neutral faces of Female_B, Male_A, Male_B, and Male_C. The second row is the small smile expressions of Female_B, Male_A, Male_B, and Male_C. The last row is the results of transplanting the four people's small smile expressions onto Female_A's neutral face.

4.3 Analysis of Smile Expressions

In this thesis, five people are set as a database in analyzing human smile expressions. In our expression database, we choose three expressions: neutral face, small smile face and big smile face. Four people's faces are scanned by a 3D laser scanner and another online face from the web source [25].

We are able to conclude from our database by parsing motion vectors and comparing different model pairs. The motion vector is formulated by the Equation 3.8.

The motion vectors are basically parsed in two ways: the orientation and the length. The lengths are analyzed in three directions: absolute length, vertical length and horizontal length. We actually using 3D motion vectors, but we project all 3D vertex points onto two 2D image panel ((x, y) coordinates panel and (y, z) coordinates panel) which makes the visualization much clearer than in 3D image space. In length calculation, for instance, (x_n, y_n) is a point from the neutral face and (x_e, y_e) is a point from small smile or big smile expression.

$$\begin{aligned} &\text{(a) Orientation} = \arctan\left(\frac{y_e - y_n}{x_e - x_n}\right) \\ &\text{(b) Absolute Length} = \sqrt[2]{(x_e - x_n)^2 + (y_e - y_n)^2} \\ &\text{(c) Vertical Length} = |y_e - y_n| \\ &\text{(d) Horizontal Length} = |x_e - x_n| \end{aligned} \tag{4.1}$$

4.3.1 The Analysis of the Orientation of Motion Vectors

The first option of analysing the orientation of the motion vector is to partition all the motion vectors into four quadrants in 2D space. For a given motion vector (for example, MV (x, y)):

- If x>=0 and y>=0, the motion vector locates in the first quadrant with blue color.

- If x<0 and y>=0, the motion vector locates in the second quadrant with red color.
- If x<0 and y<0, the motion vector locates in the third quadrant with green color.
- If x>=0 and y<0, the motion vector locates in the fourth quadrant with black color.

Figure 4.4 gives an example of allocating the motion vectors in four quadrants. To consider the numerical precision limits, we subtract and add 3 degrees for each quadrant's upper and lower bounds.

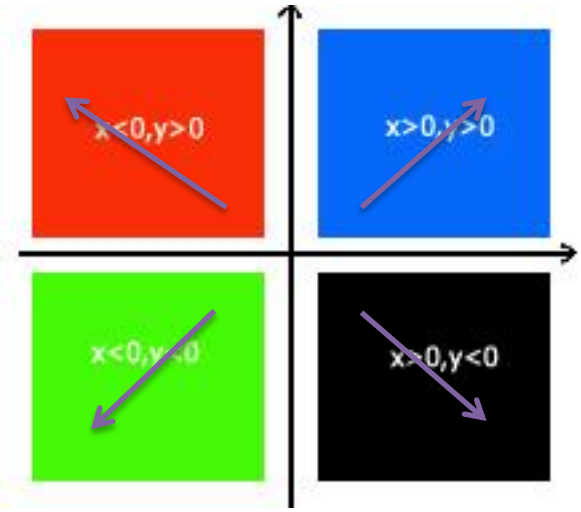


Figure 4.4: Illustration of allocation of a given motion vector (the purple arrows).

When mapping the points of input face surface onto the generic model, it always happens to have error differences. Hence, we exclude the upper forehead part of each mode in the database. We mainly focus on the central surface of the face like cheeks and jaws. The left column of Figure 4.5 is the display of big smile expressions for the five people while the rightmost column of Figure 4.5 shows the small smile expressions. From Figure 4.5:

- Both in small smile and big smile expressions, most motion vectors lies inside the fourth quadrant.

- For the big smiles of Female_B and Female_A whose faces are Asian female style, the amount of the motion vectors locate in the fourth quadrant is fairly the same as the amount of the motion vectors in the third quadrant. However, for the other three male persons, only a few motion vectors are in the third quadrant.
- For each of the five people, the number of motion vectors of big smiles in the first quadrant is almost equal to the number in the second quadrant.
- The above mentioned rules of big smiles are also suitable for small smiles. Most motion vectors concentrate in the fourth quadrant. The third quadrant has the second largest amount. The second and the first quadrants share the same amount of motion vectors.

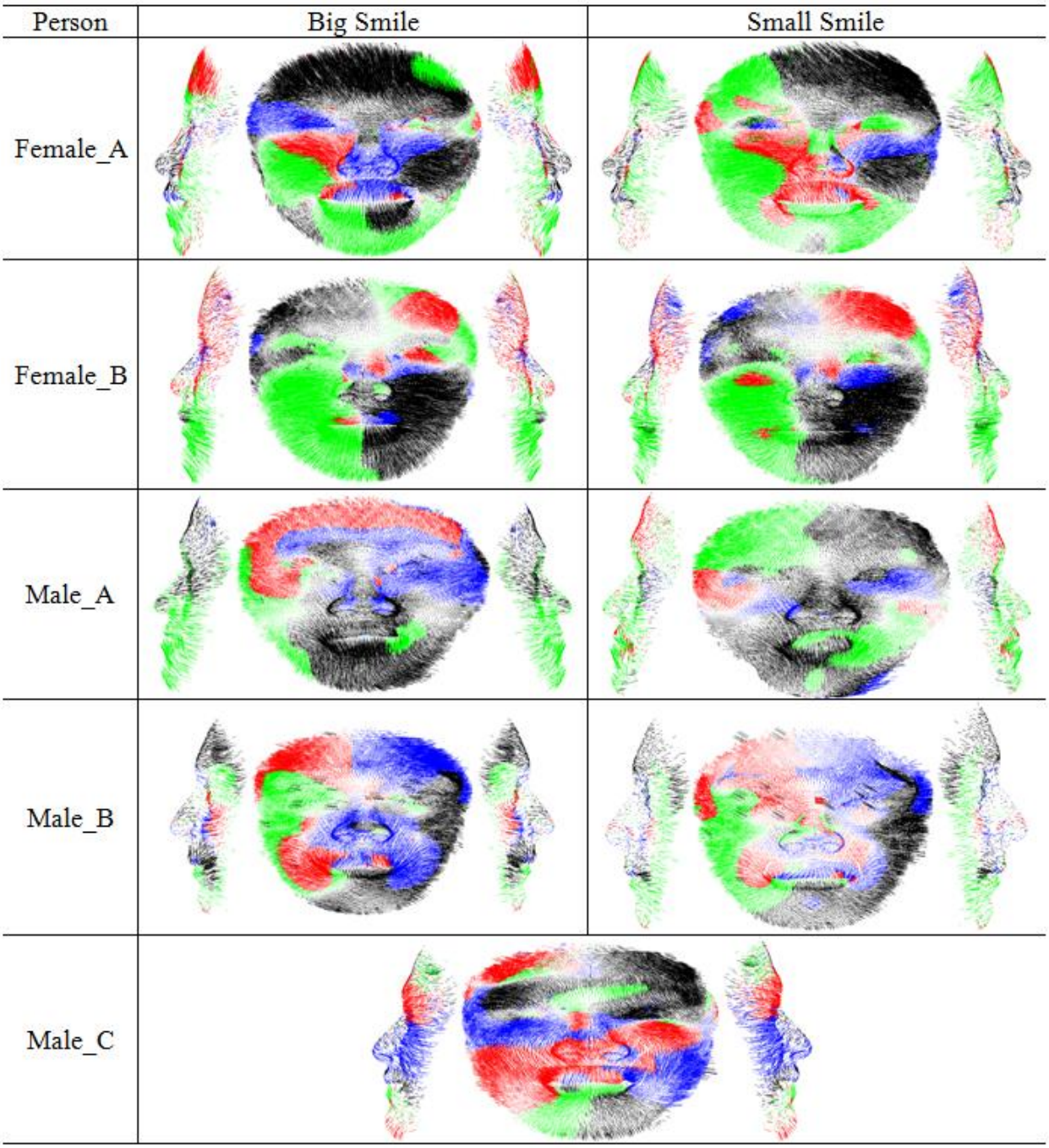


Figure 4.5: Partition MV of big smiles and small smiles into four quadrants (each face has one front and two side views).

