ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Design of Novel Algorithms for 3D Human Head Modeling and Expression Synthesis

Chen Chen

NANYANG TECHNOLOGICAL UNIVERSITY
03 JUL 2008
LIBRARY

School of Computer Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2008

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Acknowledgements

I would like to express my deepest and most sincere gratitude to my supervisor, Professor Edmond Cyril Prakash, Ph.D. His wide knowledge and his logical way of thinking have been of great value for me. His vision has led my improvement during the Ph.D research period. His understanding, encouraging and personal guidance have provided a good basis for the present thesis.

I would also like to gratefully acknowledge Professor Cai Wentong, Ph.D, Professor Shi Daming, Ph.D and Professor Foo Say Wei, Ph.D for their valuable advice and friendly help during my first year oral examination. Their extensive discussions around my work and interesting explorations in operations have been very helpful for this study.

I would warmly thank Dr Liu Qi, Mr Wang Yimin, Mr Evgeny Markin, Mr Liu Qiang, Mr Quah Chee Kwang, Mr Neo Ming Chiang Desmond, Dr Liu Wei Guo, Mr Gerrit Voss, Mr Yu Yu, Miss Zhang Juan Juan and many colleagues I have great regard. I wish to extend my warmest thanks to all those who have helped me and made in-depth talk with my work in the School of Computer Engineering of Nanyang Technological University. They have provided a lot of great idea for the survey, design and implementation of my research, their immerse help and friendship have a remarkable influence to me.

I owe my loving thanks to my father Chen Yin Chun, my mother Zhang Feng

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Zhu. They keeps encouraging me during my research abroad. Without their support and understanding it would have been impossible for me to finish this work. My special gratitude is due to my brothers, my sisters and their families for their loving support.

The equipment support of CAMTech Lab and Game Lab in Nanyang Technological University is gratefully acknowledged.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Contents

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

List of Figures

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

List of Tables

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Abstract

The efficiency and robustness to synthesize personalized facial expression is a challenging part in research of computer facial animation. To overcome this challenge, two problems have to be addressed.

Facial Shape Modeling: An accurate face shape model is required to mimic the 3D geometry of real human face surface. Accurate modeling is a tedious work and we need to accelerate the modeling process.

Robust Physical Simulation: Efficient dynamic deformation on facial skin driven by face anatomical structure is a key solution to synthesize realistic facial expressions. The deformation modeling should support real-time performance with the ability to represent skin elastic properties.

In this thesis, we propose an adaptive system for personalized 3D human shape modeling and expression synthesis.

To generate personalized face shape from 3D range scan data, landmarks compliant to MPEG-4 Facial Definition Parameters (FDP) are initially labeled on both template model and the range scan data as priori knowledge. The deformation from the template model to the target head is through a multi-level training process. Both general Radial Basis Function (RBF) and Compactly Supported Radial Basis Function (CSRBF) are applied to ensure the fidelity of the global shape and

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

face features. Situations with defective scanned data are also discussed.

A physically-based facial expression synthesizer with a novel algorithm for designing parametric muscle model is presented. The facial expression synthesizer is partitioned into four general activities: construction of volumetric soft tissue layer, design of underlying volumetric facial musculature, ensemble of muscles to synthesize the facial expression and an efficient and stable numerical integration for dynamic skin deformation. Techniques that support these activities are presented and seamlessly integrated to validate the facial expression synthesizer. Adaptation of animation component is considered so that the adapted skin shape can be automatically transferred into an animated head.

In summary, the contribution of this thesis has produced novel techniques for efficient and robust synthesis of personalized facial animation.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 1

Introduction

With the development of CAD/CAM software, modeling of mechanical objects becomes easier than ever before. On the other hand, non-mechanical objects remain a problem both in shape modeling and motion control. Usually, the shape of a non-mechanical object is irregular and asymmetric. A typical shape modeling method for such objects is free form deformation, which requires tedious work from an expert artist or modeler. When non-affine transformation is applied, the control parameters should be reset for each model. As one of the representative object, we discuss animated human face modeling in this thesis.

Many research efforts have been focused on the achievement of realistic representation of human face since the pioneer work of Parke [94]. However, the irregular shape of the head, the complex facial anatomical structure and various facial tissue behavior make it still a formidable challenge in computer graphics. The challenge of modeling an animated human face can be described from several parts as follows:

Shape modeling: To create a human head model for a specific person is a tedious task because of the variety of facial features and appearance. The development of the 3D range scanner helps to capture the shape information

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

of complex object. But holes or gaps may appear due to the variant reflective properties on facial surface, overlapped or folded surfaces produced by merge procedure may result in visual artifacts. For human head, hairy surface can not be appropriately recognized by laser scanner. Lips are not separated and eyes are not recognized as an individual part of human head. These noisy and incomplete data increases the processing difficulty.

Expression synthesis: Differing from body animation, the human facial expression can not be synthesized by skeleton-skinning techniques. Point-based approach does not help a lot either because it does not provide enough controls for the skin dynamics. The facial expression is driven by muscle contractions. The muscle forces are applied on the interior layer of soft tissue. The soft tissue has a multi-layer structure, which consists of properties such as viscoelasticity, incompressibility and nonlinearity. These properties result in dynamic stress-strain behavior. The various thickness of skin tissue at different regions also increases the complexity of the dynamics. Because of the dynamic behavior, bulge and wrinkles can be noticed among expressions. Facial motion from one expression to another can not be simply described as linear interpolation between two expressions, either.

Photo-realistic rendering: To mimic human face in the real world under specific light condition, advanced rendering technique and shading model should be applied on the face model. These techniques include texture mapping, normal mapping, shadow casting and ray tracing, etc. The challenges here include high-resolution full head texture creation, an appropriate shading model to represent the reflection and refraction on the skin and high-efficiency model to represent skin bumps.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Real-time performance: Based on the aforementioned challenges, it can be understood that to create an accurate dynamic personalized human expression synthesizer with photo-realistic skin appearance, comprehensive computational model will be employed. However, getting an interactive updating rate is a practical requirement for applications such as virtual surgery training, video conference and human-computer interactive virtual avatar. This leaves us challenge to balance between the interactive performance and synthetic quality. Taking advantage of the modern hardware such as Graphic Processing Unit (GPU), Ray-tracing Processing Unit (RPU) and Physics Processing Unit (PPU) and merging into a practical system is also a research topic.

1.1 Applications

There is sufficient proof of the great vital force of computer facial animation and its extensive application. In the next decade, computer facial animation will be indispensable.

Entertainment: There are increasing interest of putting virtual characters in films and videos (Shrek Series, Final Fantasy VII: Advent Children 2005, etc.,). Even in movies acted by real human beings, computer-aided facial morphing is a highly demanded technique. Computer games always try to mimic a "real" world in purpose of player experience. In movie industry, high quality realistic human motion is normally the essential concern. Oppositely, in game industry real-time performance is the key factor to consider.

Virtual Surgery: Surgeons expect to get the realistic human brain model so they can make an anticipation to reduce the risk of surgery to a relatively lower point. An intern can get trained on the virtual surgery system before they

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

do the real operation. These kinds of systems always require an accurately rendered facial model and the anatomically correct tissue reaction. Volume rendering technique is normally used to display different structure in human head. Koch [67] designed a virtual surgery model using Finite Element Model (FEM). Maciel et al. [78] also modeled the biological behavior of soft tissue using FEM to assist orthopedic diagnosis and surgery planning.

Video Conference: Traditionally, in video conference system, we need to transfer the whole video data of the face. This requires a stable high-rate broadband. In a computer-synthesized animation system, the animation control parameters will be transferred so that a narrow-band video conference system can be expected. Eisert [30] has developed an advanced video conference system using MPEG-4 which encodes high quality head and shoulder motion sequences at a bit rate about 1-kbit/s. The low bandwidth requirements of the system makes it easier to integrate wireless devices such as PDAs, cellular phones and notebooks into a video-based communication system. Similarly, Fedorov et al. presented their system in [36].

Lip Reading: For the physically challenged, such as the deaf people, lip reading is important to communicate with normal people. Even for normal people, lip reading will enhance the comprehension of sentence in a noisy environment. For example, ViSiCAST [53] is a project that translates English text to the motion of virtual avatar and provides services for deaf citizens.

Education: Text-to-speech talking head is also suitable for kids and adults to learn how to speak a special word in a sentence. Furthermore, using motion capture mechanism a comparison between the trainee and the source can be achieved and comment will be made. Virtual Human Interface [51] is a

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

product designed by Digital Elite Inc. which delivers information to the end user by photo-realistic animated characters. It supports *What, Where, How* forms of communication paradigm and can be used for interactive education, communication and marketing.

1.2 Research Motivations

As described at the beginning, there are so many challenges existing in synthesizing animated human head. Among these challenges, we are trying to solve some of these problems. Our motivation comes from the following two directions.

Exact Shape & Exact Animation Adaptation: The increase in complexity of programmable facial animation systems has emphasized the need to develop high-level generic synthesis tools that allow more efficient exploration of the comprehensive face shape and animation space while adapting the effects for personalized face both in model and animation synthesis. Much of the previous work, however, neglects the reuse of model shape information as well as the animation information. In addition, a new face model may not be readily animated, because of two reasons. First, the model synthesis may not be complete; Second, the animation specification for any arbitrary head is not straightforward. In traditional approaches, this necessitates manual or expert intervention as well as painstakingly laborious process in face modeling and face animation. Our aim in this thesis is to present a new technique to perform face shape synthesis, targeting the rapid synthesis of animated faces. Further, the proposed technique should exploit multiple contexts to achieve efficient adaptation.

Accurate Muscle Behavior without Real-time Performance Degradation:

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The physically-based approach on facial expression synthesis has been improved a lot since the original work of Waters [117]. Both multi-layer [109] and single layered [59] mass-spring model has been proposed, finite element models [67, 108] can drive the volumetric model reconstructed from medical images. For the muscle representation, simplified parametric models [74, 57] and precise volumetric muscle [108] has been proposed, several approaches for muscle contraction evaluation [110, 33, 34, 102, 105] have also been discussed. However, the balance between the performance and reality has seldom been achieved. We are trying to realize the anatomical accuracy with plausible efficiency. The anatomical accuracy refers to the physical fact of the skin properties such as visco-elasticity and is subjected to muscle actuation. The efficiency refers to interactive rates for animation synthesis. So dynamic deformation model with proper complexity is desired. We also expect to provide more flexible control for the muscle contraction. Thus, our task is to design an easy to use tool set for the muscle definition, muscle registration and muscle activation setting.

1.3 Thesis Contributions

According to our research motivation, we have contributed in the following parts:

Compactly supported RBF for facial shape modeling: We propose a facial modeling system using our novel multi-level adaptation procedure. We model a template head model with physical structure to synthesize the various facial expression. To make the template model suitable for different facial geometry, landmarks compliant to MPEG-4 Facial Definition Parameters (FDP) are labeled on this model. Then another personalized 3D head with different

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

geometrical information is labeled with a subset of these feature points. An iterative multi-level training process is performed to get a transformation function from the template model to the personalized model. We use **Compactly Supported Radial Basis Function (CSRBF)** to assure the local details of the adapted model. A curvature-based searching scheme is also used to find the feature points in the iterative process.

Error metrics for shape adaptation: Based on the iterative adaptation procedure, we sequentially design the error metrics which can be directly applied for the similarity between two faces. Our contribution is to develop and demonstrate an error metric that can be directly applied for the similarity between two faces. Furthermore, we wanted a metric that can be applied to sub-regions of the face (eyes, nose, mouth and the ear) which has not been reported before in the literature. The adaptation results from defective scan data and sensitivity of FPs registration are discussed.

High performance physically based facial expression synthesis: We combined multiple elements including skull, muscle, skin into a physically-based expression synthesizer. We added area constraints and volumetric constraints to simulate the incompressible property of soft tissue. For the distance constraints, we add nonlinear coefficient to mimic the stress-strain curve. We use Rational Bezier curves to represent facial muscle fibres. This approach provides simple muscle definition methods but also considers the flexibility of muscle shape. To distinguish the voluntarily (active) and non-voluntarily (passive) muscles, we use anchor points to connect the control points of muscles. This form provides better shape variation possibility for non-voluntarily muscles such as orbicularis oculi and orbicularis oris, which has never been reported in any MSD approaches.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

1.4 Thesis Organization

The thesis proceeds as follows. Chapter 2 gives a brief literature review of techniques on facial shape modeling and facial expression synthesis. Chapter 3 surveys the anatomical structure of human face and the underlying biomechanics. Chapter 4 explains adaptation procedure we proposed for personalized head modeling. Chapter 5 describes the error metrics for the adaptation results. Chapter 6 presents the techniques for rapid animated head generation and real-time facial expression synthesis using physically-based approach. In Chapter 7 we present case studies of our adaptive face model. Chapter 8 concludes the thesis and lists the scope of future work.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 2

Literature Review

2.1 Introduction

In this chapter we review the significant techniques on human face synthesis. Research on computer generated human face has started for more than 30 years ago. We briefly describe the history in Section 2.2. The facial shape modeling techniques is investigated in Section 2.3, followed by review of expression synthesis in Section 2.4. The related work on photo-realistic appearance rendering is discussed in Section 2.5. Then we describe several work which is most close to our research in Section 2.6.

2.2 Timeline of Research on Face Shape and Animation

The beginning of research on synthetic human face can be traced back into the 1970s. Parke [94] approximated a human face with polygonal mesh, at that time a polygonal skin containing about 250 polygons defined by about 400 vertices was thought sufficient to achieve a realistic face.

Blanz and Vetter [9] in 1999 proposed a method to interpolate geometry and

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

texture data from database with 200 heads. All faces in the database were pre-processed by removing the back face, removing the neck and shoulder and a normalization to brought all the face in the same orientation and pose. Using this method they successfully rebuilt individual human head automatically from 2D inputs. The blend shape-based approach was later used for facial expression generation [100]. The verbal expressions and emotional expressions were blended together to create 78(6x13) key expressions in their approach. Similarly, Vlasic et al. [116] created multi-dimensional face model so that multiple factors such as expression, age and individual are decomposed individually and they can be blended together to create a large variety.

Another type of approach called template based approach is also very popular in recent years. A successful sample using radial basis function for template-scan data adaptation was proposed by Kähler et al. [59]. They initialized 60 landmarks on the skin and 22 landmarks on the skull. These landmarks were used to solve the transformation equation from prototype model to scanned range data. By recursively refining the landmarks they increased the similarity between the deformed prototype model and the scanned head. There are also several similar approaches [71, 124]. Zhang et al. [125] used a template model to adapt head from stereo images. Instead of increasing the landmarks, their approach chose an optimization framework which automatical computes the displacement of vertices in the template model by fitting the displaced vertices with their stereo depth maps. Their face model with 16K vertices could be tracked at speed of less than 1 minute per frame. Sifakis et al.'s work [105] created an adaptive volumetric structure for a personalized head from MRI with 30K surface triangles and 370k deformable tetrahedra. To rebuild a personalized head from their prototype takes days.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

In 1981, the work of Platt and Badler [98] presented an integrated system for simulation of face using muscle control, their work is the pioneering research on muscle modeling and structure of human face in computer graphics area. Waters [117] started the development of parameterized facial muscle process in 1987, the parameterized muscle is not dependant on the model topology and provides a more general approach to model the primary facial expression. Two types of muscles were created: linear muscles that pull and sphincter muscles that squeeze. Kalra et al. [61] simulated pseudo-muscles of the face using free form deformation (FFD). The face is arranged in three dimensional cubic lattices and deformed by manipulating control points in their approach. Terzopoulos and Waters [110] estimated muscle contractions from deformable contour in a 256x256x8 image sequences. The estimation result was used to drive their tri-layered soft tissue model with 960 elements with approximately 6500 springs at interactive rate on a SGI workstation. Kähler and his colleagues [57] proposed a parametric segmented muscle presentation applying on a single layer tissue layer, the representation form of muscle was refined in [59]. Their muscle is an improvement form of Waters' linear and sphincter muscle, with the feature of piecewise-linearity. Their grid-based definition interface provides better shape flexibility for the sphincter muscle. Sifakis et al. [105] created s tetrahedral volume for each facial muscle from MRI using the technique proposed by Teran et al. [108]. They work created highly detailed geometry of muscle, tendon and skeleton geometry with over 500k tetrahedra. Their simulation was based on FEM with the capability to handle the invertible or degenerated finite elements.

Terzopoulos and Waters [109] proposed a three-layer mesh of mass points based on anatomical structure of facial tissue. They were the first ones to apply dynamic mass-spring system to facial modeling. They also discussed the dynamic behavior

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

of mass spring system to mimic the stress-strain curve of skin tissue. Lee et al. [74, 75] introduced constraint forces to prevent muscles and fascia nodes from penetrating the skull. They demonstrated the automatic construction of individual head models from laser scanned range. The development of MSD-based approaches continues in 2000s. Kähler and his colleagues [57] proposed a parametric segmented muscle presentation applying on a single tissue layer. Their simulation achieved an interactive rate (5fps) on SGI O2 workstation. This approach was later used in their adaptive face model [59] and skull-based reconstruction [58].

Koch and his colleagues [67] applied the thin plate FEM approach for craniofacial surgery simulation. A facial region with 3.1k triangles took around 30 minutes for computation. The FEM based simulation provided higher realism and better stability than the MSD model. Later approaches using FEM method include [66, 15, 105].

Williams [121] proposed the first performance-driven facial animation approach in 1990. In his approach visible markers on the actor's faces are tracked to deform the mesh. This idea of tracking performer's behavior and applying the expression on a synthetic model has developed since then. Two well-known marker based approaches are *making faces* by Guenter et al. [46] and *re-animating faces in video* by Blanz et al. [8]. Essa et al. [33, 34] proposed a control-theoretical approach to analyze physical parameters through actors' performances, which is so called performance-based physical simulation on the human face. Basu et al. [102, 7] also proposed a performance-based approach to mimic the lip motion.

Ekman and Friesen introduced *the Facial Action Coding System (FACS)* [31] in 1978. The FACS represents anatomic notation of the expressive abilities of the face, it defines the actions of face by small visible units in the region of lips, cheeks, and eyebrows. Other similar works include Thalmann et al.'s Abstract Muscle

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Action (AMA) [113] and Kalra et al.'s Minimum Perceptible Action (MPA) [60]. In 1996, ISO started standardizing MPEG-4 specification. The facial animation is specified in MPEG-4 standard by Facial Definition Parameters (FDPs) [2] and Facial Animation Parameters (FAPs) [3]. FDPs define 10 groups with 84 head feature points and FAPs define 2 levels with 66 facial actions. This standard gives the guide to create MPEG-4 compatible facial animation player. The applications of MPEG-4 can be found in [32, 90, 71, 36, 39].

In the speech driven facial animation area, Lewis and Parke [77] used linear predictive coding for limited speech analysis to automate the lip-synched speech animation creation. Bregler et al. [11] proposed the first video resync facial animation system in 1997. The other speech driven facial animation approaches include [16, 95, 35, 90, 17]. Sifakis et al. [106] used inverse muscle activation estimation framework [105], translated a motion captured training set of speech examples (phonemes and visemes) into muscle activation signals. Speech animation was generated from their anatomy-based FEM framework [105] by inputting trained muscle activation signals according to the analyzed speech signals. Their approach provides the ability to interactive with character in any physically describable way which is not available in image-based and shape-based approaches.

The timeline of related work on shape modeling and animation is shown in Table 2.1, Table 2.2.

Table 2.1: Timeline of 3D Human Face Shape Modeling

	before 1990	1990–1995	1996–2000	after 2000
First Face	Parke [94]			
Template Adaptation		Lee et al. [75]	Lavagetto et al. [71]	Noh and Neumann [124] Kähler et al. [59] Zhang et al. [125] Sifakis et al. [105]
Blend shape			Blanz & Vetter [9]	[100, 56, 125, 87] Vlasic et al. [116]

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Table 2.2: Timeline of 3D Human Facial Animation Synthesis

	before 1990	1990–1995	1996–2000	after 2000
Muscle Model	Platt and Badler [98] Waters [117]	Kalra et al. [61] Lee et al. [74, 75]		Kähler et al. [57, 59] Sifakis et al. [105]
MSD model		Terzopoulos and Waters [109] Lee et al. [74, 75]		Kahler et al. [57, 59] Teschner et al. [112]
FEM model			Koch et al. [67, 66]	Choe et al. [15] Sifakis et al. [105]
PDFA		Williams [121]	Essa et al. [33, 34] Guenter et al. [46] Basu et al. [102, 7]	Blanz et al. [8] Sifakis et al. [105]
Face Action	FACS [31] AMA [113]	MPA [60]	MPEG-4 FAPs [3] MPEG-4 Apps [32, 90, 71]	MPEG-4 Apps [36, 39]
Speech Animation	Lewis and Parke [77]	Cohen et al. [16]	[95, 35, 90, 17]	

2.3 Facial Shape Modeling

2.3.1 Shape Representation

It is very hard to isolate the facial model representation methods from the deformation methods. The geometric representation is associated with the deformation method. For example, to approximate the physics of facial deformation after surgery, a non-linear, globally C^1 continuous finite element model of the facial surface, which is based on triangular polynomial shape functions similar to the modeling paradigm, is employed [67]. The representation technique also depends on the requirement of the application. For example, the medical application needs volume data and highly physical realism rather than the characters in games which only require surface data. In some other cases, cinematic visual effect is all that matters and the issue will be on convenient motion control and achieving higher resolution smooth mesh.

Polygon model: Polygon model is supported by current graphics hardware and several deformation methods are associated with following polygon techniques.

- Keyframe interpolation

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

- Point-based deformation
- Blend shape deformation
- Physically-based deformation
- Free-form deformation

Parametric model: For smooth face, a variety of B-spline patch method can be used. The spline method, even though it follows the skin surface, ignores the underlying anatomy, yet can model a smooth and flexible surface. But in current graphics architecture the spline representation still have to be resampled into a polygonal model. There are several types of splines being used.

- B-Splines
- NURBS
- Hierarchical Splines
- T-spline

Implicit surface model: The implicit surface model is useful when the target object is well-regulated so that it can be represented by a mathematical formula. The advantage of implicit surface is that the geometry information is re-organized by parameters and the formula. This will largely reduce the storage cost of the model. However, interacting with implicit surfaces is difficult, that says, it is difficult to apply transformation on part of an implicit surface. Therefore, it is only suitable to use implicit surface to represent a static shape model because of its storage efficiency. For a deformable model, the implicit surface representation should be transferred into other forms.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Layered Mass-Spring Model: Instead of surface representation, some applications require volumetric representation. Layered Mass-Spring model is a light-weight approach to represent the volume of flesh in the face. A Mass-Spring model is a special case of a particle system. Same as a general particle system, each point in the model has its own characterizations such as velocity, mass and acceleration. Being different from a general particle system, the points are connected with each other by the damping spring. Therefore the dynamic system of Mass-Spring model is called Mass-Spring-Damper (MSD) System. The dynamic properties of a MSD system is determined by the stress-strain relationship of the damped spring. To represent the volume, multi-layer Mass-Spring model should be used. Neighboring layers are connected together by damping springs.

FEM model: The three-dimensional FEM can handle greater details and more complex characterizations of the volumetric materials than a layered Mass-Spring model. In FEM, the object is subdivided into a number of small elements. The finite element modeling can be viewed as an improvement of particle system. Different size, and different properties may be assigned to elements depending on their location within the object. The constraints such as shape smoothness, collision problem and object properties such as incompressibility must be handled in FEM. Since FEM aims to simulate more complex characterizations of the object, normally the computational cost is higher than the MSD model. Currently it is still not suitable for real-time animation on a consumer PC.

ATTENTION: The Singapore Copyright Act ap o the use of this document. Nanyang Technological University Library

2.3.2 Personalized Human Face Geometry

A method called blend shape or shape interpolation [101, 122] is widely used on animation control and facial expression synthesis. Blanz and Vetter [9] create a face shape and texture database. A parametric morphable head is interpolated by linear combination of the face model in the database. The parameters of the head model are detected by their novel method to track the corresponding features from multiple images. Some other approaches based on blend shape for facial expression and animation can be found in [56, 125, 87]. A new improvement in blend shape approach is achieved by Vlasic et al. [116]. In their work, multi-dimensional face model is represented as the tensor product of vertex information, expression type or viseme (visual representation of acoustic unit). Any imported multi-dimensional model is applied with a *singular value decomposition (SVD)*, and the result is a core tensor with some tensor product weight. The advantage of this approach is that if there is data missing in one dimension or multiple dimension, the missing data can be *guessed* or interpolated from other similar core tensor. They also provided a solution to extract the blending weights from 2D images. It is obvious that blend shape approach is not limited on the geometric domain. The same idea can be directly applied to the color domain. But since shape blending is based on a finite number of samples in the database and their linear combination, the successful rate of the reconstruction depends on the size of the database.

There is another method which is called template-based approach. The template-based approach contains one or several template models, estimates the difference between the source model and the template, and then deforms the template to mimic the source model. The benefit from template-based approach is that template based model is system specific which can contain a lot of pre-defined information. The information can be muscle structures, region definition and hierarchical

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

layered tissue-bone structures for facial geometry. This solves the problem that in usual case, only surface data can be obtained from range scanner or photos. The medical images such as MRI can provide volumetric information of interior structure, but that would require more expensive equipment and not suitable for regular usage. In Escher et al. [32], a cylindrical projection is applied on the generic face model first to interpolate any missing feature points, then their Dirichlet Free Form Deformation (DFFD) method is employed to generate the deformation of head model. This allows for volume deformation and continuous surface interpolation. Kähler et al. [59] trained a global Radial Basis Function (RBF) network by the pairs of Feature Points (FPs) from the 3D scanned data and 3D generic model. The generic model was thus smoothly transformed to the shape close to the scanned model.

2.3.3 Facial Feature Modeling

Lips are one of the important features of human face. Some physiological-oriented models have been proposed to describe muscle and tissue of the lip. Nevertheless, the high complexity of these models makes performance-driven lip tracking a generally difficult task. Basu et al. [103] proposed a 3D lip model both applicable to analysis and synthesis of lip motions. This model is based on a 3D polygonal surface. The motion of some points is statistically learned from video, while the whole surface stiffness is regularized by a finite element method. The 3D model position is evaluated so that its 2D projection fits the area of the lip estimated by color analysis. Badin et al. [6] based on MRI and video images, obtained 3D lip, tongue and face model. The tongue model is from MRI data, and lip and face are reconstructed from frontal and side view of images. Their approach improves the cylindrical tube lip model proposed by Beautemps et al. [18] by adding sev-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

eral control parameters. Principal Component Analysis (PCA) was performed to extract the control parameters from 3D data.

Eyes are another feature on the face. During face-to-face conversations, the eyes exhibit conversational turn-taking and agent-thought processes through gaze direction, saccades, and scan patterns. Anatomically, the eyes gain their appearance from a layered and highly complex structure that is difficult to model and render with conventional techniques. A statistical method [72] and texture synthesis techniques [76, 24] have been proposed to render a realistic eye appearance and movement. Lee et al. [72] implemented an eye movement model combining an eye tracking system with an empirical model. Their eye model can reflect the dynamic movement from the tracking video. Lefohn et al. [76] merged several semi-transparent textures together to create realistic iris appearance. Deng et al. [24] analyzed the eyelid movement, the head motion and speech to synthesize a simple gaze-blink eye model. The model can generate performance-driven natural eye movement to fulfill the requirement of computer games.

2.4 Facial Expression Synthesis

2.4.1 Performance Driven Expression Synthesis

Due to the variety of facial expressions and thousands of personalized features, to build a life-like facial animation system for a specific person needs a lot of tuning of control parameters. This difficulty leads to the performance-driven animation. Performance-driven animation captures the motion data from real human beings and uses the data to drive the virtual character's motion. Real-time interactive animation can be observed by the actor with their individual motion and expressions. The key point is to track the feature point and convert it to control parameters

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

accurately. Normally the tracked 2D or 3D motion data are transformed or filtered to generate the motion data for a specific animation system. Temporal control is important in performance-driven animation to improve the visual effect. The motion data of performance driven animation for body parts can be captured using Data Glove, instrumented body suit, video or laser-based motion tracking system. For facial animation, video or laser-based motion tracking system is obviously more useful.

Direct Marker Tracker: The simplest way to detect the motions from video sequences is to track the markers placed directly on the face of a target person. The task becomes tracking bright spots on a dark field. Williams [121] is the pioneer of this approach. The idea is to track expressions of a live performer with special makeup and then map 2D tracking data onto the surface of a scanned 3D face model. Since then, colored markers painted on the face or lips are extensively used to aid in tracking facial expressions or recognizing speech from video sequences [15, 46, 65, 85, 64, 97]. Moubaraki et al.'s work [85] synthesizes the color texture, real-time wrinkles and shape deformation. Bump mapping techniques are used to synthesize the wrinkles. In Kishino et al's approach [64], tape markers are attached to facial muscles; magnetic sensors and data glove are also used for detecting the head movement. Pighin et al. [97] used marker-based capture to accurately reconstruct the movement of a sparse set of 3D point on the face surface. The marker movement used in facial model reproduced coarse facial dynamics. The recent work by Choe et al. [14], which employed hand-generated muscle actuation basis models both to estimate muscle actuation parameters and to produce 3D facial animation achieved excellent result. Their work requires a set of 27 facial markers applied on an actor's face. But for robust estimation, a time consuming fully-fitted model of the actor must be produced first.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Snakes: Snakes [62], or deformable minimum-energy curves, are widely used to track intentionally marked features used in tracking facial features. The recognition pattern of snakes is primarily based on color samples and edge detection. A snake is basically the numerical solution of a first-order dynamic system. Human face has many feature lines and boundaries that a snake can track. Terzopoulos and Waters tracked contour features on the eyebrows and lips for automatically estimating the muscle contraction parameters from a video sequence. These tracked parameters were later used in their physically-based face model [75]. Goto et al. [44] tracked the lip and eye separately using an energy-based line fitting algorithm which is similar to snakes. Individual probability databases are used for mouth shape combination and eye features (pupil, eyebrow, eyelid) combination.

