國立交通大學

資訊科學與工程研究所

碩士論文

應用於人臉動畫的表面細紋重建技術之研究

Estimating Facial Details by Space-time Shape-from-shading for 3D Animation

研究生：羅永盛

指導教授：林奕成　教授

中華民國 九十六 年七月

應用於人臉動畫的表面細紋重建技術之研究

Estimating Facial Details by Space-time Shape-from-shading for 3D Animation

研究生：羅永盛　　Student：Yung-Sheng Lo

指導教授：林奕成　　Advisor：I-Chen Lin

國立交通大學

資訊科學與工程研究所

碩士論文

A Thesis

Submitted to Institute of Computer Science and Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

July 2007

Hsinchu, Taiwan, Republic of China

中華民國九十六年七月

應用於人臉動畫的表面細紋重建技術之研究

研究生：羅永盛　指導教授：林奕成 教授

國立交通大學

資訊工程與科學研究所

摘要

在本篇論文中，我們提出一個可應用於三維動畫中之人臉表面細紋的重建技術。我們利用傳統的動態捕捉技術來產生出基本的三維人臉，並使用我們所提出來的改良 shape-from-shading 方法來增進細紋部份。我們主要是針對在複合空間中，利用空間與時間的相關性來重建三維資訊。 此利用有時間與空間相關性的 shape-from-shading 方法，可以得到一個較為準確的結果，並且可以減少因為瞬時的擾動所造成的雜訊現象。 為了能夠避免 shape-from-shading 中固有條件不足之情況，我們利用漸進最佳化的方式去得到三維的幾何資訊以及 Phong 反射模型 上的光影參數。在人臉動畫方面，我們會採用多邊型細分及法向量差圖的方式，將所擷取之表面細紋與人臉模型結合。利用本論文之方法，將可呈現出具細紋且更擬真之三維人臉動畫。

關鍵字：動態捕捉，時間及空間限制條件，shape-from-shading，法向量圖

Estimating facial details by space-time shape-from-shading for 3D animation

Student: Yung-Sheng Lo Advisor: Dr. I-Chen Lin

Institute of Computer Science and Engineering
National Chiao Tung University

ABSTRACT

In this thesis, we propose a facial details estimation approach for 3D animation. We utilize a motion capture technique to create the primitive 3D model and apply our novel shape-from-shading (SFS) for facial details. We exploit a space-time hybrid domain for recovering the 3D information. Space-time shape-from-shading can resolve the noise and ambiguity and get a more reliable result. In order to avoid the intrinsic ill-condition, an optimization method is proposed to approximate both the 3D face geometry and Phong reflectance properties. On rendering framework, the height and normal maps of facial details can be acquired from the image. While combing facial details with fundamental face model, our system can synthesize more detailed facial animation.

Keywords: motion capture, space-time constraints, shape-from-shading, normal maps

Acknowledgements

First of all, I would like to thank my advisor, Dr. I-Chen Lin, for his guidance in the past two years. Also, I appreciate all members of Computer Animation & Interactive Graphics Lab for their help and comments. Finally, I am grateful to my family for their support and encouragement.

Contents

List of Figures

Chapter 1
Introduction

1.1 Background

Nowadays, 3D characters and the virtual scene have been popularly utilized in various kinds of media. Usually animator has to manually adjust key poses of 3D models in order to vividly animate the characters; however, generating key poses is still a labor-intensive work for animators. To speed up the production of 3D animation motion capture (mocap) techniques are popularly utilized. An actor is usually asked to wear the specific clothing with conspicuous markers on specific features. While tracking these markers, the dynamic variations of features can be recorded. Although mocap data can efficiently drive the 3D models but there are still subtle portions, such as wrinkles or creases, whose variations are much smaller than the markers' size. These details are difficult to be acquired by mocap techniques but are critical for facial animation. Therefore, we propose extracting the facial details by 3D reconstruction method.

In computer vision and graphics, 3D reconstruction and surface reflectance estimation from the images have been a classic and essential problem for a long time. For 3D reconstruction, various algorithms have been proposed. Stereo triangulation is the most typical method to calculate the 3D positions. But the pixel correspondence of un-textured regions is usually ambiguous. Structured light system, using a camera and a laser scanner to acquire 3D data, is another popular method. The correspondence is more reliable but it requires quite high-resolution devices for facial details.

Other methods for shape reconstruction are shape-from-shading (SFS) and photometric stereo. Most of the existing methods belonging to SFS and photometric stereo are under the assumption of the Lambertian reflectance model. Photometric stereo recovers the shape from multiple images in a fixed view direction, but different light source directions. Using photometric stereo for dynamic objects will require expensive high-speed cameras and a light dome. A SFS technique recovers the surface shape from the variation of intensity in the image. Most SFS techniques apply a single light source and assume in the simple lighting condition. Hence the problem of pixel correspondence will be avoided. But it will face another problem, sensitivity to noise. Therefore, in this thesis, we adopt the shape-from-shading (SFS) technique with space-time optimization to reconstruct the 3D details.

1.2 Framework

The goal of this thesis is to enhance the feature-driven animation with facial details. The proposed framework can be divided into two parts. Off-line process reconstructs the 3D detailed surface and estimates the reflectance parameters. Online process is about the rendering of facial animation. Fig.1 shows a flowchart of our system.

For 3D reconstruction, Zhang et al. [1] produced the space-time structured light system which can capture the dynamic variations on a face. But the details, such as wrinkles on the forehead which can be easily observed, are difficult to be acquired by the structured light system. Photometric stereo is another method that can be used. Although the details can be acquired accurately, the experiment devices for dynamic scenes are expensive. To acquire reliable data for facial animation, we have to record at least eight images illuminated from different light directions per 1/30 second.

In this thesis, I applied the SFS technique to reconstruct the 3D shape. SFS with a single image could avoid the errors resulting from the inaccurate pixel correspondence and the facial detail can be acquired simply. However a simple SFS with the lambertian model is error-prone. It can easily be tainted by input and digitization noise. Therefore, I applied an optimization method for recovering the 3D shape to obtain more accurate results.

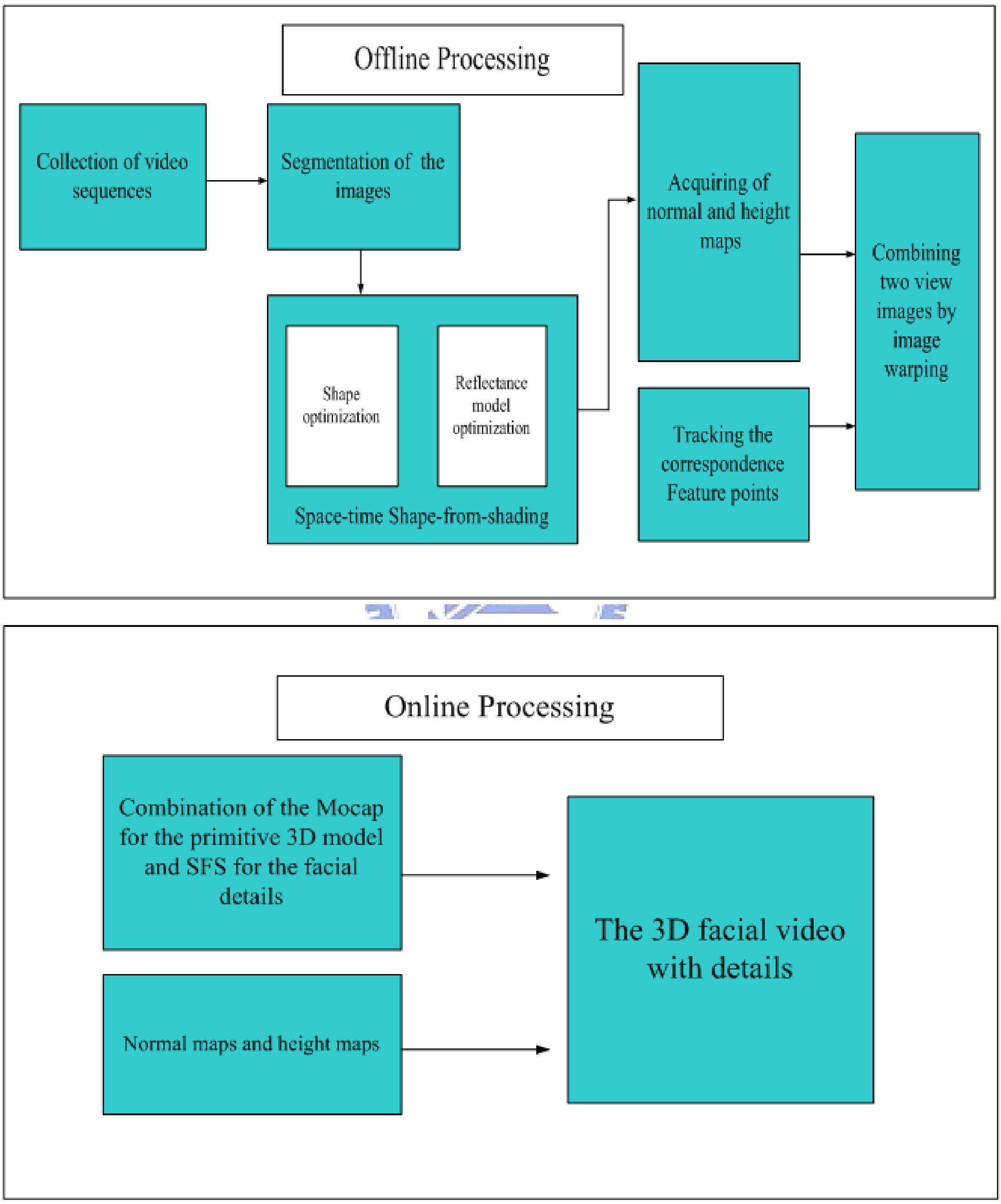

Offline Processing
Collection of video sequences
Segmentation of the images
Shape optimization
Reflectance model optimization
Space-time Shape-from-shading
Acquiring of normal and height maps
Tracking the correspondence Feature points
Combining two view images by image warping
Online Processing
Combination of the Mocap for the primitive 3D model and SFS for the facial details
Normal maps and height maps
The 3D facial video with details

Figure 1: The flowchart of our system

SFS can capture relative undulation on the subtle areas, such as wrinkles, because it is sensitive to intensity variation. But absolute geometry properties, such as depth, are unreliable. On the other hand, the motion capture technique can capture the features reliable. Therefore in our thesis we propose combining the motion capture (MoCap) technique for the primitive 3D model and SFS for facial details.

In computer graphics, Bidirectional Reflectance Distribution Function (BRDF) is widely used to represent the reflectance model of the human face. Most of the works for human BRDF acquisition assumed the 3D geometry is known.

In fact, human faces are composed of multi layers, including the oil layer, epidermis and dermis. Skin reflectance can be explained by a specular component at the oil-air interface and a diffuse reflectance component due to subsurface scattering. Weyrich et al. [2] proposed a practical skin reflectance model whose parameters can be robustly estimated from measurements. However, reflectance model is complex and the experiment devices are expensive.

On the other hand, Phong model is widely used in computer graphics. It consists of simpler parameters and is more efficient for parameter estimation. In this thesis, we apply the Phong model and approximate its parameters through an optimization method. The proposed optimization method minimizes the difference between the captured real image and the synthetic image. Furthermore additional constraints such as smooth constraints or integrated constraints can be used to find a more reliable result.

Since each image is computed independently, it can lead to produce the noise caused by digitization etc. Solving the temporal coherence can reduce these noises. Besides, the temporal constraint will also be added by tracing the variation of pixel fame-by-fame.

1.3 Contributions

This thesis is aimed at simultaneous recovery the 3D shape and surface reflectance parameters of a non-lambertian object. Including both the spatial and temporal coherence can improve the reliability if reconstruction and produce more smooth animation. The main contributions of this thesis are as follow:

1. An advanced SFS is proposed where the spatial and temporal coherences are utilized to acquire the more accurate and reliable results.
2. Estimation of surface reflectance model of real persons by the optimization method.

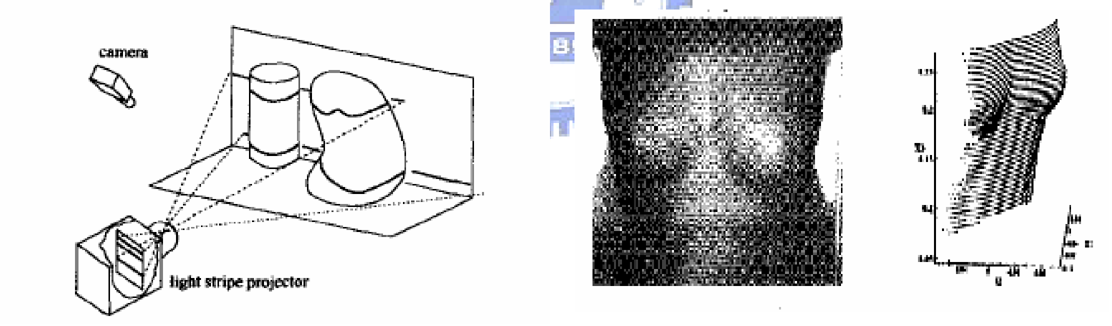
1.4 Organization

This thesis is organized as follows. Chapter 2 introduces related works about the shape and reflectance model reconstruction. Chapter 3 proposed an optimization method to solve the objective function for acquiring the 3D information; various issues such as initialization, spatial and temporal constraints are also discussed. Charter 4 proposed the framework of facial animation. Charter 5 gives the experiment result from real data and synthetic ones. Charter 6 presents the conclusions and the future work.