The first option of the visualization gives us the information of which range the motion vectors locate the most or the least. Then, we want to know which range the motion vectors

changes the most or the least. Hence, we make continuous colormap visualization for the orientation of the motion vectors. The formula of calculating orientation has been shown in equation 4.1. For a given MV (x, y):

- If the orientation is equal to 0^{o} or 360^{o} ,the color of the motion vector is pure red.
- If the orientation is between 0^{o} and 90^{o} , the color of the motion vector changes gradually from red to yellow.
- If the orientation is between 180^{o} and 270^{o} , the color of the motion vector changes gradually from red to yellow.
- If the orientation is equal to 90^{o} and 270^{o} , the color of the motion vector is pure yellow.
- If the orientation is between 90^{o} and 180^{o} , the color of the motion vector changes gradually from yellow to red.
- If the orientation is between 270^{o} and 360^{o} , the color of the motion vector changes gradually from yellow to red.

Figure 4.6 displays the color map changing with different orientation angles.

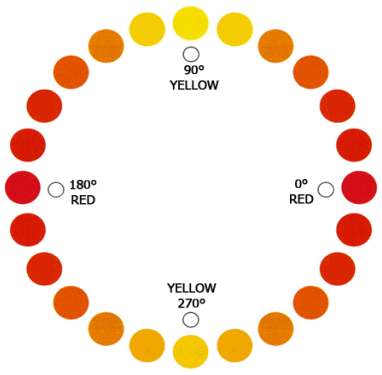


Figure 4.6: The color map of the orientation.

The left column of Figure 4.7 is the color map of orientation in big smile and the right column of Figure 4.7 is the color map of orientation in small smile. By making comparison of different faces:

- For Asian style faces (Female_A, Female_B, Male_A), when they are expressing big smile, except the forehead, the chin and lower lip move in straight vertical direction and the rest of the face tend to move horizontally. But, the moving direction of Female_B's upper lip is totally in the opposite of the movements of the other two people.
- For Caucasian style faces (Male_B), it does not have fixed rules. For Male_B, the region around the bridge of the nose and his upper lip are moving in vertical direction. The other parts of his face are moving horizontally.
- For big smile, Male_C is the only person whose cheeks, mouth and chin move vertically among the five people.
- For small smile expressions, Female_A and Female_B smiles similarly. Most of the face moves horizontally since almost the whole surface is pure red which means the orientation is either around zero degree or 180 degrees. The main difference between these two smiles is the moving directions of the lower and upper lips. Female_A's upper lip moves vertically while this happens on Female_B's lower lip.
- Male_B's small smile is an exception of these four people since both his upper and lower lips move mostly in vertical direction. Apart from that, no pure red can be seen in the figure which means most of his motion vectors move along the vertical direction.

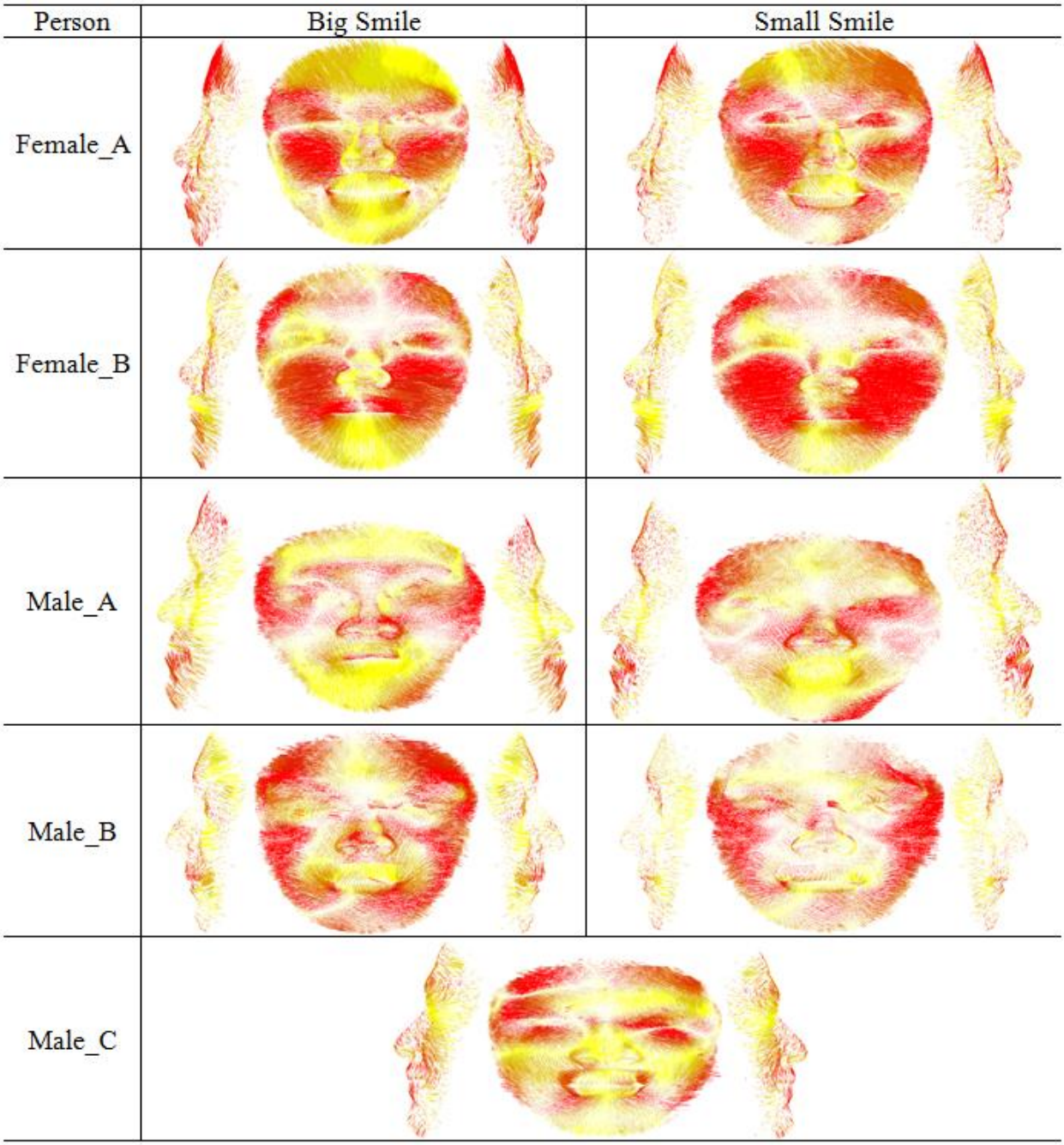


Figure 4.7: Orientation of big and small smile expressions ((each face has one front and two side views).

4.3.2 The Analysis of the Absolute Length of Motion Vectors

The analysis of the length of motion vectors are done in three ways, similar to what we have already described at the beginning of this section. First, we visualize the absolute length of the motion vector using equation 4.1 (b).

Assumedly all the motion vectors are unity normalized from 0 to 1, for a given MV (x, y):

- If the absolute length of this MV is around 0, the color is almost yellow;
- If the absolute length of this MV is close to 1, the color is nearly pure red.

Figure 4.8 shows the visualization of absolute length for the models in the database. Compared to the general rules of orientation, the rules of absolute length of big smile expression are distinctive for each person.

- For big smile, Male_A's lower lip moves the most. The part under the lower lip moves fairly long but no longer than the displacements of his lower lip. Apart from these two regions, the rest of the facial motion vectors hardly moves in absolute length.
- For Female_B's big smile, we can easily tell that most of her face is moving in a large step. Her cheeks and mouth and chin are almost pure red. Consequently, when she laughs or feels happy, the movements of her face are big.
- For Female_A's big smile, the cheeks and mouth move a lot while her other parts of the face merely move.
- For the trending movements of Male_B's and Male_C's big smiles, there is no specific region that shows which part moves the most. Almost the whole surface is moving at an average level since the visualization is full of orange.

- However, when they perform small smile, Female_A, Female_B and Male_A smile similarly. All their cheeks move a lot considering the whole region of the face. Male_B does not perform a very passionate small smile as the whole face looks still due to the pure yellow color.