Tracking errors accumulated over long image sequences and the numerical integration becomes unstable. Occasionally, snakes will lose the contour it is tracking. A reliable tracker should allow a snake to re-initialize when large error accumulation occurs.

Marker Independent: The use of markers is not always practical. To track the markers, markers must be seen every moment, this restricts the head movement, so it is more attractive to recognize features without markers. Approaches that employ facial trackers without special markers are [70, 33, 22, 43].

Optical Flow: Optical flow performs natural feature tracking and therefore obviates the need of marking on the face. Combining with spatio-temporal normalized correlation measurements [19], dynamics of facial expression can be captured together with the temporal characterization.

Optical flow in image is produced by motion rather than the shape [34]. A low-level processing of the pixel is performed. It is a direct representation of facial

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

animation after smoothing is performed on the motion field.

Optical flow is very sensitive to the noise and can not be applied on rapid motion. But Essa and Pentland's work [34] which employs a multi-grid optical flow coupled with a physical-based representation seems promising. Their method captures spatial and temporal pattern in sufficient details. Optical flow processing is used as the basis of measurement of facial motion. Facial features are tracked on all the images in the database initially, a coarse-to-fine algorithm is then used to estimate the flow vector. The 2D motion estimation is mapped onto a prototype 3D model using a spherical mapping and the muscle actuation is obtained using a control theoretical approach. A continuous time Kalman filter (CTKF) is employed to mimic the noise-free dynamics and to recursively correct the estimation. Since CTKF is a recursive optimal least-square estimation technique, this control framework is free from the rapid motion problem of optical flow and the estimation is well-proportioned to the difference between observation and prediction.

2.4.2 Facial Action Representation

2.4.2.1 FACS

Ekman and Frisen's FACS [31] describes all the visible facial movements as movements of facial muscles, jaw/tongue derived from analysis of facial anatomy. The primary goal in developing FACS was comprehensiveness, a technique that could measure all possible, visible and discriminable facial movement. Action Unit (AU) is the base measure unit in FACS. FACS includes 44 base AUs. Combination of different AUs generates various expressions. For example, combination of AU1 (Inner brow raiser), AU6 (Cheek raiser), AU12 (Lip corner puller), AU14 (dimpler) generates happy expression.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Because FACS is based on facial anatomy, it is very suitable for physically-based muscle simulation approach. This kind of approach overcomes the two difficulties of correspondence and light changing/variation that happens during interpolation and image morphing. Hence, it becomes a very popular approach. Physical muscle or pseudo muscle models are used to describe the dynamic properties and behaviors of skin, tissue, bone and muscle system. FACS works as a higher level control system over the muscle movement.

Although the authors of FACS intend to describe all facial movements as FACS AU combination, it still limits to visible facial movement, the invisible movements like muscle tonus, vascular and glandular changes produced by automatic nerve system are not measured. Another problem is that AU descriptions are static and passive, the contraction, the maintaining and the relaxing period of the muscle movement is not described in FACS. Lacking of temporal information remains a problem.

2.4.2.2 MPEG-4 FAP

MPEG-4 specifies a face model in its neutral state, a set of facial features (FPs) on the neutral face as reference point and a set of FAPs, corresponding to a set of facial actions deforming from the neutral face. Deforming the face model with specified FAP values in each time-frame generates the animation sequences. FAP values are defined in face animation parameter unit (FAPU). FAPU are computed from spatial distances between major facial features on the model in its neutral state. MPEG-4 defines 84 feature points on a neutral face, the purpose of FP definition is to provide spatial reference to define FAP. The 84 FPs are divided into 10 groups based on the features and regions.

FAPs are also closely related to muscle actions. There are 68 parameters which

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

are categorized into 10 groups. The group 1 is called high level FAP which includes visemes and expressions. The groups 2-10 are considered as low-level parameters which precisely describe how much FPs should be moved for a specific amplitude.

FAPs are not only suitable for realistic human face animation but can also be used for cartoon-like animation. Facial AU can also be defined by FAPs. Pandzic [92] proposed facial animation framework for web applications, a very simple geometry mesh and FAPs are used to control the animation, a facial motion cloning approach is also proposed. Because FAPs include viseme which is the corresponding visual state of phoneme, speech-driven animation can be synthesized. Hong et al. [50] used a neural network to analyze the temporal audio and video data and convert FAPs to Motion Unit Parameters (MUPs) and generate the real-time speech-driven face animation.

FAPs provide a new description for facial action, and is specially suitable for web applications. However, it just considers the spatial control and the problem of temporal control still exists.

2.5 Facial Appearance Acquisition

2.5.1 Analytic Illumination Model

Simple Phong illumination model cannot properly represent the skin reflectance because of the translucency and multi-layer properties of skin. For human skin rendering, Bidirectional Reflectance Distribution Function (BRDF) [88] is the simplest rendering model. Marschner et al. [80] measured spatially uniform BRDFs on the foreheads of different gender, race, and age, the uniform specular BRDFs was later augmented with a spatially-varying albedo texture [79]. Researchers have also tried to estimate parameters of uniform analytic BRDF models from

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

photographs [9, 93, 40]. One of the basic limitation of BRDF is that the model ignores the subsurface scattering which makes skin look soft.

Hanrahan and Krueger [47] described scattering and absorption parameters of material using the Henyey-Greenstein function and compute scattering with Monte Carlo simulation. Their model is suitable for appropriate biological material such as skin. So they were the first to model subsurface scattering in skin. To avoid expensive Monte Carlo simulations, Stam [107] developed an analytic approximation to multiple subsurface scattering for skin with a rough surface. Though these models approximate the subsurface scattering, they still simulate skin reflectance using BRDF in which light enters and exits at the same location.

The light scattering from two locations can be simulated by Bidirectional Subsurface Scattering Reflectance Distribution Function (BSSRDF). But the BSSRDF model is eight-dimensional which is hard to estimate. Jensen et al. [55] were the first to propose a practical low-dimensional analytic approximation for the BSSRDF of homogeneous materials using diffusion theory. Their model can effectively simulate materials behavior such as color bleeding within materials. Optical properties of translucent materials such as milk, marble and skin can be recovered from their approach. Jensen and Buhler [54] developed fast rendering methods for this approximation. Their implementation traverses an octtree and evaluate the radiant exitance flux of voxels. The hierarchical structure reduce the cost of summing the contribution from all irradiance samples and reserve the important illumination. Donner and Jensen [27] introduced a multi-layer diffusion approximation for rendering skin. However, it is difficult to measure the parameters of each layer in the multi-layer translucent material. Weyrich et al. [120] proposed a novel analytic model which decomposes large amount of measured skin data into a spatially-varying analytic BRDF, a diffuse albedo map, and diffuse subsurface

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

scattering. The subsurface scattering is measured using the technique introduced in Jensen et al.'s work [55]. The decomposition enables statistical analysis of the variations within subject age, gender, skin type and other external factors.

2.5.2 Image Based Technique

Compared to analytic illumination model, image-based models can create highly realistic representations for human faces because they directly capture effects such as self-shadowing, color bleeding, inter-reflections and subsurface scattering. Kurihara and Arai [69] demonstrated an example of creating faces from a single canonical face at the beginning of 1990s. The model independent image based approach was improved by Pighin et al. [96] in 1998 by using images from multiple views to reproduce faces under static illumination conditions, their approach combined 2D morphing with 3D transformations of a geometric model. Recently variations of light condition can be applied on static face [42, 20], expressions [48] and real time performance [119]. Debevec et al. [20] present a process for creating realistic, relightable 3D face models by mapping image-based reflectance characteristics onto 3D scanned geometry. In order to change the viewpoint, they use colorspace analysis to separate the image data into specular and diffuse components that can be extrapolated to new viewpoints. While their method does consider the aggregate behavior of subsurface scattering, they do not model a specific diffusion parameter. Thus, their method cannot produce correct subsurface scattering effects for closeup light sources or high-frequency spatially varying illumination. Borshukov and Lewis[10] combined an image-based model, an analytic surface BRDF, and an image-space approximation for subsurface scattering to create highly realistic face models for the movie industry. Sander et al. [104] developed a online (real-time) version of this method [10] on ATI graphics cards. To obtain similar result

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

of translucency on the ears, they increase the kernel size (pixel density) when compute the light map on the ears.

The timeline of related work on facial appearance acquisition is presented in Table 2.3.

Table 2.3: Chronicle of Facial Appearance Acquisition

	before 1990	1990–1995	1995–2000	after 2000
Analytic Model	BRDF [88]	Hanrahan and Krueger [47]	Marschner et al. [80, 79]	Stam [107] BSSRDF [55, 54, 27, 120]
Image based method		Kurihara and Arai [69] Lee et al. [75]	Pighin et al. [96] Various Light [42, 20]	Borshukov and Lewis[10] Various light [48, 119] Sander et al. [104]

2.6 Our Observation of Synthetic Human Face

In this thesis, we report a complete system to synthesis a personalized facial expression in the following four aspects:

- Accurate & stable method for face adaptation of shape
- Multi-level adaptation for coarse-to-fine shape adaptation
- Physically-based method for adaptive animation model using skin definition and muscle model
- High quality face surface rendering

Therefore, besides a general review of research topics on computer facial animation, we emphasis our observation on related techniques.

2.6.1 Template Based Facial Geometry Modeling using RBF

One type of function called Radial Basis Function is identified as a powerful and numerically stable method to solve scattered data interpolation problems. The

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

RBFs has a variational nature [49] which supplies a user with a rich palette of types of basis functions. RBF-based approach can be template-based or implicit-surface based. Template-based approach mainly focus on special usage whereas implicit surface-based approach will not lose the generality.

In most cases, RBF is trained to represent an implicit surface [13, 26, 25, 84]. The advantage of this method is that after the training procedure, only the RBF function and the radial centers, rather than the scattered, noisy point cloud need to be stored. So it saves a lot of space during the data storage and transfer. RBFs can be local or global. The global RBF is useful in repairing incomplete data, but usually it needs some sophisticated mathematical techniques [13]. Carr et al. [13] employed a fast multi-pole method to solve a global RBF function. Their approach also uses a greedy method to reduce the number of radial centers they need to store. On the other hand, local, compactly supported RBF leads to a simpler and faster computational procedure. But this type of RBFs are sensitive to the density of scattered data. Therefore, a promising way to combine the advantages provided by locally and globally supported basis function is to use the locally supported RBF in an hierarchical fashion. A multi-scale method to fit the scattered bumpy data was first proposed in [86], and recent approaches [38, 52, 89] also address this problem.

The power of RBF in head reconstruction is proved in [71, 123, 59]. Noh et al. [123] employed a Hardy multi-quadrics as the basis function and trained their generic model for performance driven facial animation. Since their approach only tracked about 18 corresponding points, the computational cost is relatively low and real-time facial animation was synthesized. But the low number of corresponding points does not ensure the fidelity of the deformed model. Kähler et al. [59], on the other hand, use a higher resolution template to fit the scanned range data. A

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

feature mesh was used to search more corresponding points between the template and scanned data in their iterative adaptation steps. The feature mesh is also refined during each step. Our work uses the same concept of feature mesh. But the level of detail is not in their consideration, thus the only way to represent local detail is to add more corresponding points, which is relatively expensive. We use a hierarchical adaptation process to solve this problem. That says, in the initial stage we use global adaptation and change to a local adaptation by the increment in number of radial centers. Lavagetto et al. [71] also proposed a compactly supported RBF in their experiment and use a hierarchical fashion for the model calibration in their MPEG-4 Facial Animation Engine. But the result of their CSRBF is still not convincing enough as depicted in Fig. 2.1). From our observation, their approach has the following problems:

- The number of FPs are constant during their adaptation, which is around 80 points based on MPEG-4 standard. The problem here is the small set of FP pairs which are not sufficient enough to represent the face features. We combine feature mesh and curvature based searching method for FP pair expanding.

- Their low-resolution model (less than 1000 polygons in frontal face) degrades the adaptation result. And it results in a low quality expression synthesizer. We use a higher resolution face model with more than 3000 polygons in the frontal face. This not only provides a better detail to the adapted shape, but also provides more features in expression generation (wrinkles).

- Though they are using CSRBF with hierarchical fashion, their function only supports 4 levels of iterations with a constant number of FPs, which did not help a lot. We choose a classic CSRBF function without restriction on the

number of iteration count and restrict on the number of FP pairs.

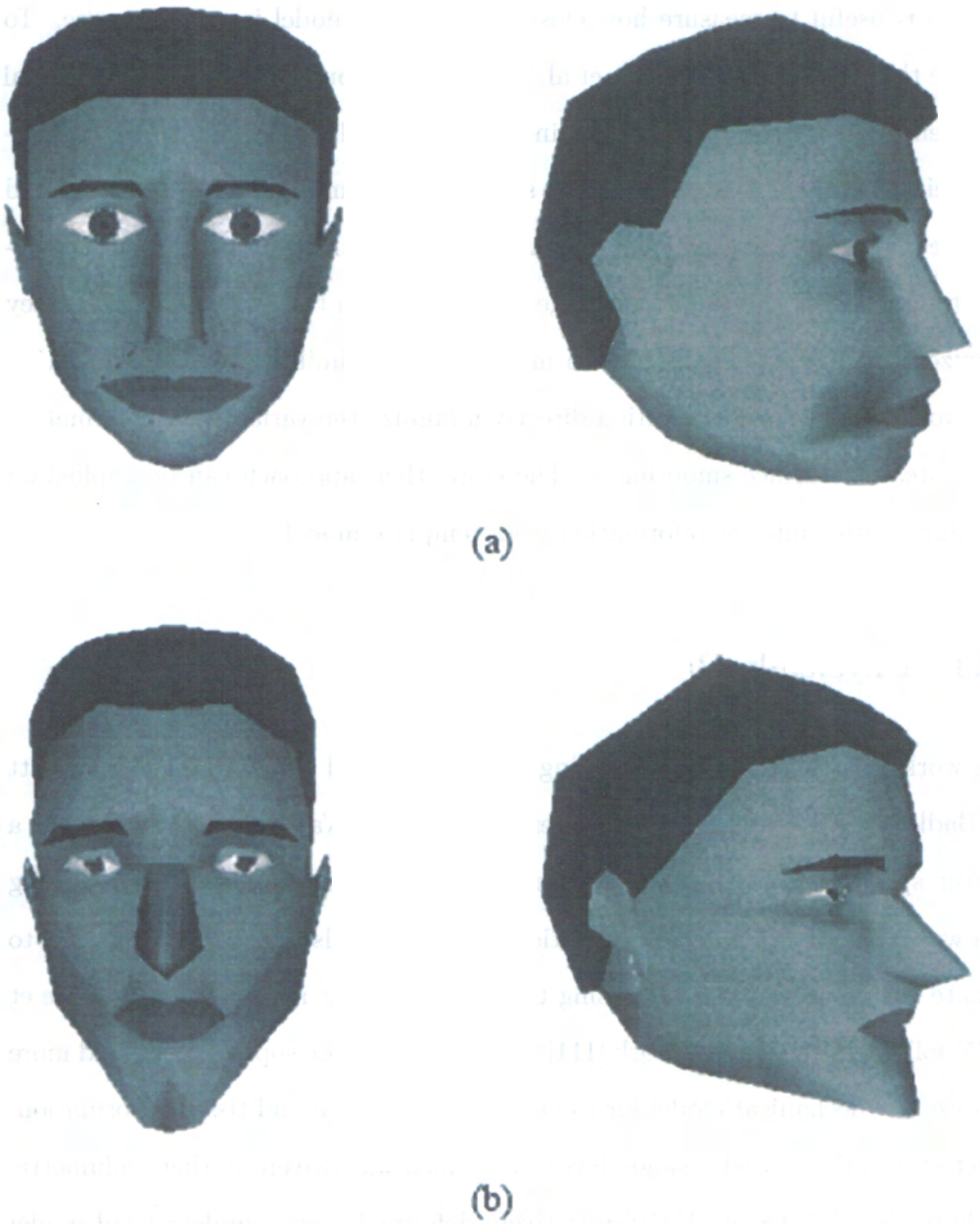


Figure 2.1: Human face adaptation using CSRBF proposed by Lavagetto et al. [71]

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.6.2 Error Metrics for Template Adaptation

Error metrics are not only to validate the effectiveness of the proposed approach, but also is useful to measure how close the adapted model is to the source. To measure the human face, DeCarlo et al. [21] used a corpus of anthropometric facial measurements to model the variation in face shapes. Marschner et al. [79] regularized their fitting process using a surface smoothness term. Allen et al. [5] combined distance error, smoothness error and marker error together to estimate the similarity between the template and the range data. Based on the error estimation, they optimized their affine transformation matrix using a similar energy-minimization framework. But their optimization directly minimize the variation of deformation itself instead of surface smoothness. Therefore, their approach can be applied on hole filling with sufficient information from template model.

2.6.3 Physically-Based Facial Animation

Prior work in 3D facial animation using physically-based approaches include Platt and Badler [98], and Waters [117]. Terzopoulos and Waters [111] constructed a tri-layer spring mesh based on anatomical structure of soft tissue. This spring mesh was deformed by muscle contraction. This paper also described a solution to estimate the muscle contraction using the active contour method (Snake). Lee et al. [75] followed the previous work [111] but proposed more sophisticated and more accurate bio-mechanical model for the muscle contraction and tissue deformation. Kähler et al. [57] created a single layer skin which was driven by their volumetric muscle model. Kähler et al. [59] iteratively deformed their template head model into the shape of the scanned head data. The physical simulation model on the template mesh was transferred to the adapted head so that expression on the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

personalized head could also be synthesized. They described their subsequent work in Kähler et al. [58] where facial surface information can be restored from the profile of a skull, with the adaptive anatomical structure [57], they can reanimate the dead people.

Another physically-based approach employs the Finite Element Method (FEM) for facial deformation, which generates highly accurate deformation result. Koch et al. [67] applied FEM simulation to anticipate the effect of facial surgery. An emotion editor was presented in Koch et al. [66] where the face model was created from the Visible Human Dataset [114] and the deformation followed the work of Koch et al. [67]. Choe et al. [15] used a 2D linear quasi-static FEM model to compute the deformation of skin surface and estimated the muscle actuation profile from performance by optimizing the distance between the tracked markers and the simulated features. But their lack of anatomical structure suffered from both the loss of physical accuracy and the loss of nonlinearity. Sifakis et al. [105] followed previous work in Teran et al. [108], reconstructed the flesh mesh, skeleton and muscular structure from the Visible Human Dataset [114] to construct a detailed head model. They used a nonlinear approach of the constitutive model and FEM model. Though the method is more complicated than the blend shape method, it has the ability to simulate a wider range of skin behaviors. Their framework can automatically recover the muscle activation parameters, head pose parameters and jaw kinematics from marker-based video analysis by least-square optimization. Because the estimation is to resolve the physical simulation parameters rather than factors which directly influence the facial shape, it also minimizes the problem of noise marker data. The flesh model contains around 370k tetrahedral elements and the average processing time of simulation is 8 minutes per frame on a Xeon 3.06Ghz CPU. Hence, it is currently not suitable for real-time performance.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.7 Findings

In this chapter we have presented literature review on topics that are directly related to face model representation, face adaptation, face animation as well as realistic face rendering. From our survey the findings are:

- Rapid face shape adaptation is still a problem which needs to be solved.
- For high quality face adaptation, we need to develop new and robust error metrics.
- A comprehensive and fast system to combine adaptation for personalized animation and realistic rendering is also a challenge that needs be addressed in our work.

2.8 Summary

We have presented the literature review and our findings. A more detailed anatomy based human face analysis is presented in next chapter.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 3

Anatomy of Human Head

3.1 Introduction

State-of-the-art facial models for computer animation attempt to represent the geometry, photometry, deformation, motion, etc., of the various organs and features associated with the face, as well as with the rest of the head and neck. Typically, the models are designed to produce meaningful facial images [12]. To create a physically realistic facial model, we try to understand the anatomical structure. The human head can be viewed as a layered structure, the significant layers include the skin, the muscles, the skull and the brain. Since our focus is on facial expressions, we do not consider the brain as part of our head model. In the chapter, we will analyze these structures and their properties.

3.2 Skin

3.2.1 Layered Soft Tissue Structure

Soft tissue is a collective term for almost all anatomical structures. The term *soft* is used when compared to bones.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Soft tissues are mainly composed of different types of polymeric molecules embedded in a hydrophilic gel called ground substance [41]. A basic structural element of facial and other soft tissues is collagen, which amounts up to 75% of dry weight. The remaining weight is shared between elastin, actin, reticulin and other polymeric proteins.

The direction of collagenous bundles (connective tissue) in the skin determines lines of tension, known as the Langer's Lines [23]. The arrangement of fibrous structures in the skin is individual, which, in particular, reflects in the individual wrinkles of the skin.

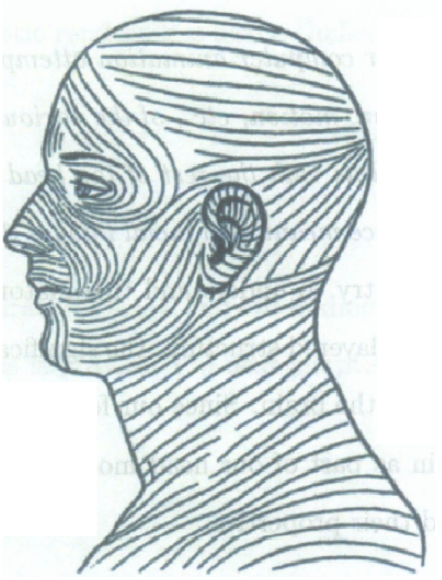


Figure 3.1: Langer's Lines [82]

The facial tissue consists of anatomically and microbiologically distinctive layers including epidermis, dermis and the underlying subcutis (hypodermis).

The outermost layer, the epidermis, isolates our body from dust, germs and harmful ultraviolet rays from sun. Epidermal layer consists of a basement mem-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

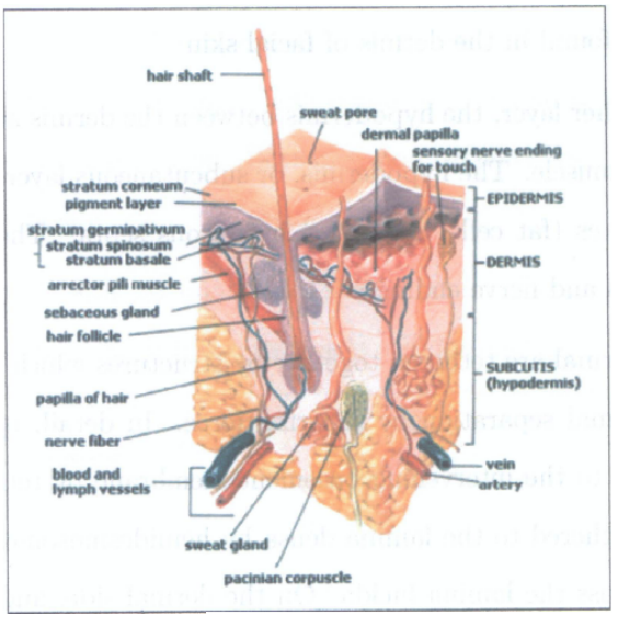


Figure 3.2: Skin Layer [1]

brane lined with basal cells that divide, flatten, and work their way outward to form the waterproof outer layer of dead cells, called the stratum corneum. Interspersed in the basal layer are melanocytes, which synthesize pigment, or melanin, which gives skin its color. Epidermis is devoid of vessels. The terms *thin* or *thick* refer to the thickness of the epidermal layer. Thick skin covers on the palms, fingertips or the soles of the feet. Thick skin lacks follicles, sebaceous glands and arrector pili muscles. In contrast, thin skin is over most of the body, contains hair follicles, sebaceous glands and arrector pili muscles. The skin on the face is thin and hairy.

The second, larger layer is called dermis. The dermis is composed of collagen, elastic and connective tissue fibers. The dermis is divided into two strata. The thin upper papillary dermis is the main agent of dermis function. From here dermis supplies nutrients to the epidermis and regulates temperature. The thicker deeper reticular dermis consists of thick dense network of collagen and reticular fibers.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Striated muscles are found in the dermis of facial skin.

There is also another layer, the hypodermis between the dermis and the underlying fascia covering muscle. The hypodermis, or subcutaneous layer, is composed of groups of adipocytes (fat cells) separated by fibrous septa. The hypodermis contains sweat glands and nerve endings.

Epidermal and dermal are tethered together by structures which minimize the risk of dermal-epidermal separation by shearing force. In detail, epidermis and dermis are connected to the intervening basement membrane by tethering fibers. The basal cells are tethered to the lamina densa by hemidesmosomes from which anchoring proteins cross the lamina lucida. On the dermal side, anchoring fibrils of type VII collagen attach the lower surface of the lamina densa to collagen fibers in the dermis; fibrillin microfibrils attach it to elastic fibers in the dermis. Also the dermal and hypodermal layers are joined together by fatty tissues.

3.2.2 Skin Properties

Biomechanics combines the field of engineering mechanics with the fields of biology and physiology. In the area of tissue biomechanics, numerous experimental and theoretical studies have been carried out in recent years [41, 81, 91]. Generally, soft tissue exhibits *non-homogeneous, anisotropic, quasi-incompressible, non-linear viscoelastic* material properties. The following parts will describe these phenomena.

Non-homogeneity, anisotropy: Cells, intracellular matrix, fibrous and other microscopical structures compose the soft tissue. Therefore, the mechanical properties of living tissues vary everywhere within the tissue. The main objects of mechanical modeling are spatial distribution of material stiffness and the organization of fibrous structures such as collagen and elastin, which have some preferen-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

tial orientation in the skin. The dependence on coordinates along the same spatial direction is called non-homogeneity. The material is called anisotropic when its property depends on the direction. Both non-homogeneity and anisotropic property exist on facial tissue.

However, there is still no quantitative data about these properties so that modeling facial tissue with these properties is still impractical.

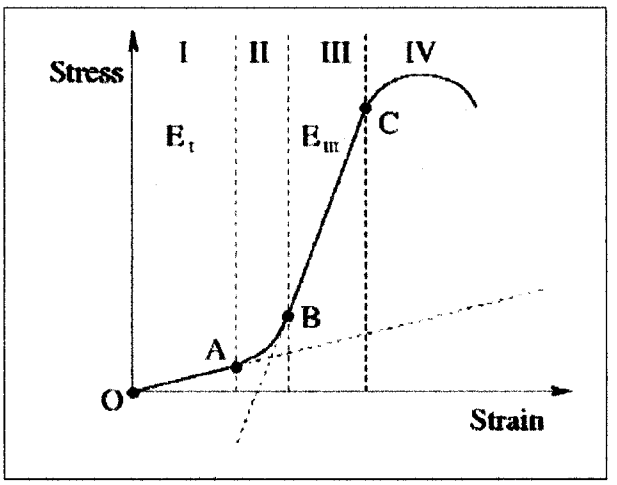
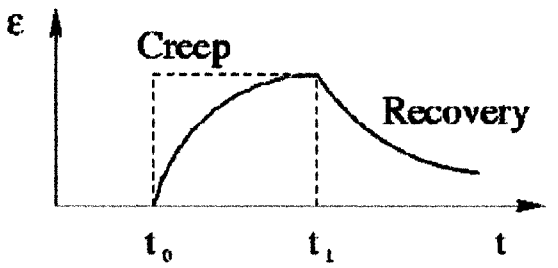


Figure 3.3: Non-Linear Stress-Strain Curve for soft tissue [41, 91].

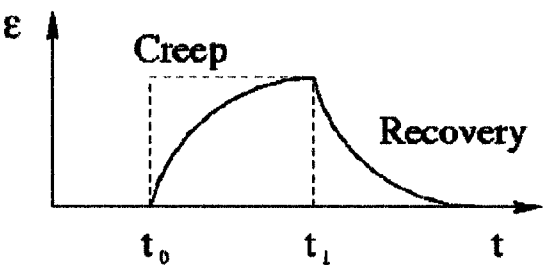
Non-linearity: The stress-strain relationship, the so-called constitutive equation of the skin and other soft tissue is non-linear [63]. The non-linear stress-strain curve of soft tissue is shown in Fig. 3.3. At low strain (Phase I), the response of soft tissue is linear; at average strains (Phase II), the straightening of collagen fibers occurs and the tissue stiffness increase; at high strains (Phase III), all fibers are straight so that the reaction becomes linear again. By applying even larger stress, material destruction occurs. There are no quantitative data about stiffness coefficients and the critical strains for bilinear approximation of facial tissue. However, these parameters may vary from person to person and depend on multiple factors such as age and gender.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Viscoelasticity: The response on tissue stress depends on its history of deformation, and this phenomenon is called viscoelasticity. This means the stress is a function of both the strain and the strain rate, which is time-dependent. Viscosity is a fluid property and elasticity is a solid property. Therefore, viscoelasticity combines both fluid and solid properties. There are several characteristics of time-dependent behavior, which are *Hysteresis, Stress Relaxation or Recovery, Creep, Incompressibility and Preconditioning.*



(a) response to a viscoelastic fluid



(b) response to a viscoelastic solid

Figure 3.4: Creep and Recovery with constant stress σ applied at time t_0 and removed at time t_1 [91].

- **Hysteresis:** The history of strain affects stress, which means that the stress-strain relations in the loading and unloading processes are different.

- **Stress Relaxation:** It refers to the reduction in the force opposing a deformation held constant in time. A viscoelastic solid will completely recover.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

For a viscoelastic fluid, a residual strain will remain in the material and complete recovery will never be achieved [see Fig. 3.4].

- **Creep:** This refers to the exponentially increasing strain over time under a constant load [see Fig. 3.4].

- **Preconditioning:** The phenomena that same load results in different deformation responses.