Chapter 2
Related work

2.1 3D reconstruction

For 3D reconstruction based on stereo triangulation, Du Q. [3] proposed a projective calibration method for a structured light system. This method used 4 known non-coplanar sets of 3 collinear world points and computers 3D points that fall onto each light stripe plane. Such active stereo system partially solves the problem by the pixel correspondence but it requires very high-resolution devices for the object detail. Fig. 2 shows the structured light system with a camera and a projector. The rough 3D shape can be recovered correctly but the skin detail is difficult acquired.



(a) (b)

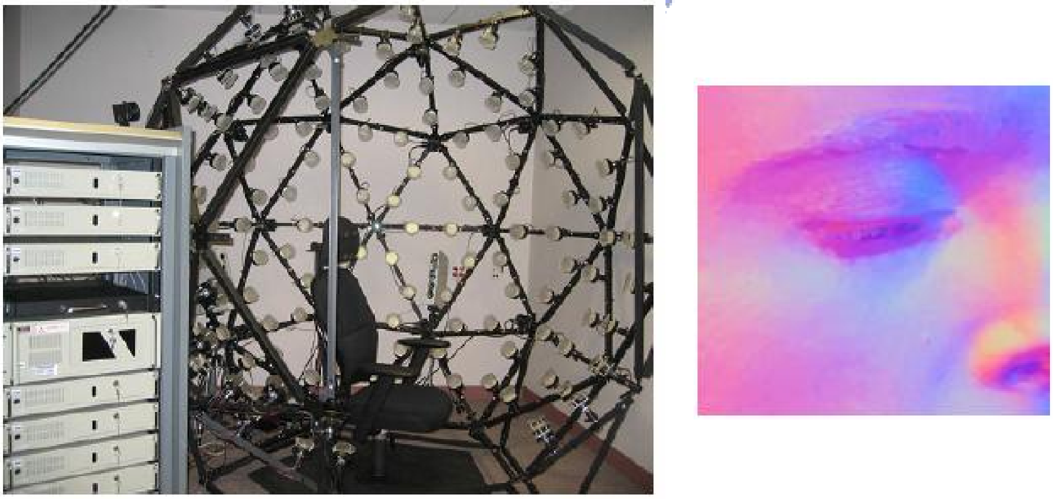
Figure 2: (a) The structured light system (b) input image and the shape reconstruction [3]

In the context of shape recovery from the images, there are two main researches: photometric stereo and shape-from-shading. Photometric stereo is usually conducted in a fixed view direction with different light source direction. According to the Lambertian model, it can estimate the surface normal by a least-square solution.

$$\begin{bmatrix} L_1 \\ L_2 \\ L_3 \end{bmatrix} \begin{bmatrix} N_x \\ N_y \\ N_z \end{bmatrix} = \begin{bmatrix} I_1 \\ I_2 \\ I_3 \end{bmatrix} \tag{1}$$

where *L, N, I* denote the light source, surface normal and intensity.

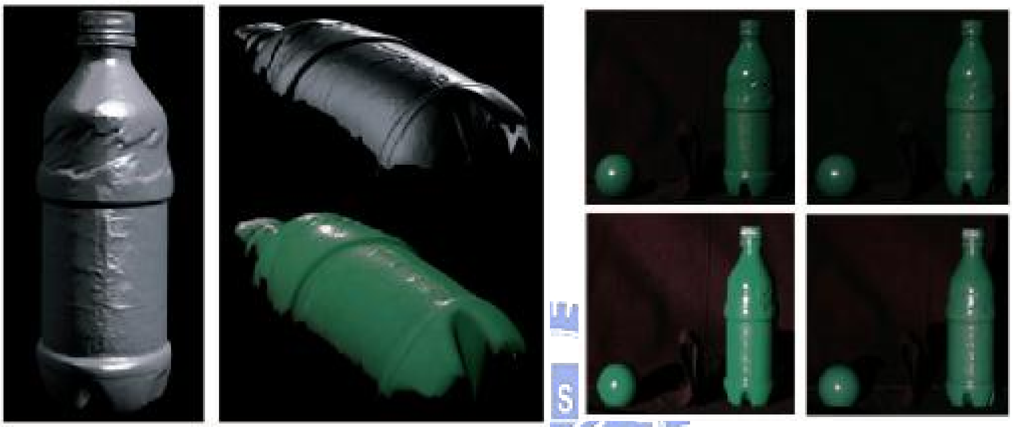
Generally, to acquire reliable normal by a photometric stereo method, we will need more than eight images of different light directions. For capturing dynamically moving objects, we have to exploit a high speed camera to record images at more than 240 fps. In Weyrich's [2] research, they produced a face-scanning dome consists of 16 digital cameras, 150 light sources and a commercial 3D face-scanning system. This experiment can acquire the facial detail accurately but it is a high-cost method. Fig. 3 shows the experiment and the acquiring normal maps.



(a) (b)

Figure 3: (a) The face-scanning dome (b) Normal maps [2]

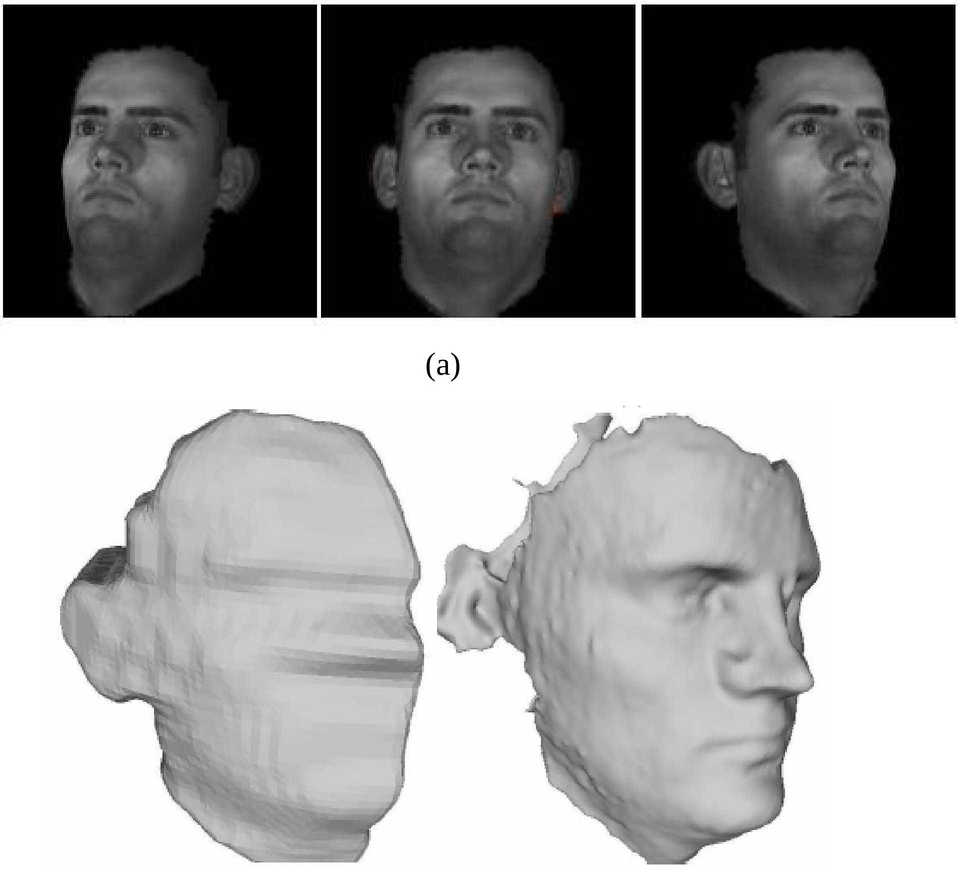
M. seitz et al. [4] proposed the example-based photometric stereo method. It introduced orientation-consistency to reconstruct the surface normal from the reference images. The object details can be obtained by this method but the constraints like the distant lighting, no cast shadows and materials etc. will limit the selection of the reference images. Fig. 4(a) shows the bottle result of the 3D recovery. Fig 4(b) illustrates images with different light source directions.



(a) (b)

Figure 4: (a) The result of the 3D reconstruction (b) Reference images [4]

Another method of 3D recovery, Vogiatzis et al. [5] proposed a novel multi-view stereo reconstruction method. This method uses the visual hull as the initial volumetric shape and optimizes the global solution by applying the graphic-cuts. It can provide a view-dependent 3D surface and handle the occlusion problem. Fig 5(a) shows the multi-view input images. Fig 5(b) shows the visual hull generated form silhouettes of the face. Fig 5(c) is the reconstructed 3D face by their method.



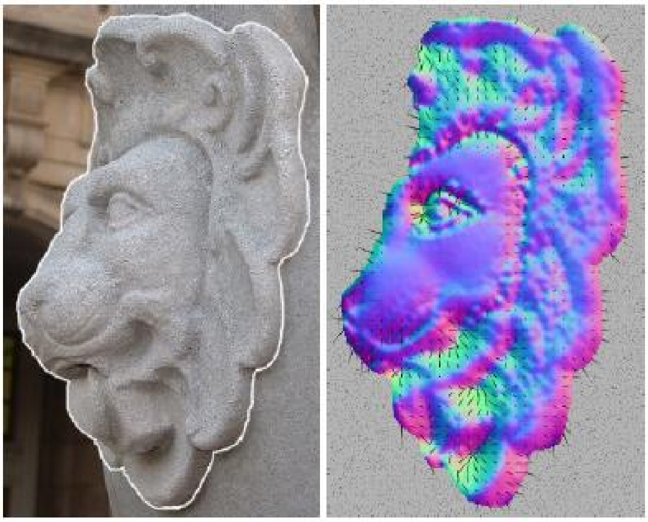
(a)

(b) (c)

Figure 5: (a)input images (b)the visual hull (c)The reconstructed 3D face [5]

2.2 Shape-from-shading

SFS can avoid the error results from the inaccurate pixel correspondence and it is a low-cost method. Fang et al. [6] adapted Horn’s [7] approach and simply utilized Lambertian reflection model to extract the normal map from a single image. This approach spends less time and doesn’t need expensive experiment. However, this approach is error-prone due to the simple Lambertian assumption and input noise etc. Manually adjustments are required for post processing. Fig. 6 shows the input image and the result normal map.



(a) (b)

Figure 6:(a) the input image (b) a normal map [6]

On the other hand, SFS is difficult for real data due to its intrinsic ill-condition. In interactive modeling, Zeng at al. [8] proposed a global solution of continuous surface. Users input surface normal on specific feature points and the system refines the surface variations to the whole face. This method applied a Fast Marching Method (FMM) to speed up the SFS technique and efficiently solved its ambiguity by human assistance. Fig. 7 shows the flowchart of interactive shape-from-shading. Users input the normal and automatically segmentation the image into local patches. Local surfaces are then reconstructed individually

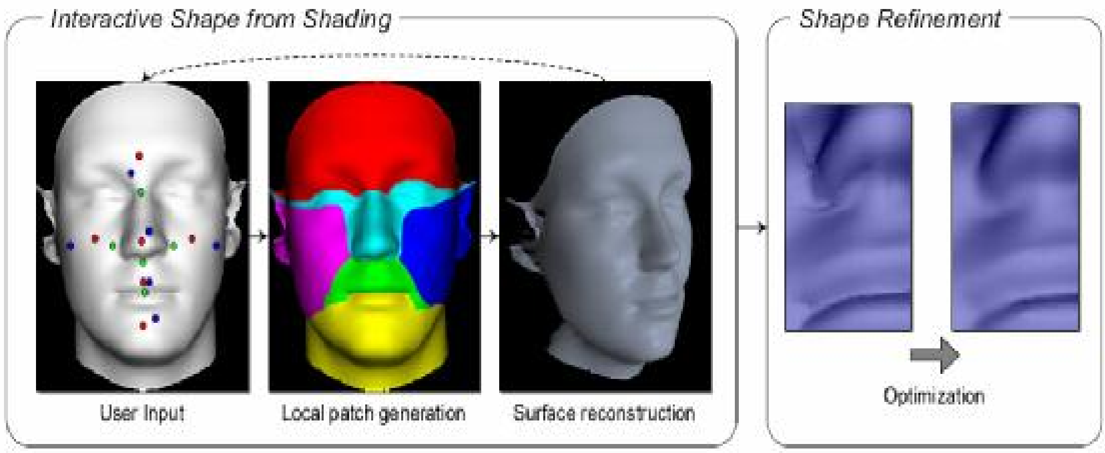
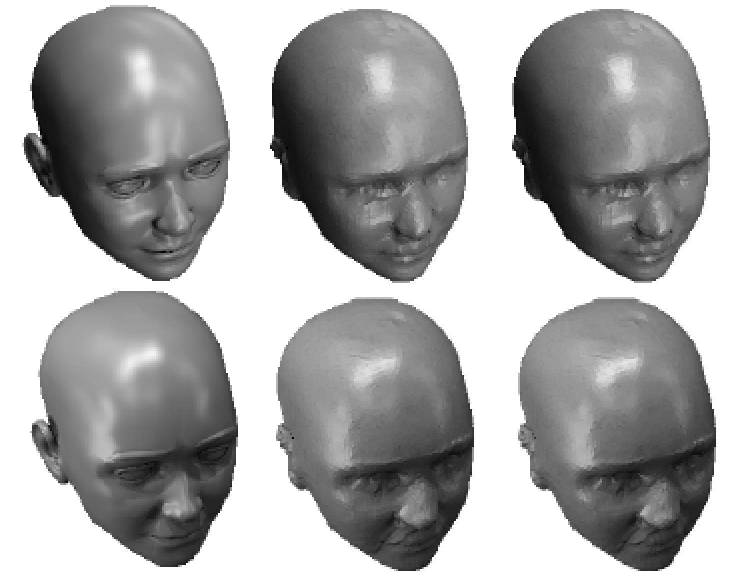


Figure 7: interactive shape from shading. [8]

Ahuja et al. [10] proposed an optimization method for obtaining the shape and reflectance parameters on a static model with non-lambertian model. This method initiated the reflectance parameters from different scales and further refined the estimation with multi view information to acquire the reliable result. However this method doesn't yet apply their approach to real persons. Fig. 8 shows the comparison of the input image and the synthesis image. The ground truth image is the enfant data sets composed of 12 views placed around a computer-rendered human head model with a single material. This method applies the Richardson extrapolation which makes the specualr part a little bit sharper.