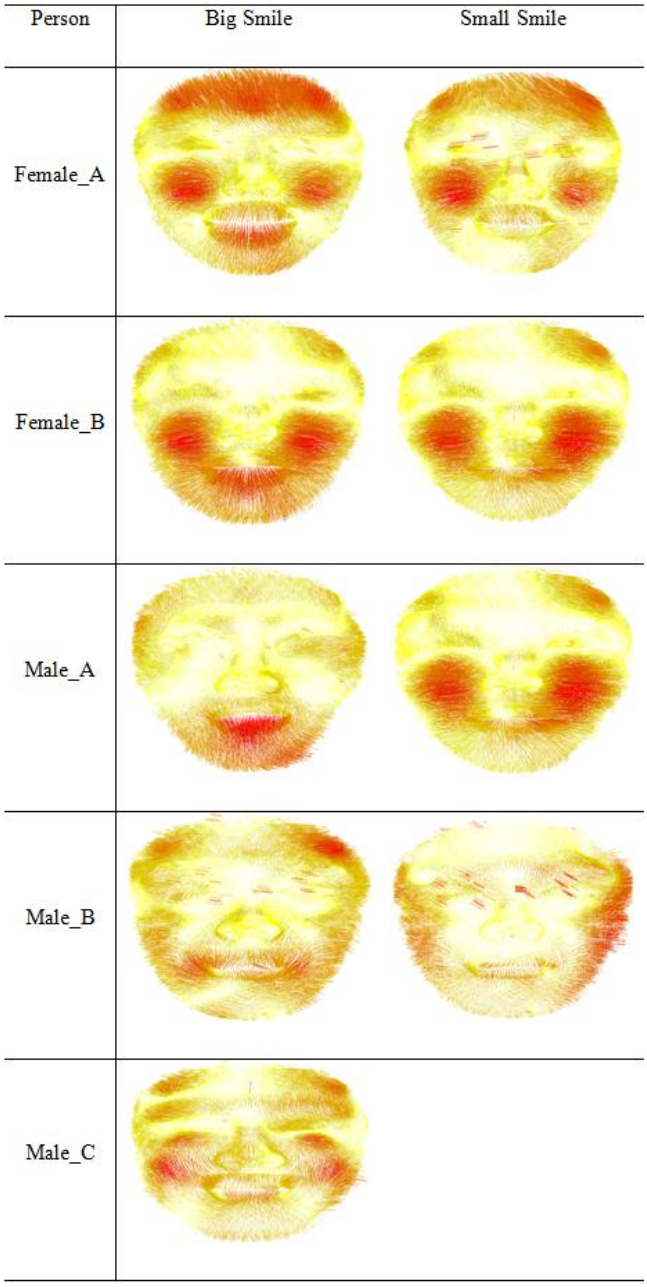


Figure 4.8: Absolute length of big and small smile expressions.

4.3.3 The Analysis of the Vertical and Horizontal Length of Motion Vectors

The other two branches of analyzing the length of motion vectors are vertical inspection and horizontal inspection. The equation of computing vertical and horizontal length can be seen in equation 4.1 (c) and (d). The strategy of visualization for vertical and horizontal lengths is the same for absolute length. Assumedly all the motion vectors are unity normalized vertically (or horizontally) from 0 to 1, for a given MV (x, y):

- If the vertical (or horizontal) length of this MV is around 0, the color is almost yellow;
- If the vertical (or horizontal) length of this MV is close to 1, the color is nearly pure red.

Figure 4.9 is the visualization for vertical length of big and small smile expressions. The conclusions are obvious:

- For big smile, Asian style Female_A, Female_B and Male_A have very much similar movement. All of their lower lips shift absolutely vertically while other parts are hardly move in vertical direction. For Male_C, his cheeks shift the most in vertical direction.
- Caucasian style Male_B accents different ways of big smiling. For Male_B, the part that shifts a lot is the region around the upper lip.

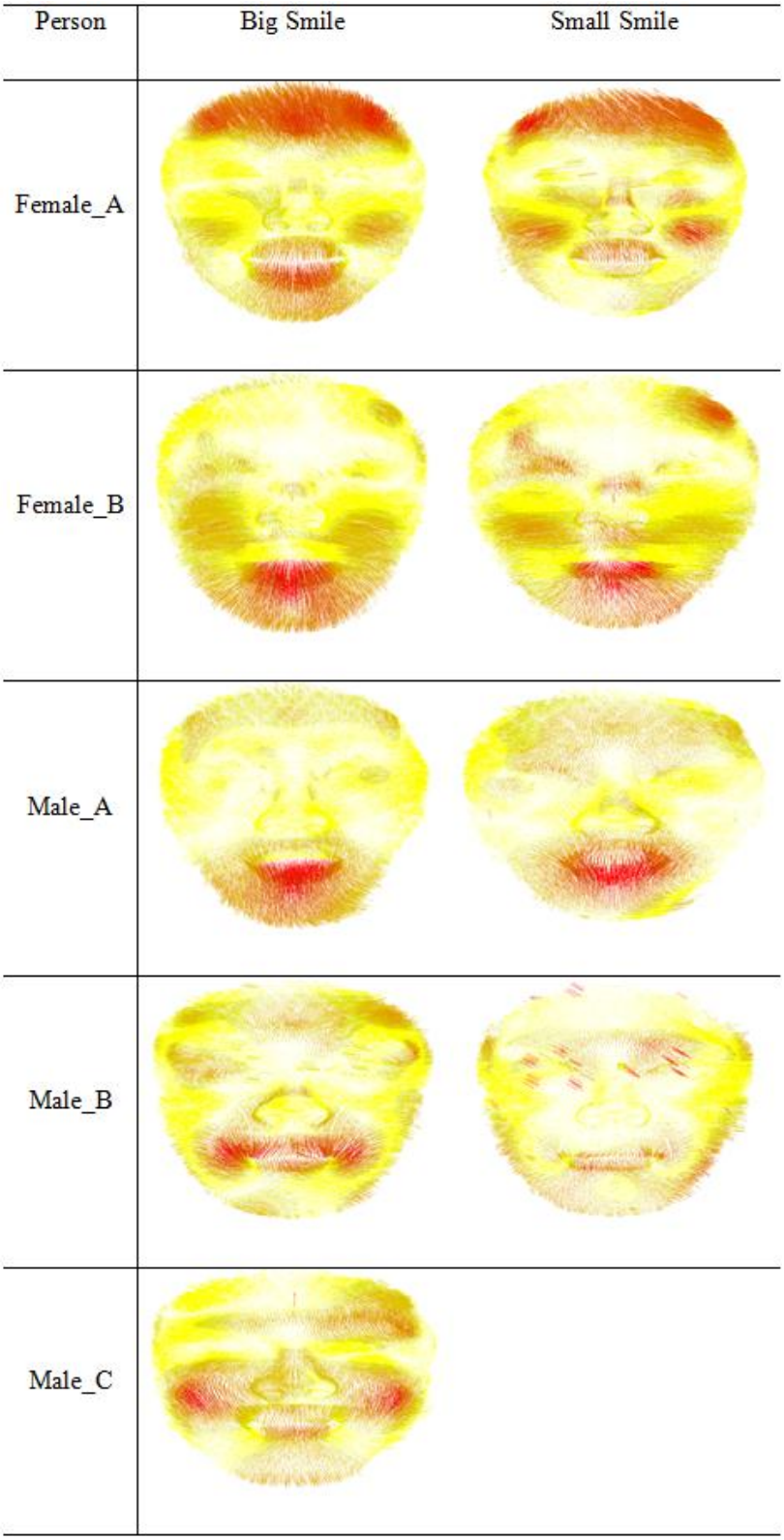


Figure 4.9: Vertical length of big and small smile expressions.

Figure 4.10 is the visualization for horizontal length of big and small smile expressions. The results follow:

- For big smile, the lengths of Female_A's cheeks movements are close to 1 in horizontal direction. Female_B's cheeks and the regions under the cheeks are moving at their most. The middle parts of these two females are totally yellow that means those regions hardly move horizontally.
- For big smile, the central region of Male_B's face keeps still. The region along the face moves at the mid-level as the color is almost orange (between pure red and pure yellow). Male_C's cheeks shift a bit while the rest remain still.
- For big smile, Male_A's face does not change a lot in horizontal direction.
- For small smile, both Female_A's and Female_B's cheeks smile horizontally a lot. Female_A's right bottom region moves horizontally at the mid-level. Female_B's two cheeks are the only regions that move when performing small smile while the rest of her face keeps still.
- For small smile, Male_B's face and Male_A's face shift very much the same as their big smile movements. Both of them are not moving in the horizontal direction.