- **Incompressibility:** If the volume of a material remains unchanged by deformation, it is called incompressible. Generally, soft tissue contains both incompressible and compressible ingredient. Tissues with high proportion of water, for instance the brain or water-rich parenchymal organs are usually modeled as incompressible material, while tissue with low water proportion are assumed quasi-incompressible. Since fat cells and ground substance are composed mostly of water, we assume that the soft tissue as incompressible. For the volume to remain incompressible, while the tissue is compressed from one direction, the material inside will be forced out perpendicular to the direction. This makes sense to the bulge effect where that muscle contracts and hard tissue remains immovable, so skin elements are forced to the outside direction of the face.

3.3 Facial Muscle

Facial muscles are attached to the skin tissue by short elastic tendons at many places in the fascia, and fixed to the facial skeleton at a few points. Muscle force propagates through facial soft tissue when the muscles contract, this force which is looked upon as external force to the skin which produces various facial expressions.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Therefore muscle movement is the major influential factor to facial expression and muscle should be carefully modeled.

3.3.1 Muscle Anatomy

The main macroscopic property of muscle is its ability to contract. In accordance with the sliding filament theory, this macroscopic property is effected through the shortening of muscle fibers that consists of two main contractile proteins (actin, myosin) and several other protein structures building a separate morphological unit, the so-called sarcomere [41]. The shortening of the serially arranged sarcomeres develop a tension along their longitudinal axis. Sarcomeres represent the lowest level of the complex hierarchical structure of muscles. The muscle fibers group themselves into bundles known as fascicles, which are enveloped by fascia. Each fascia group represents different heads of a muscle. Within the fascia, the muscle fibers can have an orientation characterized by the angle they make with respect to the tendons they attach to. The fiber arrangement is known as the muscle pennation.

3.3.2 Biomechanics

The arrangements of muscle fibers determine the direction of the produced forces. Since tendon and muscle are attached together to create a functional unit of force generation and transmission, they are often referred as a collective biomechanical structure, called musculotendon unit.

The tendon portion of the skeletal musculotendon unit attached to the bone at two different sites known as the origin and the insertion. Facial muscles have several differences with other skeletal muscles. In particular, some muscles are attached to soft tissue or other muscles. In this case, the tendon portion attaches

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

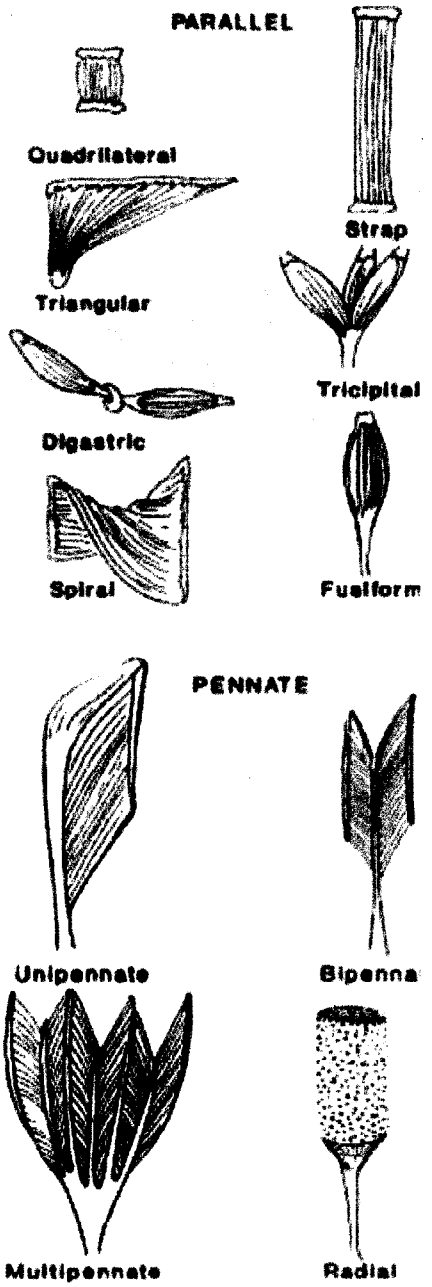


Figure 3.5: Pennation patterns of skeletal muscles (from [81])

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

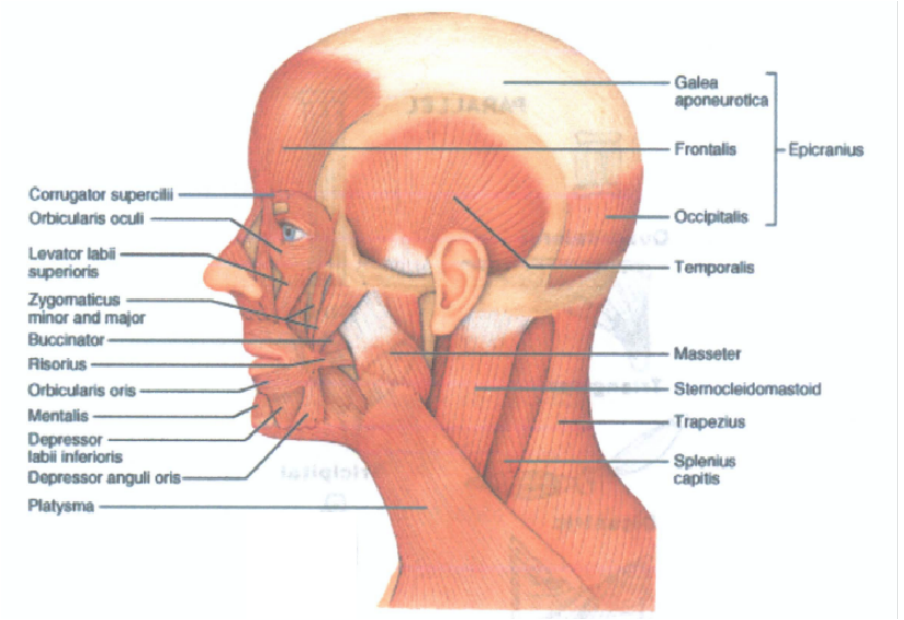


Figure 3.6: Facial Muscles [4]

to the bone is called *the origin* and the portion connected to soft tissue is *the insertion*. Fig. 3.6 shows a set of the muscles of the face. Several muscles may work together to influence a same region. These muscles interweave with one another and the individual effect functions exist in each group. These muscles may or may not always be independently controllable.

3.4 Incorporation of Anatomical Analysis into Our Design

The anatomical analysis helps to strengthen our design in various ways:

- Skin Modeling: We represent the skin as a multi-layer shape model (described in detail in Section 6.2).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

- Skin Deformation: In our approach, as described in this chapter, we attempt to simulate the deformation of the skin using a non-linear computational model (described in detail in Section 6.4).

- Muscle Modeling: The analysis of the shape of the muscle (Section 3.3) and the understanding of the origin and insertion helps to specify the muscles in the human face for our simulation. The muscle definition and registration are further described in detail in Section 6.3.

- Muscle Deformation: As analyzed in Section 3.3.2, both the individual muscle biomechanics is important as well as the group of muscles for synthesizing a specific expression. The formulation is described in Section 6.4.1.

- Skull Shape Modeling: The human cranium and the facial bones are the foundation for the soft tissues of the face and head. The skull is also starting point of several facial muscles (called attachment). We use the skull for muscle control point registration (Section 6.3). It can also be used as a constraint during the simulation because of the skull impenetration property. Compared to the fixed *cranium*, the *mandible* a movable bone in the head. The mandible, or jaw bone rotates the bottom face in a non-uniform fashion, we describe the jaw rotation in Section 6.4.3.

3.5 Summary

This chapter focused on a detailed analysis of the anatomical characteristics of the human face. We also analyzed the biomechanics aspects of the skin tissue and muscles. As explained in the previous section, we attempt to formulate these anatomical properties in our simulation. In the next three chapters we will explain how do we construct our adaptive head model with anatomical realism.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 4

Template-based Adaptation for Facial Shape Modeling

4.1 Introduction

In this chapter, we propose an approach to generate personalized human head from range scan data. Landmarks compliant to MPEG-4 Facial Definition Parameters (FDP) are initially labeled on both template model and any target human head model as prior knowledge. The deformation from the template model to the target head is through a multi-level training process. Both general Radial Basis Function (RBF) and Compactly Supported Radial Basis Function (CSRBF) are applied to ensure the fidelity of the global shape and facial features.

For traditional facial shape modeling, we need a skilled modeler to spend a lot of time on building the model from scratch. With the availability of range scanners, the shape information is easily captured in a few seconds. Fig. 4.1 shows a scanned face from a Minolta Vivid 700 lazer scanner. But the scanned model still suffers from the following problem.

- **Shape Problem:** From the range scanned data, the smoothness of the reconstructed data is still not complete. Holes or gap may appear during the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Figure 4.1: Range scan data from Minolta Vivid 700 Laser Scanner

merge procedure of two scanned data from different views. Overlapped or folded surfaces produced by the merging procedure results in visual artifacts. One particular problem in facial data acquisition by range scanning method is that hairy surface can not be appropriately recognized by the scanner.

- **Manual Editing:** Facial shape is not a totally continuous close surface. It contains some feature parts such as lips, eyes and nostril. In a neutral face, the mouth is closed, the eye gaze direction is towards the front and the nostrils are not visible. The range scanner does not have the capability to detect these features, so tedious manual editing effort such as lip contour separation is still required.

- **Animation Ready:** Even as the precision of the scanner is increasing, modeling the portion of head other than the face can be solved by scanning a head with very short hair or wearing some special head cover, the scanned data is still not animation ready. For an animated head model, the interior

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

deformation engine has to be set. The engine can be totally anatomy-based, or geometry-based. Different approaches have different requirements, the more complex the engine is, the more parameters we need to define on our obtained model, before it is deformable.

Based on the above analysis we understand that it is not an easy work to get an animated head directly from range scan data. Therefore we create a conversion tool to convert an arbitrary 3D head from laser scanner or other sources into an MPEG-4 compliant head with high fidelity to the original input data but still at a relatively rapid speed. To accomplish this objective, we model a template anatomy-based head embedded with skin, muscles and skull, the model is ready to generate impressive facial expressions. Given an input 3D mesh, we adapt our template model to the input data, with all the anatomy structure, thus the adapted head has the appearance of the input head and it is fully animated.

This chapter describes the adaptation procedure. The adaptation is achieved by Radial Basis Function (RBF), in which we propose a multi-level adaptation process to increase the shape fidelity between the input data and our deformed template model. During the iterative process we proposed a curvature-based feature points searching scheme, which works fine in our system.

4.2 Basic form of RBF Function

Radial Basis Function (RBF) has been proved very useful in surface reconstruction from scattered, noisy, incomplete data. A general RBF equation is defined as

$$f(x) = \sum_{i=1}^{m} c_i \phi(||x - p_i||) \tag{4.2.1}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

m	Number of centers
c_i	RBF weight
p_i	RBF center
ϕ	Basis function
x	N-dimension input vector
$\|\|x - p_i\|\|$	vector norm of x to p_i

The function is named as *Radial Basis* because it is a circularly-symmetric function centered at a single point p_i. The main advantages of RBFs include their compact description, ability to interpolate and approximate sparse, non-uniformly spaced data, and analytical gradient calculation.

Given $r = ||x - p_i||$, there are some very popular RBFs listed below:

Gaussian	e^{-cr^2}
Hardy Multiquadric	$\sqrt{r^2 + c^2}$
Inverse Multiquadric	$\frac{1}{\sqrt{r^2+c^2}}$
Bi-harmonic	r
Tri-harmonic	r^3
Thin plate Spline	$r^2 \log r$

4.3 Initialization of Adaptation

The adaptation problem can be formulated as follows: given a set of feature points p_i and q_i, which is the corresponding feature points of the template model and the scanned model, we want to find a transformation function so that

$$q_i = f(p_i) \tag{4.3.1}$$

Before we actually apply the adaptation, there are two steps which we need to process.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.3.1 Feature Points on Human Head

The MPEG-4 standard defines a set of parameters for calibration of the face model, which is called Facial Definition Parameters (FDPs). The parameters can be used either to modify the geometry of the model available in the decoder [71, 32], or to encode this information with the model as a priori knowledge for animation control [68, 71]. The FDPs correspond to the relevant facial features. MPEG-4 standardizes 84 feature points, which is subdivided into 10 groups based on the content they represent, Fig. 4.2 shows the position of these points. We use these points as the initial facial features on the face model.

4.3.2 Range Data Calibration

Another task is to calibrate the scanned face surface so that the calibrated scanned data will have the same orientation and proportion as to the template model. The same orientation ensures the success in semi-automatic feature points registration on both models. The same proportion makes the animation control parameters (skin thickness magnification, stiffness coefficients and damping coefficients in our case) on the adapted model easy for tuning as it will be close the the template model.

The problem can be expressed mathematically as follows. Given a template model P and a scanned data Q, if we want to fit the scanned data to the template model, each vertex q of the Q mesh should be transformed by

$$q^* = SR(q - q_c) + q_c + T \qquad q \in Q \tag{4.3.2}$$

where S is the scaling matrix, R is the rotation matrix and T is the translation vector, q_c is the rotation and scaling center of Q.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

In the above equation $T = p_c - q_c$, where p_c is the corresponding rotation center of P, so Eq. (4.3.2) becomes

$$q^* = SR(q - q_c) + p_c \tag{4.3.3}$$

Because the scanned head model is always incomplete, so it is hard to determine the exact center. We first pick 8 most obvious feature points for the further calibration, which are the bottom of head (FDP 2.1), top of head (FDP 11.1), outer eye corners (FDP 3.7, 3.12), outer lip corner (FDP 8.3, 8.4) and ears (FDP 10.9, 10.10). Denote the FDP point $u.v$ as $P_{u.v}$, the rotation center is defined as

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{P_{8.3}+P_{8.4}+P_{3.7}+P_{3.12}}{4}{}_x \\ \frac{P_{8.3}+P_{8.4}+P_{3.7}+P_{3.12}}{4}{}_y \\ \frac{P_{10.9}+P_{10.10}}{2}{}_z \end{pmatrix} \tag{4.3.4}$$

To get the orientation matrix, we compute the axis rotation matrix R_z, R_y and R_x in sequence.

Firstly, $\vec{l} = \overrightarrow{(P_{3.7} - P_{3.12}) + (P_{10.9} - P_{10.10}) + (P_{8.3} - P_{8.4})}$, is considered as the transformed x axis of the scanned mesh. We use We compute the R_z and R_y from $\vec{l}$. In experiment, we first project $\vec{l}$ into the XY plane and get R_z. After that, we rotate $\vec{l}$ with R_z so that the axis is on the XZ plane. Thus, we can get the rotation matrix R_y. Because it is hard to find a vertical axis on human face based on facial anatomy, we take $\vec{m} = \overrightarrow{(P_{3.7} + P_{3.12}) - (P_{8.3} + P_{8.4})}$ as a reference axis. The axis $\vec{m}$ is first rotated by R_z and R_y, then projected into the YZ plane. A corresponding axis $\vec{m'}$ on the template mesh is found using the same method. R_x can be computed from the internal angle between $\vec{m}$ and $\vec{m'}$.

The axis $\vec{m}$ and $\vec{m'}$ are also used as the reference axis for proportion recovery, so that we get the scale factor s. After we get all the unknowns, Eq. (4.3.3) is

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

applied on all the vertices of the scanned data, so the scanned data is calibrated to the same spatial domain as to the template head.

4.4 RBF Adaptation on Human Head

To get the transformation from FPs on template model to the scanned model. Least-square fitting method can also be employed. However, the result of least-square fitting is normally smooth everywhere which is not suitable for full-featured human head surface. Though there is approach such as Fleishman et al. [37] which can be used to construct sharp features. It is still the nature of RBF to form a local, feature-sensitive and radius-influenced adaption network. In this section we introduce two adaptation approaches using RBF network. We choose a simple regular RBF in the first approach and propose a Compactly Supported RBF (CSRBF) in the second one.

4.4.1 Approach #1: Regular RBF Adaptation

By replacing Eq. 4.3.1 with the RBF form, we can get

$$q = f(p) = \rho(p) + \sum_{i=1}^{m} c_i \phi(||p - p_i||) \tag{4.4.1}$$

The kernel of RBF is a basic function $\phi(||p-p_i||)$). The choice of basis function is based on experiment experience. Thin-plate spline is for fitting smooth functions of two variables, Gaussian function is mainly used for neural networks, the multi-quadric function is used for fitting to topographical data. The bi-harmonic function is a good choice for fitting functions of three variables and also minimizes the bending energy for the deformation [28], which is exactly what we need.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

In addition, constraints $\sum_{i=1}^{m} c_i = 0$ and $\sum_{i=1}^{m} c_i^T p_i = 0$ are used to remove the affine effect during the transformation [59, 96].

In practice, Eq. 4.4.1 becomes

$$q_i = f(p_i) = \sum_{j=1}^{m} c_j \phi(||p_i - p_j||) + Rp_i + t \tag{4.4.2}$$

where m is the number of the landmarks, $R \in R^{3\times3}$ and t are the parameters of $\rho(x)$.

By denoting B, P and Q as follows,

$$\begin{aligned} B &= (q_1 \quad \cdots \quad q_m \quad 0 \quad 0 \quad 0 \quad 0)^T \in R^{(m+4)\times3} \\ P &= \begin{pmatrix} \phi(||p_1 - p_1||) & \cdots & \phi(||p_1 - p_m||) \\ \vdots & \ddots & \vdots \\ \phi(||p_m - p_1||) & \cdots & \phi(||p_m - p_m||) \end{pmatrix} \\ Q &= \begin{pmatrix} p_1^T & 1 \\ \vdots & \vdots \\ p_m^T & 1 \end{pmatrix} \in R^{m\times4} \end{aligned}$$

we get a linear system of equations in the form of $\Lambda X = B$ with

$$\begin{aligned} A &= \begin{pmatrix} P & Q \\ Q^T & 0 \end{pmatrix} \in R^{(m+4)\times(m+4)} \\ X &= (c_1 \quad \cdots \quad c_m \quad R \quad t)^T \in R^{(m+4)\times3} \end{aligned}$$

Since error exists during the multiple view scanned data registration, we add an error coefficient ρ to reduce the scatter effect of noisy data. The bigger the ρ, the smoother the result of adaptation will become. But the detail of the face will be reduced.

Adding the error coefficient, the matrix A becomes A^*, where

$$A^* = \begin{pmatrix} P - \rho I & Q \\ Q^T & 0 \end{pmatrix} \in R^{(m+4)\times(m+4)} \tag{4.4.3}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Solving this linear system, we put all the vertices of the template model into the function, then generate the new position of the point.

$$q = f(p) = \sum_{j=1}^{m} c_j \phi(||p - p_j||) + Rp + t \tag{4.4.4}$$

4.4.2 Approach #2: Compactly Supported RBF Adaptation

The approach #1 works in our system, but we still want a better result for the local details of our model. Simply increasing the corresponding number of points does not solve the problem and the error during the corresponding points registration will lead to more manual work, so we propose our second approach of our adaptation method, which is based on CSRBF. The Eq. 4.3.1 now becomes

$$q_i = f(p_i) = \sum_{j=1}^{m} \Psi(p_i) = \sum_{j=1}^{m} (g_j(p_i)I + c_j)\phi_\sigma(||p_i - p_j||) \tag{4.4.5}$$

where m is the number of the feature points, I is the identity matrix, $\phi_\sigma(r) = \phi(r/\sigma)$, $\phi(r) = (1-r)_+^4(4r+1)$ is the Wendlands compactly supported RBF [118], where

$$(1-r)_+^4 = \begin{cases} (1-r)^4 & if\ 0 \le r \le 1 \\ 0 & elsewhere \end{cases} \tag{4.4.6}$$

σ is its support size, and $g_j(x)$ and $c_j \in R^3$ are the unknown functions and coefficients we need to solve. The functions $g_j(x)$ and c_j are solved in the following two-step procedure.

- At each point p_i we define a function $g_i(x)$ such that its zero level-set $g_i(x) = 0$ approximates the shape of scanned data in a small vicinity of p_i.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

- We determine the coefficients c_i from Eq. 4.4.5

$$\sum_{j=1}^{m} c_j \phi_\sigma(||p_i - p_j||) = q_i - \sum_{j=1}^{m} g_j(p_i) I \phi_\sigma(||p_i - p_j||) \tag{4.4.7}$$

Eq. 4.4.7 leads to sparse linear equations with respect to c_j.

To solve the unknown function $g_i(x)$, for each point p_i, we determine a local orthogonal coordinate system (u, v, w) with the origin of coordinates at p_i such that the plane (u, v) is orthogonal to the normal of p_i and the positive direction of w coincides with the direction of the normal. We approximate the template in the vicinity of p_i by a quadric

$$w = h(u, v) \equiv Au^2 + 2Buv + Cv^2 \tag{4.4.8}$$

where the coefficients A, B and C are determined via the following least-squares minimization

$$\sum_{j=1}^{m} \phi_\sigma(||p_j - p_i||)(w_j - h(u_j, v_j))^2 \rightarrow min \tag{4.4.9}$$

After we get A, B and C, we can set

$$g(x) = w - h(u, v) \tag{4.4.10}$$

The support size σ describes the level of detail we want to get from the adaptation, the bigger σ is, the better the template will fit the global shape of the incomplete scanned data, but it will obviously slow down the adaptation speed and requires more iterative steps of adaptation.

We use the preconditioned bi-conjugate gradient method [99] with an initial guess $c_j = 0$ to solve linear equations of Eq. 4.4.7.

Notice that we can also add the unknown R and t used in approach #1.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.5 Multi-level Adaptation

Based on Section 4.3.1, we choose the FDPs as the initial feature points in our adaptation procedure. But for detailed head feature adaptation, the small number of feature points is not enough. Therefore an iterative multi-level adaptation procedure with increasing FPs is proposed. Manually specifying too many corresponding features is tedious and impractical, thus automatic FP pair registration is introduced in this section. We also describe how to get a coarse-to-fine result using the CSRBF approach by dynamically setting the radius coefficient σ in this section.

4.5.1 Feature Mesh Expansion

The task here is to find more FP pairs between the two models, the k^{th}-level new point set should be merged into the $k-1$ level point set. A feature mesh method is proposed in Kähler et al. [59]. Basically the idea is that the existing feature points build up a triangle mesh, which is called feature mesh. In each iterative step, each triangle in the mesh is subdivided at its barycenter, a ray is cast at this point along the normal of the triangle to intersect with the surface of both the source data and the target data. So a new pair of FPs are found and 3 new feature triangles are created by splitting the triangle at the barycenter. Fig. 4.3 shows the feature mesh we used and the first adaptation result using RBF approach in our system.

4.5.2 Curvature-based FP Searching

The feature mesh method can be considered as an average surface data sampling technique since it samples surface data according to its own topological structure.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

But if a specific region of the face is not covered by the feature mesh, then the shape of this area is not controlled by any local points in this area, which means the feature mesh should always be carefully defined. On the other hand, average sampling means that all the regions are the same to the feature mesh, detailed information is only obtainable by increasing the mesh subdivision count, which is not so useful to specific features in minor region, for example, the boundary of the eyes.

We solve this problem by analyzing the properties of the scanned mesh itself and propose a mean curvature based feature searching scheme. The curvature is an important property in geometry. We use the mean curvature as metric for FP searching because:

- In 2D case, the value of curvature represents the inverse of radius of osculating circle at a specific point of the curve.
- Considering it in 3D situation, the curve becomes surface. There are two principal directions at any point on the surface, where the values of curvatures at such directions are *maximal* and *minimal*.
- The two principal directions are perpendicular to each other.
- If we consider the mean curvature $\overline{\kappa} = \frac{\kappa_{max}+\kappa_{min}}{2}$, The *bigger* value of $\overline{\kappa}$, the *smaller* the sum of radius of osculating circles at two principal directions.
- Position with small radius of osculating circle on the surface can be considered as representative point.

For a triangle mesh, Meyer et al. [83] has proposed a solution, the property of each vertex can be considered as spatial average around this vertex. Thus, by using

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

this spatial average, they extend the definition of curvature for a discrete mesh. We use this method for our purpose, the basic idea is explained in Section 4.5.2.1.

To show the validity of our method, we tested the approach not only on one specific scanned data, but also on our template head. We can see that in Fig. 4.4a, vertices with the largest mean curvature congregate in the area of facial features. It should be noted that the largest curvature occurs at the boundary of the model when we apply this method to the scanned data, In Fig. 4.4c, the top region of the head shows the problem. But this can be easily solved by a simple bounding box method or some boundary detection technique. In Fig. 4.4c we apply a bounding box from left to right eye horizontally and from top of the head to the bottom of the mouth vertically, we can see facial features such as eyes, nose and lips are filled with the newly-detected vertices. Given a vertex, we take it as a new feature point on the scanned data and searching the corresponding points on the template model using the ray-surface intersection method along the vertex normal.

4.5.2.1 Mean Curvature Computation on Discrete Mesh

For a vertex whose neighbor triangles are all non-obtuse, the Voronoi region is defined as:

$$\mathcal{A}_{Voronoi} = \frac{1}{8} \sum_{j \in \mathcal{N}_1(i)} (\cot \alpha_{ij} + \cot \beta_{ij}) ||x_i - x_j||^2 \tag{4.5.1}$$

Where $\mathcal{N}_1(i)$ is the 1-ring neighborhood of the vertex, see Fig. 4.5.

Extending this to arbitrary mesh, a new surface area for each vertex x, denoted $\mathcal{A}_{Mixed}$, is defined, see Fig. 4.6.

Now we can calculate the mean curvature normal operator $K(x_i)$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

$$K(x_i) = \frac{1}{\mathcal{A}_{Mixed}} \sum_{j \in \mathcal{N}_i} (\cot \alpha_{ij} + \cot \beta_{ij})(x_i - x_j) \tag{4.5.2}$$

We can easily get the mean curvature value κ_H by taking half of the magnitude of Eq. (4.5.2).

4.5.3 Hierarchical Adaptation

Obtaining the new set of FP pair on both the scanned data and template model in Section 4.5.1 and Section 4.5.2, we can use all these corresponding points in the single-level adaptation for adaptation approach #1. We can also use these points in a hierarchical fashion for the adaptation approach #2 which we described in Section 4.4.2.

After we obtain the point set p_i^k and q_i^k in k^{th} level, we can recursively determine the transformation function, from Eq. 4.4.5 we get

$$q_i^0 = f^0(p_i^0) \tag{4.5.3}$$

where $f^0(x)$ is the first function we have solved with the initial corresponding points. The k^{th} $(k = 1, 2, \ldots)$ level function is trained as follows

$$q_i^k = f^k(p_i^k) = f^{k-1}(p_i^k) + o^k(p_i^k) \tag{4.5.4}$$

and $o^k(x)$ is called the offsetting function

$$o^k(x) = \sum_{j=1}^{m^k} [g_j^k(x) I + c_j^k] \phi_{\sigma^k}(||x - p_i^k||) \tag{4.5.5}$$

$o^k(x)$ has the form used in single-level adaptation, the local approximation $g_j^k(x)$ is determined by least square fitting and the coefficients c_j^k are solved by

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

the following linear equations using the same preconditioned biconjugate gradient method [99]

$$o^k(p_i) = q_i - f^{k-1}(p_i) \tag{4.5.6}$$

The support size σ^k is defined by

$$\sigma^k = \frac{\sigma^{k-1}}{2} \tag{4.5.7}$$

4.5.4 Comparison between CSRBF and RBF Approach

There are several pros and cons between CSRBF and RBF function. The main advantage of RBF is that it supports global adaptation, this feature is quite useful when the number of feature points is low compared to the number of vertices of target mesh. Thus, at the beginning stage of adaptation, RBF is simpler and much more useful than CSRBF, though CSRBF can still be used in such situation if the radius coefficient σ is set large enough. But once the density of FPs gets high, we consider using CSRBF alone for feature adaptation. Our template model contains 3349 vertices, so in the experiment we set 500 as the threshold between low density and high density of FPs. It is also unreasonable to consider the FP on the forehead will influence the vertices on lips. Another problem is about the computational complexity. With the increasing of number of FPs, the size of the matrix becomes bigger. In RBF case, because of the global feature, each element in the matrix is not zero, which means in each training iteration step, we need to solve a high-order non-sparse linear equation system. A LU Decomposition costs $O(N^3)$ operations. To solve a system of linear equations $AX = B$ where $A = LU$ costs $O(N^2)$ operations. Instead, the nature of CSRBF enables a sparse linear system,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

which reduces the computational time and complexity. We present the adaptation results of RBF and CSRBF in Fig. 4.7. The top row is the adaptation result of Combination approach and the bottom row is the RBF approach. The first two adaptation results are the same because the number of FPs has not exceeded 500.