(a) (b) (c)

Figure 8: (a) input images with different light directions (b) Synthesis images using estimated shape and Phong parameters (c) Synthesis images with extrapolated Phong parameters. [10]

2.3 spatial and temporal constraints

To help stabilize the iterative SFS algorithm, constraints such as smoothness, integrability and intensity gradient constraints will be used to obtain the reliable result.

In contrast to static scans, Zhang et al. [1] proposed a space-time structured light approach to capture the dynamic facial variation. They also presented a keyframe interpolation technique to synthesize video frames and a controllable face model. The points sampled by the structured light system can be easily constructed in video hyper-volume. Fig. 9 shows the facial animation with temporal constraints. This method use two depth maps from two view points to warp a template face model by estimating feature correspondence.

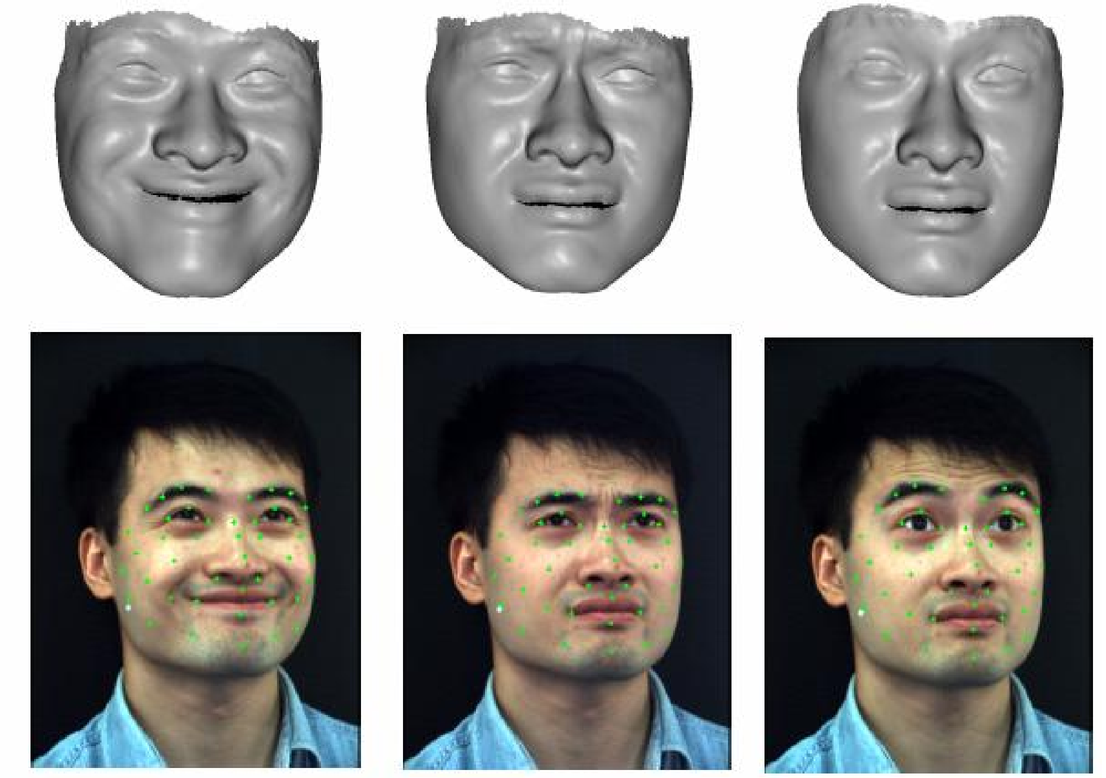


Figure 9: Using both depth maps and optical flow to produce the facial animation. [1]

Fang et al. [11] proposed a video editing system that allows user can apply a time-coherence texture to surface patches and use the RotoTexture synthesis technique for texture mapping. The temporal smooth constraint is applied to reduce the visual noise. Fig. 10 shows the ability to map an image onto a surface in a video.

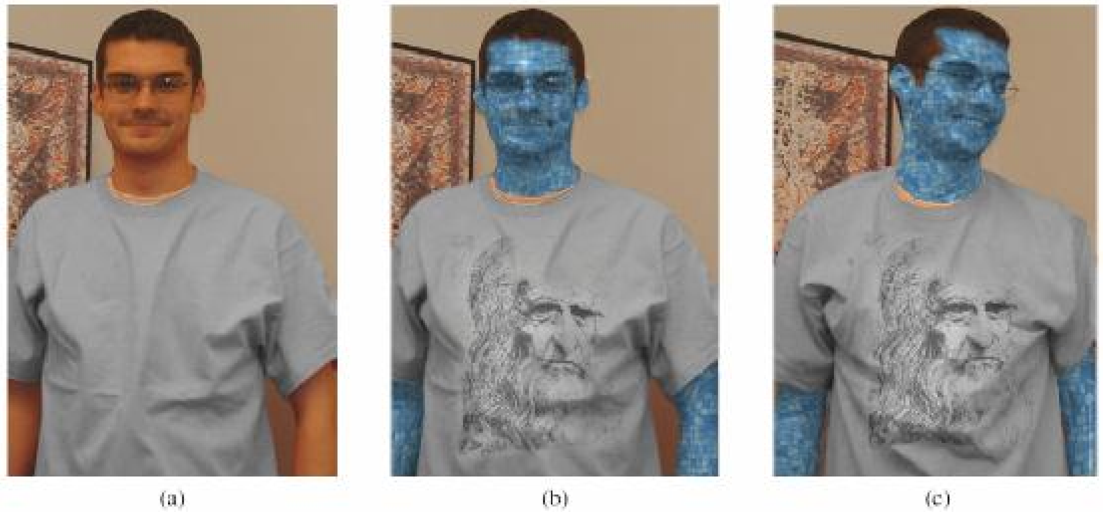


Figure 10: (a) an image from the input video sequence (b) replacing the skin with blue tile (c) Synthesis the time-coherence texture. [11]

Chapter 3
Facial Detail Estimation

Our thesis is to enhance the mocap technique with the additional facial details. The primitive 3D model is estimated by a feature-point-driven face deformation and the facial details are acquired by 3D recovery techniques. For 3D recovery, shape-from-shading can avoid the pixel correspondence problem and does not need lots of image inputs such as those in photometric stereo methods. Therefore, in this thesis, we adopt a novel shape-from-shading technique to reconstruct the facial details.

3.1 Preprocessing Framework

Generally, Shape-from-shading assumes that the target surface composed of only one material. For this reason, we must divide the input image into segments. Before segmentation, all of the expression images are automatic warped to the neutral image. Fig. 11 shows the mask images assigned by users. We prefer areas with more wrinkles such as the forehead, glabella, left and right cheek.

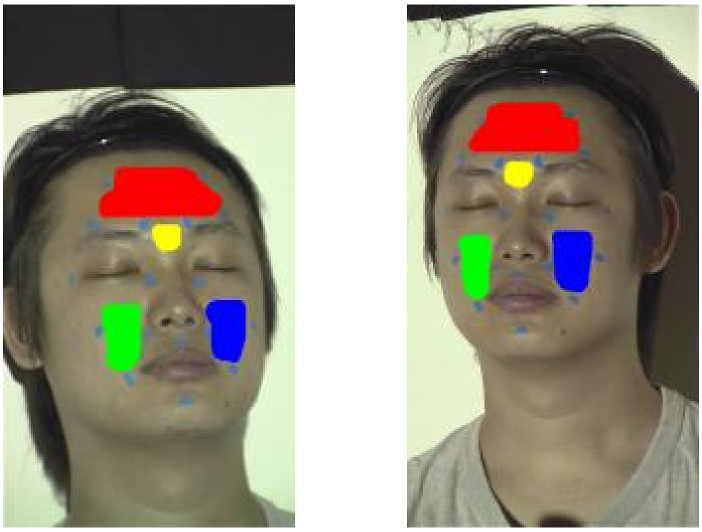


Figure 11: The mask images of left and right view.

3.2 Problem formulation

Our proposed reconstruction method for details is based on the shape-from-shading technique and space-time constraints. It recoveries the 3D shapes from a sequence of video according to the surface intensity variation and the spatial and temporal coherence. In order to solve the intrinsic shape-from-shading ill-condition, we adopt the optimization method to recover the 3D data.

Let V denotes the set of shape parameters. In order to decrease the dimensions of the cost function, we represent the 3D data in terms of height maps, and therefore, we only have to optimize the z values. Hence the set of shape parameters V can be defined as:

$$V=\{z_i\} \ \forall i=1\sim \text{ point number} \tag{2}$$

Our reflectance model is Phong model, since it is widely used in the computer graphics and parameters can be efficiently acquired. Given a light source L and the surface normal N, the Phong reflection model can be written as:

$$I=|L|(K_d(L\cdot N)+K_s(e\cdot r)^a) \tag{3}$$

, where K_d and K_s are the diffuse and specular coefficients and a is the Phong exponent term. The vector e denotes the eye direction and r is the reflection vector at the mirror direction of the light source with respect to the surface normal.

Let R denotes the reflectance parameters which can be defined as:

$$R=\{Kd, Ks, \alpha\} \tag{4}$$

Our cost function C can be defined as the sum of the square error between the input real image I and synthesis image S,

$$C(V,R)=\sum_{i}(I_i-S_i(V,R))^2 \quad i=1\sim number\ of\ \ pixel \qquad (5)$$

Fig. 12 demonstrates the flow chart of the minimization of the objective function. The synthesis data will be iteratively refined .

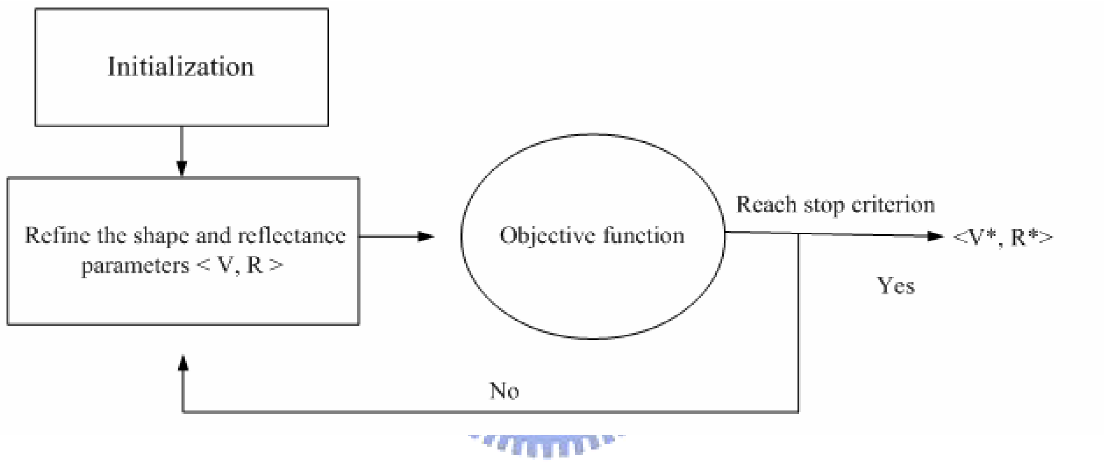


Figure 12: The flowchart of the minimization of the cost function

In other words, our goal is to find the shape and reflectance parameters $<V,R>$ that will minimize the cost function:

$$<V^*,R^*>=\arg\min\{C(V,R)\} \qquad (6)$$

3.3 Iterative Framework

Our cost function has two sets of parameters, the shape parameters V and reflectance model parameters R. The parameters R is the global parameters for all the triangle facets but the parameters V represent only local geometry. If we refinement these two parameters into a single optimization procedure, we should choose the proper scales of the two kinds of parameters to balance the effects. In order to avoid unbalance condition, we optimize these two parameters separately to obtain the more accurate result. The flow chart of the optimization method is showed below,

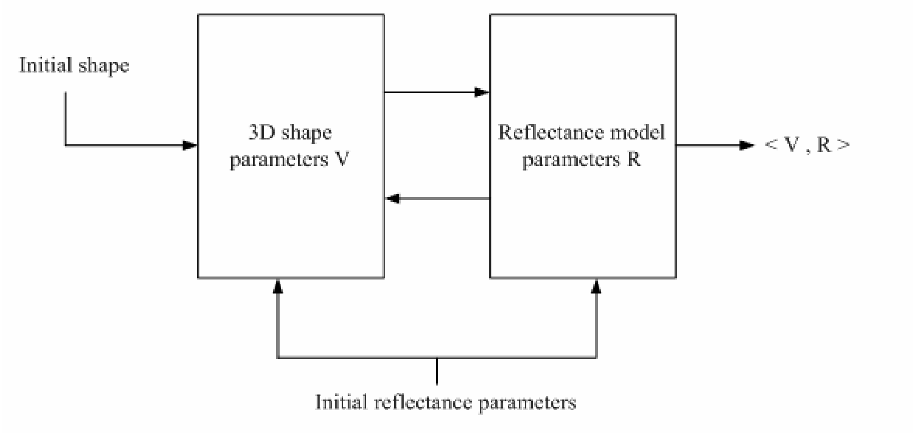

Initial shape
3D shape parameters V
Reflectance model parameters R
< V , R >
Initial reflectance parameters

Figure13: Flow Chart of the optimization algorithm

First, we assign the initial shape and reflectance parameters to this system. Different initial conditions will influence the optimization results. We can adjust the parameters manually or apply the batch work. Afterward, the reflectance parameters will fix and the shape parameter V is refined.