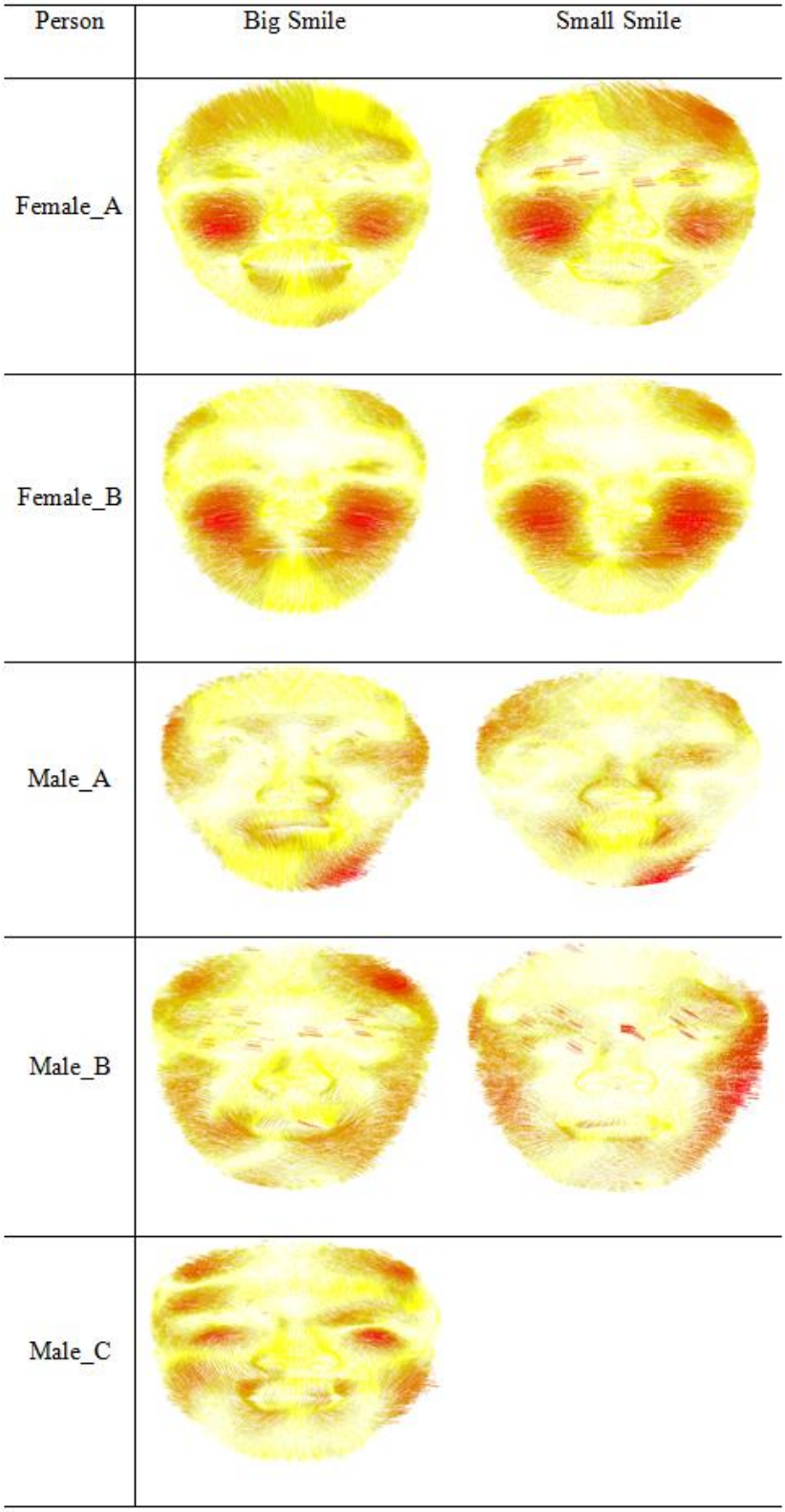


Figure 4.10: Horizontal length of big and small smile expressions.

4.4 Analysis of Facial Muscles

Each facial expression connects with different muscle nerves. Only the frontalis muscle independently moves when performing expressions. In this thesis, we do not take frontalis into consideration. According to the descriptions of the facial muscles involved in each of the emotions Darwin considered universally, we draw out two from the list: joy (big smile) and happiness (small smile). According to the FACS, the muscles flex when expressing small or big smile are: zygomatic and orbicularis. In our work, we mainly inspect four kinds of facial muscles: zygomatic, orbicularis oris, depressor, and masseter. These four kinds of facial muscles are mainly credit to the movements of different smiles. The strategy of analyzing the facial muscles is the same for the whole face: orientation and absolute/vertical/ horizontal length observations.

4.4.1 The Analysis of the Orbicularis Oris Facial Muscle

The first muscle we are going to observe is orbicularis oris. The orbicularis oris muscle lies in the lips encircling the mouth. Figure 4.11[5] visualizes the muscle on the face. Orbicularis oris is also very important for analyzing the facial expression since all the smiles need the mouth to move.

[5] https://www.studyblue.com/notes/note/n/facial-muscles/deck/8122103

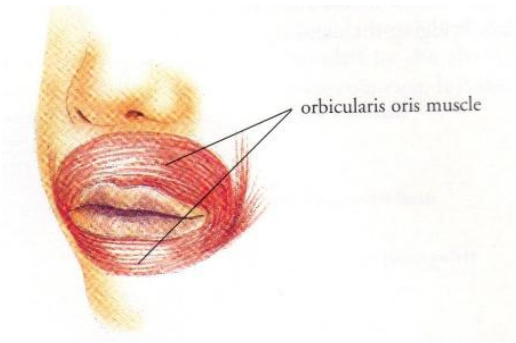


Figure 4.11 : The orbicularis oris facial muscle.

Figure 4.12 is the visualization of the big and small smile expressions of the orientation of orbicularis oris muscle.

- When performing big smile, five people illuminate different manners.
    - The rotation of Female_B, Male_B and Male_C is highly symmetrical. We can draw a perpendicular line along the middle line of the muscle. For Female_B, the upper part of the muscle rotates mainly along the x axis. Male_C's upper and lower lips move straightly up or down. His mouth corners are stretching along the horizontal line. However, Male_B's upper part rotates straightly vertically. Female_B's lower part rotates quietly vertically. Both Male_C's upper and lower parts have rotation in vertical direction.
    - The color of Female_A's muscle and Male_A's muscle is almost totally yellow. This hints that the orbicularis oris muscle for Female_A and Male_A rotates a lot along the y axis.
- When performing small smile, the rules of all the models in the database is similar to the rules of performing big smile. But there are big differences in Female_A's smile and Male_B's smile. For Female_A, the upper lip moves straightly up and her lower

lip moves around the horizontal line. For Male_B, part of his lower lip and his right mouth corner rotate along the y axis while the rest is moving along the x axis.

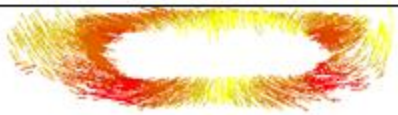
Person	Big Smile	Small Smile
Female_A		
Female_B		
Male_A		
Male_B		
Male_C		N/A

Figure 4.12 : the orientation for the big and small smile expressions of orbicularis oris muscle.

Figure 4.13 and Figure 4.14 are the length visualization for the big and small smile expressions of orbicularis oris muscle.

- When performing big smile:
    - Female_A's lower lip moves the longest while her mouth corners are moving very little. Her motion vectors mainly move in vertical direction.
    - Female_B's whole mouth moves at a long distance. Her upper lip and mouth corners shift horizontally. Most of her lower lip is moving strictly downward.

- Male_A's absolute length and vertical length visualization are the same. His lower lip moves vertically.
- Male_B's whole mouth shifts at a large pace. His upper lip moves a lot in vertical direction while his lower lip moves in horizontal direction.
- Male_C's upper and lower lips are both moving in vertical and horizontal directions. His average color is orange which his motion vectors are moving at the mid-level among the database.

- The conclusions of the muscle movements of the small smile are similar to the conclusions of the big smile. There are still a few differences:
  - Male_B and Female_A are moving the least while Male_A moves the most among the database.
  - Female_B's lower and upper lips are moving vertically while her mouth corners are moving horizontally.
  - Male_A's mouth corners are moving both in vertical and horizontal directions while his lower and upper lips are moving up and down.
  - Male_B's and Female_A's mouth muscles are shifting at a short distance.

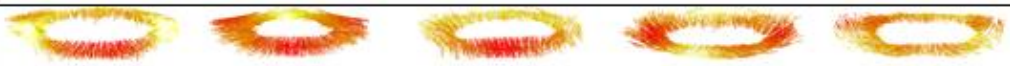
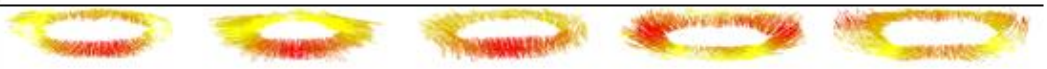
Big Smile	Female_A	Female_B	Male_A	Male_B	Li
Absolute					
Vertical					
Horizontal					

Figure 4.13: The length visualization of big smile of the orbicularis oris muscle.