From Fig. 4.7 it can be seen that both of the approach get the same global shape. But from the side view shown in Fig. 4.8 we can notice the feature difference at the nose bridge.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

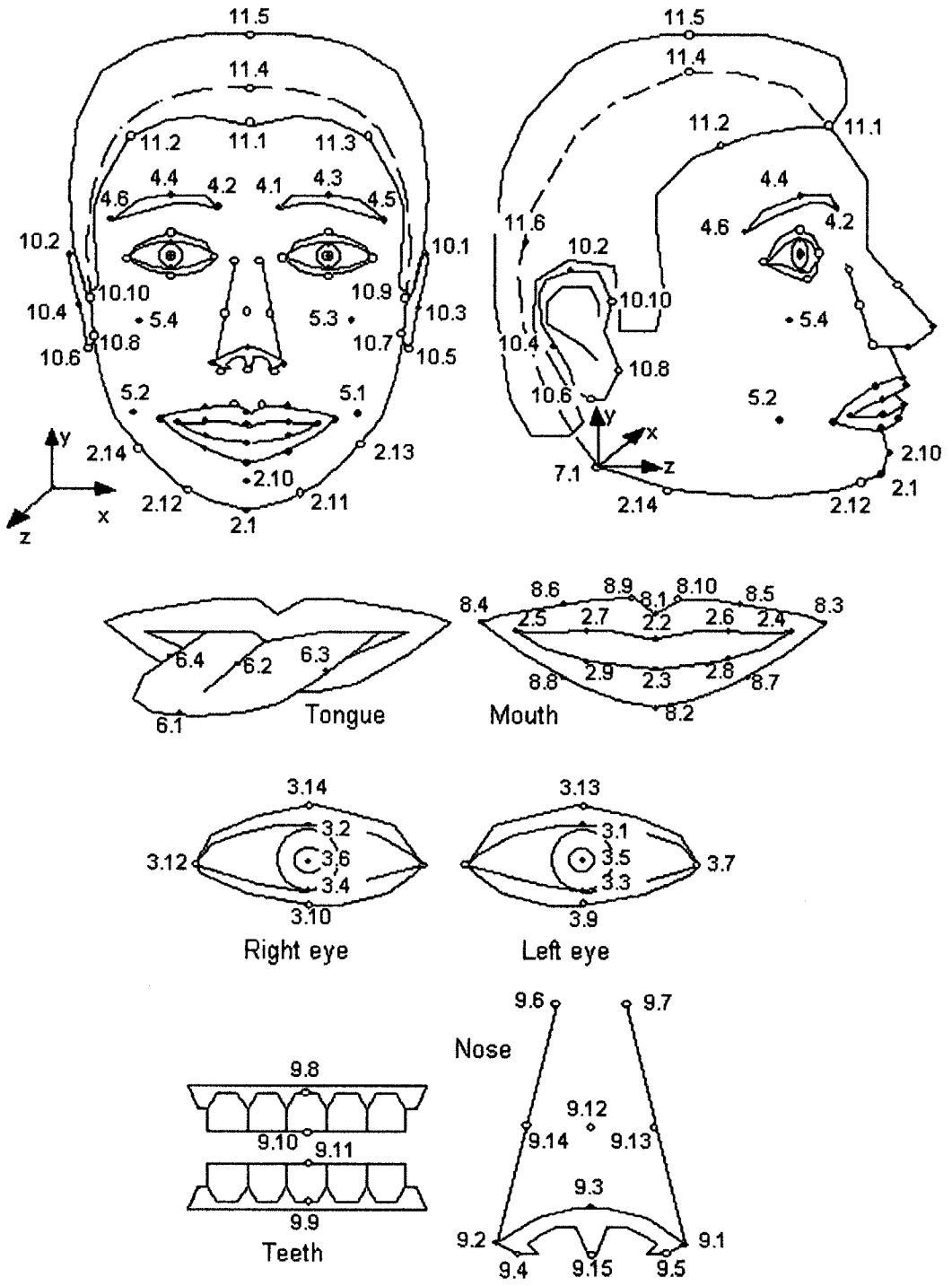


Figure 4.2: MPEG-4 Facial Definition Parameters

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

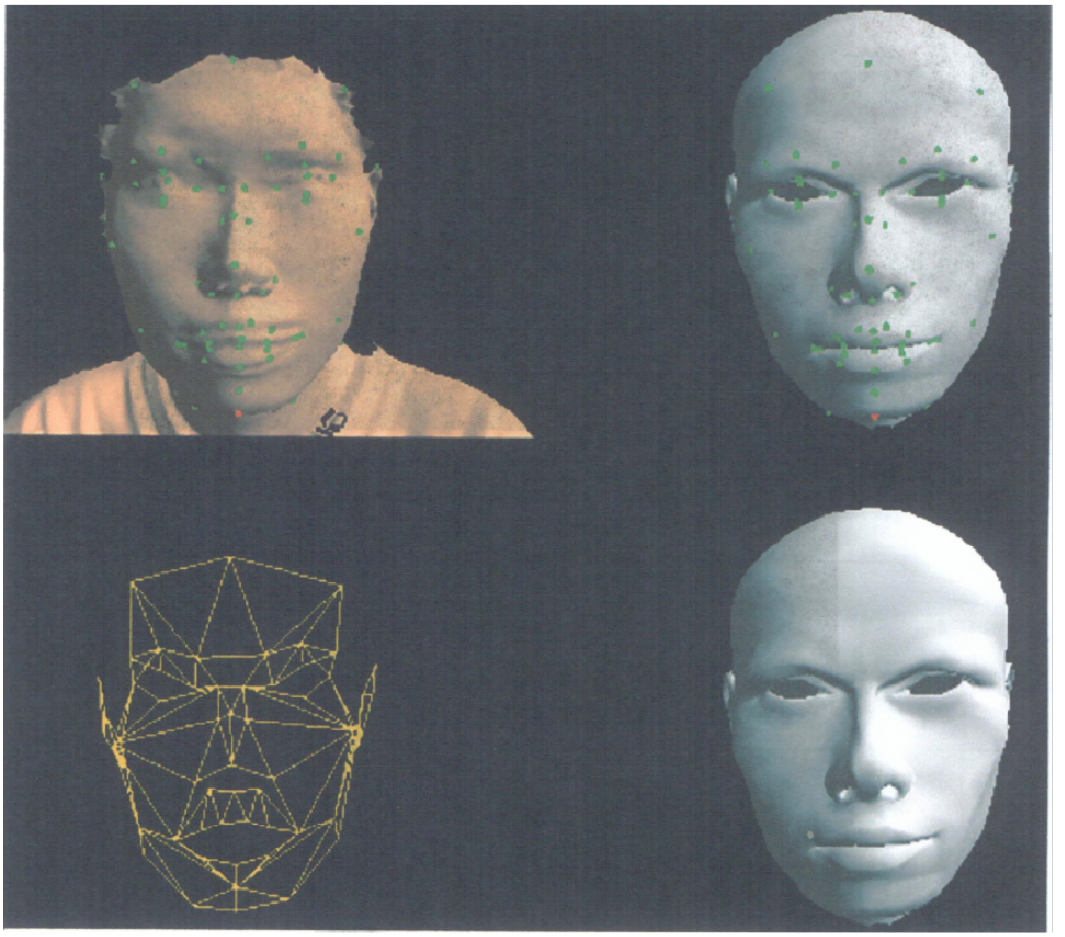


Figure 4.3: Initial adaptation using approach #1 with subset of MPEG-4 feature points. top-left: the input scanned data; top-right: the adapted head model; bottom-left: the initial feature mesh; bottom-right: the head model for test in the iterative process.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

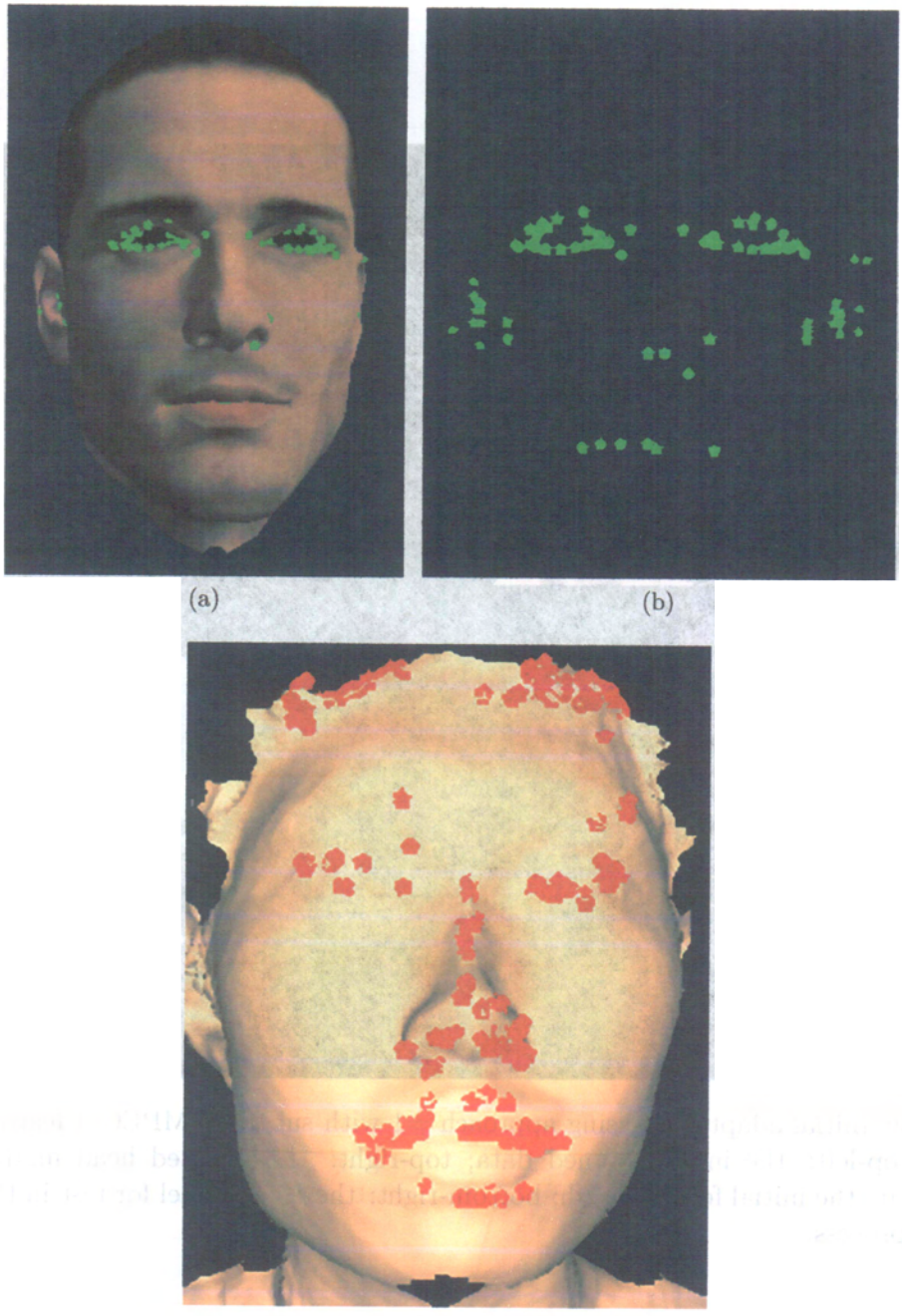


Figure 4.4: Curvature-based point searching

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

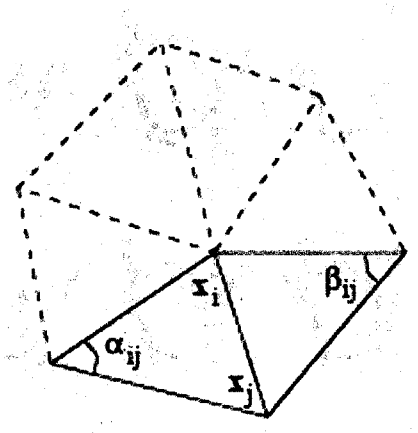


Figure 4.5: 1-ring neighbors and angles opposite to an edge

$\mathcal{A}_{Mixed} = 0$

for each triangle T from the neighborhood of x

if T is non-obtuse // Voronoi safe

//add Voronoi formula (see equation4.5.1)

$\mathcal{A}_{Mixed}+=$ Voronoi region of x in T

else // Voronoi inappropriate

//add either area(T)/4 or area(T)/2

if the angle of T at x is obtuse

$\mathcal{A}_{Mixed}+=$ area(T)/2

else

$\mathcal{A}_{Mixed}+=$ area(T)/4

Figure 4.6: Pseudo-code for region $\mathcal{A}_{Mixed}$ in arbitrary mesh

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

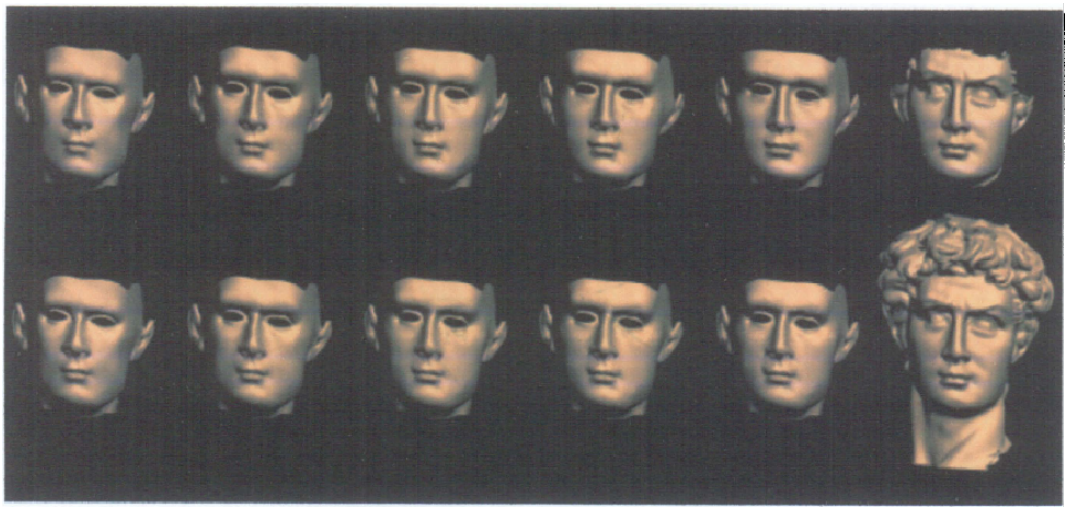


Figure 4.7: Comparison of adaptation results between combination approach and RBF approach. Top row: Combination approach; Bottom row: RBF approach.

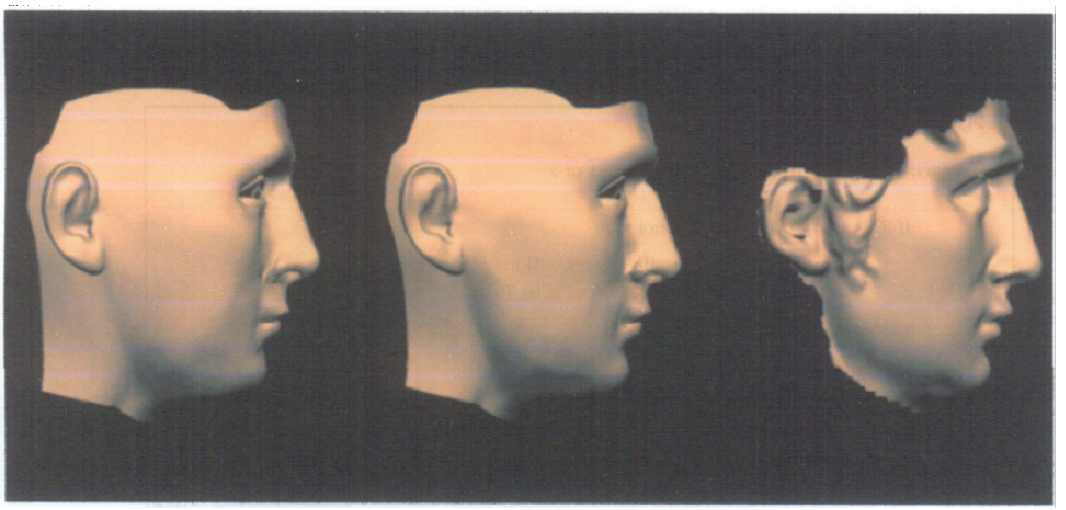


Figure 4.8: Side view comparison of adaptation results between combination approach and RBF approach. Left: RBF approach; Middle: Combination approach; Right: Scanned data.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 5

Measurement of Performance of Template-based Adaptation

5.1 Introduction

In this chapter we present error metrics to estimate the similarity between the adapted model and scanned range data. The sensitivity of our method to the defective data is also discussed in this chapter. The error metrics validate the usefulness of RBF-based coarse-to-fine approach.

In Chapter 4 we have described the procedure to personalize a template head from the range scanned data. Our approach applies to the following problems caused by directly using the range scanned data:

- Because of human breathing, the head keeps moving during the scanning procedure. The small movements cause the blur result in the range scanned data.
- Because of the lighting condition, limited view range and merging from different views, holes and gaps may appear on the scanned data.
- Because of the color of hair and eyes, these parts cannot be captured by laser

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

scanner.

During the adaptation, there should be a method to measure how close the synthetic face is to a real human face. We have designed error metrics to measure the difference between the synthetic head and the target scanned face. We also discuss several factors that affect the adaptation results in this chapter.

5.2 Error Estimation of Adaptation

The whole adaptation is an iterative process and we want to optimize the result. Thus, we evaluate the quality of the adaptation using two error functions: (1) *Distance error* and (2) *Marker error.*

5.2.1 Distance Error

The first criterion is that the adapted template surface should be as close enough as possible to the input surface. Since the vertex of the template and the one of the target model is not one-on-one mapping, we define a data objective term E_d as the square of the minimum distance from the vertex of template to the scanned range data.

$$E_d = \frac{\sum_{i=1}^{n} w_i dist^2(x_i, Q^*)}{n} \tag{5.2.1}$$

where n is the number of vertices of the deformed template, x_i represents one of the vertices, Q^* is the calibrated scanned mesh, w_i is the weight term to control the influence of the data in different regions.

To compute the distance from x_i to Q^*, the problem becomes compute the distance from the point x_i to its projection $x_{i,j}^*$ on triangle T_j^* in Q^* [29], where

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

$x^*_{i,j}$ is in the boundary of triangle T^*_j. There is chance that the ray starting from x_i along with its normal intersect more than one time with the scanned data. However, only the minimum distance is recorded. A constraint that the nearest distance does not exceed 1/10 of the max width of the template head is also added to avoid wrong registration.

The weight term w_i is useful when there are holes and areas with poor quality on the scanned model (eg. the ears). If one vertex x_i cannot match any of the surface of the scanned data, which means there is a hole, we set the w_i to be zero. In areas with low quality, we provide interactive tools for the user to specify areas and the influence coefficient (eg. in our experiment, w_i of the vertices on the ears is set to be 0.2 due to the low quality in this area). This makes the distance error estimation to be fair enough.

5.2.2 Marker Error

The distance error is capable of estimating the similarity of the two models. However, sometimes we want to estimate the corresponding relationship between the two models, in that case we place some corresponding markers of recognizable features on both the template and the target mesh. These markers will not participate in the training process of the adaptation, but we can compute the distance between them to check if the transformation makes the markers getting closer or not. The marker error E_m is represented as

$$E_m = \sum_{i=1}^{m} ||u_i - v_i||^2 \tag{5.2.2}$$

where u_i and v_i are corresponding markers on the template and the target model, and m is the number of markers.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

5.2.3 Combined error

Our complete objective function E is the weighted sum of the two functions:

$$E = \alpha E_d + \beta E_m \tag{5.2.3}$$

Specifying corresponding markers on both models is usually not accurate, generally the weight of the distance function should be higher than the one of the marker function. The summed E is computed in each iterative procedure, and when we get a local minimum E, the adaptation is complete.

5.2.4 Adaptation Results with Error Estimation

We display 2 adaptation results from two different people in Fig. 5.1 and Fig. 5.2 to validate our approach. The error estimation results are provided in Table 1 and Table 2, count 0 presents the distance between original template and scanned data.

Figure 5.1: Adaptation Result: the right end is the original scanned data.

5.3 Scanned Data Error

The RBF adaptation is basically an interpolation method. The final adaptation results depend on the successful rate of feature point pair registration on both mod-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Table 5.1: Error estimation of the adaptation process shown in Fig. 5.1

Iterative count	Distance Error	Marker Error	Combined Error
0	5.98340	27.78620	8.76960
1	1.28463	1.30477	1.41511
2	1.16637	0.57937	1.22431
3	1.07375	0.46750	1.12050
4	1.00459	0.477152	1.05231
5	0.92843	0.480015	0.97643

Figure 5.2: Another adaptation result

Table 5.2: Error estimation of the adaptation process shown in Fig. 5.2

Iterative count	Distance Error	Marker Error	Combined Error
0	9.47471	44.91340	13.96605
1	3.27452	2.13032	3.48755
2	2.66195	1.63096	2.82505
3	1.72435	1.40743	1.86509
4	1.05364	1.40157	1.09380
5	0.85435	1.40288	0.99464

els. Thus the qualities of both models determine the final adaptation results. The quality of template mesh is normally controllable compared to the noisy scanned data from real human being. In this section, we analyze the influence of defective scanned data on the final adaptation result.

5.3.1 Few Errors

Due to the light condition and muscle movements of human face, there are always blurry noise on the scanned face. Complex features such as nostril and ear also

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

cause holes and gaps. Therefore, it is very difficult to get a perfect model on the human face using a laser scanner. We show the original scanned data from 3 different views in Fig. 5.3. We notice that the scanned data have noise on its cheeks, ears and nostrils, but the noise does not affect the initial registration of FPs. The adaptation result is shown in Fig. 5.4. We also present the nose area feature of final adaptation result from Fig. 5.4 in Fig. 5.5. To compare the features from different model, another adapted nose from a statue model is also included in Fig. 5.5. The error measure results is listed in Table 5.3.

Figure 5.3: Original Scanned Model

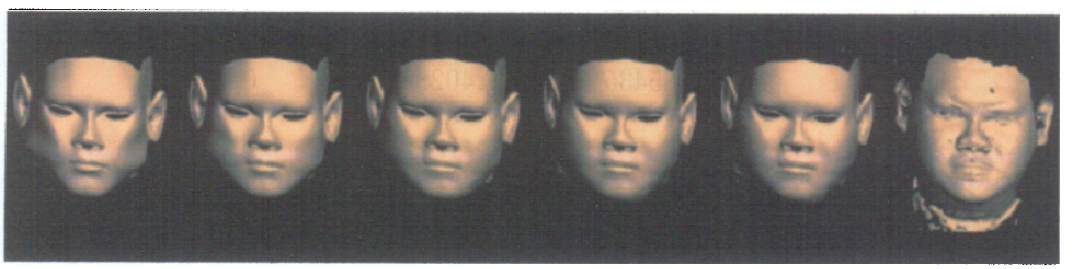


Figure 5.4: Adaptation results from scanned data with small holes.

5.3.2 Hopeless Scanned Data

Very bad scanned data is normally unacceptable. To show the result, we still use the same scan model but manually add a lot of holes on the mesh. The

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

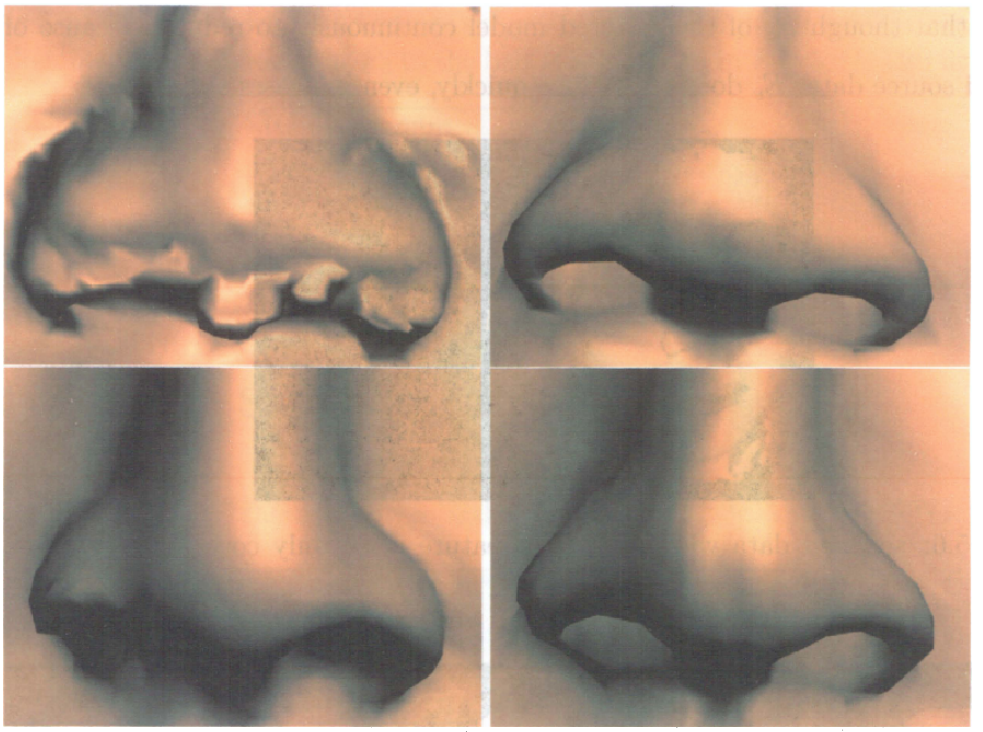


Figure 5.5: Correctly adapted nose. Left: scan data; Right: adapted result; top: real human face; bottom: statue model.

Table 5.3: Error estimation of the adaptation process shown in Fig. 5.4

Iterative count	Distance Error	Marker Error	Combined Error
1	12.37691	23.82554	14.75946
2	3.58563	2.01921	3.78755
3	2.34867	1.40874	2.48954
4	1.45029	1.20338	1.57063
5	0.83574	1.15396	0.95114

modified artificially hopeless mesh can be seen is Fig. 5.6. The adaptation results are presented in Fig. 5.7. We can notice that due to the lack of information on the right cheek and right side of the nose, these parts do not adapt correctly when compared to Fig. 5.4. A close look at the nose can be seen in Fig. 5.8. Vertices by vertices distance error E_v between adapted model in Fig. 5.7 and Fig. 5.4 is listed in Table. 5.4, the distance error E_d to the bad scanned data in Fig. 5.6 is also listed.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Notice that though E_d of the adapted model continuously to reduce, because of the bad source data, E_v does not reduce quickly, even increases somehow.

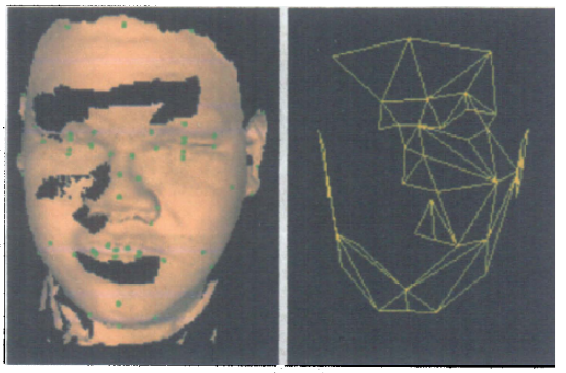


Figure 5.6: scanned data with holes, the feature mesh only cover partial of the region

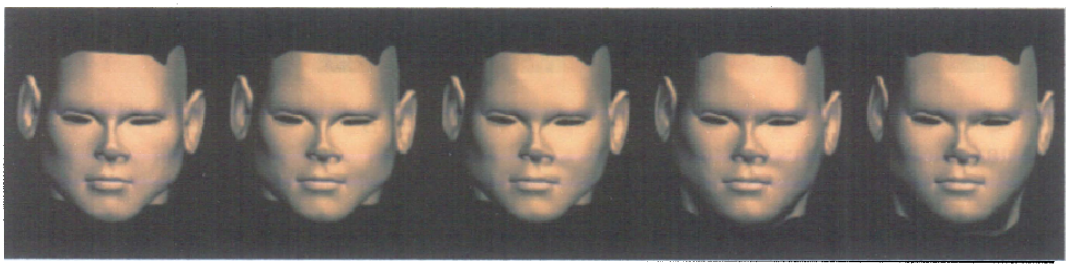


Figure 5.7: Adaptation with bad scanned data (Right side of the cheek and nose are not correct).

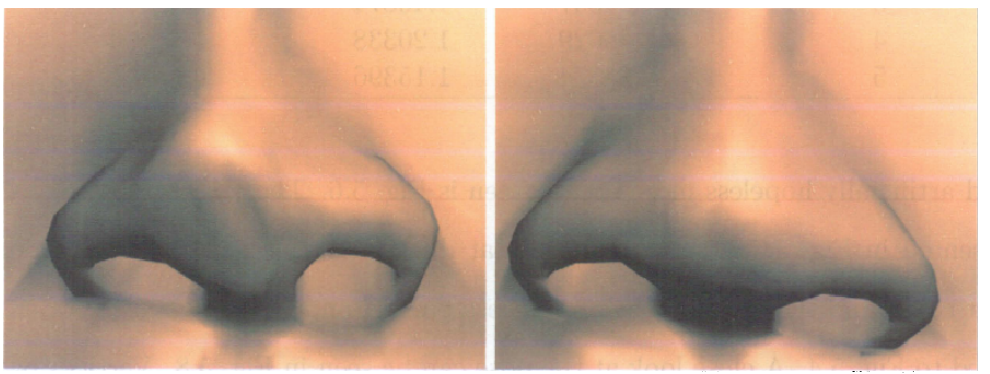


Figure 5.8: Close look of adapted result from model with large holes shown in the left compared to the adapted from model without holes (right).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Table 5.4: Vertices by vertices distance error E_v between adapted model in Fig. 5.7 and Fig. 5.4, the distance error E_d to the bad scanned data in Fig. 5.6 is also listed.

Iterative count	E_v	E_d
1	0.50893	3.88023
2	0.31167	3.03789
3	0.29328	2.39156
4	0.25176	1.52077
5	0.27487	0.92901

5.4 FPs on Scanned Data

Based on Section 4.5, there are two searching schemes for FP expansion. The first one uses a feature mesh defined by the small number of initial FPs, the second one uses a curvature-based metric to search for the most representative point on the surface. Both methods use ray-surface intersection strategy to register the new pair of FPs on source and target meshes. Because of the principle of RBF, the successful rate of FP registration also directly affects the adaptation result.

5.4.1 Missing FPs

Sometimes initial FP set cannot be defined on the scanned data due to the holes. In this case, the feature triangles related to the unregistered FPs are removed from the feature mesh. As a result, the area covered by these feature triangles will not adapt correctly using feature mesh expansion scheme. In this situation, if the missing FPs are crucial ones, the curvature based searching scheme introduced in Section 4.5.2 will help to compensate the FP registration. In Fig. 5.9 the FPs on the nose are removed from the scanned model to show the case when holes happen on crucial features on the scanned data. The adaptation results are presented in Fig. 5.10. Similarly, we compare the nose area between the correct adapted data and the adapted result without FPs registered on nose area. It can be noticed from

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Fig. 5.11 that the shape of the nose still remains correct without twist happening in Fig. 5.8. The scale is smaller compared to the correct adapted model because of the less number of FPs in this area. Vertices by vertices distance error E_v between adapted model in Fig. 5.10 and Fig. 5.4 is listed in Table. 5.5, the distance error E_d to the scanned data in Fig. 5.3 is also listed. Since there is no lost information, E_v and E_d both continuously decrease, though the speed of this adaptation is slower than the adaptation with complete FP registration.