In order to optimize the initial shape more reliably, we just apply the diffuse model in the first phase. specular terms will be included on the following phases. After optimizing the shape parameters, the reflectance parameters R will be refined while the optimized parameter V is fixed. We will repeat these steps until the cost value is small than the threshold.

3.4 Parameters optimization

We treat the minimization of the cost function as a non-linear least square problem and solve its solution by the conjugate gradient method.

According to the conjugate gradient, the cost function should be differentiated by the variable of shape parameters V,

$$\frac{\partial C}{\partial V} = \frac{\sum_{i}(I_i - S_i)^2}{\partial V} \tag{7}$$

Expanding the function becomes

$$\frac{\partial C}{\partial V} = \sum_{i}\left(\frac{I}{\partial V} - \left(\frac{K_d(L \cdot N)}{\partial V} + \frac{K_s(e \cdot r)^a}{\partial V}\right)\right)^2 \tag{8}$$

The $I/\partial V$ can be approximated using the observed image gradient and the $\frac{K_d(L \cdot N)}{\partial V}$ term should transform the surface normal to the position representation. Fig 14 shows the normal estimated by the cross product. Given a surface, we can approximate the surface normal by using the cross product of the two vectors. In order to reduce the error caused by the approximation, normal estimated by other pairs of tangent vectors should be included for a more reliable result.

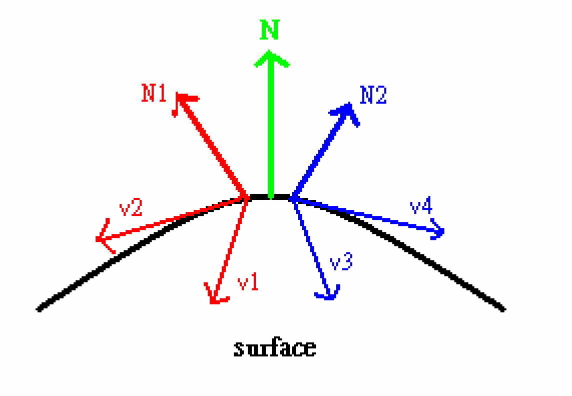


Figure 14: the surface normal N

The $\frac{\partial K_s (e \cdot r)^a}{\partial V}$ term can also be evaluated in the same way, but the reflection vector should be represented in term of *N*, *L*:

$$r = 2N - L \tag{9}$$

In order to acquire more accurate results, we only adjust the z component to reduce the dimension of the optimization function.

Our reflectance parameters $R = \{K_d, K_s, a\}$ can also be computed by finite differencing as the shape optimization. In general, we normalized the parameters of Phong model and assumed $K_d + K_s = 1$.

3.5 Space-time shape-from-shading

In this section, we present a method to recovery time-varying normals and height maps from the two synchronized video streams. The first step is to find correspondence between images. The traditional stereo matching algorithms find the correspondence pixel on the left and right images by searching the similar intensity or color. For a pixel in the left image, it will search more than one pixel with similar intensity may becomes candidates in the right image. To resolve this ambiguity, we can use the small matching window to obtain the local value under the specified threshold. In this thesis, we adopt a simple adaptive image warping method to combine the two synchronized video streams. More precisely, given two video streams $I_L(x, y, t), I_R(x, y, t)$, a time-varying normal map can be acquired by the above optimization method and adaptive image warping method.

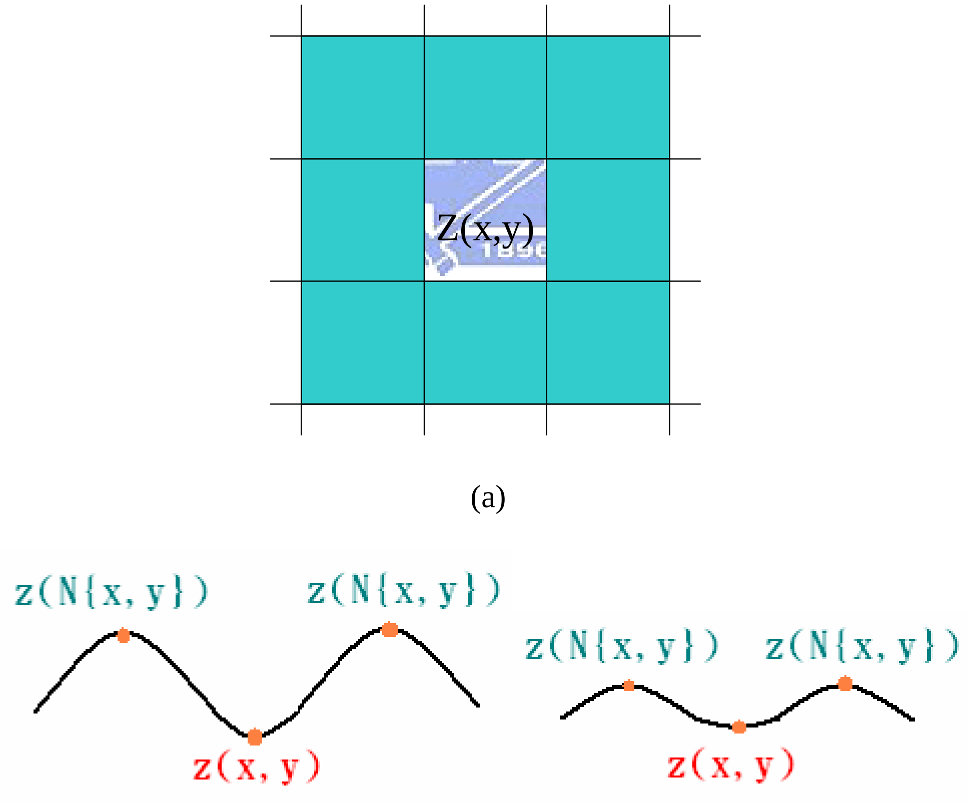
It will produce too much noisy, since every time-varying normal map is acquired independently and temporal flicker as shape changes discontinuously form this frame to the next. In order to solve this problem, we should add the spatial and temporal constraints to obtain the more reliable result.

3.5.1 Spatial constraints

In this thesis, we propose the space-time shape-from-shading to recover the 3D data. To help stabilize the iterative shape-from-shading and to obtain the reliable result, we use a spatial constraints as

$$Spatial = \iint [z(x,y) - z(Neighbor\{x,y\})]^2 \, dxdy \tag{10}$$

where Neighbor{x.y} denote the neighbor pixels. Fig. 15(a) shows the neighbor pixels (blue color). Adding the spatial constraint will smooth the height value.



(a)

(b)

(c)

Figure 15: (a) the neighbor pixels (b) The original height value (c) Adding the spatial constraint to get the reliable result.

In order to find the proper scale of this constraint, we add a weighted value b to adjust the result. Transforming the input image to the frequency domain, the high-frequency area will get the high ratio weighted value. We can additionally add the intergrability constraint to obtain a unique minimum,

$$Intergrate = \iint z(x,y)^2 dxdy \tag{11}$$

where z(x,y) is the height field component.

3.5.2 Temporal constraints

In this thesis, the facial details are acquired from video sequences. These data have the quality of temporal coherence. If we just adopt the spatial constraint to optimize the parameters, there are still noises such as temporal flicker as shape changes. In order to reduce these noises, we need to exploit the temporal coherence.

To include the cost function in the temporal domain, most researches track the time-variation of each pixel. However it will produce other noise cased by the pixel correspondence and the occlusion problem. To avoid these conditions, our temporal constraint is applied to the variation of the same position of the hyper domain.

Fig 16 shows the temporal constraints in the hybrid domain. We track the time-varying surface normal (red arrow) at the same position (green point) to avoid the occlusion problem.

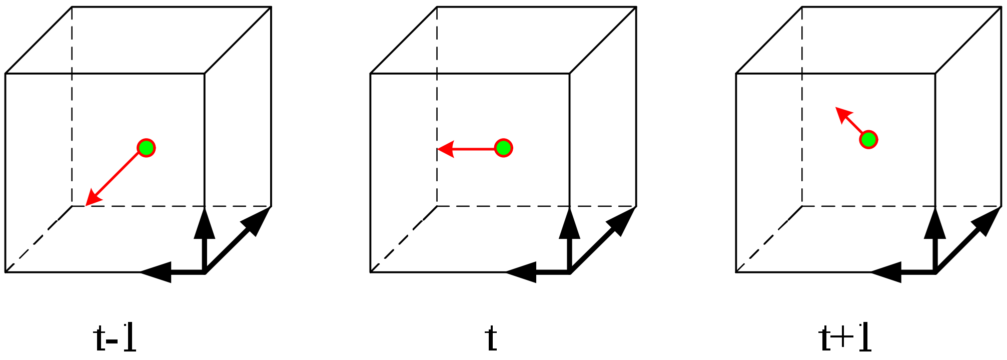


Figure 16: The temporal constraints in the hyper domain. We track the variation of the same surface to avoid occlusion problem.

Apply optimization the whole video sequence is extremely time-consuming. In order to speed up this algorithm, we only apply the optimization to a small segment of video at one time. When sweep the windows from the start to the end, we get a pseudo-optimization result. This approach will dramatically reduce the processing time and obtain more stable result. Our temporal constraint is

$$temporal(w,t)=\iint\sum_{i}^{|w|}(N(x,y,t)-N(x,y,t_i))^2dxdy \tag{12}$$

where $|w|$ denote the window sizes and N(x, y, t) denote the surface normal.

Therefore our cost function becomes:

$$C(V,R,w,t)=\iint[\,(I-S(\mathrm{V},\mathrm{R},\mathrm{t}))^2+spatial+temporal(\mathrm{w},\mathrm{t})\,]dxdy \tag{13}$$

3.6 Normal Difference and Height Map

In the above subsection, we describe how to extract surface normals from video sequence. Nevertheless, when we applied the space-time shape-from-shading (SFS) technique based on the uniform-skin-color assumption, some defective normal occur. Color variations on human skin, acnes, scars etc. may also make the image gradients changes dramatically. Hence, we propose using normal difference map to alleviate defects.

We select a neutral face from our video sequence. Then the normal map can be estimated by the space-time shape-from-shading. The normal difference map can be calculated by subtraction of the normal map of expressed face to the normal map of neutral face.

$$NDM = NM_{\exp} - NM_{neu} \tag{14}$$

where NDM is the normal difference map, $NM_{\exp}$ is the normal map of novel expression and NM_{neu} is the normal map of the neutral face. Figure 17 shows the normal difference map.