Small Smile	Female_A	Female_B	Male_A	Male_B
Absolute				
Vertical				
Horizontal				

Figure 4.14: The length visualization of small smile of the orbicularis oris muscle.

4.4.2 The Analysis of the Zygomaticus Facial Muscle

The Zygomaticus facial muscle consists of two parts: zygomaticus major muscle and zygomaticus minor muscle. The whole zygomaticus muscle is the facial muscle of anterior cheek extending to upper lip. When expressing smile, zygomaticus major muscle draws the mouth superiorly and posteriorly. The zygomaticus minor muscle extends from the cheek bone and tracks along the lateral face and then deeps into the outer part of the upper lip. Figure 4.15[6] shows where the whole zygomaticus muscle is located on the face.

[6] http://quizlet.com/24446743/ch9-gross-anatomy-of-skeletal-muscles-flash-cards/

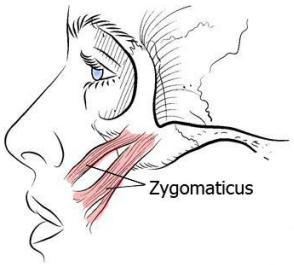


Figure 4.15 : The zygomaticus facial muscle.

Figure 4.16 is the visualization for the big and small smile expressions of the orientation of orbicularis oris muscle.

- When performing big or small smile, the five people appear differently.
    - Female_B's and Female_A's cheeks are almost red which indicates that all the muscles are moving horizontally. However, in Male_C's condition, the angle appears to be around 90 or 270 degrees as the overall color is totally yellow. Male_C's muscle tends to move straightly upward or downward. For Male_B and Male_A, the upper half muscles are moving along the x axis while the lower half muscles are dropping down.
    - Based on the above-mentioned conclusions, it is obvious that there are no firm generalizations of the expression of smiling. The smile-related muscles can move vertically, horizontally, or half vertically and half horizontally.

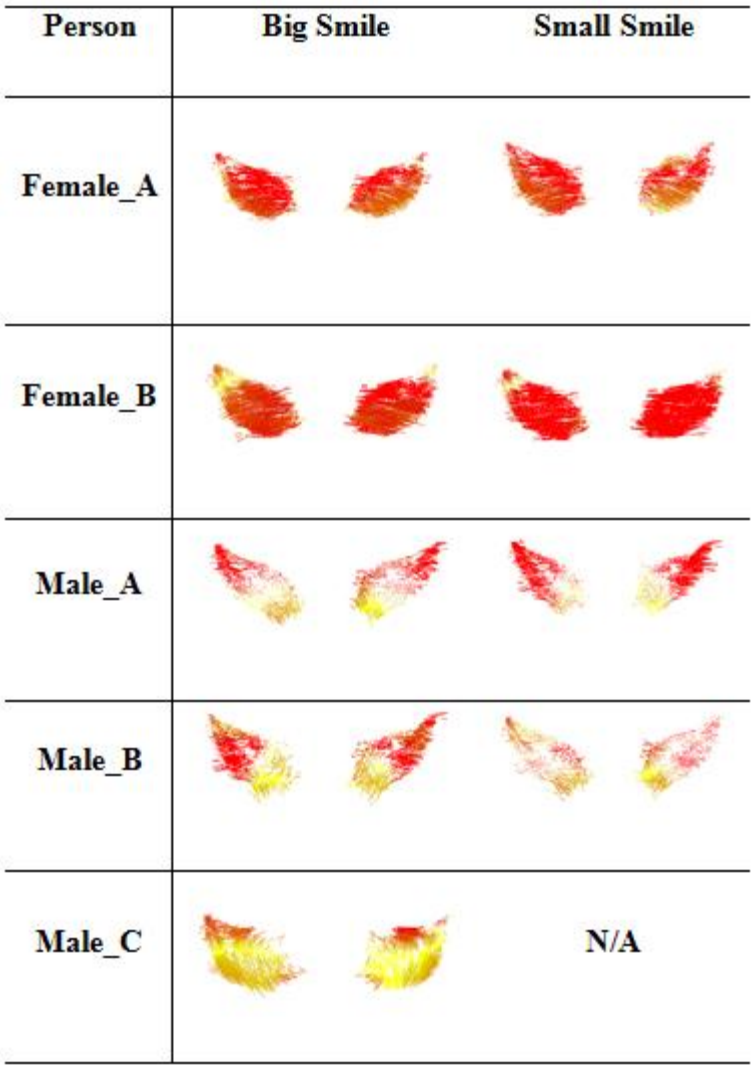


Figure 4.16 : The orientation visualization of big and small smiles of the zygomaticus muscle.

Figure 4.17 and Figure 4.18 are the length visualization for the big and small smile expressions of the orientation of orbicularis oris muscle.

- When performing big or small smile:
    - For absolute length, Male_A has the lightest color. The color of Female_B's and Female_A's is almost red. The color of Male_C and Male_B is orange, but the Male_C's is darker than Male_B's. Interpret the color information related

muscle: Male_A's muscle does not move a lot; Female_A's and Female_B's muscles are moving the most among the five people; Male_C's movement is at the mid-level pace; Male_B's muscle shifts longer than Male_A's.

- For vertical length, the four people except Male_C are totally yellow. Only Male_C's muscle has movement in vertical direction.
- For horizontal length, Female_B wins the most motion vectors in the horizontal direction. Female_A also moves a lot horizontally. Male_A again has little pace in horizontal direction. Male_C tends to move straightly up and down. Male_B's outer muscle moves longer than the inside muscle.
- For Male_A, when he smiles big, the (absolute, vertical and horizontal) lengths of his zygomaticus muscle are the least among the five people database.

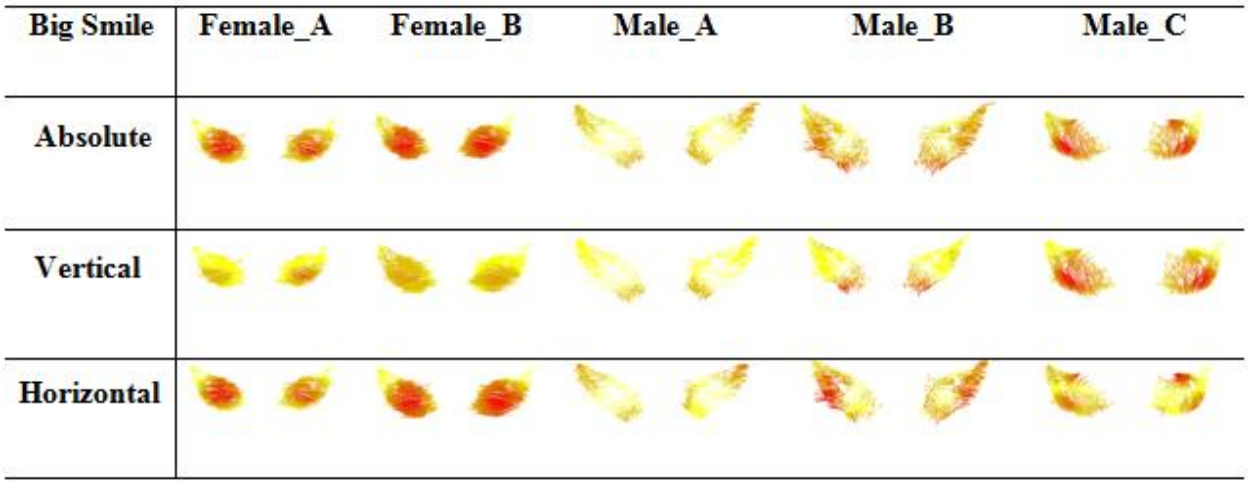


Figure 4.17 : The length visualization of big smile of the orbicularis oris muscle.