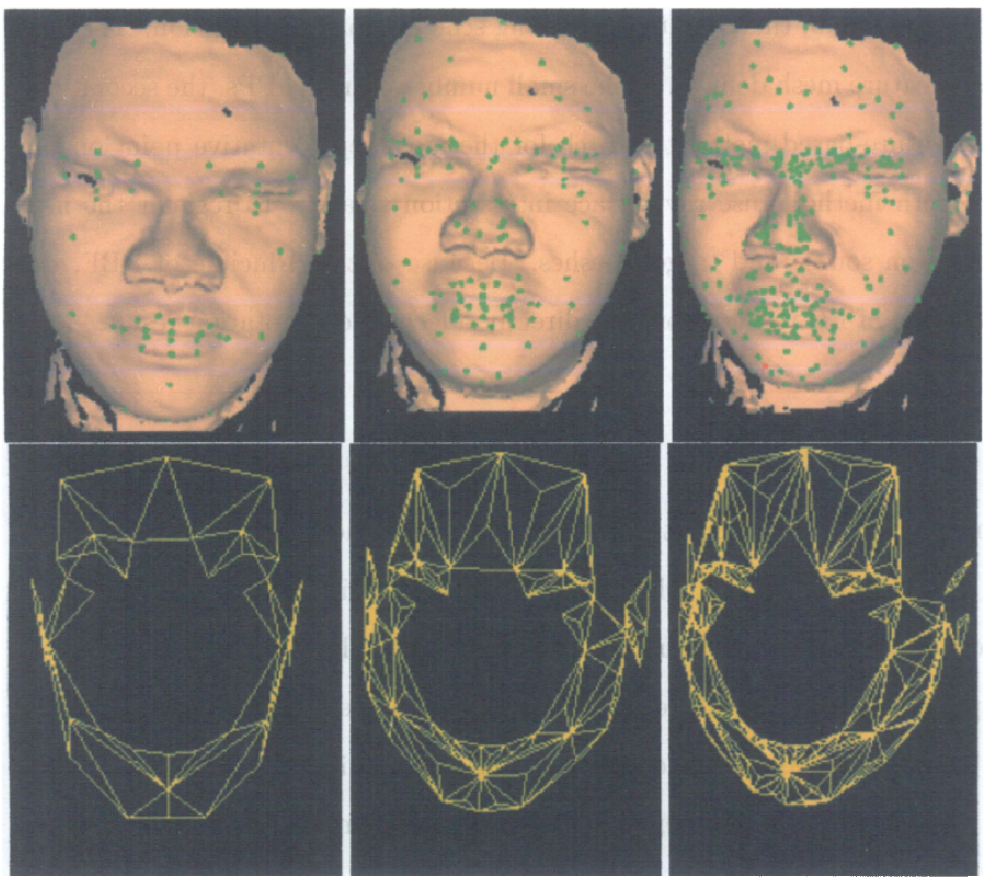


Figure 5.9: The FPs on the nose are removed from the scanned model to show the case when holes happen on crucial features on the scanned data.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

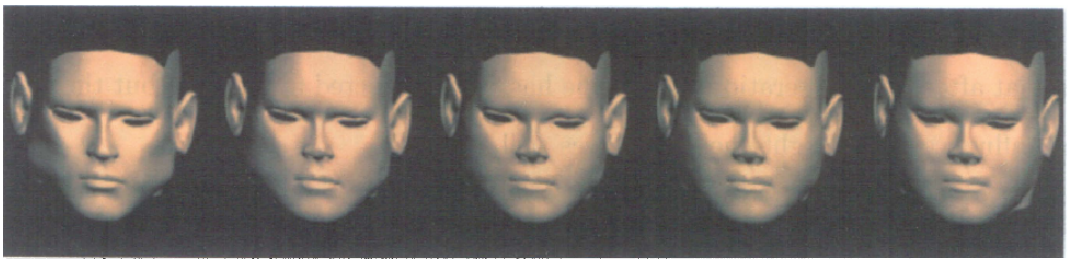


Figure 5.10: Adaptation results from model shown in Fig. 5.3 with crucial FPs missing initially (No noticeable error). The

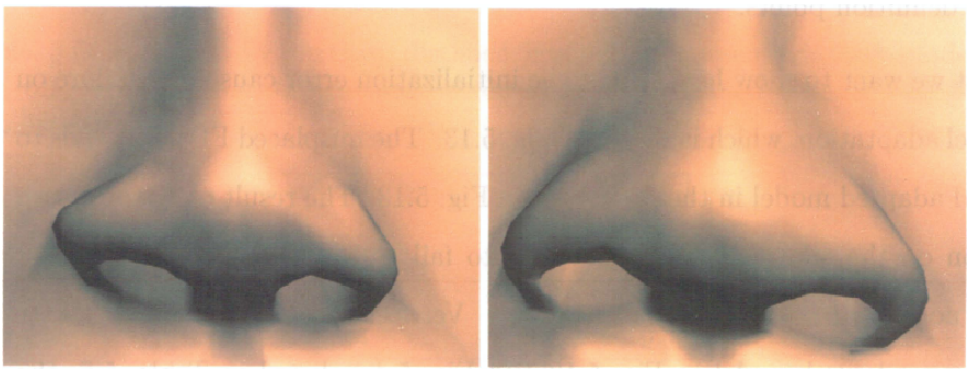


Figure 5.11: Close look at nose region. Left: crucial FPs missing initially; right: adapted nose from without any missing FPs.

Table 5.5: Vertices by vertices distance error E_v between adapted model in Fig. 5.10 and Fig. 5.4, the distance error E_d to the scanned data in Fig. 5.3 is also listed.

Iterative count	E_v	E_d
1	0.72474	3.76162
2	0.53256	3.04127
3	0.41675	2.45386
4	0.32138	1.80473
5	0.25760	1.36453

5.4.2 Misplaced FPs

In another case, misplaced FPs cause the failure of adaptation. This is because wrong weights would be obtained from the training procedure. Once these weights are applied to the vertices of template model, it leads to incorrect displacement.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Fig. 5.12 shows an example of bad FP registration. The results in Fig. 5.13 indicates that after several iterations the shape has been corrected a little bit, but the initial influence at the right cheek and eyes still exists.

In Fig. 5.12 we use the red circle to indicate the misplaced FP pairs. For example, on the cheek, the FP 5.3 and 5.4 on the scanned model are closer to the nose than these points on the template model. The same thing happens on the eyebrow definition points.

What we want to show here is that the initialization error causes the failure on the model adaptation, which is shown in Fig. 5.13. The misplaced FP pairs leads to the failed adapted model in the left image in Fig. 5.13. The result of feature mesh expansion on the adapted model continues to fail and after iterations, artifacts are still noticeable on the adaptation result. Vertices by vertices distance error E_v between adapted model in Fig. 5.13 and Fig. 5.4 is listed in Table. 5.6, the distance error E_d to the scanned data in Fig. 5.3 is also listed. Because E_d is the sum of minimum distance from adapted model to the normalized scanned model, it always reduce during the adaptation. However, is the case of misplacing FPs, the E_v remains at a high level.

Figure 5.12: FP initialization error.

How is minor FP errors influence our results? In our test, we collect 20 sets

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

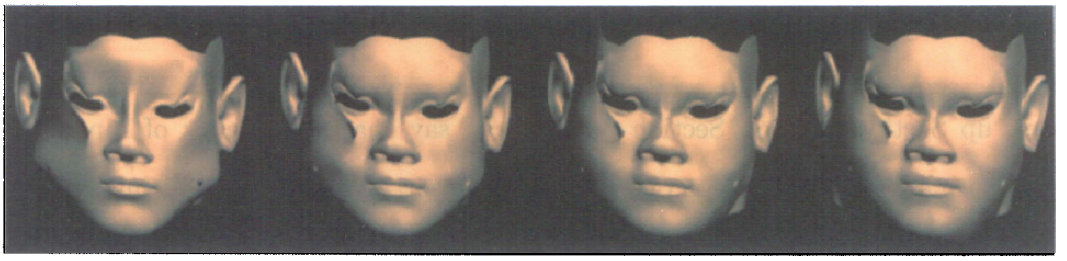


Figure 5.13: Adaptation results with bad FP initialization (Artifacts can be seen in the cheek & eyes).

Table 5.6: Vertices by vertices distance error E_v between adapted model in Fig. 5.13 and Fig. 5.4, the distance error E_d to the scanned data in Fig. 5.3 is also listed.

Iterative count	E_v	E_d
1	1.35675	5.47411
2	0.94273	3.93518
3	0.75998	2.65632
4	0.76822	1.92807

of FPs with standard deviation $\sigma_{E_{FP}}$ on the same model and calculate standard deviation σ_{E_v} and σ_{E_d}. The results is shown in Table 5.7. It can be concluded that when the registration of FPs follows the feature definition on the human face, the difference between results are small.

Table 5.7: Measurement on variance of adapted models from 20 sets of initial FPs with $\sigma_{E_{FP}} = 0.03291$ on the same model.

Iterative count	σ_{E_v}	σ_{E_d}
1	0.08090	0.22011
2	0.07252	0.15252
3	0.07438	0.09429
4	0.06837	0.04185
5	0.07015	0.02388

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

5.5 Summary

To sum up Section 5.3 and Section 5.4, we can say that the quality of range scanned data and the registration of FPs are two important factors influencing the adaptation. It can be concluded that:

- Scanned data with less noise but correct registration on FPs will not affect the adaptation quality.
- Very bad scanned data will result in loss of information, this can not be recovered by RBF adaptation, but with correct hole filling strategy, the problem can still be solved.
- The incomplete FP initialization will affect the completion of feature mesh, but with the compensation of curvature based FP searching, feature area on the face can still be covered, so the influence is not so much.
- The error in initial FP pair registration will cause the failure of the adaptation, so these FPs should be selected precisely and carefully.

We have already shown in Section 5.2.4 that based on our hierarchical adaptation framework, our template model can get a close fit to the normalized scanned data. However, due to the resolution limitation of Minolta Vivid 700 laser scanner and the biomechanical nature of human face (continuous face surface), the geometric data from the scanner is already not accurate enough. The Minolta merge software applies a smooth filter when data from different views is merged together. So it is trivial to quantify the error introduced in the adaptation stage. Scanning from multiple views at the same time with higher resolution scanners will reduce the error introduced in the scanning stage and grant us a better comparison between the scanning target and the adapted model.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 6

High Performance Physically-based Facial Expression Synthesis

6.1 Introduction

Our adaptive template model is not just to shape a personalized head. Furthermore, it can be used for expression generation. In this chapter, we introduce the facial expression synthesizer with volumetric soft tissue structure and parametric musculature.

Starting with a single layer trianglular mesh for the human face, we generate a tetrahedral mesh by radially extruding (inside) the skin to obtain a 3D skin layer for the desired thickness. This tetrahedral mesh represents the volumetric structure of soft tissue. We choose the tetrahedron as the basic unit since it is the simplest volumetric representation in polygonal form and it provides both stability as well as efficiency.

We choose a parametric Bezier curve to represent the fiber of facial muscle. Once define these fibers, we further subdivide the muscles into segments. These segments are represented by ellipsoidal structures, which mimic the real muscu-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

lature. We also consider the physical phenomena that muscles simultaneously activate at several specific points on the face, especially at the lip and eye corners. Thus we mark several specific anchor points where the muscle insertions or intermediate control points converge. So the position of the anchor points moves according to the interpolation of its related muscle control points. In this case, the related control points of the parametric muscle fibers are displaced so that all of these muscles apply force onto the soft tissue.

We utilize the technique introduced by Teschner et al. [112] to compute the potential energy of the soft tissue, which results in the deformation of the tetrahedral mesh. The method has been proved to be stable, easy to apply and a relatively large time step can be used in the numerical integration procedure. Since this method can be applied on various types of objects, we take into account the special elastic properties of the soft tissue and add the nonlinear stiffness coefficient when computing energy of the distance preservation.

The significant benefit of our approach is that we realize the anatomical accuracy with plausible efficiency. We also combine a single muscle activation with its related muscles. Moreover, the parametric curve based muscle fiber structure provides a flexible control for synthesizing facial expressions.

6.2 Volumetric Skin Layer Creation

Our work starts with triangulated face model wherein eyes and teeth structures have been isolated, as our interest centers on the dynamic deformation occurring on facial skin. The task is therefore to create physically-based control elements. The human skin on the face is soft but incompressible. To simulate these properties, we model the skin as a tetrahedral mesh.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Table 6.1: Skin thickness on various regions

Forehead	0.901mm
Eyelid	0.569mm
Cheek	1.237mm
Chin	0.891mm

We are interested in the frontal face region where the skin mesh is assigned a region code. We limit the computation to the frontal region alone, since there is no non-rigid deformation at the back. Given an arbitrary triangle of the frontal face, we create a prism-like element by scaling along the normal of each vertex radially inside the head. In the human body, the skin thickness varies at different regions. Lee and Hwang's reported data [73] is used as our reference to set skin thickness.

A region is a collection of triangles, the scaling length of the vertex on the boundary of neighboring regions is the average thickness of these regions.

We take special care in the lips and nose regions, which are involute regions in our model. The problem of self-intersection occurs when using the method described above. But we notice that the lip is solid, so the topology of the these regions is predefined. In order to avoid the re-definition of regions for every head, the method to transfer the physical structure from one person to another proposed by Kähler et al. [59] can be used.

The prism-like element is then subdivided into tetrahedrons. We choose the tetrahedron as our basic element for the reason given below. Firstly, as a volume representation, a tetrahedron reduces the instability or degeneration problem of thin-plate representation. Secondly, compared to simpler volume representations, such as spheres, the tetrahedron provides larger degrees of freedom in deformation, but is computationally efficient compared to more complex representations, such

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

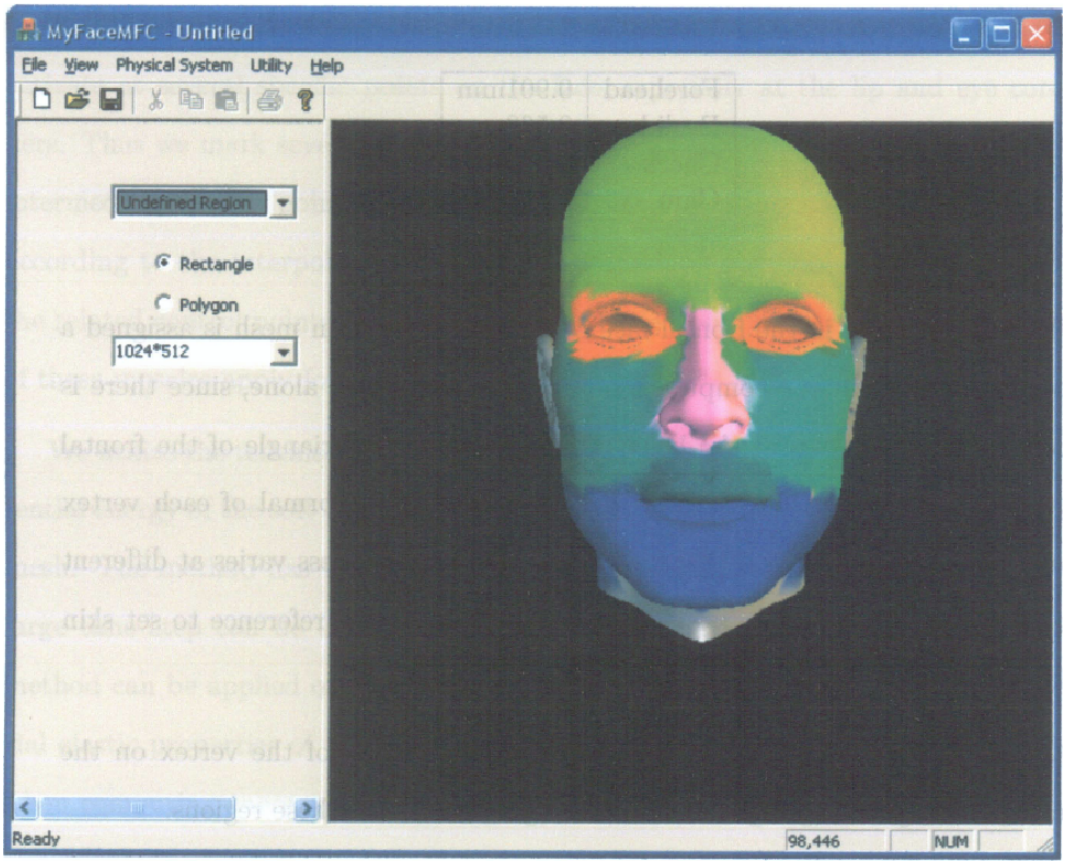


Figure 6.1: Define facial regions in GUI

as the hexahedron. Fig. 6.2 shows how we operate with any arbitrary prism-like element.

6.3 Parametric Muscle Definition

Face muscles activate human face movement. In accordance with Gray's Anatomy [45], we integrate 14 different types of muscles which are the major sources for facial expression into our system as listed in Table 6.2. The muscles are identified by numbers. Muscle no. 1 to muscle no. 12 are referred as *linear muscles* in the rest of the paper. Traditionally these muscles are defined as line segments or piecewise

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

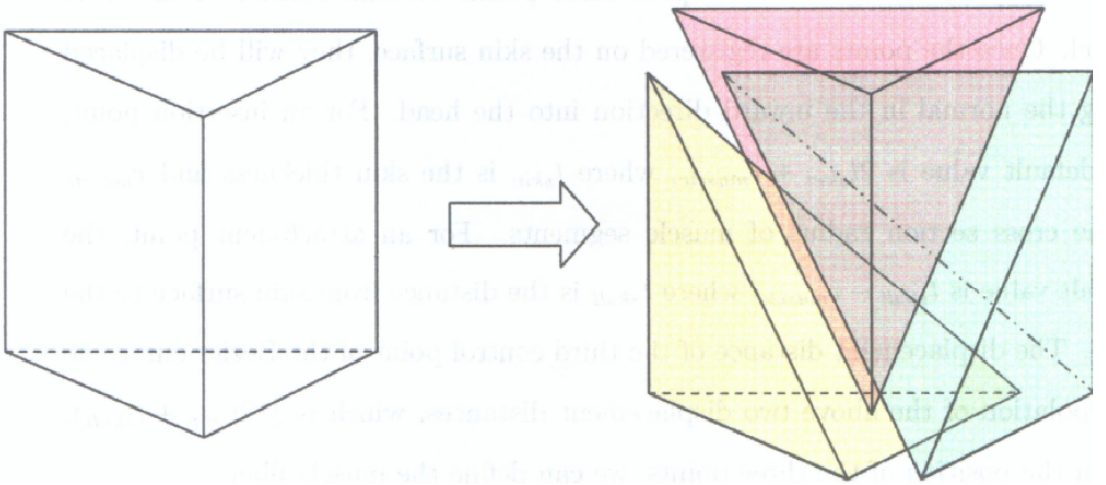


Figure 6.2: Division of prism structure into three tetrahedrons

line segments, which eases the process of computation but limits the flexibility. To simplify the deformation control of muscles, we choose Bezier curve(s) to represent all the muscles. Since there is no twist in a linear muscle, we choose a *regular* (with all weights equal to 1) quadratic Bezier curve to represent the fiber. For *Orbicularis Oculi* and *Orbicularis Oris*, we choose 4 quadratic rational Bezier curves to represent each muscle. This is because rational Bezier curves can be used to represent circles which regular Bezier curves can not.

Table 6.2: The muscles used by our system

No.	Muscle Name	No.	Muscle Name
01	Zygomatic Major	08	Frontalis Inner
02	Zygomatic Minor	09	Frontalis Major
03	Depressor Anguli Oris	10	Frontalis Outer
04	Risorius	11	Nasalis
05	Mentalis	12	Levator Anguli Oris
06	Quadratus Labii Superioris	13	Orbicularis Oris
07	Corrugator Supercilli	14	Orbicularis Oculi

A quadratic Bezier curve has 3 control points. For a linear muscle, two end points can be viewed as the insertion and attachment points. To insert a new

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

linear muscle, the user needs to pick three points on skin surface of the head model. Once the points are registered on the skin surface, they will be displaced along the normal in the inward direction into the head. For an insertion point, the default value is $2t_{skin} + r_{muscle}$, where t_{skin} is the skin thickness and r_{muscle} is the cross section radius of muscle segments. For an attachment point, the default value is $t_{skull} - r_{muscle}$, where t_{skull} is the distance from skin surface to the skull. The displacement distance of the third control point of the Bezier curve the interpolation of the above two displacement distances, which is $\frac{1}{2}(2t_{skin} + t_{skull})$. Given the position of the three points, we can define the muscle fiber.

The *Orbicularis Oculi* and *Orbicularis Oris* is defined as non-voluntary muscles, that says, they cannot actively control their contractions. In the muscle definition stage, we have to define the controls points of non-voluntary muscles first. The displacements of these control points are set equally as $\frac{1}{2}(2t_{skin} + t_{skull})$. The related insertion points of the linear muscle (voluntary muscle) will be registered too. The other control points of linear muscles (voluntary muscles) should also be registered later. The registration is done using an interactive user interface, please refer to Fig. 6.3.

After we define the muscle fiber, we consider the volume of the muscle. Given a Bezier curve, we perform an uniform subdivision. The number of segments is determined by the arc length of the curve. We subdivide the *Zygomaticus Major* into 5 segments, the number of segments of other muscles are scaled accordingly. After the subdivision, we connect the two end points of each segment as the x-axis of an ellipsoid. The direction of y-axis is the perpendicular line to the x-axis in the plane determined by three control points of the curve, z-axis direction can be determined by $x \times y$. The length of the x-axis is adjusted to be 5% more than the distance between the two points to prevent tangency between muscle segments.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

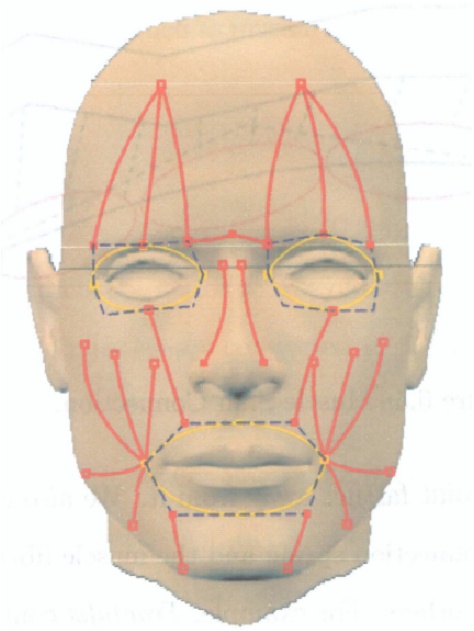


Figure 6.3: Voluntary and non-voluntary muscles are connected together at anchor points

The radius in y, z direction is normally $2d_m$, but at the two end segments of the muscle this value is scaled down based on the fact that muscle tapers at the ends.

The muscles connecting with the skin is represented by linking mass points in the inner layer after the tetrahedralization to the muscle segments on ellipsoidal surface. The search algorithm is listed in Fig. 6.4.

```
for each muscle i
    for each bottom layer mass point x in the region the muscle is related to
        if d(x, i) < λ and θ(x, i) < φ
            connect the muscle with the mass point
```

Figure 6.4: Algorithm for muscle skin connection

In Fig. 6.4, $d(x, i)$ is the distance from a mass point to its nearest surface point on the muscle segment ellipsoid. The parameter λ constrains the distance from the muscle to the skin mass point. The idea is that the muscle contraction will not

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

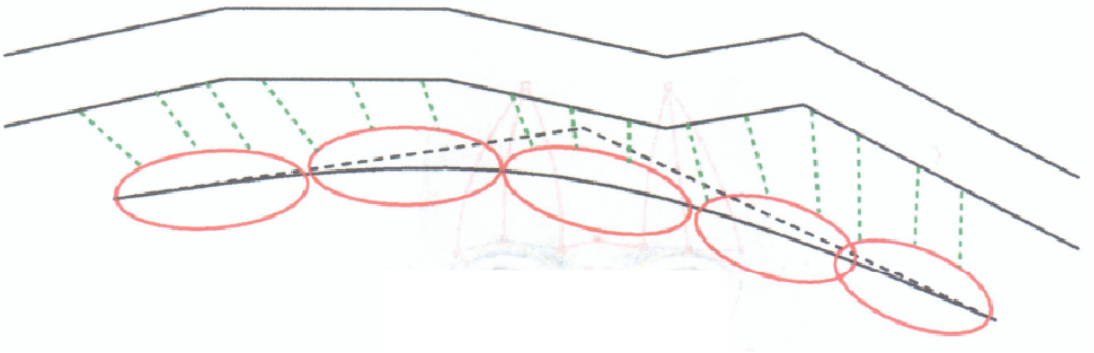


Figure 6.5: Muscle Skin Connection.

directly apply on a mass point farther away from it. We also constrain the inner angle θ between the skin connection spring and the muscle fiber, as some muscles control a wider range than others. For example, *Frontalis* controls a wider range on the forehead which was previously modeled as sheet muscle while *Depressor Anguli Oris* affects the lip corner area alone. The surface point on the ellipsoid is recorded with its surface parameters (u, v). Fig. 6.5 describes the aforementioned connection relationship between the muscle and the mass points. The muscle is subdivided into segments, each segment is represented by an ellipsoid. The mass point on the bottom layer of the tetrahedral mesh is searched for its shortest distance to the muscle volumetric surface and registered to the muscle.

6.4 Deformation Model

Although FEM generates highly realistic result for physical simulation, it does not meet real-time performance requirements. On the other hand, mass-spring model provides a relatively cheap solution between performance and accuracy. Teschner et al. [112] proposed a fast mass-spring model for physical simulation. We extend this by applying the special physically realistic parameters for human skin to construct the deformation model. As a non-rigid deformable object, human

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

skin deforms when external force such as muscle stretch force is applied and interior potential energy changes.

6.4.1 External Force

Ensemble muscle activation is the major source of facial expression. In Section 3.3 we have described how the muscles are connected to the nearby mass points. The muscle force is applied along these connections to the skin.

The muscle connected with anchor points requires an interpolation computation. We first compute the displacement of control points on each muscle according to their contraction ratio. The position of anchor points is computed as interpolation of its related control points. The new positions are then updated to the corresponding control points of the muscle to get the final shape of updated Bezier curves. The parametric ellipsoids representing muscle structure are updated and so do the muscle spring connecting to the skin surface.

Once a linear muscle activates, we compute the axis length of each muscle segments. Considering the fact that muscle volume is incompressible, if the radius of x-axis decreases when muscle contracts, the radius of y-axis and z-axis increase and vice versa. With the recorded surface parameters (u, v), the new position on the ellipsoidal surface can be computed. The pair of mass point $p_{skin}(i)$ on the skin tetrahedral mesh and the point $p_{muscle}(i)$ on the ellipsoidal surface define a distance constraint. The final computation equation for the muscle force is

$$f(p_{skin}(i), p_{muscle}(i)) = k_m \cos(p_{skin}(i), p_{muscle}(i))(|p_{muscle}(i) - p_{skin}(i)| - l_0)(\overrightarrow{p_{muscle}(i) - p_{skin}(i)}) \tag{6.4.1}$$

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

where k_m is the stiffness of muscle spring, and l_0 is the original length of the spring.

6.4.2 Internal Force

Considering the constraints C, the potential energy derived from C is

$$E = \frac{1}{2}kC^2 \tag{6.4.2}$$

where k denotes the stiffness coefficient. The structure of the right side is the form of elastic potential energy, which is defined as a result of deformation of an elastic object. The potential energy is zero if the object is undeformed. For skin deformation, as we discussed in Chapter 3, skin has the property of elasticity. The external forces from facial muscles cause the deformation of layered skin. The internal forces derived from the potential energies which try to preserve the object shape are then applied to the mass points. These forces balance the external force from muscles and also cause the displacement of mass points which are not directly connected by the muscle. If we denote each mass point as $p_{skin}(i)$, then the force at $p_{skin}(i)$ can be derived as

$$F_i = -\frac{\partial}{\partial p_{skin}(i)}E = -kC\frac{\partial C}{\partial p_{skin}(i)} \tag{6.4.3}$$

The negative sign indicates that the force derived from the energy always tries to reduce the energy of deformation.

Since the basic deformation element is a tetrahedron, we consider three types of potential energy and their corresponding forces. The first potential energy $E_{distance}$ preserves the distance between all pairs of mass points, which are the edges of the tetrahedron. It can be considered as a spring. Denoting the original distance

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

between a pair of mass points $(p_{skin}(i), p_{skin}(j))$ as D_0, then the energy between these two points is given by

$$E_{distance}(p_{skin}(i), p_{skin}(i)) = \frac{1}{2} k_{distance}(\frac{|p_{skin}(i) - p_{skin}(i)| - D_0}{D_0})^2 \qquad (6.4.4)$$

where $k_{distance}$ is the distance stiffness coefficient.

For mass-spring model, damping significantly improve the stability of numerical integration. Thus a damping coefficient is to compute $F_{distance}$.

$$F_{distance} = \frac{\partial F_{distance}}{\partial p_{skin}(i)} - k_{damping} v_i \qquad (6.4.5)$$

where $k_{damping}$ is the damping constant and v_i the velocity of the mass point.

As we have introduced in Section 3.2.2, the stress-strain relationship of human skin is not simply linear, the distance stiffness $k_{distance}$ is not a constant. We therefore employ a nonlinear equation [126] as given below

$$k_{distance}(p_{skin}(i), p_{skin}(j)) = (1 + \left[\frac{|p_{skin}(i) - p_{skin}(j)| - D_0}{D_0}\right]^2) k_0 \qquad (6.4.6)$$

where k_0 is the basic stiffness constant.