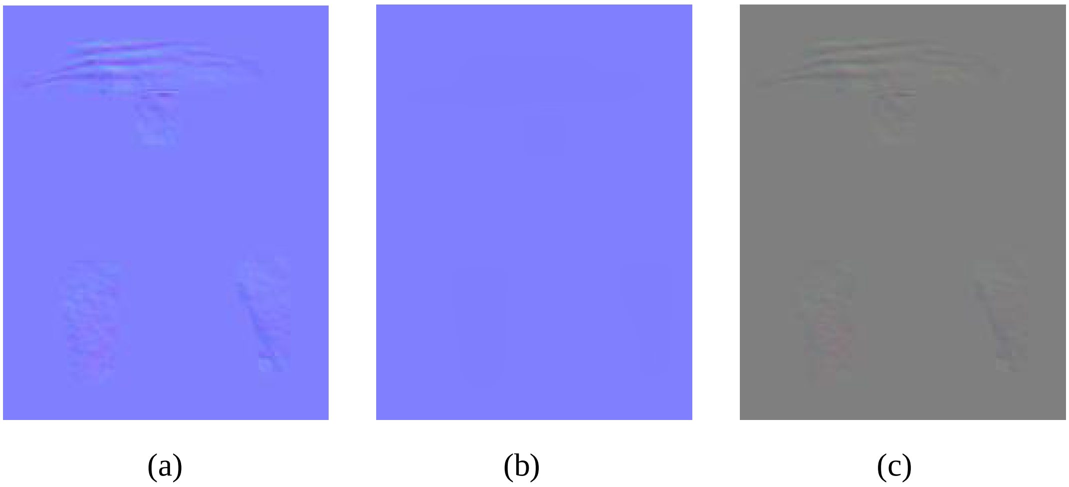
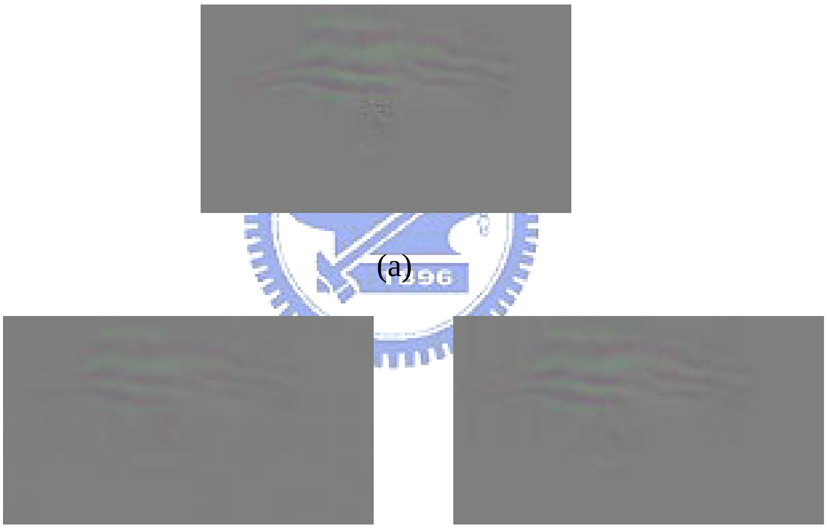


Figure 17 (a) the expression normal map (b) the neutral normal map (c) the difference normal map

3.6.1 Post Processing of Normal Difference Maps

Due to the error caused by pixel alignment, input noise, and digitization, etc, our normal difference maps have some unavoidable estimation errors. We utilize an adaptive Gaussian filter to reduce noise problem. Around the wrinkle region, we select a 3-by-3 filter mask. In other regions, a larger mask filter will be used. This procedure can remove the noise effectively and make the normal difference map much smoother. We also apply the adaptive bilateral filter to reduce these noises. The bilateral filter can retain the details more accurately. Figure 18 shows the result where the adaptive Gaussian filter and bilateral filer are applied.



(a)

(b) (c)

Figure 18 (a) The difference normal map(the facial details of forehead) before filter (b) After apply Gaussian filter (c) After apply bilateral filter

3.6.2 Height Maps

Our method can reconstruct the z value of the 3D shape. We estimate the minimum and maximum height value of the optimized shape to create the height map.

$$Height(x, y) = \frac{z(x, y) - \min(z(x, y))}{\max(z(x, y)) - \min(z(x, y))} \tag{15}$$

While combing facial details with the primitive face model, our system can synthesize more detailed facial animation. Fig 19 shows the height map (The facial details of forehead , glabella and cheeks).

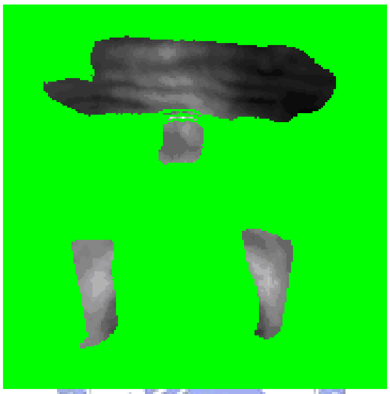


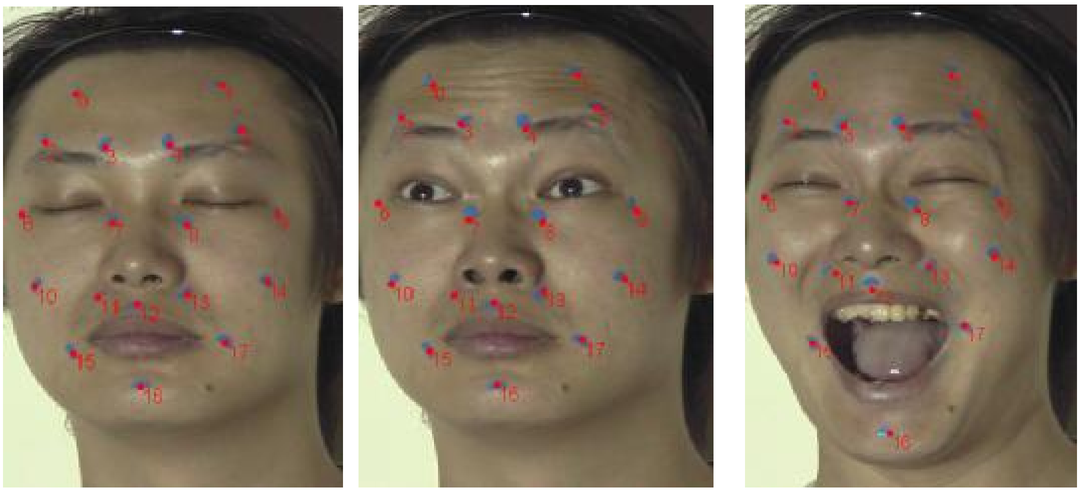
Figure 19: the height map.

Chapter 4
Feature Point Driven System

In the previous chapters, we described how to acquire the facial details from the video sequence. In our research, we utilize a motion capture technique to create the primitive 3D model and apply a novel shape-from-shading for facial details. In order to acquire the fundamental 3D model, we use stereo triangulation for feature points in two views. While morphing a generic model according to these feature points, we can get an approximate geometry.

4.1 Tracking the Correspondence Feature Points

Before creating the primitive 3D model, we need to track the correspondence feature points and use the stereo triangular algorithm to reconstruct the 3D shape. First, users need to assign the correspondence feature points $\{p_0, p_0'\}$ on the neutral image. We apply the simple block matrix method to search the feature points $\{p_i, p_i'\}$ with similar intensity or color on the nearby area. But it will find more than one similar intensity pixel and lead to a flicker result. For this problem, we apply the simple filter like Gaussian filter to smooth the searching result. Fig 20 shows the tracking result of the block matrix method.



(a) (b) (c)

Figure 20:(a) the natural feature points (b-c) Using the block matrix method to track the feature points

4.2 Reconstruction of Primitive Model

The reconstruction of primitive 3D model can be produced by the camera calibration and stereo triangulation algorithm.

Given the projection matrices M and M', the corresponding feature 2D points p and p’, we can rewrite the equation p=MP and p’=M’P as:

$$\begin{cases} p\times MP\equiv 0 \\ p'\times M'P\equiv 0 \end{cases} \Leftrightarrow \begin{pmatrix} [p_x]M \\ [p'_x]M' \end{pmatrix} P\equiv 0 \tag{16}$$

, where P denotes the reconstructed 3D points. This is an over-constraint system and we can solve P easily by a linear least-square method.

4.3 Deformation of 3D primitive model

After we get the 3D position of the calibration, we can use the geometry to deform the 3D primitive head model. First, users have to select a set of corresponding pairs $\{\mathbf{p}_i, \mathbf{q}_i\}$, where p_i is the feature point position of our synthesizing expression and q_i is the corresponding point position on the generic model. Once the displacement of each feature point $\mathbf{u}_i = \mathbf{q}_i - \mathbf{p}_i$ was calculated, we use scattered data interpolation $S(\mathbf{p})$ to estimate the displacement of other vertices on the original mesh. We adopted the radial basis function as:

$$\mathrm{S}(\mathbf{p}) = \sum_i c_i f\left(\left\|\mathbf{p} - \mathbf{p}_i\right\|\right) + M\left(\left\|\mathbf{p} - \mathbf{p}_i\right\|\right) + t \qquad (17)$$

where f is radial symmetric basis function, and c_i are displacement coefficients, and M, t are affine terms. To determine c_i, M and t, we solve a set of linear equations that includes interpolation constraints $\mathbf{u}_i = S(\mathbf{p})$. We use $f(\mathbf{r}) = e^{-\mathbf{r}/32}$.

When deforming the 3D primitive model, we need to divide the 3D face model into sub-regions such as the forehead, the nose and the mouth…etc. After applying RBF functions locally to deform the sub-region, we can produce primitive 3D animation Fig. 21 shows the deforming result using the local RBF. The detailed facial animation will be introduced in section 5.3.

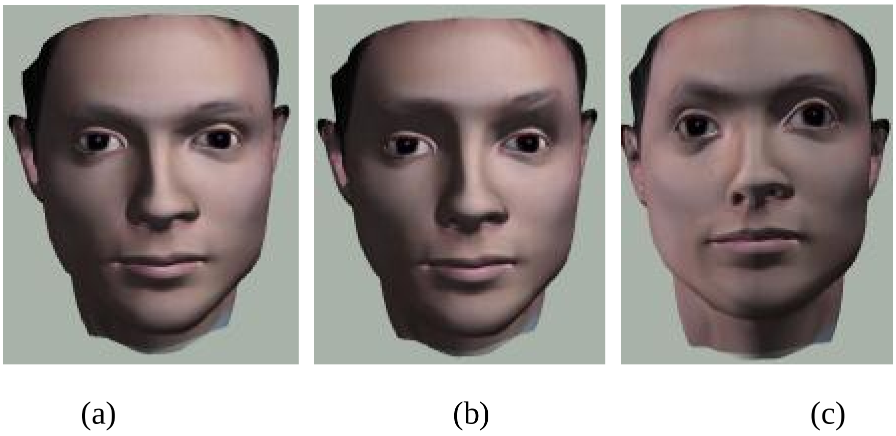


Figure 21: (a)the netural face (b) Using local RBF functions (c) global RBF

Chapter 5
Experiment and Result

In this chapter, we will describe our experiment and show our result. At the beginning, we introduce the experiment of the input video sequence and analyze the optimized result. Then, we will show the synthetic results where the facial details are included.

5.1 The Experiment of Input Video Sequence

In our system, we use two synchronized video streams to create the difference normal and height maps. In order to acquire the more accurately facial details, our input images are taken under an illumination-controlled environment.

We set a projector as the single light source. Our input data are the two synchronized high-definition video (HDV, 1280*720 pixel resolution) and the frame per second (FPS) is set to the 30 frames per second. Fig 22 shows the two different views of video data.

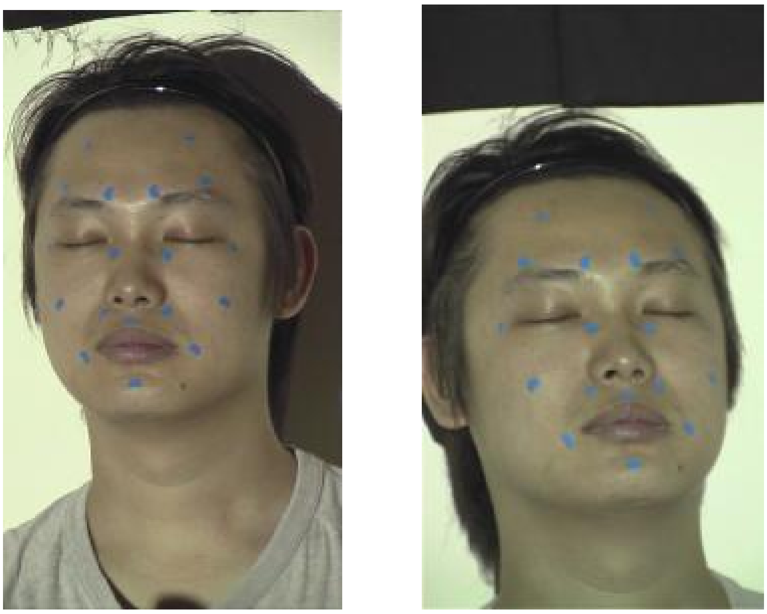


Figure 22: the two different views of video data. (natural face)

We put a set of markers (as shown in Fig 21, 18 markers) on the actor’s face but avoid placing markers on regions of wrinkles or creases.

5.2 The Result of Space-time SFS

In our research, we apply a novel space-time shape-from-shading to reconstruct the 3D shape. We utilize an optimization method to solve the ill-condition of shape-from-shading. This method can optimize the space and reflectance parameters to minimize the cost function. Fig 23 shows the chart of optimizing the motion of raising the forehead, bending actor’s brows and smile. The cost value is obviously decreasing and we stop the optimization until the variation of the parameters is less than the threshold.