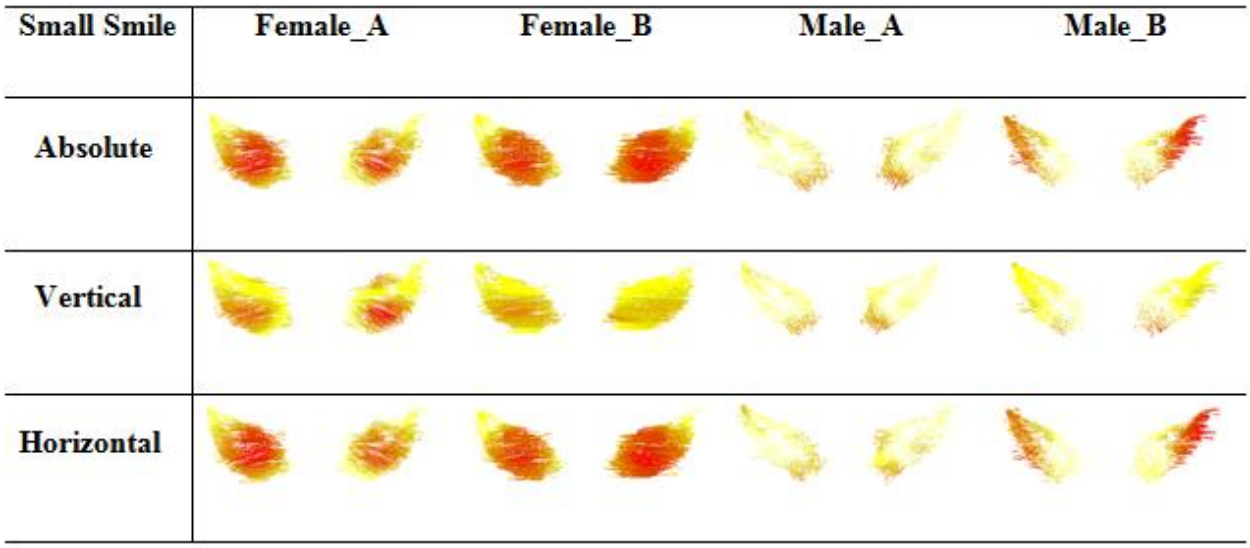


Figure 4.18 : The visualization of small smile of the zygomaticus muscle in orientation.

4.4.3 The Analysis of the Masseter Facial Muscle

The masseter muscle plays a major role in food chewing. The shape of masseter looks very much like a parallelogram. It connects to the mandible and the cheekbone. The masseter is the main muscle which pulls the mandible upward. Figure 4.19[7] visualises the place of the masseter muscle on the face.

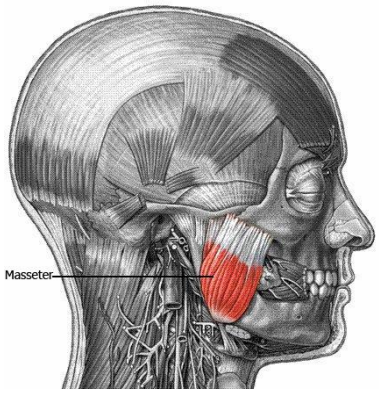


Figure 4.19 : The masseter facial muscle.

[7] https://www.studyblue.com/notes/note/n/muscle-origin-insertion-action/deck/6114393

Figure 4.20 is the visualization for the big and small smile expressions of the orientation of masseter muscle.

- When performing big smile, there are three types of moving,
    - The motion vectors of Male_B's master muscle are clinging to the x axis.
    - Male_A's and Male_C's movements which are similar to their right parts are moving horizontally and their left parts are inclined to shift between vertically and horizontally.
    - Female_A's and Female_B's movements are symmetrical. For Female_B, the whole movement is at a mid-level pace. For Female_A, the middle regions of her left and right masseter muscles rotate along the y axis while the rest are moving along the x axis.
- When performing small smile,
    - Male_B's majorly shifts along the x axis. The movements of Female_B are close to the x axis.
    - Female_A's rotation is between the x axis and y axis.
    - Male_A's tendency of orientation is close to the y axis.

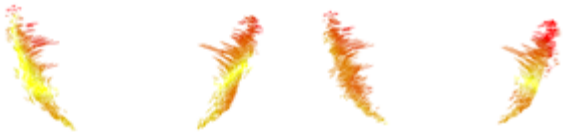
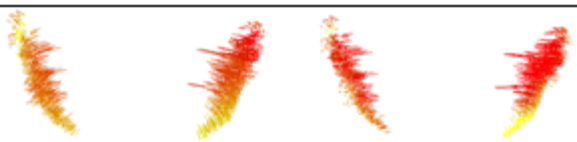
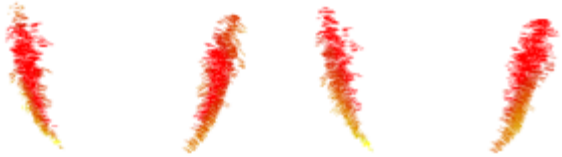
Person	Big Smile	Small Smile
Female_A		
Female_B		
Male_A		
Male_B		
Male_C		N/A

Figure 4.20 : The orientation visualization of big smile of the masseter muscle.

Figure 4.21 and Figure 4.22 are the length visualization for the big and small smile expressions of zygomaticus muscle.

- When performing big smile:
    - For Male_B, the whole movement is at a large pace and the horizontal movement is the major direction.

    - For Female_A, her muscle does not move very much. The upper inside of the masseter muscle moves longer than the rest of her muscle.
    - For Female_B, the movement is at the mid-level pace. The whole muscle is moving almost in the same distance.
    - For Male_C, the movement of his muscle is similar to that of Female_B. But his muscle moves mainly horizontally.
    - Male_A's lower muscle shifts longer than the inside muscle. The muscle close to his mouth moves a lot.
- When performing small smile:
    - Male_A hardly moves his lateral cheek.
    - For Female_B and Female_A, their muscles are moving at the mid-level both in the horizontal and vertical directions.
    - Male_C's major moves in the horizontal direction and his lateral cheek extends from inside to outside when he smiles.

Big Smile	Female_A	Female_B	Male_A	Male_B	Male_C
Absolute					
Vertical					
Horizontal					

Figure 4.21 : The length visualization of big smile of the masseter muscle.

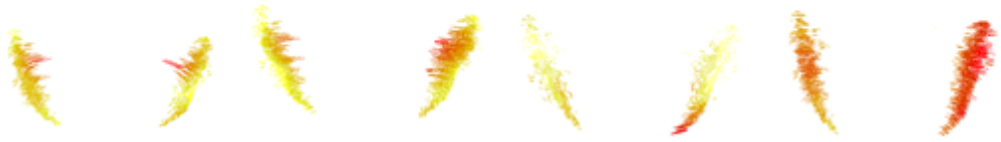
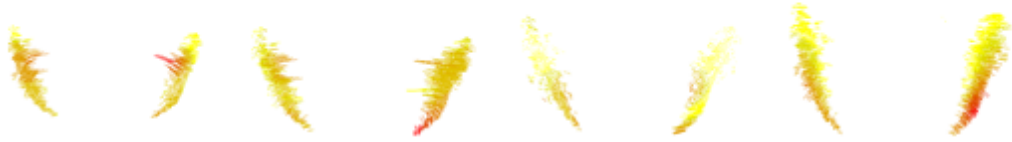
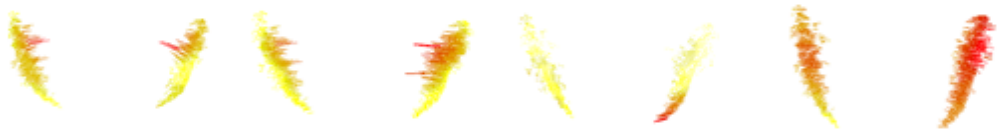
Small Smile	Female_A	Female_B	Male_A	Male_B
Absolute				
Vertical				
Horizontal				

Figure 4.22 : The length visualization of small smile of the masseter muscle.

4.4.4 The Analysis of the Depressor Anguli Oris Facial Muscle

The depressor anguli oris muscle is attached to the edge of the lips. When the mouth is moving, it pulls downs or pushes up the movements of jowls. The depressor anguli oris muscle consists of two muscles, one on each side of the mouth. Figure 4.23[8] displays the place of the depressor anguli oris muscle on the face.

Figure 4.23 : The depressor anguli oris facial muscle.

Figure 4.24 is the orientation visualization for the big and small smile expressions of depressor anguli oris facial muscle.

- When performing big smile, there are three types of moving.
    - The orientation of Male_B and Male_C are symmetrical. Male_B's two depressor anguli oris muscles are expending externally and the orientation sticks to the x axis. Male_C's upper part muscles rotate along x axis while his lower muscles rotate around the y axis.

[8] http://www.healthline.com/human-body-maps/depressor-anguli-oris-muscle

- Female_B's right orientation is generally smaller than her left muscle. Both her two lower depressor anguli oris muscles orientate bigger than the upper parts.
- Female_A's mandible rotates almost. But her right muscle's orientation is smaller than her right orientation.
- Male_A two sides rotate differently. His right side is around the x axis while his left side is around the y axis.