The second potential energy E_{area} tries to preserve the shape of each face of the tetrahedron. The energy is generated from the surface area change of each triangle. Consider a triangle constructed by $(p_{skin}(i), p_{skin}(j), p_{skin}(k))$ and the initial area as A_0. E_{area} can be computed as

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

$$E_{area}(p_{skin}(i), p_{skin}(j), p_{skin}(k)) = \frac{1}{2} k_{area} \left[\frac{\frac{1}{2}|(p_{skin}(j) - p_{skin}(i)) \times (p_{skin}(k) - p_{skin}(i))| - A_0}{A_0} \right]^2 \quad (6.4.7)$$

where k_{area} is the area preservation stiffness coefficient.

The third potential energy E_{volume} considers the set of four mass points of each tetrahedron, representing the volume change. For a tetrahedron represented by 4 vertices $(p_{skin}(i), p_{skin}(j), p_{skin}(k), p_{skin}(l))$, with initial volume V_0, E_{volume} is computed as

$$E_{volume}(p_{skin}(i), p_{skin}(j), p_{skin}(k), p_{skin}(l)) = \frac{1}{2} k_{volume} \left[\frac{\frac{1}{6}(p_{skin}(j) - p_{skin}(i)) \cdot ((p_{skin}(k) - p_{skin}(i)) \times (p_{skin}(l) - p_{skin}(i))) - V_0}{V_0} \right]^2 \quad (6.4.8)$$

where k_{volume} is the volume preservation stiffness coefficient.

From the above mentioned equations, forces F_{area} and F_{volume} are computed.

Based on our template model and experiment, we are using the following stiffness coefficient settings.

6.4.3 Jaw Kinematics

When opening the mouth, the lower jaw rotates around a horizontal axis passing through the mandibular condyles which are located at the rear extreme of the jaw bone. We set the maximum rotation as 60 degree here, considering he lower jaw is loosely constrained to the upper jaw. The mandible bone can also rotate 5 degrees

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Table 6.3: Settings for skin stiffness coefficient

Item	Value
Mass m	0.5
$k_{damping}$	0.002
$k_{distance}$ on outer layer	15
$k_{distance}$ on inner layer	10
$k_{distance}$ cross layer	5
k_{area}	7.5
k_{volume}	7.5

around a vertical axis starting from the midpoint of the two condyles. These two axes are described as u-axis and v-axis in Fig. 6.6 and the axis $w = u \times v$.

In the deformation model, the jaw movement is modeled as a rigid transformation that is applied on all the muscle control points, skin mesh mass points and anchor points for multiple muscles.

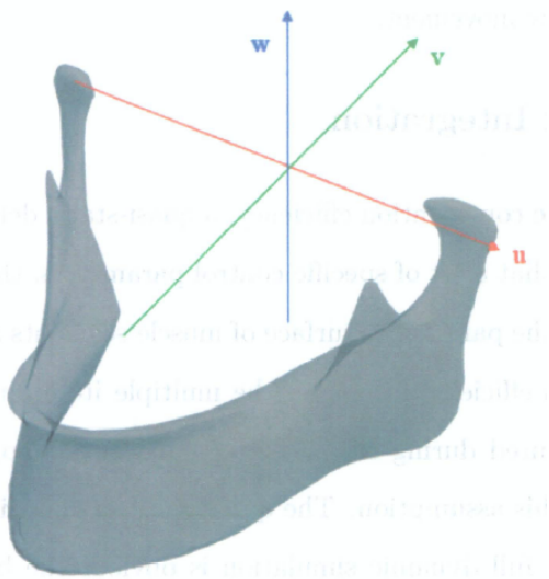


Figure 6.6: Jaw deformation model: mandible is allowed to rotate around axes u, v

The rotation amount of control points, mass points and anchor points is not uniform, they are variant according to the distance in u direction to the midpoints

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

between two condyles. The points in the middle of lower lip are rotated by the same amount as the jaw rotation. The rotation amount decreases where the points are closer to the corners of the mouth. The decrease rate follows a square root curve so that the mouth shape will remain smooth. For an arbitrary point x_i, the rotation amount is

$$\theta(x_i) = \frac{1}{3}\theta_{rot} + \frac{2}{3}\theta_{rot}\sqrt{\frac{|x_i - x_{center}|}{|x_{lipcorner} - x_{center}|}} \tag{6.4.9}$$

where θ_{rot} is the jaw rotation angle.

The corners of the mouth are rotated by one third of the jaw rotation. The points close to the upper lip corner are also pulled down with jaw rotation.

We set the lip corner, upper lip, lower lip and bottom face as the regions that are influenced by the jaw movement.

6.4.4 Numerical Integration

In order to maximize the computation efficiency, a quasi-static deformation model is defined. This entails that a set of specific control parameters, the control points of the muscle fiber and the parametric surface of muscle segments attain the static state instantly. By this efficient technique, the multiple iterations of refinement of ellipsoid are not required during the dynamic simulation loop, as would have been the case without this assumption. The quasi-static assumption is mainly for performance purpose, a full dynamic simulation is obvious the better choice for animation synthesis due to the continuity in muscle activations. But due to the fashion in which muscle activations translate to the deformation of skin tissue, even the muscle activations are not smooth, the simulated deformation is smooth.

After setting the requisite control parameters, we simulate the dynamic behav-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ior by computing the position and velocity of the mass point based on the initial position, velocity, internal forces and external force at the mass point, and a time step Δt. The Verlet algorithm [115] proposed in the context of physical simulation of general mass-spring systems has the advantage of high accuracy with a local discretization error of $O(\Delta t^4)$, so it allows for relatively large time step. The Verlet algorithm requires the positions and force at time t, and positions at the previous time $t - \Delta t$ to calculate the new position:

$$\begin{aligned} x(t+\Delta t) &= 2x(t) - x(t-\Delta t) + \Delta t^2 \frac{F(t)}{m} + O(\Delta t^4) \\ v(t+\Delta t) &= \frac{x(t+\Delta t) - x(t-\Delta t)}{2\Delta t} + O(\Delta t^2) \end{aligned} \tag{6.4.10}$$

The entire simulation procedure is explained in Fig. 6.7. Our interactive user interface is shown is Fig. 6.8.

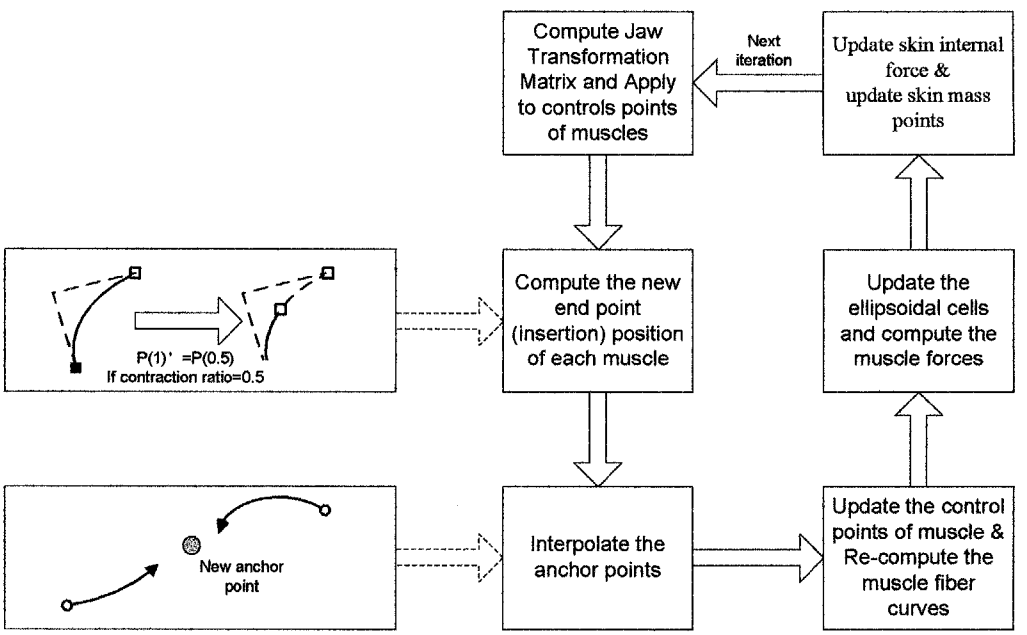


Figure 6.7: Process flow for dynamic facial expression synthesis

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Figure 6.8: Interactive user interface for expression generation

6.5 Expression Generations

In this section, we present some results of our dynamic simulation and the synthesized facial expression after we applied the muscle activation.

The object in Fig. 6.9 and Fig. 6.10 is a tetrahedral mesh generated from a 50×50 rectangle mesh. In Fig. 6.9 we show various deformations under the constraint of different inner angle between the muscle fiber and muscle skin spring. Fig. 6.10 shows the deformations with different muscle activation ratio. The stiffness of the skin and damping coefficient remain the same in these experiments. We denote the activation ratio as σ, time step as t, inner angle as θ.

Finally, we present several facial expressions generated by our physical simu-

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

lation in Fig. 6.11-6.15. The muscle activation parameters are listed in Table 6.4. The animation sequences of the five expressions are shown in Fig. 6.16.

We are using texture map in the rendering here. A 400x400 bitmap with UV map is provided by the Minolta Vivid 700 laser together with the geometric data. After the adaptation, each vertex from the template model is mapped onto its closest primitive (triangle) on the scanned data. We then compute the weights of the mapped point within the triangle and apply the weights to the UV coordinates. Therefore, each vertex of the adapted template model will obtain its UV coordinate to the scanned texture image.

Table 6.4: Table of muscle activation parameters of the facial expression (0 = no contraction, 1 = 100% contraction)

muscle no.	01	02	03	04	05	06	07	08	09	10	11	12	u-axis Jaw Rot.
smile	0.3	0.1	0	0	0	0	0	0	0	0	0	0.2	0
laugh	0.7	0.2	0	0.5	0	0.2	0	0	0	0	0	0.1	$\pi/6$
surprise	0	0	0	0	0.3	0.3	0	0	0.2	0	0	0	$\pi/6$
angry	0	0	0.3	0.2	0	0.5	0.4	0	0	0.3	0.4	0	$\pi/12$
sad	0	0	0.5	0	0	0	0.6	0.4	0.1	0	0	0	0

6.6 Adaptive Physical Structure

Based on Chapter 4, it can be understood that arbitrary 3D head mesh can be calibrated and adapted into a personalized head. Furthermore, we adapt the physical structure introduced in this chapter into the personalized shape mesh, this greatly reduces the workload on adjusting the skin parameter.

As we have introduced in Section 6.2, the skin volume is generated from single layer triangle skin mesh. After the surface shape is deformed during the shape adaptation, the tetrahedralization procedure remains the same.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Muscle is defined by its control points. Since we are defining muscles by specifying control points on the skin surface, the face index and the barycenter coordinate can be recorded. During the shape adaptation process, the topological structure of the template mesh has never changed, hence it is reasonable to reuse the face index and barycenter coordinate to define the muscle.

A *region* is a collection of triangles assigned to it, each triangle is assigned to one region. Region is modeled as a constraint of the muscle contraction. This property of each triangle does not change during the shape deformation, so the region information is still available for the adapted model.

The eyeballs, iris, eyelids are imported as individual rigid parts of the head. They are transformed according to their related markers. We describe the transformation function of left eye here; the functions of the others are very similar. The left eye is related to the neighboring FDP 3.7, 3.9, 3.11, 3.13. Both point position from template model and scanned data can be easily obtained since these points are obvious face features. As a rigid transformation, we only consider the uniform scale, rotation and translation. We get scale factor δ from

$$\delta = \frac{||P^s_{3.7} - P^s_{3.11}|| \times ||P^s_{3.13} - P^s_{3.9}||}{||P^t_{3.7} - P^t_{3.11}|| \times ||P^t_{3.13} - P^t_{3.9}||} \tag{6.6.1}$$

where s represents the scanned data, t represents the template model.

To compute the rotation matrix, we assume that the vector $\overrightarrow{P^s_{3.7} - P^s_{3.11}}$ represents the transformed x axis and vector $\overrightarrow{P^s_{3.13} - P^s_{3.9}}$ represents the transformed y axis. The template head is in standard Cartesian coordinate. So the problem becomes the computation of the rotation matrix between the two coordinate systems, which is a very basic graphics problem (see Section 4.3.2).

After obtaining rotation Matrix R_{rot}, we can compute the new center position

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

of the eye balls. First the center position of the eye ball of the template φ^t is computed, then the center point of $P^t_{3.7}$, $P^t_{3.9}$, $P^t_{3.11}$, $P^t_{3.13}$ is considered as a reference point ψ^t, thus we get a vector $\overrightarrow{\gamma^t}$, where

$$\overrightarrow{\gamma^t} = \psi^t - \varphi^t \tag{6.6.2}$$

Using the same idea on the scanned model, we get a reference point ψ^s, finally the new center position of left eye ball $\varphi^{t'}$ is computed as

$$\varphi^{t'} = \psi^s - \delta R_{rot}\overrightarrow{\gamma^t} \tag{6.6.3}$$

Now given any vertex of the left eye from template model, the new position is

$$p(i)' = \delta R_{rot}(p(i) - \varphi^t) + \varphi^{t'} \tag{6.6.4}$$

Fig. 6.17 and Fig. 6.18 show two expressions synthesized after shape adaptation. Compared to Fig. 6.11- 6.15 it can be observed that the shape is completely different but we can still generate expressions with acceptable quality.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University

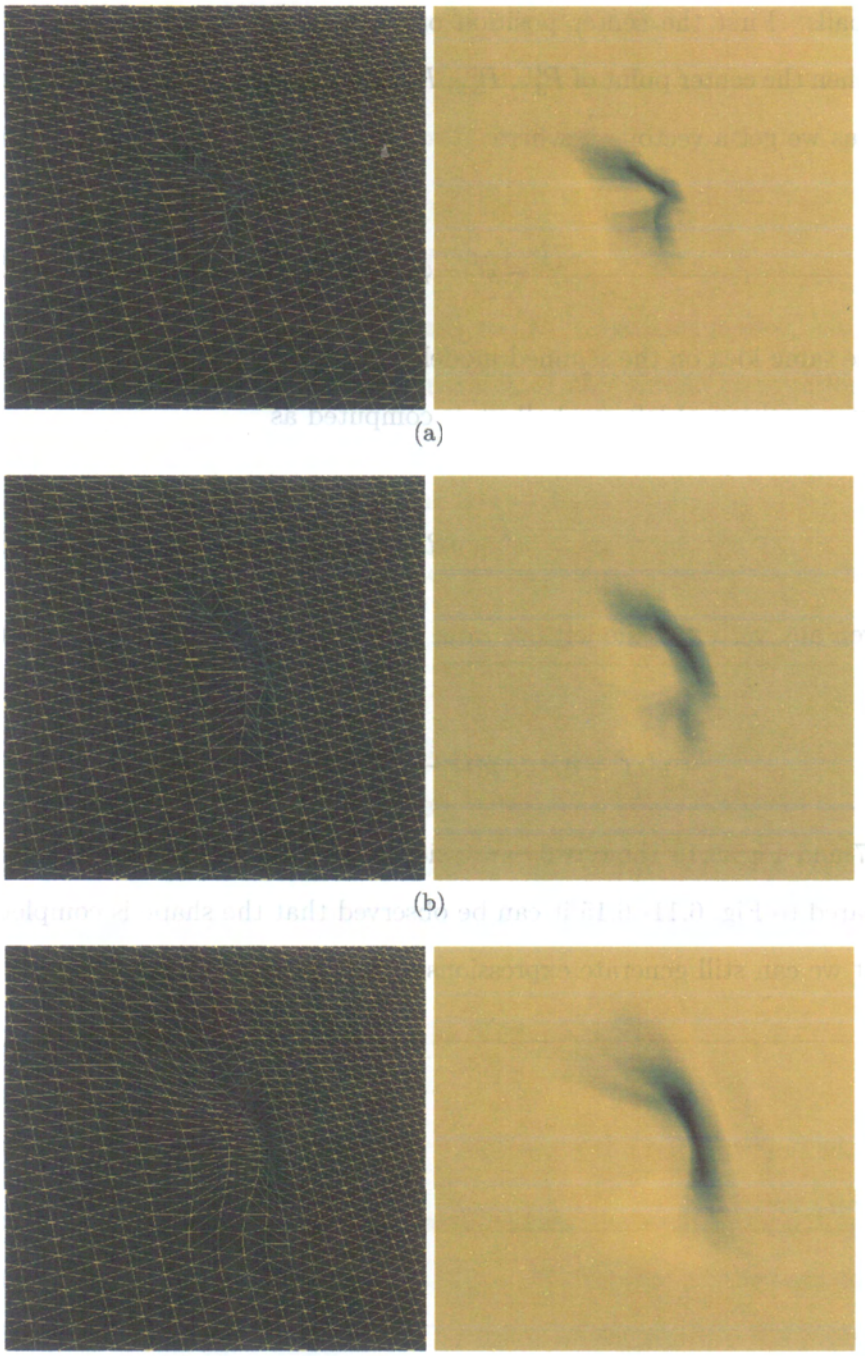


Figure 6.9: Single muscle activation force applies to volumetric mesh generated from a 50x50 triangular mesh. $\sigma = 0.5$, $t = 0.002$, constrained muscle connection length

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

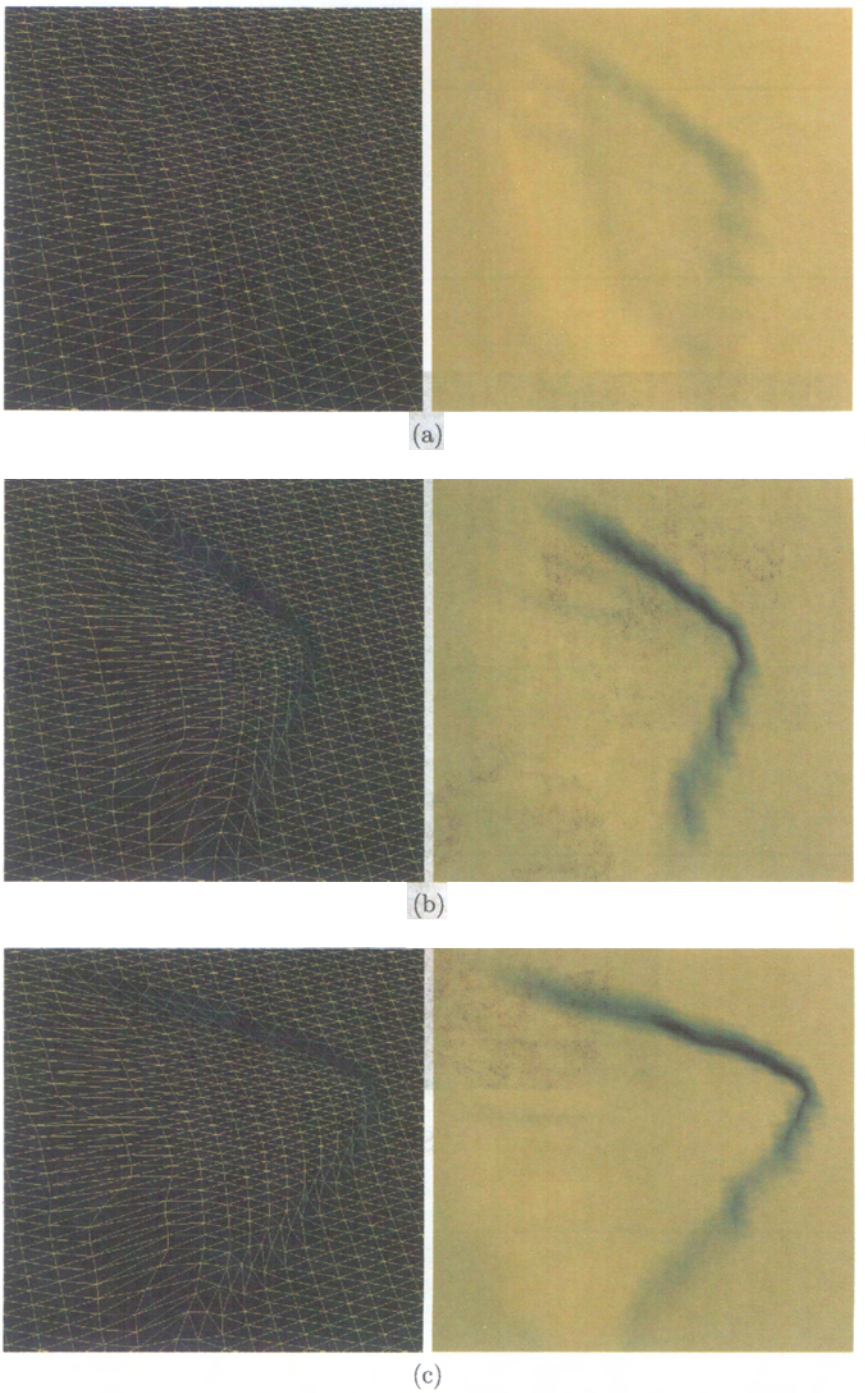


Figure 6.10: Single muscle activation force applies to volumetric mesh generated from a 50x50 triangular mesh. $\theta = \pi/4$, $t = 0.002$, unconstrained muscle connection length

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

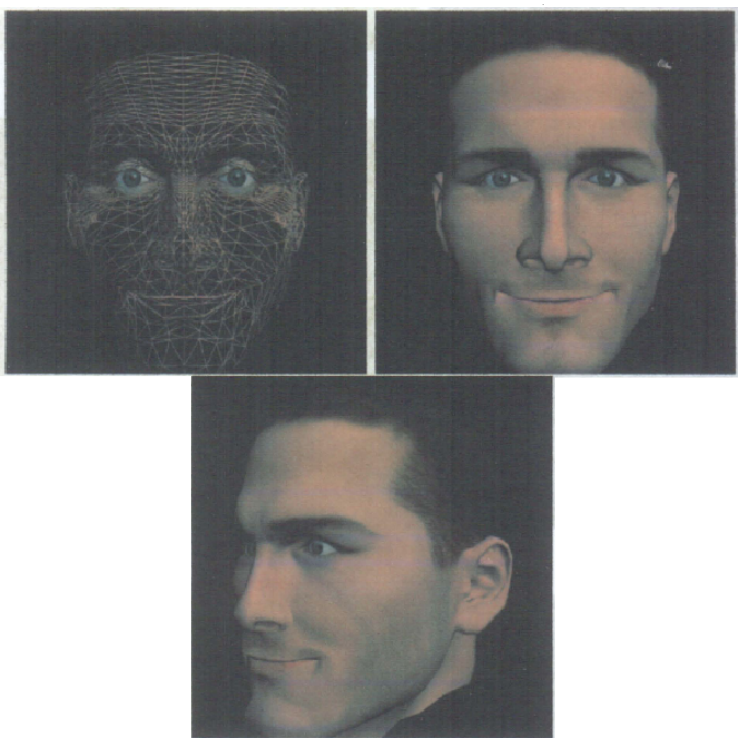


Figure 6.11: Smile

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

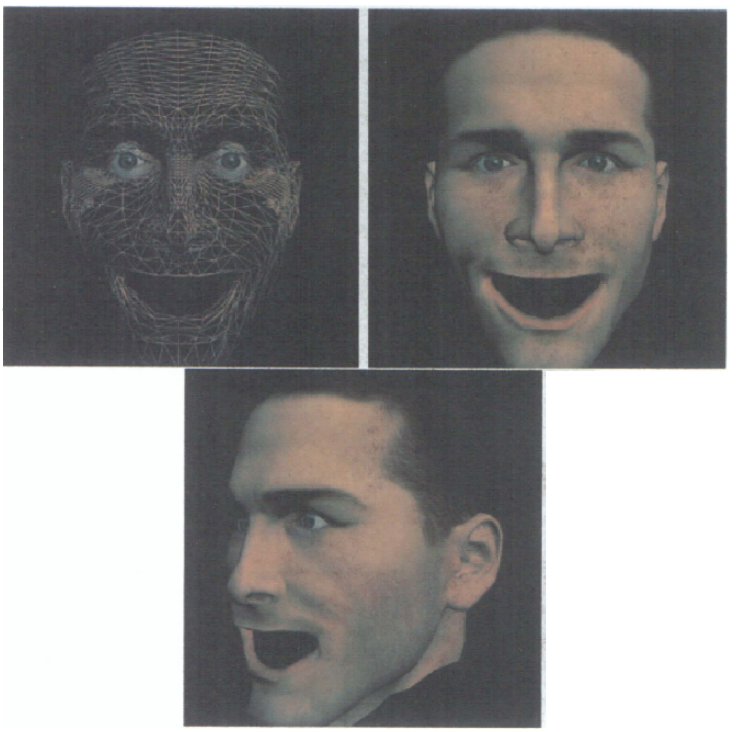


Figure 6.12: Laugh

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

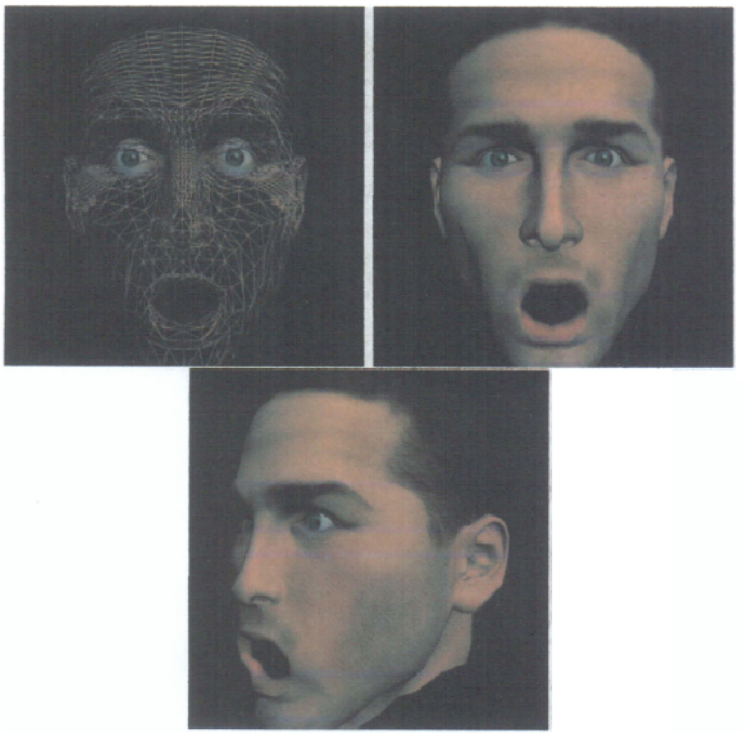


Figure 6.13: Surprise

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

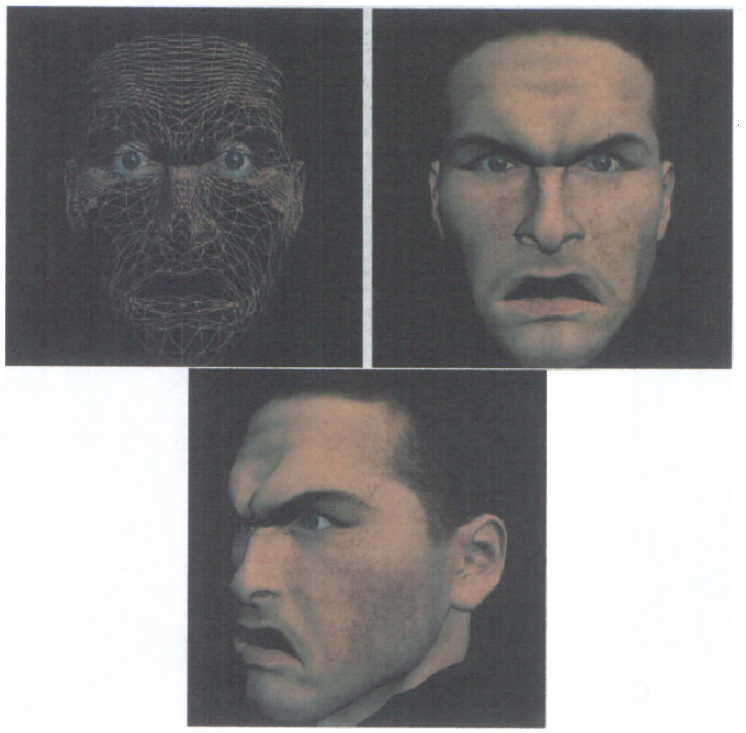


Figure 6.14: Angry

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

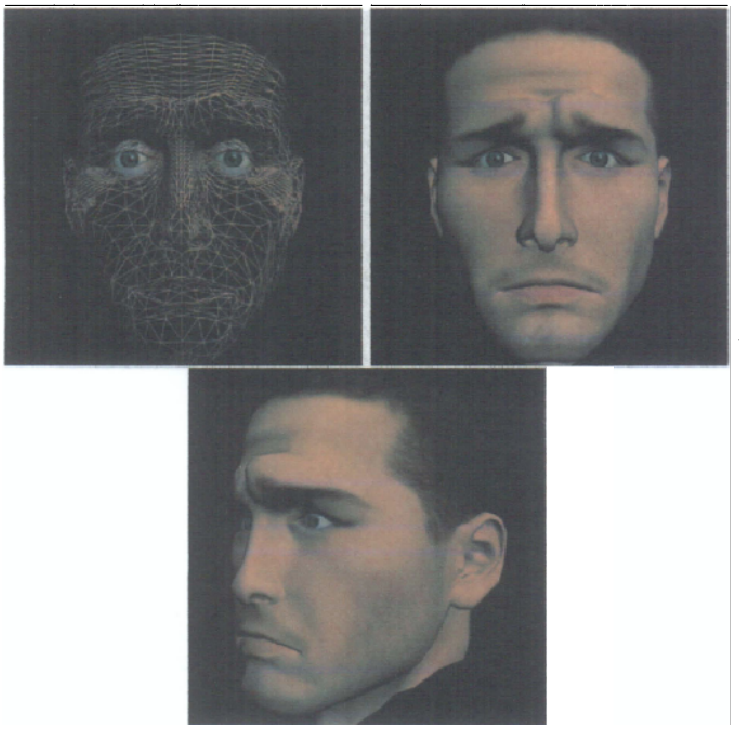


Figure 6.15: Sad

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

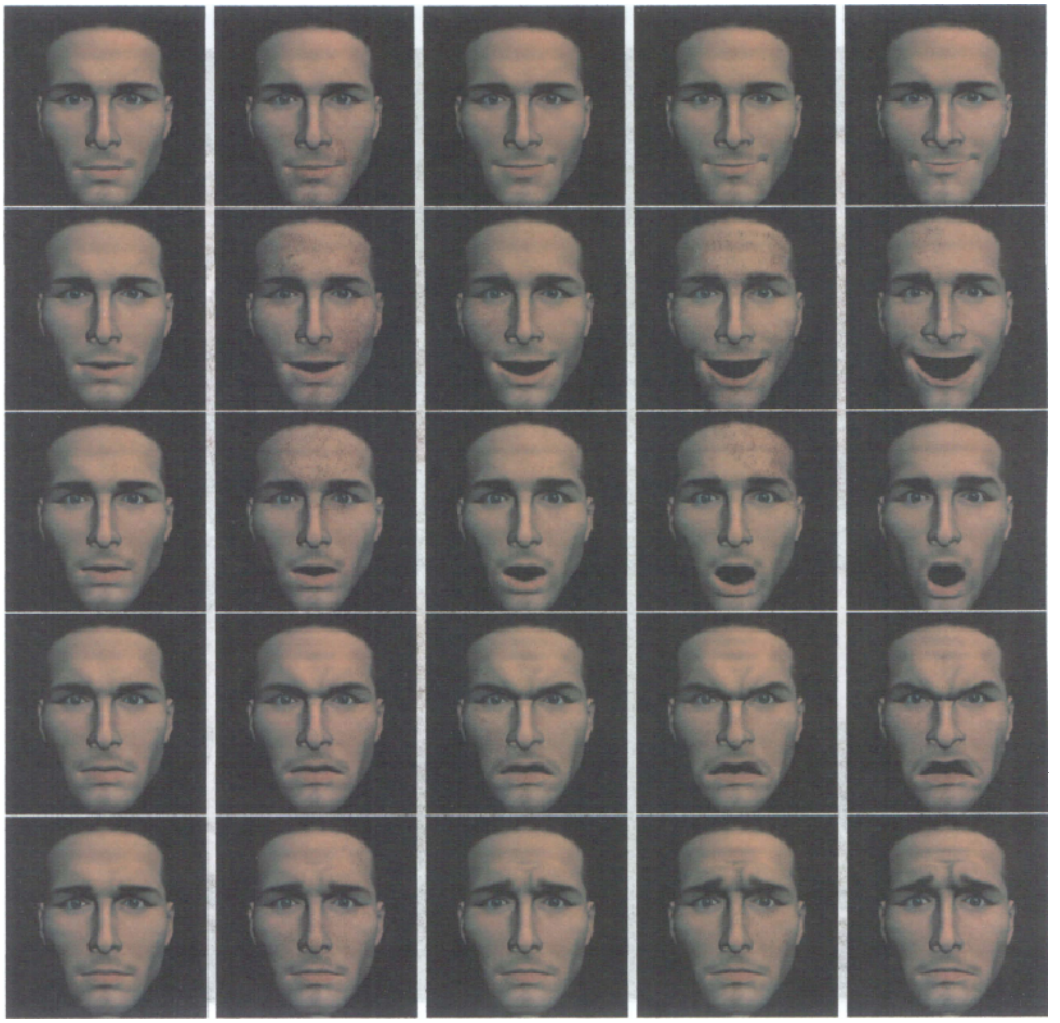


Figure 6.16: Animation Sequences of Expressions: smile, laugh, surprise, angry and sad