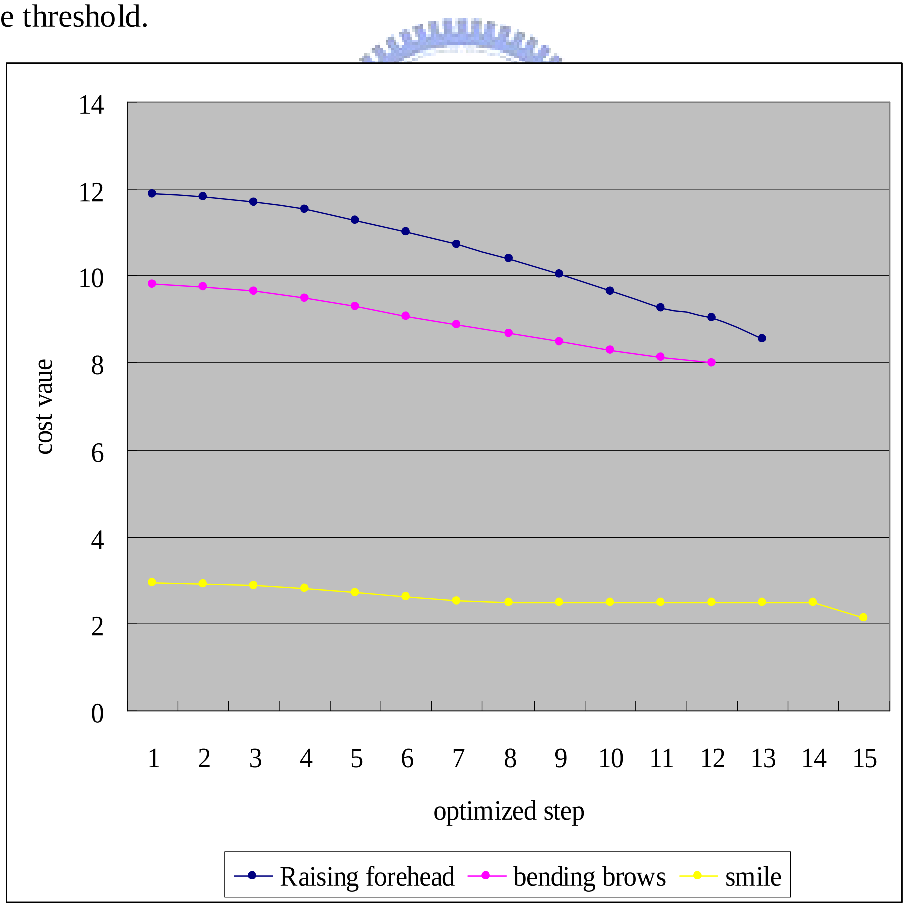
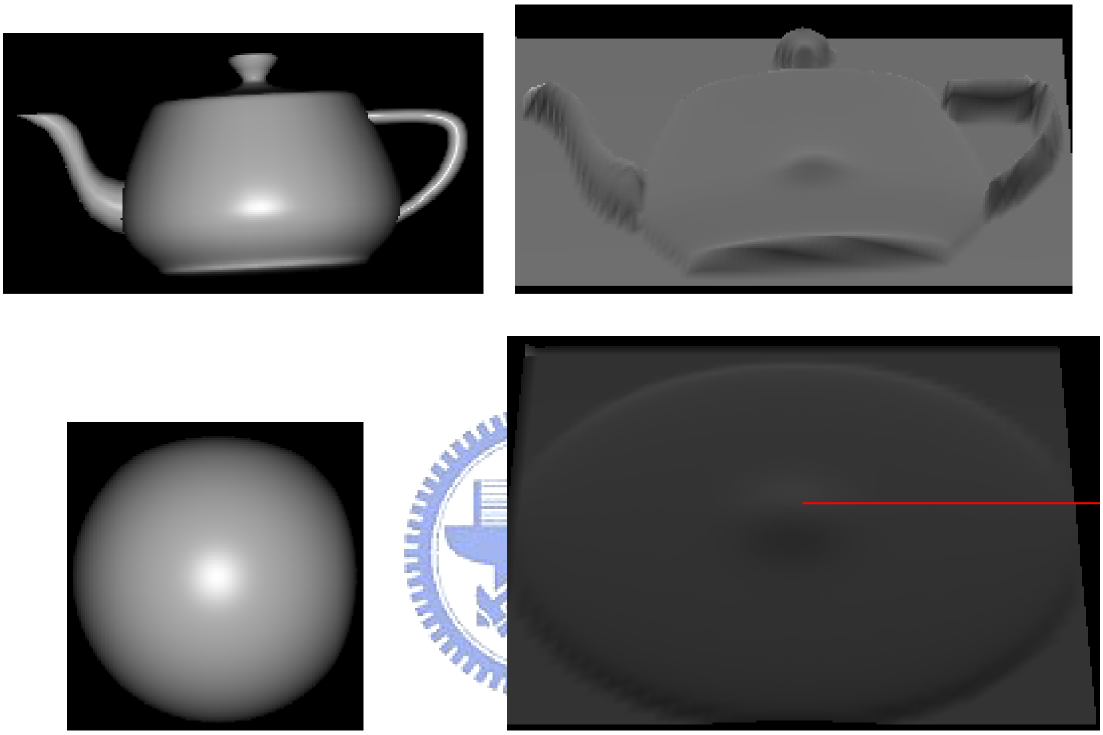


Figure 23: the progress of optimization of the cost function

We also optimize the reflectance parameters for the synthesized images. Fig 24 shows the result of the optimized reflectance parameters. We set the initial shape as the flat shape and the reflectance parameters Kd=0.5, Ks=0.5, alpha=15. The optimized result is close to the accurate value.

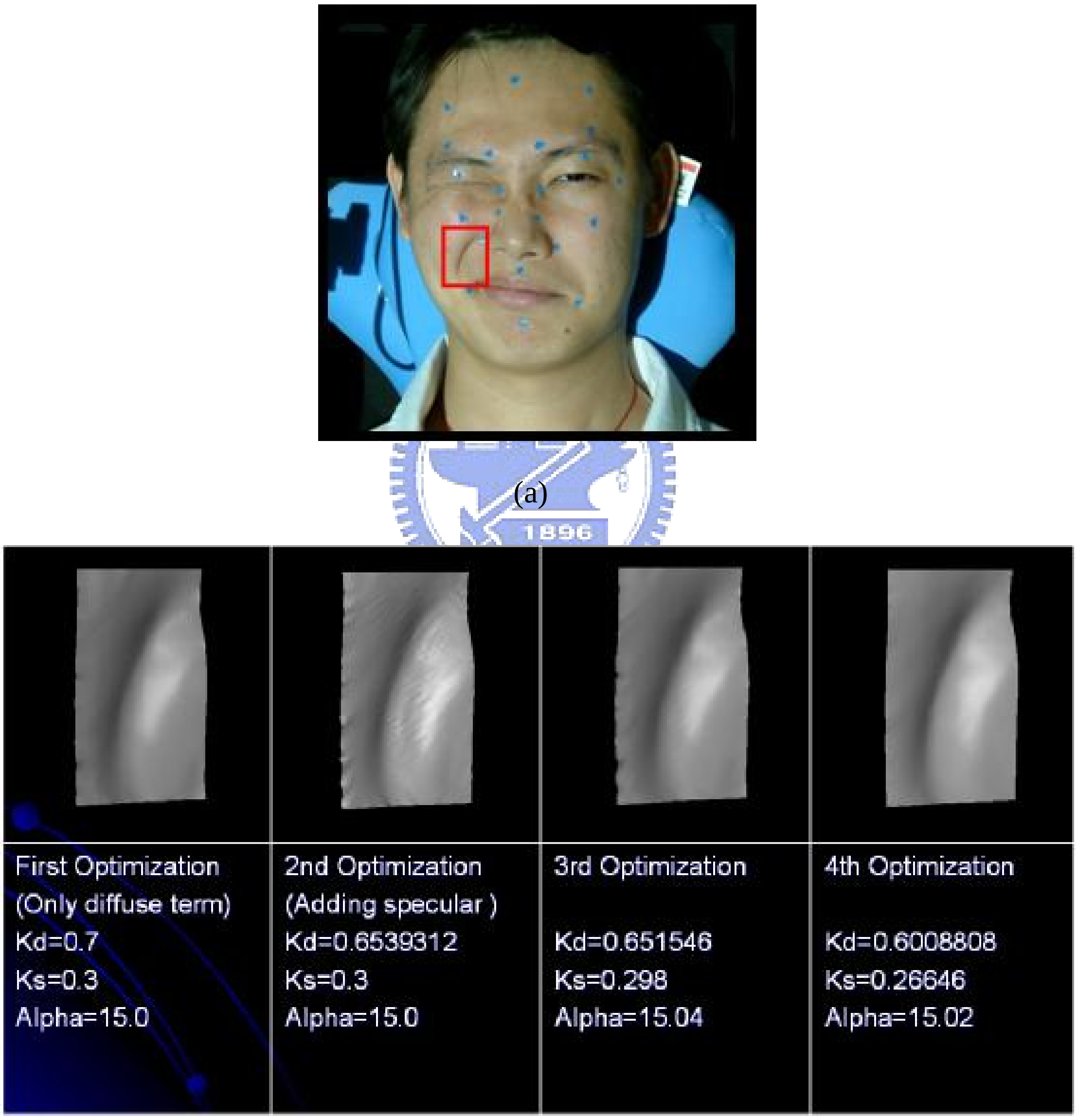


(a) (b)

Synthesized data				the recovery reflectance data			Normalized	
	Kd	Ks	Alpha	Kd	Ks	Alpha	Kd	Ks
teapot	0.7	0.3	15	0.536298	0.153753	15.001	0.7772	0.2228
ball	0.7	0.3	15	0.28751	0.09362	15.004	0.7543	0.2456

Figure 24: the result of the optimized reflectance parameters

Fig 25 shows the progress of optimization phases. In the first phase, we just optimize the diffuse term to get the more accurately initial shape. After the second phase, specular term will include to be optimized. Adding the spatial constraint will smooth the optimized result. The optimization method will stop until the variation of reflectance parameters is small tan the threshold. Another result of shape recovery is show on Fig 26.



(a)

(b)

Figure 25: (a) the wrinkle of the input image (b)the optimized phase

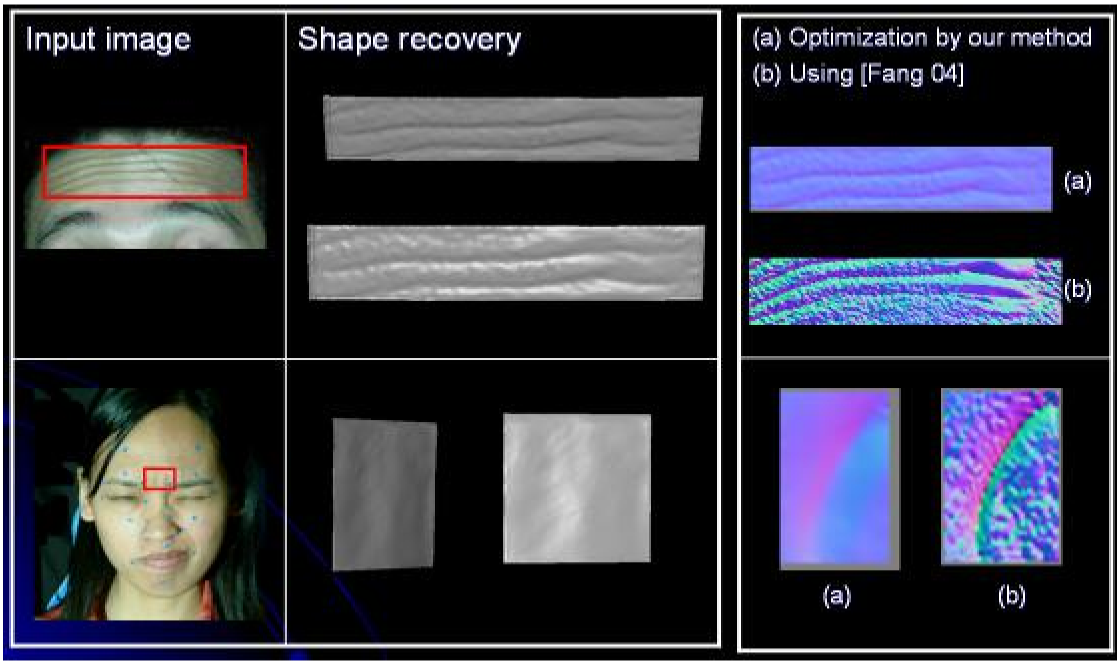


Figure 26: results of shape recovery(forehead and glabella)

5.3 The Synthesized facial details

The 3D head model used in this thesis has 6078 vertices and 6315 polygons. Every vertex has a predefined normal vector. We need to separate the region to apply the local RBF functions. These sub-regions include the forehead, nose and mouth…etc. In order to apply the height map on the face model, we subdivide the polygons and utilize the difference normal map to render the synthesized data. Figure 28 shows the subdivision result for the real-time rendering. Fig 29-34 shows the facial details using normal difference and height map are added on the face model.