- When performing small smile, all the four people's orientation clings to the x axis. Only Male_A's left side muscle rotates along the y axis. The orientation of Male_B is just opposite of the orientation of Female_B. While Female_B's left side closely sticks to the x axis while this happens on Male_B's right side. The orientation of Female_A's two side muscles is symmetrical.

Person	Big Smile	Small Smile
Female_A		
Female_B		
Male_A		
Male_B		
Male_C		N/A

Figure 4.24 : The orientation visualization of big smile of the depressor anguli oris muscle.

Figure 4.25 and Figure 4.26are the length visualization for the big and small smile expressions of depressor anguli oris facial muscle.

- When performing big smile:
    - Male_C's muscle almost has no movement. The pure yellow visualization shows his depressor anguli oris does not shift. For Male_B, the visualization of his absolute length and the horizontal length is the same. Male_B's major moves horizontally. Male_A's left muscle moves vertically while his right side shifts half horizontally and half vertically. The movements of Female_B's

smile drop up and down vertically. Male_B's upper and lower parts of the muscles are moving at a large distance while his middle parts move the least.

- When performing small smile:
    - All of the four peoples' muscles move very small. For Female_A, her muscles do not move at all in any length branch. Female_B's right bottom is the only part which moves at a big pace. Male_A's left side moves vertically and his right side moves horizontally. Male_B's muscles shift both vertically and horizontally.

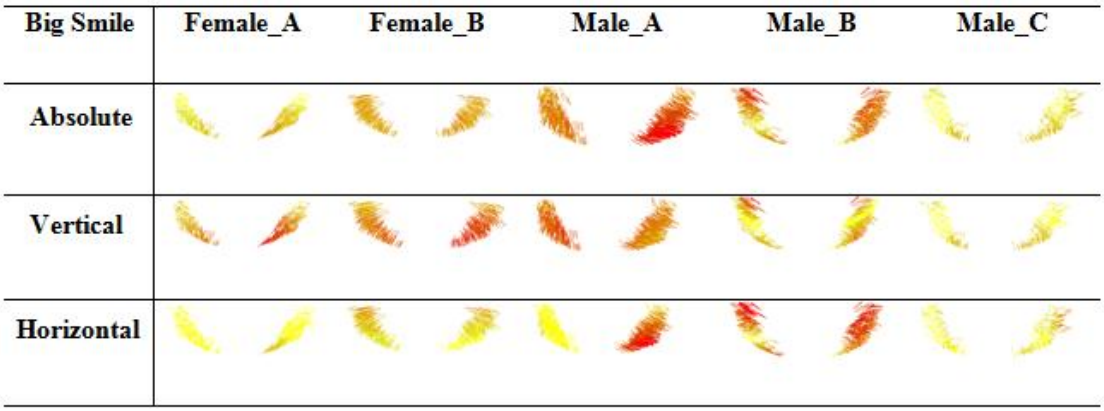


Figure 4.25 : The length visualization of big smile of the depressor anguli oris muscle.

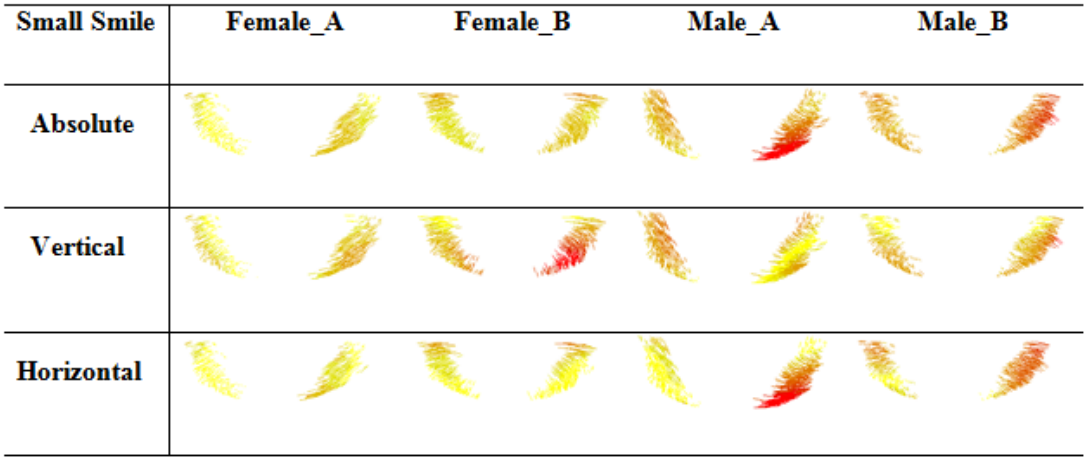


Figure 4.26 : The length visualization of small smile of the depressor anguli oris muscle.

Chapter 5.Conclusion

Face is the essential organ of sensing and feeling. Almost all the emotions are made by the muscles underneath the face surface. Our aim is to analyze how differently people are smiling and to find out a way to generate more natural and various expressions in 3D facial animation. Not only we use real human 3D face data as inputs in our face reconstruction system, but also proposing a new way of face movement visualization and facial muscle movement visualization.

In this thesis, we reconstruct 3D scanned faces of live people using our face reconstruction system. Then we extract different people's expression motion vectors so that we could do expression cloning on other persons' faces. At last, we visualize and analyze the extracted motion vectors of the whole face surface as well as the facial muscle units.

We point out our main contributions and outlook the future research.

5.1 Contribution

Based on the work we have done, our main contributions can be summarized as follows:

- A framework to analyze facial expressions starting from scanning real-life people in 3D and recreating face models to comparing normalized expression vectors among different subjects.
- Analysis of human smiles through the whole faces using 3D motion vectors.

- Our way of visualizing the smile expression is by parsing expression motion vectors mathematically. We inspect the orientation and the absolute/vertical/horizontal lengths of the expression motion vectors.

- Analysis of the related muscles of smiling using 3D motion vectors.
  - The main facial muscles related to smiling are orbicularis oris, masseter, depressor and zygomaticus.
  - When people express the big smile, the orbicularis oris and zygomaticus muscles move the most both in orientation and length. When people perform small smile, orbicularis oris muscle again shifts the most since it circles the mouth.

We also make improvements for system realization as follows:

- We made enhancement for faster and more stable reconstruction process for scanned 3D faces.
  - We apply tree data structure to dissolve the long time processing. When performing dense surface points searching, the time can last for hours if only using iterative searching method. When applying k-d tree method, the time consumption is shortened in thirty minutes.
  - The use of k-d tree data structure makes fewer defects on the recreated face than the use of overall searching method so that the following complex refinement steps are briefed only to one Single Laplacian Smoothing.

- Introducing a stable and fast way of transferring facial expressions among individual subjects.
    - As our 3D face models are restructured to share the same geometrical structure, the only mathematical work in extracting facial expressions is to do subtraction between expressional faces and neutral faces.
    - After then, facial expression cloning is finished by placing facial expression motion vectors onto another person's neutral face.

5.2 Future Research

Even though we have proposed a novel way of analyzing facial expressions by using real 3D human face data, which is captured by the 3D laser scanner, and we analyze two kinds of smile, there is still space where we could deepen our research and develop our existing methodology. We may take the following considerations into account for our future research.

- To improve the visual effect of the head model, we can add texture and color information on the obtained face model.
- Our feature points are located manually, and in future work, we could think of a way of collecting them automatically.
- We would like to expand our database having more sample faces.
- With bigger database, analysis-based classification such as gender/age/ethnic group/characters/facial shapes for whole and individual muscles can be done.
- Facial markers glued on a face can be used to locate individual muscle more precisely.

- Individual facial muscles can be recognized by analyzing motion vectors. In this thesis, the muscles are identified manually.
- If we have enough facial expression samples, we could categorize which expression or expressions belong to what kind of emotion. Plus, we could expand this thesis not only in computer science field but also in psychology field.

References

[1] Volker Blanz and Thomas Vetter, "A morphable model for the synthesis of 3D face," in *Computer Graphics Proceedings SIGGRAPH*, 1999, pp. 187-194.

[2] Won-Sook Lee and Nadia Magnenat-Thalmann, "Generating a population of animated faces from pictures," in *Proc. IEEE International Workshop on Modelling People (ICCV'99 Workshop mPeople)*, 1999, pp. 62-69.

[3] Jongmoo Choi, Gerard Medioni and Yuping Lin, "3D face reconstruction using a single or multiple views," in *ICPR*, 2010, pp. 3959-3962.

[4] Unsang Park and Anil K. Jain, "3D face reconstruction from stereo video," in *The 3rd Canadian Conference on Computer and Robot Vision*, 2006, pp. 41-48.