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

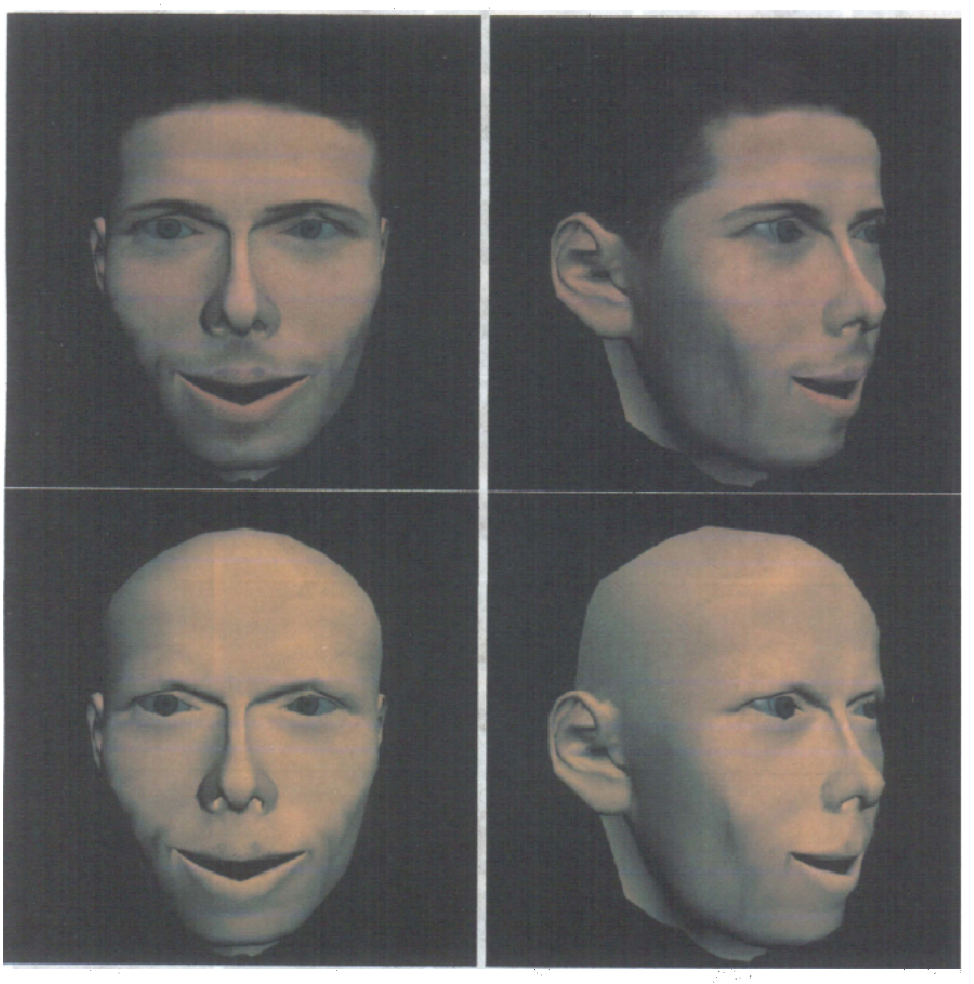


Figure 6.17: A neutral adapted head with mouth opening

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

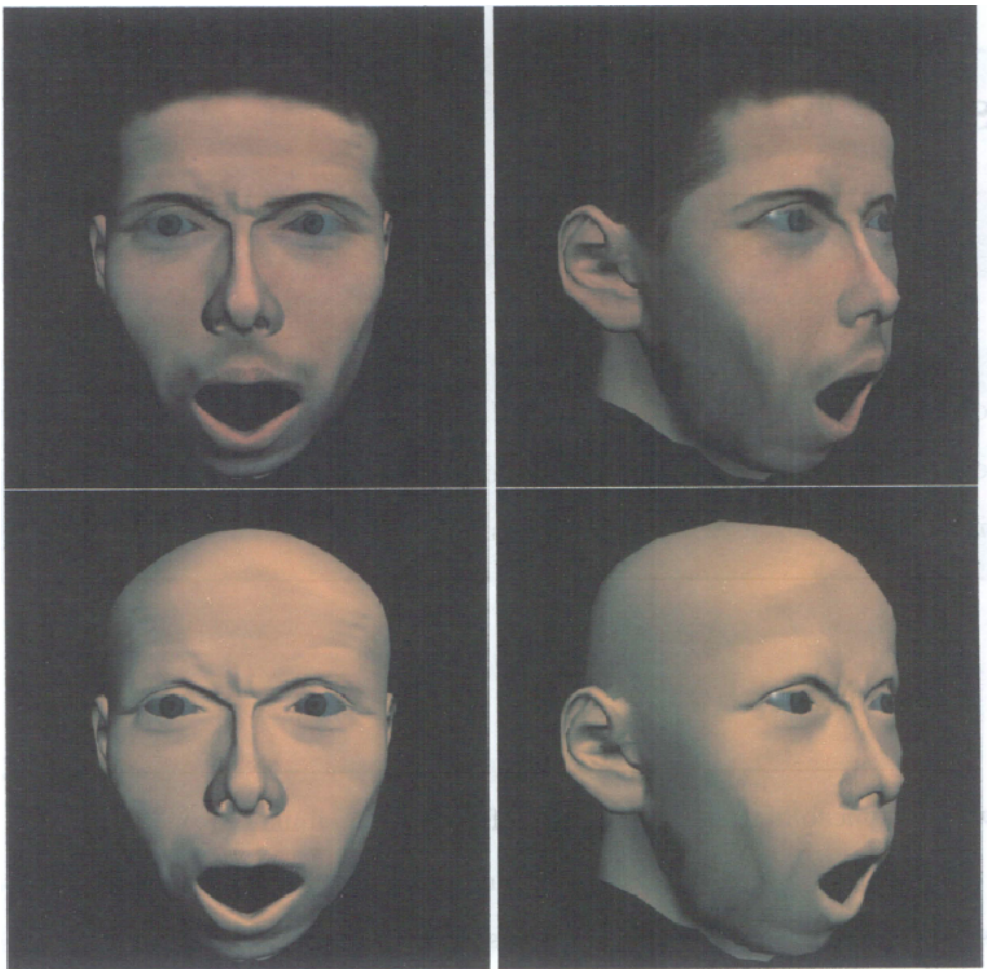


Figure 6.18: A surprise expression with scowl

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 7

Case Studies

7.1 Introduction

In this chapter we discuss two applications of our work. Section 7.2 presents some examples about face morphing between different people. Section 7.3 shows the simulation results by applying the same physical parameters to different people.

7.2 Face Morphing

We show the applicability of our method for face cloning similar to the work done by Noh [124]. The face morphing means a smooth transition of appearance from one person to the other. Inherited from our shape adaptation procedure the face transition can be easily implemented because we use the same template model without any problem. That means the topology of the models remain the same. Only the positions of vertices change among models. The adapted models can also be used as a blend shape library. Fig. 7.1 presents 3 examples of face morphing. The results prove that our approach can achieve clone motions on adapted model with noticeable feature difference.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

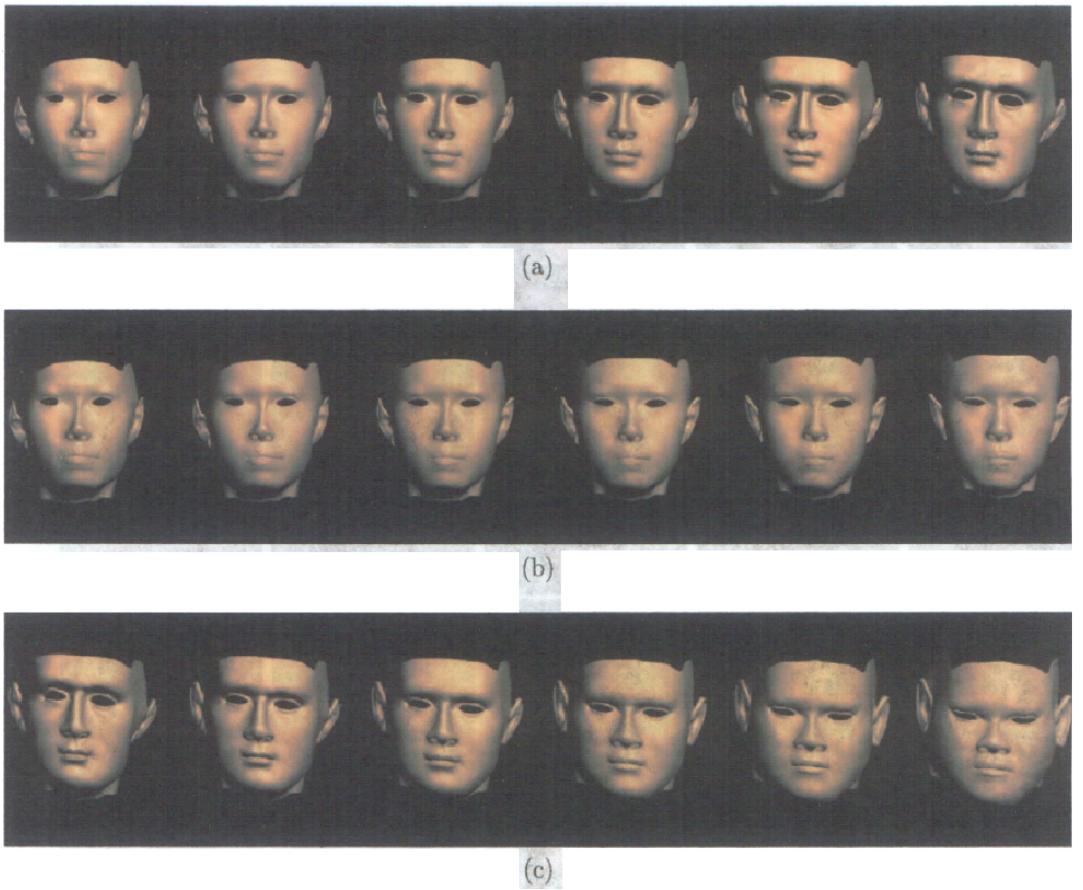


Figure 7.1: Face morphing examples

7.3 Motion Retargetting

Motion retargetting means applying the same motion to different characters. For facial animation, the problem is normally solved by point-based approach. During our adaptation procedure, the physical structure is automatically transferred into the target model (scanned data). We also mentioned that the adapted model is still in the same scale as the original template model, so we can use the physical control parameters directly. Fig. 7.3, Fig. 7.4 and Fig. 7.5 presents motion retargetting examples of the same animation sequence with physical control parameters listed in Table 6.4. The real human texture is applied on the adapted model to improve

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

the realism. For comparison, we show the expression from template model again in Fig. 7.2. The results prove that our approach can achieve clone motions on adapted model with noticeable feature difference.

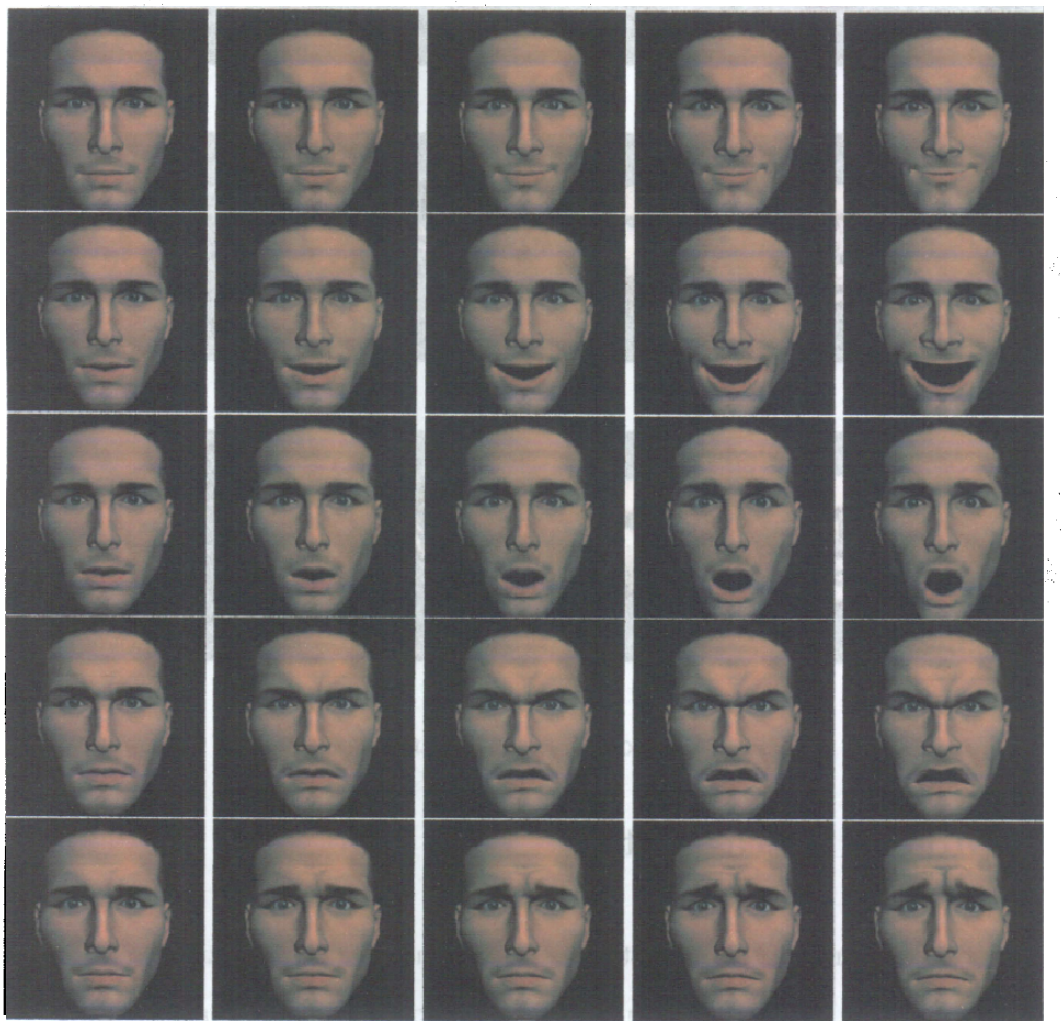


Figure 7.2: Animation Sequences of Template

7.4 Discussion

This discussion highlights some of the common issues and solutions for research presented in this thesis:

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Figure 7.3: Motion Retargetting Example 1

- **Significance of error metrics for shape adaptation - for face and sub-face regions:** Our contribution is to develop and demonstrate an error metric that can be directly applied for the similarity between two faces. Furthermore, we wanted a metric that can be applied to sub-regions of the face (eyes, nose, mouth and the ear) which has not been reported before in the literature. We have adapted the approach of Allen et al. [5] used for fitting high-resolution template meshes to detailed human body range scans with sparse 3D markers. Allen et al. [5] focus on the whole body, whereas

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

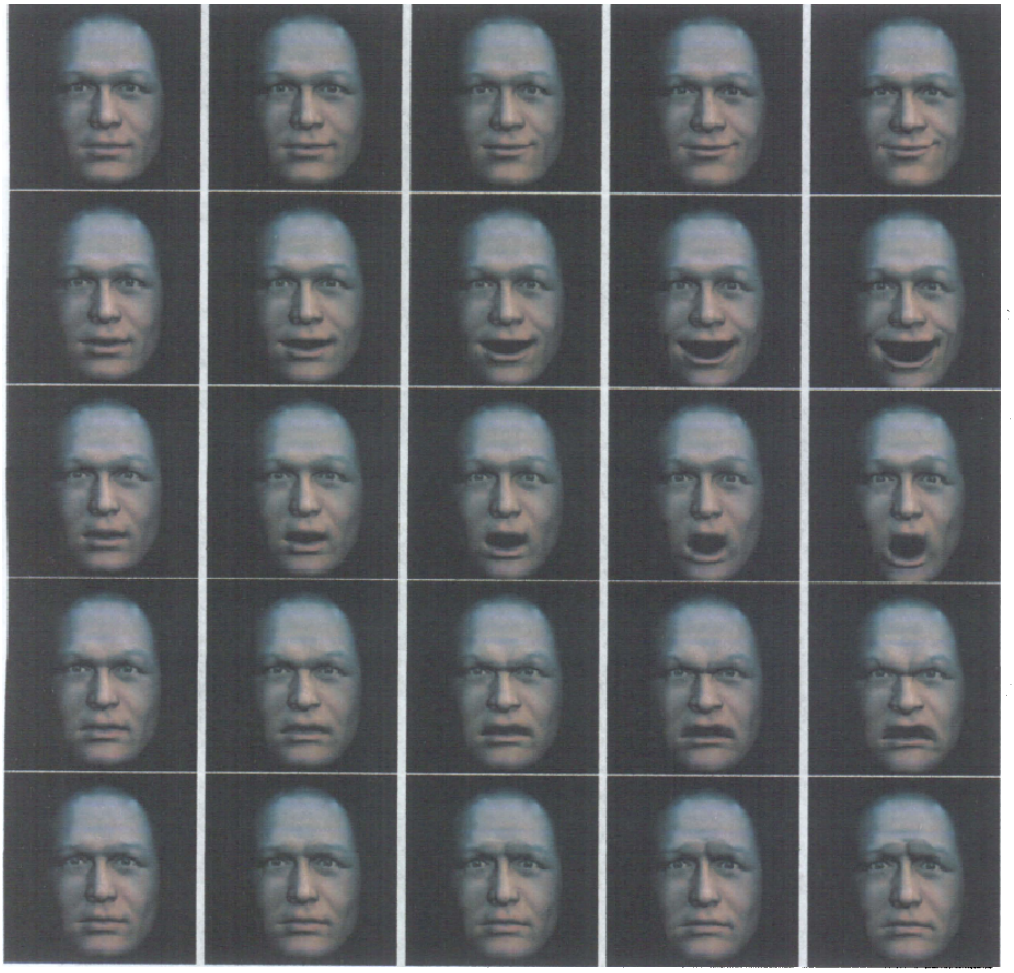


Figure 7.4: Motion Retargetting Example 2

we demonstrate the use of fitting for the face and sub-regions.

- **Significance of high performance physically-based facial expression synthesis:** The high performance is the key contribution. It might appear that in the proposed high performance facial expression synthesizer, we combined multiple elements including skull, muscle, skin into a physically-based approach. However we have significantly improved the elements in following ways.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Figure 7.5: Motion Retargetting Example 3

– *Novel constraints:* To simulate soft tissue of facial skin, the original multi-layer mass-spring model proposed by Lee [74] only considers distance constraints. We added area constraints and volumetric constraints to simulate the incompressible property of soft tissue. For the distance constraints, we add nonlinear coefficient to mimic the stress-strain curve shown in Fig 3.3. These improvements provide a more realistic skin simulation with volumetric Though similar approaches have been used in [112] and [126], we have not seen any reports combined

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

these factors to successfully simulate facial skin behavior.

- *Control of Voluntary and Involuntary facial muscles:* We use Rational Bezier curves to represent facial muscle fibres. This approach provides simple muscle definition methods but also considers the flexibility of muscle shape. Both linear muscle and sphincter muscle are defined by curves. A muscle curve is subdivided into segments. By defining these segments, we simulate the muscle volume which interacts with the skin volume. With this representation, we contribute that voluntary muscles and involuntary muscles are connected together at anchor points. The placements of control points belonging to the involuntary muscles, such as orbicularis oculi, are controlled by voluntary muscles. Therefore, we solve the problem of how to control contraction of involuntary muscles, which has not been discussed by any other reports before.

- How much CPU time does the fitting require? How much human time goes into the fitting processes? How many individual face models have been made in this thesis and how much variation is there in the processing, the parameter tuning, and the time required to create them? The human time in the fitting process is mainly used on FP initialization and removal of incorrect FP pair. The CPU time in the adaptation is mainly used on the solving of linear system of equations for the RBF weight coefficient computation. A LU Decomposition costs $O(N^3)$ operations. To solve a system of linear equations $\Lambda X = B$ where $\Lambda = LU$ costs $O(N^2)$ operations. Our adaptation is currently on a single template model. But we have tested our face adaptation on eight different scanned faces without much additional tuning for individual heads. Once a scanned head is adapted, the adapted head can be used as a reference template for adapting new scanned data. More template models

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

with measurable properties are suggested for possible extension of this thesis and for future work of our research.

- The curvature analysis varies widely from face to face, even scan to scan of the same face, or eye to eye of the same person and scan. How useful is this data given such variations? How sensitive are our results to variations of these points? Based on the curvature analysis, we are able to easily detect the features on surface with lowest fairness. This tool provide a better feature tracking method and keep the number of FPs, which are used RBF adaptation, in a lower number than the feature mesh expansion method. We admit the variation of curvature on different model. But our experiment shown in Fig. 4.4 indicates that the FPs detected with high curvature mainly appear at eyes, lips, nose, which are our main consideration during the adaptation.

- The proposed method meticulously divides facial expressions into anatomical parts, though there are a great number of approximations and assumptions. How does each part of the model contributes the overall precision. How does the error in the skull shape affect the overall precision, for example? There are several layers that influence the end output results. We have focused our attention on the variation in the face mesh (skin) and the variation in muscle (for deformation). The skull acts as a scaffolding and sits beneath the facial skin. Variation of the skull as pointed out by the examiner is an important parameter. This may lead to variation in the face shapes. One possible use for a variable skull model would be in applications such as forensics or archaeological face and head reconstruction from skulls. But this requires reconstruction from medical images. This also happens to the skin thickness element, which we cannot accurately measure according to our current equipments. This can be a possible extension of this thesis and

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

identified as an area for future work.

- As it is almost impossible to precisely simulate biomechanical deformations of facial shape, there are a great number of assumptions. When are these assumptions plausible? When do these assumptions do not hold? What are the limitations of the model? It is impossible to precisely simulate the facial shape. The synthesized shape can have several variations both large and subtle ones. The model in this thesis works well for small time steps, for elongation and for limited stages in compression. So it is suitable for mimicking most of the human-like expressions. The model in this thesis doesn't account for cartoon-like expressions (but can be accomplished with manual fine tuning), and doesn't account for excessive compression (when the polygons degenerate to a line or a point). Problems may also occur when wrinkles or buckling of the skin are introduced.

- The proposed model does not distinguish between voluntary and involuntary muscles. However, a human cannot control orbicularis oculi voluntarily. The skin deformation along this muscle is also difficult to calculate because of its circular shape unlike others. How is this muscle handled in the model? How is its contraction affect the overall facial expressions? Although this muscle does not show a wide difference in its shape, it plays a very important role in performing facial expressions because it is not controllable spontaneously. The muscle behavior for regions around the mouth and eyes are modeled as rational Bezier curve(s). The unique representation helps to simplify the deformation control. The examiner is correct, orbicularis oculi and orbicularis oris cannot be handled by human voluntarily. Therefore, we use anchor points to link the insertion points of voluntary muscles and the Bezier control points of non-voluntary muscles.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 8

Conclusion and Future Work

8.1 Conclusion

Great effort have been made to model realistic human face shape and synthesize nature-looking facial expression in recent years. High fidelity model and rapidness construction are the ever increasing requirements in applications such as human-computer interaction, video-conference and virtual avatars. The adaptive physically-based face model is a sufficient approach to meet up the above requirements.

In this thesis, we have described our adaptive face model. We have achieved two levels of adaptation, the coarse-to-fine personalized shape generation and adaptive anatomical structure for expression synthesis.

Face Adaptation: The shape adaption is based on a blend function of general RBF and compactly supported RBF. To obtain the advantage of both global shape fitting and local detail deformation, we use a hierarchical multi-scale approach to iteratively deform our template model to the scanned data. A curvature-based technique is introduced to capture the detail of facial features and a feature mesh based technique is employed to perform an average

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

surface data sampling.

Error Metrics: We have presented error metrics in 3D human face modeling and adaptation. The error metrics can be used in iteration control and estimate the adaptation functionality. The qualification of facial shape and face features are essential to advance the understanding of how the adaptation enhances the representation of face models. To achieve such a measurement, we introduce the distance error and marker error terms to estimate the similarity between our adaptation result and the range scan data. The estimation result validates the success of the adaptation, it also measures the distance between the scanned model and the adapted template.

Expression Synthesis: The physically-based facial expression synthesizer consists of a tetrahedralization procedure from triangle skin mesh, a novel parametric volumetric muscle structure with simultaneous activation function, and a dynamic simulation system. The dynamic system considers the volumetric skin tissue constraints, the muscle activation constraints and the jaw kinematics. Verlet integration is employed to achieve high performance computation. We have proved that this anatomically based structure can synthesize various facial expressions.

Expression Adaptation: Since the tetrahedralization and muscle definition is determined by skin geometry. These anatomical structures are adapted according to the non-rigid shape transformation. Other parts of the face are treated as rigid transformation. The dynamic soft tissue behavior is inherited during the adaption procedure so that facial expression is able to be synthesized.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

8.2 Future work

This thesis can be extended in several related directions, we find the following research topics are especially worth considering for future research work.

GPU level simulation: The latest graphics hardware provides tremendous memory bandwidth and computational horsepower, with fully programmable vertex and pixel processing unit which is quite suitable for floating point computation. GPUs are highly parallel streaming processors with both MIMD (vertex) and SIMD (pixel) pipelines. Researchers have already started using GPU for general purpose computation such as simulation of physics or for conventional computational science. Our physical simulation procedure is well suited for parallel computation, hence simulation in GPU is desirable. A possible procedure for GPU processing is proposed in Fig. 8.1.

Performance driven physically based simulation: One compelling direction would be the estimation of the physical control parameters, for more realistic facial expressions. Although user-defined facial expression can be synthesized in our system, the final expression is not easily predictable according to the dynamic procedure. On the other hand, performance driven animation, marker based or not, contains the target expression information. An optimized framework is desirable to solve the dynamic simulation procedure and recursively recover reasonable control parameters for the simulation. The combination of both performance driven animation and physical simulation will:

1. greatly improve the fidelity of reconstructed animation from performance;

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2. reduce the user interaction;
3. increase the accuracy of the control parameter selection;
4. provide a better method to validate the dynamic simulation system.