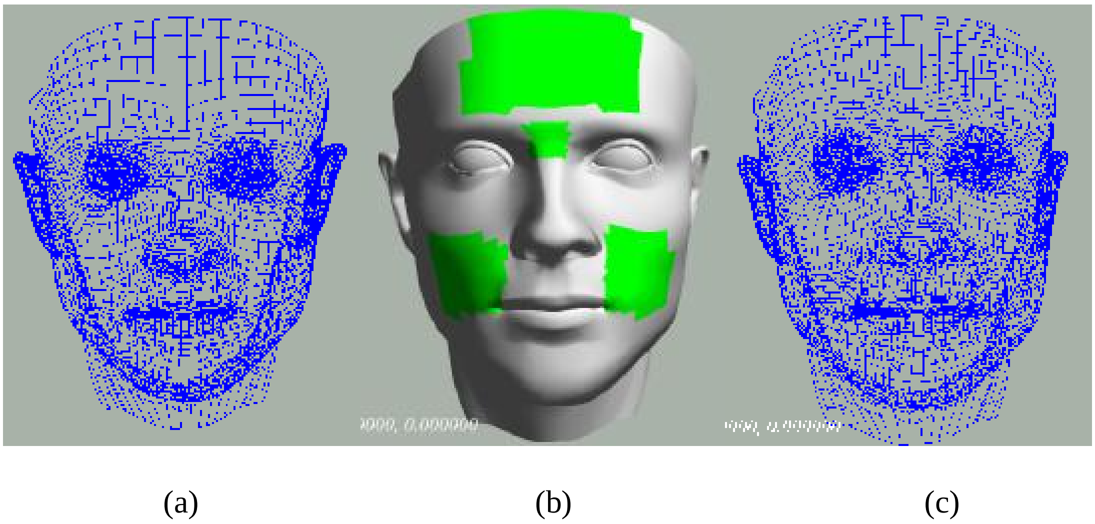


Figure 27: (a) original 3D mesh (b) subdivide area(green) (c) subdivided result

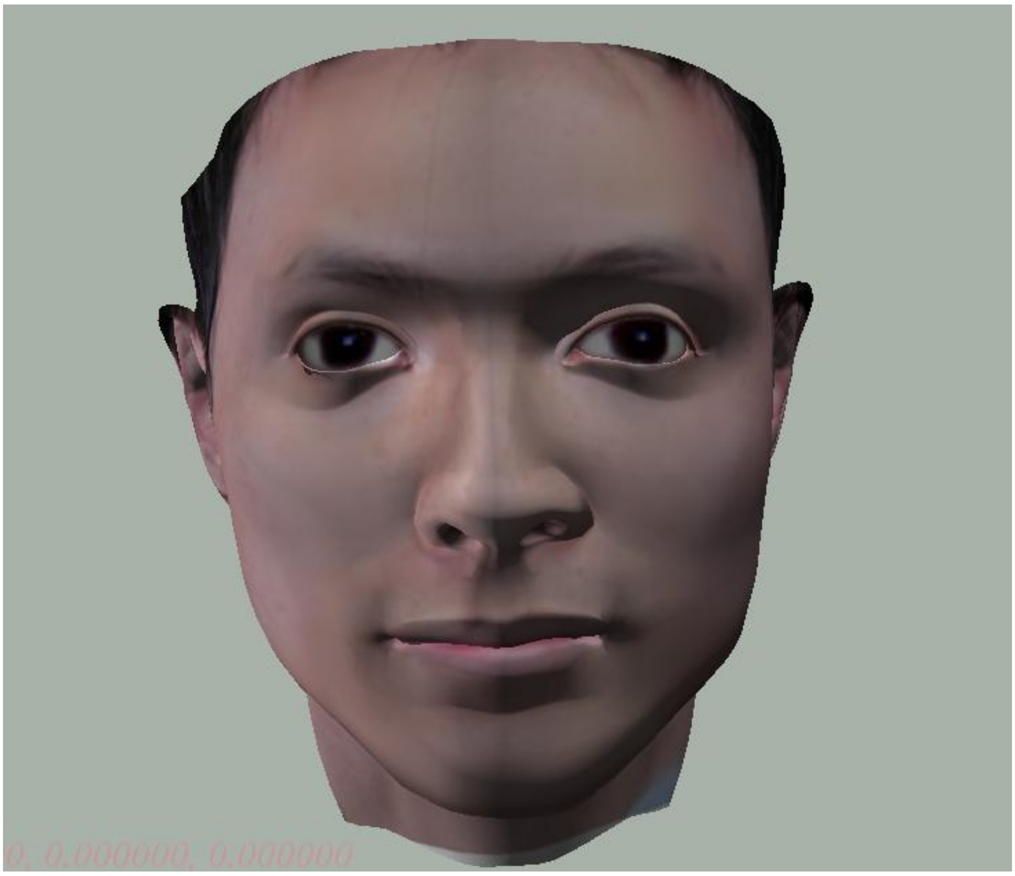
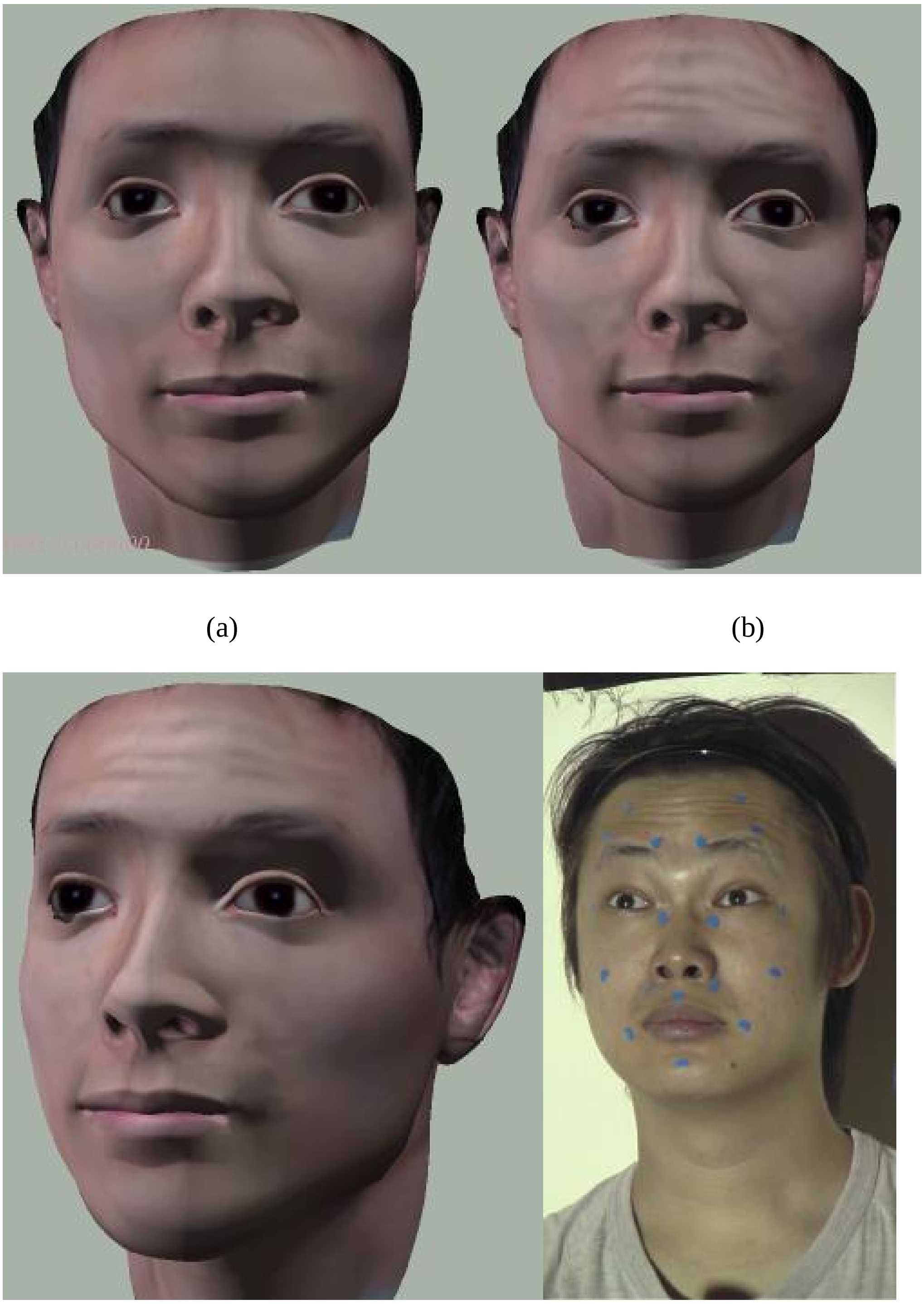


Figure 28:the natural face model



(a) (b)

(c) (d)

Figure 29(a) Raising the forehead without facial details (b-c) adding the facial details (d) the original captured image

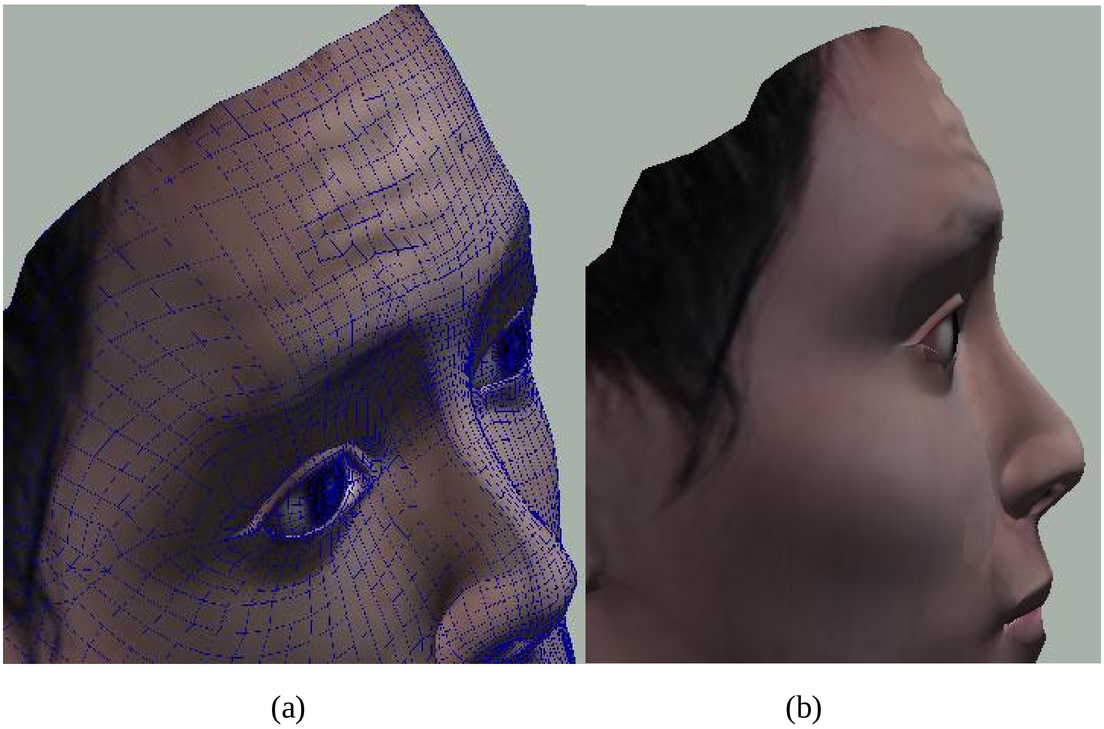

(a) (b)

Figure 30(a) Applying the height map by subdivision(b) the side view

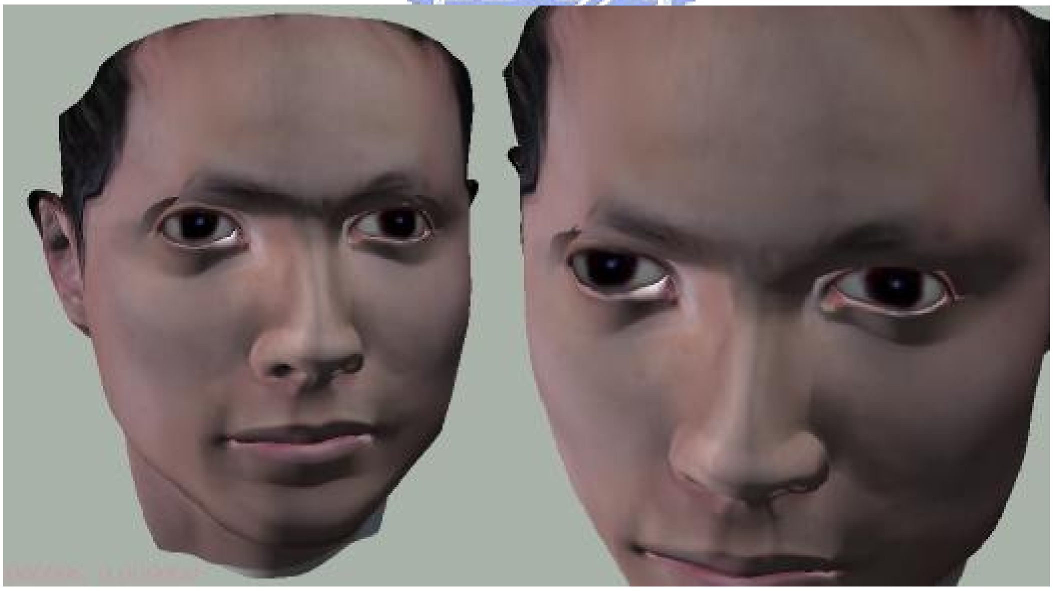


Figure 31: anger expression.