[5] Haoda Huang, Jinxiang Chai, Xin Tong and Hsiang-Tao Wu, "Leveraging motion capture and 3D scanning for high-fidelity facial performance acquisition," in *The Visual Computer: International Journal of Computer Graphics*, 2014, pp. 649-659.

[6] Pablo Garrido, Levi Valgaerts, Chenglei Wu and Christian Theobalt, "Reconstructing detailed dynamic face geometry from monocular video," in *Proceedings of the 11th European Conference on Visual Media Production*, 2014, pp. 1-10.

[7] Won-Sook Lee, Andrew Soon and Lijia Zhu, "3D facial model exaggeration builder for small or large sized model manufacturing," in *Virtual Reality*, 2007, pp. 229-239.

[8] Thabo Beeler, Bernd Bickel, Paul Beardsley, Bob Sumner and Markus Gross, "High-quality single-shot capture of facial geometry," in *ACM Transactions on Graphics*, 2010.

[9] J. Rafael Tena, Fernando De la Torre and Iain Matthews, "Interactive region-based linear 3D face models," in *ACM Special Interest Group on Computer Graphics and Interactive Techniques*, 2011.

[10] Amit H. Bermano, Derek Bradley, Thabo Beeler, Fabio Zund, Derek Nowrouzezahrai, Ilya Baran, Olga Sorkine, Hanspeter Pfister, Robert W. Summer, Bernd Bickerl and Markus Gross, "Facial performance enhancement using dynamic shape space," in *ACM Transactions on Graphics (TOG)*, 2014, Vol. 33.

[11] Jun-yong Noh and Ulrich Neumann, "Expression Cloning," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 2001, pp. 277-288.

[12] Hyewon Pyun, Yejin Kim, Wonseok Chae, Hyung Woo Kang and Sung Yong Shin, "An example-based approach for facial expression cloning," in *Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer Animation*, 2003, pp. 167-176.

[13] Kyunggun Na and Moonryul Jung, "Hierarchical retargeting of fine facial motions," in *the Eurographics Association and Blackwell Publishingr*, 2003.

[14] Jaewon Song and Byungkuk Choi, "Characteristic facial retargeting," in *Computer Animation and Virtual Worlds*, 2011, pp. 187-194.

[15] Ludovic Ductreve, Alexandre Meyer and Saida Bouakaz, "Feature points based facial animation retargeting," in *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, 2008, pp. 197-200.

[16] Yang Wang, Xiaolei Huang, Chan-Su Lee, Song Zhang, Zhiguo Li, Dimitris Samaras, Dimitris Metaxas, Ahmed Elgammal and Peisen Huang, "High resolution acquisition, learning and transfer of dynamic 3-D facial expressions," in *Computer Graphics Forum*, 2004, pp. 677-686.

[17] Lijia Zhu and Won-Sook Lee, "Facial expression via genetic algorithms," in *Proceedings of the 19th Annual Conference on Computer Animation and Social Agents*, 2006.

[18] Lijia Zhu and Won-Sook Lee, "Facial animation with motion capture based on surface blending," in *Proceedings of the 2nd International Conference on Computer Graphics Theory and Applications*, pp. 63-70, 2007.

[19] Takeo Kanade, Jeffrey F. Cohn and Yingli Tian, "Comprehensive database for facial expression analysis," in *Proceedings of the 4th IEEE international conference on automatic face and gesture recognition*, 2000, pp. 46-53.

[20] Ragini Verma, Christos Davatzikos, James Loughead, Tim Indersmitten, Ranliang Hu, Christian Kohler, Raquel E. Gur and Ruben C. Gur, "Quantification of facial expressions using high-dimensional shape transformation," in *Journal of Neuroscience Methods*, 2005.

[21] Frederick I. Parke, "Computer generated animation of faces," in *72 Proceedings of the ACM annual conference*, 1972, pp. 451-457.

[22] VICON motion capture system. http://www.vicon.com/

[23] P. Ekman and D. Keltner. "Universal facial expressions of emotion: an old controvery and new findings," in *Nonverbal communication: where nature meets culture*, 1997, pp. 27-46.

[24] L. Zhang, N. Snavely, B. Curless and S.M. Seitz. "Spacetime Faces: high-resolution capture for modeling and animation," in *ACM SIGGRAPH Proceedings*, 2004.

[25] http://grail.cs.washington.edu/software-data/stfaces/index.html.

[26] P. Ekman and W. Friesen, "Facial action coding system: a technique for the measurement of facial movement," Consulting Psychologists Press, 1978.

[27] A.J. Zlochower, "Deciphering emotion from the face: an evaluation of FACS, EMG, and copmuter-vision based approaches to facial expression analysis," unpublished manuscript, 1997.

[28] Kyung Won Chung and Harold M. Chung, "Gross Anatomy (Board Review)," Lippincott Williams & Wilkins, 2005, p. 364.

[29] Niloofar Aghayan, "Reconstruction of Complete Head Models with Consistent Parameterization," 2014.

[30] http://cyberware.com/.

[31] Micheal Kazhdan, Matthew Bolitho and Hugues Hoppe, "Poisson surface reconstruction," in *Eurographics Symposium on Geometry Processing*, 2006.

[32] T. Smith, R. S. Rana, P. Missiaen, K. D. Rose, A. Sahni, H. Singh, et al., "High bat (Chiroptera) diversity in the Early Eocene of India," in *Naturwissenschaften*, pp. 1003-1009, 2007.

[33] A.-N. Ansari and M. Abdel-Mottaleb, "3D face modeling using two views and a generic face model with application to 3D face recognition," in *Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance*, pp. 37-44, 2003.

[34] S. Berretti, A. Del Bimbo, and P. Pala, "3D Face Reconstruction from Two Orthogonal Images for Face Recognition Applications," in *International Journal of Digital Library Systems (IJDLS)*, pp. 42-58, 2010.

[35] D. Fidaleo, J.-y. Noh, T. Kim, R. Enciso, and U. Neumann, "Classification and volume morphing for performance-driven facial animation," in *International Workshop on Digital and Computational Video*, 2000.

[36] G. Barequet and S. Kumar, "Repairing CAD models," in *Visualization'97, Proceedings*, pp.363-370, 1990.

[37] M. J. Powell, "Radial basis functions for multivariable interpolation: a review," in *Algorithms for apporxiamtion*, pp. 143-167, 1987.

[38] Charles Teorell Loop Thesis, "Smooth subdivision surfaces based on triangles," Dept. of Mathematics, University of Utah, 1987.

[39] SIGGRAPH 2000 course notes, "Subdivision for modeling and animation," 2000.

[40] Lance Williams, "Performance-driven facial animation," in *SIGGRAPH '90 Proceedings of the 17th annual conference on Computer graphics and interactive techniques*, pp. 235-242, 1990.

[41] Jerome H. Friedman, Jon Louis Bentley and Raphael Ari Finkel, "An algorithm for finding best matches in loarithmic expected time," in *ACM Transactions on Mathematical Software (TOMS)*, pp. 209-226, 1977.

[42] Andrew W. Moore, "An introduction tutorial on k-d trees," computer laboratory, University of Cambridge, 1991.

[43] Gabriel Taubin, "A signal processing approach to fair surface design," in S*IGGRAPH '95 Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pp. 351-358, 1995.

[44] Gabriel Taubin, "Geometric signal processing on polygonal meshes," in the *Eurographics*, 2000.

[45] O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Rossl and H. -P. Seidel, "Laplacian surface editing," in the *Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pp. 175-184, 2004.

[46] Zhongping Ji, Ligang Liu and Guojin Wang, "A global laplacian smoothing approach with feature preservation," in the *Ninth International Conference on Computer Aided Design and Computer Graphics*, 2005.

[47] De Boor, C. "A practical guide to splines (revised version)," in *Springer*, pp. 207-214, 2001.

[48] Trevor Hastie and Robert Tibshirani, "Generalized Additive Models," in the book *Statistical Science*, pp. 297-310, 1986.

[49] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum and T. R. Evans, " Reconstruction and representation of 3D objects with radial basis functions," in the *ACM SIGGRAPH*, 2001.

[50] Dacher Keltner and Paul Ekman, "Intoduction: Expression of Emotion", in the *Handbook of Affective Sciences*, New York: Oxford University Press, pp. 411-414, 2003.

[51] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang and Matthew J. Rosato, "A 3D facial expression databse for facial behavior research", in the *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, 2006.

[52] http://mpeg.chiariglione.org/standards/mpeg-4