Multi-dimensional template and blend shape library: We are currently using one shape template during geometrical adaptation. Though great variant has been achieved with the combination of feature mesh and curvature based searching, global and local RBF interpolation, the utilization of blend shape library with different templates can optimize a closer template for the input mesh. Other parameters such as gender, race and age can be considered as weight coefficient to impact the optimization. Besides, the template should not be limited as the shape domain, textures and motion sequences can also be consider as training target.

Photo-realistic skin texture: The quality of head texture we have generated is not pleasing yet. Merging photos from multiple views still surfer from manual registration and inappropriate blending due to different light conditions. High quality texture will definitely improve the visual effect of face model. Therefore automatic approach to generate personalized texture is under investigation.

Finally, it is hoped that this thesis will provide a solid foundation and a good start for future researches and some of the future directions identified in this thesis can be further explored. It could help to uncover more secret on the modeling and synthesis of facial expressions.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

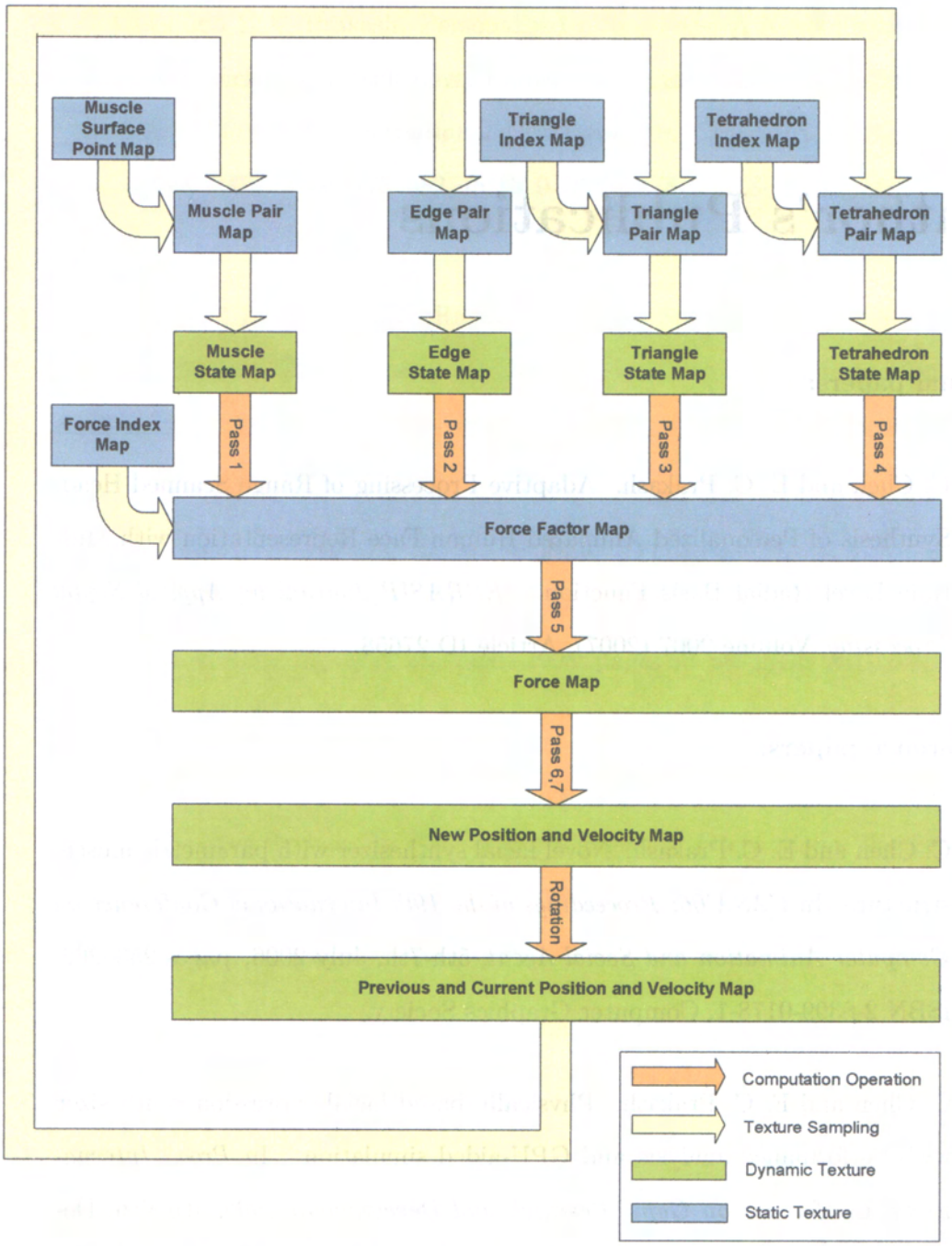


Figure 8.1: Proposed multi-pass rendering steps for facial expression simulation in GPU

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Author's Publications

Journal papers:

1. C. Chen and E. C. Prakash. Adaptive Processing of Range Scanned Head: Synthesis of Personalized Animated Human Face Representation with Multiple Level Radial Basis Function. *EURASIP Journal on Applied Signal Processing*, Volume 2007 (2007), Article ID 27658.

Conference papers:

1. C. Chen and E. C. Prakash. Novel facial synthesizer with parametric muscle structure. In *CASA'06: Proceedings of the 19th International Conference on Computer Animation and Social Agent*, 5th-7th, July 2006, pages 253–262, ISBN 2-8399-0178-1, Computer Graphics Society.

2. C. Chen and E. C. Prakash. Physically based facial expression synthesizer with performance analysis and GPU-aided simulation. In *Proc. International Conference on Game Research and Development 2006*, 4th–6th, December 2006, pages 171–175.

3. C. Chen. 3D Adaptive Personalized Facial Shape Modeling and Expression Synthesis. In *Virtual Motion Workshop*, 14th-16th December, 2006.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4. C. Chen and E. C. Prakash. Personalized Cyber Face: A Novel Facial Modeling Approach Using Multi-level Radial Basis Function. In *CW'05: Proceedings of the 2005 International Conference on Cyberworlds*, 23th-25th, November 2005, pages 475–482, ISBN 0-7695-2378-1, IEEE Computer Society.

5. C. Chen and E. C. Prakash. Rapid Normalized Shape Modeling (NSM) of Non-Mechanical Objects using 3D Digitization: Case Study of the Human Head. In *Proc. International Conference on Information and Automation*, December 15-18, 2005.

6. C. Chen and E. C. Prakash. 4D Faces: From 3D Digitized Heads to Animatable Heads. In *Proc. Cyber Games'05: International Workshop on Games Research and Development*, 28 March 2005, pages 63–68, ISBN 981-244-959-0, PEARSON Education.

7. N. M . Chiang and C. Chen. SkinIT: An Efficient GPU Based Character Skin Rendering System. In *Proc. Cyber Games'05: International Workshop on Games Research and Development*, 28 March 2005, pages 63–68, ISBN 981-244-959-0, PEARSON Education.

8. C. Chen and E. C. Prakash. Face personalization: animated facial modeling approach using radial basis function. In *Proc. IEEE TenCon 2005.*

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Bibliography

[1] Anatomy of the skin. http://en.wikipedia.org/wiki/Skin.

[2] *ISO/IEC MPEG-4 Part 1 (ISO/IEC 14496-1): Systems.*

[3] *ISO/IEC MPEG-4 Part 2 (ISO/IEC 14496-2): Visual.*

[4] Muscular system. "http://www.lhwl.lhsa.com/Site2001/Courses/Adv Bio/muscular system.htm".

[5] Brett Allen, Brian Curless, and Zoran Popovic. The space of human body shapes: reconstruction and parameterization from range scans. *ACM Trans. Graph.*, 22(3):587–594, 2003.

[6] Pierre Badin, Gerard Bailly, Lionel Reveret, Monica Baciu, Christoph Segebarth, and Christophe Savariaux. Three-dimensional linear articulatory modeling of tongue, lips and face, based on mri and video images. *Journal of Phonetics*, 30(3):533–553, July 2002.

[7] Sumit Basu, Nuria Oliver, and A. Pentland. 3d lip shapes from video: a combined physical-statistical model. *Speech Commun.*, 26(1-2):131–148, 1998.

[8] Volker Blanz, Curzio Basso, Thomas Vetter, and Tomaso Poggio. Reanimating faces in images and video. In *Proc. EUROGRAPHICS 2003*, pages 641–650, 2003.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[9] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Proc. ACM SIGGRAPH*, pages 187–194. ACM Press/Addison-Wesley Publishing Co., 1999.

[10] George Borshukov and J. P. Lewis. Realistic human face rendering for *The Matrix Reloaded.* In *Proc. ACM SIGGRAPH Conference on Abstracts & Applications (Sketch)*, pages 1–1, 2003.

[11] Christoph Bregler, Michele Covell, and Malcolm Slaney. Video rewrite: driving visual speech with audio. In *Proc. ACM SIGGRAPH*, pages 353–360, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[12] Pelachaud C., Norman I. Badler, and Viaud M. Final report to the NSF of the standards for facial animation workshop. Technical report, University of Pennsylvania, School of Engineering and Applied Science, Computer and Information Science Department, Philadelphia, PA 19104–6389, 1994.

[13] 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 *Proc. ACM SIGGRAPH*, pages 67–76. ACM Press, 2001.

[14] Byoungwon Choe and Hyeong-Seok Ko. Analysis and synthesis of facial expressions with hand-generated muscle actuation basis. In *Proc. Computer Animation*, pages 12–19, 2001.

[15] Byoungwon Choe, Hanook Lee, and Hyeong-Seok Ko. Performance-driven muscle-based facial animation. *The Journal of Visualization and Computer Animation*, 12(2):67–79, May 2001.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[16] M. Cohen and D. Massaro. Modeling coarticulation in synthetic visual speech. In *Models and Techniques in Computer Animation*. Springer-Verlag, 1993.

[17] Eric Cosatto and Hans Peter Graf. Photo-realistic talking-heads from image samples. *IEEE Trans. Multimedia*, 2(3):152–163, 2000.

[18] Beautemps D, Badin P, and Bailly G. Linear degrees of freedom in speech production: analysis of cineradio and labio-film data and articulatory-acoustic modeling. *Acoustical Society of America*, 109(5):2165–2180, 2001.

[19] T. Darrell and A. Pentland. Space-time gestures. In *Proc. CVPR*, pages 335–340, 1993.

[20] Paul E. Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. In *Proc. ACM SIGGRAPH*, pages 145–156, 2000.

[21] Douglas DeCarlo, Dimitris Metaxas, and Matthew Stone. An anthropometric face model using variational techniques. In *Proc. ACM SIGGRAPH*, pages 67–74, New York, NY, USA, 1998. ACM Press.

[22] Douglas DeCarlo and Dimitris N. Metaxas. Optical flow constraints on deformable models with applications to face tracking. *International Journal of Computer Vision*, 38(2):99–127, 2000.

[23] X. Deng. *A Finite Element Analysis of Surgery of Human Facial Tissues*. PhD thesis, Columbia University, 1988.

[24] Zhigang Deng, J. P. Lewis, and Ulrich Neumann. Automated eye motion using texture synthesis. *IEEE Comput. Graph. Appl.*, 25(2):24–30, 2005.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[25] H. Q. Dinh, G. Slabaugh, and G. Turk. Reconstructing surfaces using anisotropic basis functions. In *Proc. ICCV*, pages 606–613, Vancouver, Canada, July 2001.

[26] H. Q. Dinh, G. Turk, and G. Slabaugh. Reconstructing surfaces by volumetric regularization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(10):1358–1371, October 2002.

[27] Craig Donner and Henrik Wann Jensen. Light diffusion in multi-layered translucent materials. *ACM Trans. Graph.*, 24(3):1032–1039, 2005.

[28] Jean Duchon. Spline minimizing rotation-invariant semi-norms in Sobolev spaces. *Lecture Notes in Mathematics*, 571:85–100, 1977.

[29] David Eberly. *Distance Between Point and Triangle in 3D.* Geometric Tools, Inc.

[30] P. Eisert. Mpeg-4 facial animation in video analysis and synthesis. *International Journal of Imaging Systems and Technology*, 13(5):245–256, March 2003.

[31] P. Ekman and W. Friesen. Manual for the facial action coding system. *Consulting Psychologists Press*, 1978.

[32] Marc Escher, Igor S. Pandzic, and N. M. Thalmann. Facial deformations for MPEG-4. In *Proc. Computer Animation*, page 56, Washington, DC, USA, 1998. IEEE Computer Society.

[33] Irfan Essa, Sumit Basu, Trevor Darrell, and A. Pentland. Modeling, tracking and interactive animation of faces and heads using input from video. In *Proc. Computer Animation*, pages 68–79. IEEE Computer Society, 1996.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[34] Irfan A. Essa and A. Pentland. Coding, analysis, interpretation, and recognition of facial expressions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):757–763, 1997.

[35] Tony Ezzat and Tomaso Poggio. Miketalk: A talking facial display based on morphing visemes. In *Proc. Computer Animation*, pages 96–102, 1998.

[36] A. Fedorov, T. Firsova, V. Kuriakin, E. Martinova, K. Rodyushkin, and V. Zhislina. Talking head: Synthetic video facial animation in MPEG-4. In *Proc. the 13th GraphiCon*, 2003.

[37] Shachar Fleishman, Daniel Cohen-Or, and Claudio Silva. Robust moving least-squares fitting with sharp features. In *Proc. SIGGRAPH 2005*, pages 544–552, 2005.

[38] Michael S. Floater and Armin Iske. Multistep scattered data interpolation using compactly supported radial basis functions. *Journal of Computational and Applied Mathematics*, 73(1-2):65–78, 1996.

[39] M. Fratarcangeli and M. Schaerf. Realistic modeling of animatable faces in MPEG-4. In *Proc. Computer Animation and Social Agents*, Geneva, Switzerland, July 2004. MIRALAB, Computer Graphics Society (CGS).

[40] Martin Fuchs, Volker Blanz, Hendrik P. A. Lensch, and Hans-Peter Seidel. Reflectance from images: A model-based approach for human faces. *IEEE Trans. Visualization and Computer Graphics*, 11(3):296–305, May 2005.

[41] Y. C. Fung. *Biomechanics - Mechanical Properties of Living Tissues.* Springer-Verlag, 1993.

[42] Athinodoros S. Georghiades, Peter N. Belhumeur, and David J. Kriegman. Illumination-based image synthesis: Creating novel images of human faces

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

under differing pose and lighting. In *MVIEW '99: Proc. IEEE Workshop on Multi-View Modeling & Analysis of Visual Scenes*, page 47, Washington, DC, USA, 1999. IEEE Computer Society.

[43] Salih Burak Göktürk, Jean-Yves Bouguet, and Radek Grzeszczuk. A data-driven model for monocular face tracking. In *Proc. ICCV*, pages 701–708, 2001.

[44] T. Goto, S. Kshirsagar, and N. M. Thalmann. Automatic face cloning and animation using real-time facial feature tracking and speech acquisition. *IEEE Signal Processing Magazine*, 18:17–25, May 2001.

[45] Henry Gray. *Gray's anatomy : the anatomical basis of medicine and surgery.* Churchill Livingstone, 1995.

[46] Brian Guenter, Cindy Grimm, Daniel Wood, Henrique Malvar, and Fredric Pighin. Making faces. In *Proc. ACM SIGGRAPH*, pages 55–66. ACM Press, 1998.

[47] Pat Hanrahan and Wolfgang Krueger. Reflection from layered surfaces due to subsurface scattering. In *Proc. ACM SIGGRAPH*, pages 165–174, New York, NY, USA, 1993. ACM Press.

[48] Tim Hawkins, Andreas Wenger, Chris Tchou, Andrew Gardner, Fredrik Göransson, and Paul E. Debevec. Animatable facial reflectance fields. In *Proc. Eurographics Symposium on Rendering*, pages 309–321, 2004.

[49] S. Haykin. *Neural Networks: A Comprehensive Foundation.* Macmillan College Publishing Company, Inc., 1994.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[50] Pengyu Hong, Zhen Wen, and T.S. Huang. Real-time speech-driven face animation with expressions using neural networks. *IEEE Trans. Neural Networks*, 13:916–927, July 2002.

[51] Digital Elite Inc. Virtual human interface. http://www.digitalelite.net/Pages/DigitalElite/DEE_Technology_VHI.html.

[52] A. Iske and J. Levesley. Multilevel scattered data approximation by adaptive domain decomposition. Technical report, Technische Universität München, 2002.

[53] UK ITC, Germany IRT, UK TeleVirtual, Hamburg IDGS, Norwich UEA, Paris INT, Netherland IvD, and UK Post Office, UKRNID. Visicast. http://www.visicast.co.uk/.

[54] Henrik Wann Jensen and Juan Buhler. A rapid hierarchical rendering technique for translucent materials. In *Proc. ACM SIGGRAPH*, pages 576–581, New York, NY, USA, 2002. ACM Press.

[55] Henrik Wann Jensen, Stephen R. Marschner, Marc Levoy, and Pat Hanrahan. A practical model for subsurface light transport. In *Proc. ACM SIGGRAPH*, pages 511–518, New York, NY, USA, 2001. ACM Press.

[56] Pushkar Joshi, Wen C. Tien, Mathieu Desbrun, and Fredric Pighin. Learning controls for blend shape based realistic facial animation. In *SCA '03: Proc. the 2003 ACM SIGGRAPH/Eurographics symposium on Computer Animation*, pages 187–192, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.

[57] K. Kähler, J. Haber, and H. P. Seidel. Geometry-based muscle modeling for facial animation. In *Proc. the Graphics Interface*, pages 37–46, 2001.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[58] K. Kähler, J. Haber, and H. P. Seidel. Reanimating the dead: reconstruction of expressive faces from skull data. *ACM Trans. Graph.*, 22(3):554–561, 2003.

[59] K. Kähler, J. Haber, H. Yamauchi, and H. P. Seidel. Head shop: Generating animated head models with anatomical structure. In Stephen N. Spencer, editor, *SCA '02: Proc. the 2002 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 55–64, San Antonio, USA, July 2002. Association of Computing Machinery (ACM), ACM SIGGRAPH.

[60] P. Kalra, A. Mangili, N. M. Thalmann, and D. Thalmann. Smile: a multilayered facial animation system. In *Proc. IFIP WG 5.10*, pages 189–198, 1991.

[61] P. Kalra, A. Mangili, N. M. Thalmann, and D. Thalmann. Simulation of facial muscle actions based on rational free form deformations. volume 11, pages 59–69, 1992.

[62] M. Kass, A.P. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. Computer Vision*, 1(4):321–331, 1988.

[63] R. M. Kenedi, T. Gibson, J. H. Evans, and J. C. Barbenel. Tissue mechanics. *Physics in Medicine and Biology*, 20(5):699–717, February 1975.

[64] F. Kishino, K. Ebihara, and J. Ohya. Virtual space teleconferencing: Real-time detection and reproduction of 3D face and body images. In S. Z. Li, D. P. Mital, E. K. Teoh, and H. Wang, editors, *Recent Developments in Computer Vision*, pages 159–168. Springer, Berlin, 1995.

[65] F. Kishino, J. Ohya, H. Takemura, and N. Terashima. Virtual space teleconferencing system - real time detection and reproduction of 3D human images. In *Proc. of HCI*, pages 669–674, 1993.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[66] Rolf M. Koch and Albert A. Bosshard. Emotion editing using finite elements. *Computer Graphics Forum*, 17(3):295–302, 1998.

[67] Rolf M. Koch, Markus H. Gross, Friedrich R. Carls, Daniel F. von Büren, George Fankhauser, and Yoav I. H. Parish. Simulating facial surgery using finite element models. In *Proc. ACM SIGGRAPH*, pages 421–428. ACM Press, 1996.

[68] S. Kshirsagar, S. Garchery, and N. M. Thalmann. Feature point based mesh deformation applied to MPEG-4 facial animation. In *Proc. Deformable Avatars, IFIP TC5/WG5.10 DEFORM'2000 Workshop*, pages 24–34, 2000.

[69] T. Kurihara and K. Arai. A transformation method for modeling and animation of the human face from photographs. In *Proc. Computer Animation*, pages 45–58, 1991.

[70] Andreas Lanitis, Christopher J. Taylor, and Timothy Francis Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):743–756, 1997.

[71] Fabio Lavagetto and Roberto Pockaj. The facial animation engine: towards a high-level interface for the design of MPEG-4 compliant animated faces. 9(2):277–289, March 1999.

[72] Sooha Park Lee, Jeremy B. Badler, and Norman I. Badler. Eyes alive. In *Proc. ACM SIGGRAPH*, pages 637–644, New York, NY, USA, 2002. ACM Press.

[73] Y. Lee and K. Hwang. Skin thickness of korean adults. *Surgical and Radiologic Anatomy*, 24(3–4):183–189, 2002.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[74] Y. Lee, D. Terzopoulos, and K. Waters. Constructing physics-based facial models of individuals. In *Proc. Graphics Interface '93*, pages 1–8, 1993.

[75] Y. Lee, D. Terzopoulos, and K. Waters. Realistic modeling for facial animation. In *Proc. ACM SIGGRAPH*, pages 55–62. ACM Press, 1995.

[76] Aaron Lefohn, Richard Caruso, Erik Reinhard, Brian C. Budge, and Peter Shirley. An ocularist's approach to human iris synthesis. *IEEE Computer Graphics and Applications*, 23(6):70–75, November 2003.

[77] J. P. Lewis and F. I. Parke. Automated lip-synch and speech synthesis for character animation. In *CHI '87: Proc. the SIGCHI/GI conference on Human factors in computing systems and graphics interface*, pages 143–147, New York, NY, USA, 1987. ACM Press.

[78] A. Maciel, R. Boulic, and D. Thalmann. Deformable tissue parameterized by properties of real biological tissue. In *Proc. International Symposium on Surgery Simulation and Soft Tissue Modeling*, pages 74–87, 2003.

[79] Stephen R. Marschner, B. Guneter, and S. Raghupathy. Modeling and rendering for realistic facial animation. In *Proc. 11th Eurographics Workshop on Rendering*, pages 231–242, 2000.

[80] Stephen R. Marschner, Stephen H. Westin, Eric P. F. Lafortune, Kenneth E. Torrance, and Donald P. Greenberg. Image-based brdf measurement including human skin. In *Proc. 10th Eurographics Workshop on Rendering*, pages 139–152, June 1999.

[81] Walter Maurel. *Biomechanical models for soft tissue simulation.* Springer-Verlag, 1998.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[82] Bradley R. Wolf M.D. Wolfhair.com. http://wolfhair.com/home/.

[83] M. Meyer, M. Desbrun, P. Schroder, and A. H. Barr. Discrete differential geometry operators for triangulated 2-manifolds. In *Proc. VisMath '02*, 2002.

[84] B. S. Morse, T. S. Yoo, P. Rheingans, D. T. Chen, and K. R. Subramanian. Interpolating implicit surfaces from scattered surface data using compactly supported radial basis functions. In *Proc. Shape Modeling International*, pages 89–98, Genova, Italy, May 2001.

[85] L. Moubaraki, J. Ohya, and F. Kishino. Realistic 3D facial animation in virtual space teleconferencing. In *Proc. the 4th IEEE International workshop on Robot and Human Communication*, pages 253–258, 1995.

[86] S. Muraki. Volumetric shape description of range data using *Blobby Model*. In *Proc. ACM SIGGRAPH*, pages 227–235, 1991.

[87] Kyunggun Na and Moonryul Jung. Hierarchical retargetting of fine facial motions. *Computer Graphics Forum*, 23(3):687–695, 2004.

[88] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. Geometric considerations and nomenclature for reflectance. Technical report, Monograph 160, National Bureau of Standards(US), 1977.

[89] Y. Ohtake, A. G. Belyaev, and H.-P. Seidel. A multi-scale approach to 3d scattered data interpolation with compactly supported basis functions. In *Proc. Shape Modeling International*, pages 153–161, Seoul, Korea, May 2003.

[90] J. Ostermann. Animation of synthetic faces in mpeg-4. In *Proc. Computer Animation*, pages 49–55, Washington, DC, USA, 1998. IEEE Computer Society.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[91] Nihat Ozkaya and Margareta Nordin. *Fundamentals of biomechanics : Equilibrium, Motion, and Deformation.* Springer, 1999.

[92] Igor S. Pandzic. Facial animation framework for the web and mobile platforms. In *Proceeding of the seventh international conference on 3D Web technology*, pages 27–34. ACM Press, 2002.

[93] Sylvain Paris, François Sillion, and Long Quan. Lightweight face relighting. In *Proc. Pacific Graphics*, October 2003.

[94] F. I. Parke. Computer generated animation of faces. In *Proc. ACM'72*, pages 451–457, 1972.

[95] Catherine Pelachaud, Norman I. Badler, and Mark Steedman. Generating facial expressions for speech. *Cognitive Science*, 20(1):1–46, 1996.

[96] Frédéric Pighin, Jamie Hecker, Dani Lischinski, Richard Szeliski, and David H. Salesin. Synthesizing realistic facial expressions from photographs. In *Proc. ACM SIGGRAPH*, pages 75–84. ACM Press, 1998.

[97] Frederic H. Pighin, Richard Szeliski, and David Salesin. Resynthesizing facial animation through 3D model-based tracking. In *Proc. ICCV*, pages 143–150, 1999.

[98] Stephen M. Platt and Norman I. Badler. Animating facial expressions. In *Proc. ACM SIGGRAPH*, pages 245–252. ACM Press, 1981.

[99] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 1993.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[100] Hyewon Pyun, Yejin Kim, Wonseok Chae, Hyung Woo Kang, and Sung Yong Shin. An example-based approach for facial expression cloning. In *SCA '03: Proc. 2003 ACM SIGGRAPH/Eurographics symposium on Computer Animation*, pages 167–176, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.

[101] J. Rossignac and A. Requicha. Constant-radius blending in solid modeling. *Computers in Mechanical Engineering*, 3.

[102] Basu S., N. Oliver, and A. Pentland. 3D modeling and tracking of human lip motion. In *Proc. ICCV*, pages 337–343, 1998.

[103] Basu S., N. Oliver, and A. Pentland. 3D modeling and tracking of human lip motion. In *Proc. ICCV*, pages 337–343, 1998.

[104] Pedro V. Sander, David Gosselin, and Jason L. Mitchell. Real-time skin rendering on graphics hardware. In *Proc. ACM SIGGRAPH Conference on Abstracts & Applications (Sketch)*, New York, NY, USA, 2004. ACM Press.

[105] Eftychios Sifakis, Igor Neverov, and Ronald Fedkiw. Automatic determination of facial muscle activations from sparse motion capture marker data. *ACM Trans. Graphics.*, 24(3):417–425, 2005.

[106] Eftychios Sifakis, Andrew Selle, Avram Robinson-Mosher, and Ronald Fedkiw. Simulating speech with a physics-based facial muscle model. In *SCA '06: Proc. the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 261–270, 2006.

[107] Jos Stam. An illumination model for a skin layer bounded by rough surfaces. In *Proc. 12th Eurographics Workshop on Rendering Techniques*, pages 39–52, London, UK, 2001. Springer-Verlag.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[108] Joseph Teran, Eftychios Sifakis, Silvia S. Blemker, Victor Ng-Thow-Hing, Cynthia Lau, and Ronald Fedkiw. Creating and simulating skeletal muscle from the visible human data set. *IEEE Trans. Visualization and Computer Graphics*, 11(3):317–328, 2005.

[109] D. Terzopoulos and K. Waters. Physically-based facial modeling, analysis, and animation. *Visualizaion and Computer Animation*, 1(2):73–80, March 1990.

[110] D. Terzopoulos and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(6):569–579, 1993.

[111] D. Terzopoulos and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(6):569–579, 1993.

[112] Matthias Teschner, Bruno Heidelberger, Matthias Muller, and Markus Gross. A versatile and robust model for geometrically complex deformable solids. In *Proc. Computer Graphics International (CGI'04)*, pages 312–319, 2004.

[113] N. M. Thalmann, E. Primeau, and D. Thalmann. Abstract muscle action procedures for human face animation. *The Visual Computer*, 3(5):290–297, March 1988.

[114] U.S. National Library of Medicine. The visible human project, 1994.

[115] Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159(1):98, Jul 1967.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[116] Daniel Vlasic, Matthew Brand, Hanspeter Pfister, and Jovan Popović. Face transfer with multilinear models. *ACM Trans. Graph.*, 24(3):426–433, 2005.

[117] Keith Waters. A muscle model for animation three-dimensional facial expression. In *Proc. ACM SIGGRAPH*, pages 17–24. ACM Press, 1987.

[118] H. Wendland. Piecewise polynomial, positive definite and compactly supported radial basis functions of minimal degree. *Advances in Computational Mathematics*, 4:389–396, 1995.

[119] Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. *ACM Trans. Graph.*, 24(3):756–764, 2005.

[120] Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, Bernd Bickel, Craig Donner, Chien Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, and Markus Gross. Analysis of human faces using a measurement-based skin reflectance model. *ACM Trans. Graph.*, 25(3):1013–1024, 2006.

[121] Lance Williams. Performance-driven facial animation. In *Proc. ACM SIGGRAPH*, pages 235–242, 1990.

[122] J. Woodwark. Blends in geometric modeling. In R. Martin, editor, *Proc. The Mathematics of Surfaces II*, pages 255–297, Oxford, UK, 1987. Clarendon Press.

[123] Jun yong Noh, Douglas Fidaleo, and Ulrich Neumann. Animated deformations with radial basis functions. In *VRST '00: Proc. the ACM symposium on Virtual reality software and technology*, pages 166–174, New York, NY, USA, 2000. ACM Press.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

[124] Jun yong Noh and Ulrich Neumann. Expression cloning. In *Proc. ACM SIGGRAPH*, pages 277–288, New York, NY, USA, 2001. ACM Press.

[125] Li Zhang, Noah Snavely, Brian Curless, and Steven M. Seitz. Spacetime faces: high resolution capture for modeling and animation. *ACM Trans. Graph.*, 23(3):548–558, 2004.

[126] Y. Zhang. *3D Human face modeling for dynamic facial expression synthesis.* PhD thesis, School of Electrical and Electronic Engineering, Nanyang Technological University, 2002.