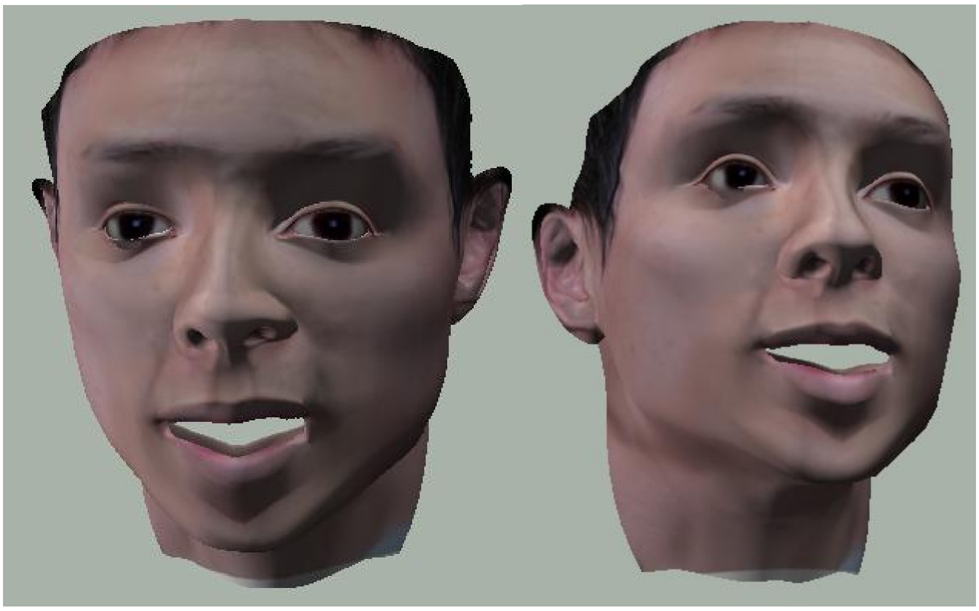


Figure 32:the facial details by opening the mouth

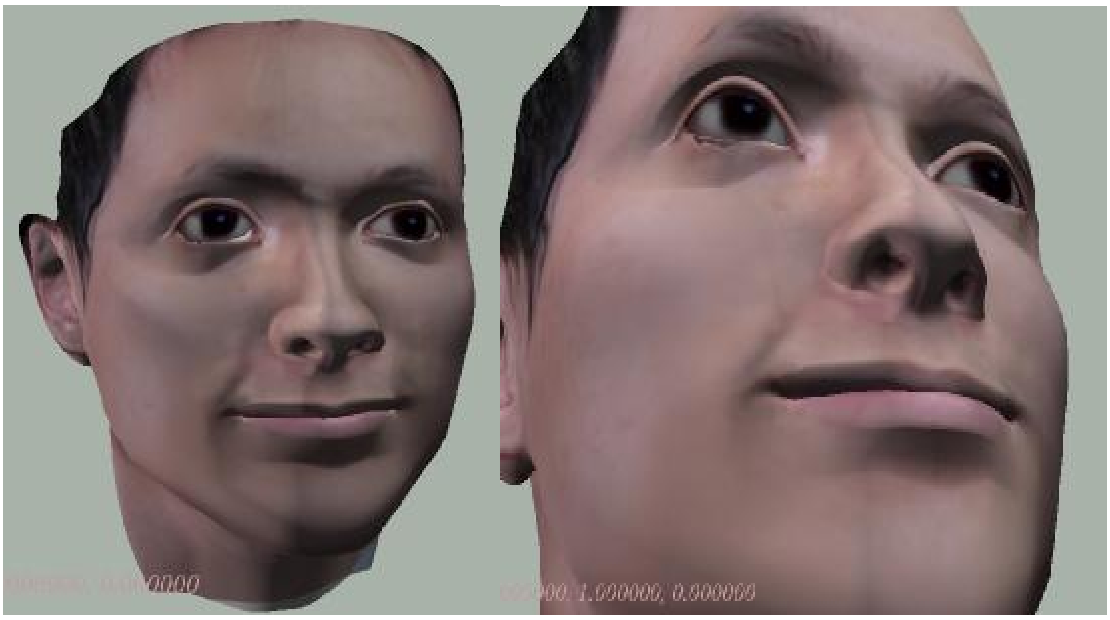
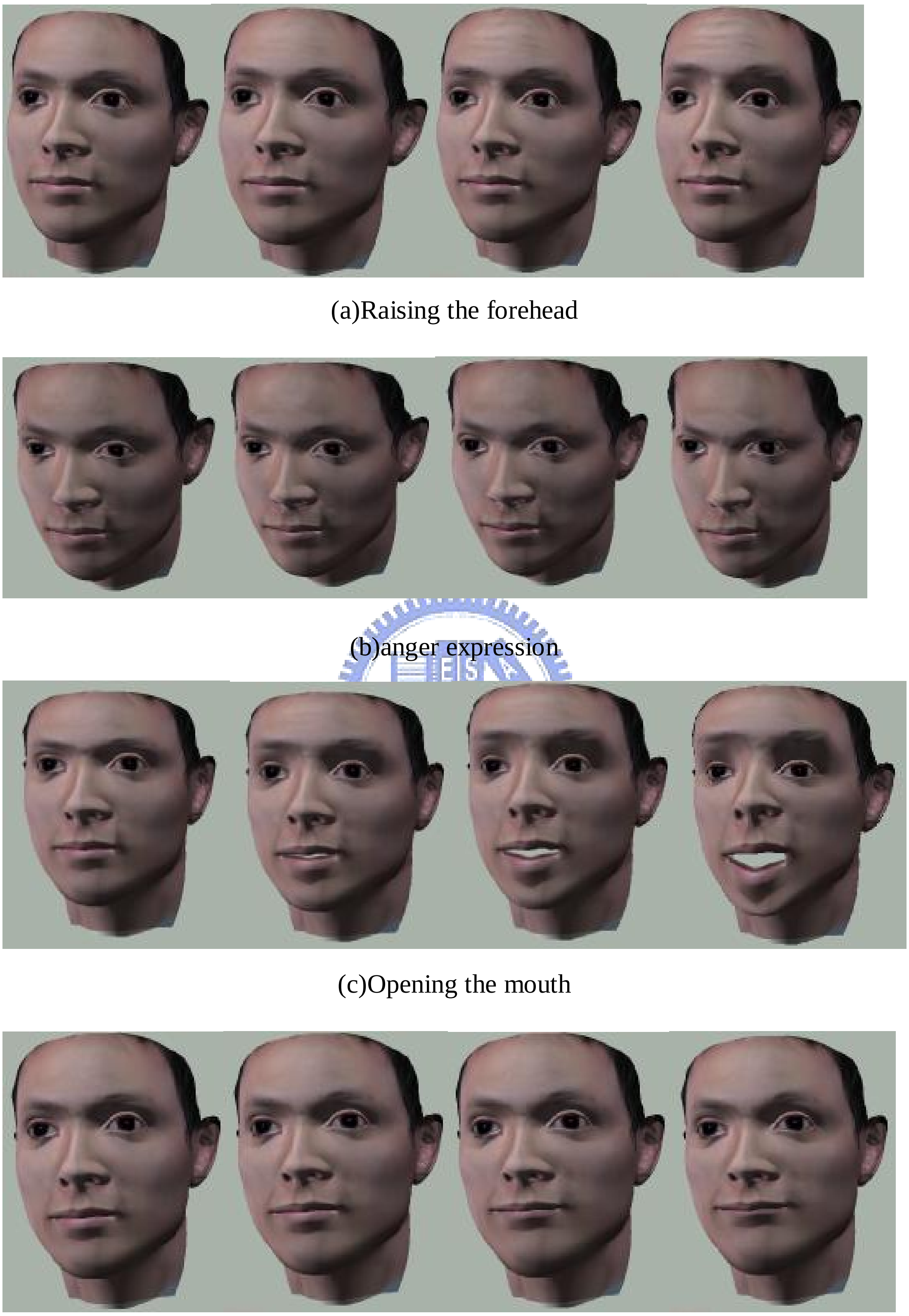


Figure 33: smile expression



(a)Raising the forehead

(b)anger expression

(c)Opening the mouth

(d)smiling expression

Figure 34: the facial animation

Chapter 6
Conclusions and Future Work

6.1 Conclusions

In this thesis, we propose a space-time shape-from-shading (SFS) to reconstruct the facial details. In order to solve the ill-condition environment, we apply the optimization method with adding spatial and temporal constraints to get more reliable results. We utilize the feature point driven system for primitive 3D face model and the estimated facial details are then combined with the face model. To render the height maps, we subdivide the primitive model according to the height values and apply normal difference maps. With the proposed method, we will get the more detailed facial animation.

Our contribution include (1) a novel space-time shape-from-shading for recovering 3D data. (2) Using an optimization method to get the more reliable results for real data.

6.2 Future Work

In this thesis, we adopt Phong model as the reflectance model. Other reflectance models such that Torrance model or BSSRDF which has more physical cues may get more accurately results. And the other hand, we can apply other numerical method such that Fast Marching Method (FMM) to speed up the optimized procedure.

References

[1] Zhang L., Snavely N., Curless B. and Seitz S.M. “Spacetime Faces: High Resolution Capture for Modeling and Animation”, Proc. ACMSIGGRAPH'04, Pages 548-558, 2004

[2] Weyrich T., Matusik W., Pfister H., Lee J., Ngan A., Jensen W. and Gross M., “Analysis of Human Faces using a Measurement-Based Skin Reflectance Model” Proc. ACMSIGGRAPH'06, Pages 1013-1024, 2006.

[3] Huynh D.Q. “Calibration of a Structured Light System: A Projective Approach”, In Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR’ 97), Pages 225-230, 1997

[4] Hertzmann A. and Seitz S.M.,”Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs”, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27 no. 8 , Pages 1254-1264 ,2005

[5] Vogiatzis G., Torr P.H.S and Cipolla R., “Multi-view Stereo via Volumetric Graph-cuts”, In Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR’ 05), vol. 2, Pages 391-398, 2005

[6] Fang, H., and Hart, J. C. “Textureshop: Texture Synthesis as a Photograph Editing Tool”, Proc. ACMSIGGRAPH'04, Volume 23, Issue 3 (August 2004), Pages 354-359, 2004.

[7] Horn, B.K. 1990. “Height and Gradient from Shading”, International Journal of Computer Vision, Vol. 5(1), Pages 37-75, 1990.

[8] Zeng G., Matsushita Y., Quan L., and Shum H.Y., ”Interactive Shape from Shading”, In Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR’ 05), vol.1, Pages 343-350, 2005

[9] Han F. and Zhu S.C. “Cloth Representation by Shape from Shading with Shading Primitives”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’ 05), vol.1, Pages 1203-1210, 2005

[10] Yu T., Xu N. and Ahuja N. ”Recovering Shape and Reflectance Model of Non-lambertian Objects from Multiple Views”, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’ 04), vol.2, Pages 226-233, 2004

[11] Fang H. and Hart J.C. “RotoTexture: Automated Tools for Texturing Raw Video”, IEEE Trans. Visualization and Computer Graphics, vol. 12, Pages 1580-1589 , 2006

[12] Yosuke B., Takaaki K., and Tomoyuki N. , “A Simple Method for Modeling Wrinkles on Human Skin”, Pacific Graphics 02, Pages 166-175, 2002.

[13] Zhang, L., Snavely, N., Curless, B., and Seitz, S. M. “Spacetime Faces: High-Resolution Capture for Modeling and Animation”, ACM Trans. on Graphics, Vol. 23, Issue 3, Pages 548-558, 2004.

[14] Zhang, Q., Liu, Z., Guo, B., Terzopoulos, D., and Shum, H. “Geometry-Driven Photorealistic Facial Expression Synthesis”, IEEE Trans. On Visualization and Computer Graphics, Vol. 12(1), Pages 48-60, 2006.

[15] Zhang, R., Tsai, P.-S., Cryer, J., and Shah, M. “Shape from Shading: A Survey”, IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 21(8), Pages 690-706, 1999.

[16] Zhu L, Lee W.-S., "Facial Expression via Genetic Algorithms", Computer Animation and Social Agents , 2006

[17] Beier, T., and Neely, S. “Feature-based Image Metamorphosis”, Proc. ACM SIGGRAPH'92, Pages 35-42, 1992.

[18] Blanz, V., Basso, C., Poggio, T., and Vetter, T. “Reanimating Faces in Images and Video”, Computer Graphics Forum 22 (3), Pages 641 - 650, 2003.

[19] Blanz, V., and Vetter, T. “A Morphable Model for the Synthesis of 3D Faces”, Proc. ACM SIGGRAPH'99, Pages 187-194, 1999.

[20] Golovinskiy, A., Matusik, W., and Pfister, H. “A Statistical model for Synthesis of Detailed Facial Geometry”, ACM Trans. on Graphics, Volume 25, Issue 3, Pages: 1025-1034, 2006.