Yung-Sheng Lo · I-Chen Lin · Wen-Xing Zhang · Wen-Chih Tai ·
Shian-Jun Chiou

Capturing Facial Details by Space-time Shape-from-shading

Abstract In this paper, we propose a facial detail estimation approach for 3D animation. First, motion capture is utilized to evaluate primitive 3D surfaces. A novel shape-from-shading (SFS) is then applied for facial details. Our SFS method, exploiting Bidirectional Reflectance Distribution Function (BRDF) properties, can extract detailed surfaces according to intensity variations. It also avoids the problem of ambiguous correspondence in stereo triangulation. In order to tackle the intrinsic ill-condition and acquire more reliable results, optimization in a space-time hybrid domain is employed to approximate both the 3D face geometry and reflectance properties. While combing facial details with a feature-driven face model, our system can synthesize more detailed facial animation.

Keywords Facial detail · Shape-from-shading · Motion capture · Facial animation

Fig. 1 Feature-point-driven facial expression without and with facial details.

1 Introduction

From TV games, movies to advertisement, 3D characters have been popularly applied to various kinds of media. However, how to efficiently generate realistic facial animation is still a challenging problem. In early years, animators have to manually adjust key poses of 3D models for vivid characters. To speed up the production, facial motion capture (mocap) techniques, recording motions of conspicuous markers on a performer's face, become one of the most practical solutions.

Y.-S. Lo, I.-C. Lin, W.-X. Zhang
Dept. of CS, National Chiao Tung Univ., Taiwan.
E-mail: yanzuzu@caig.cs.nctu.edu.tw
E-mail: ichenlin@cs.nctu.edu.tw
E-mail: starsin@caig.cs.nctu.edu.tw

W.-C. Tai, S.-J. Chiou
Chunghwa Picture Tubes, LTD., Taiwan.
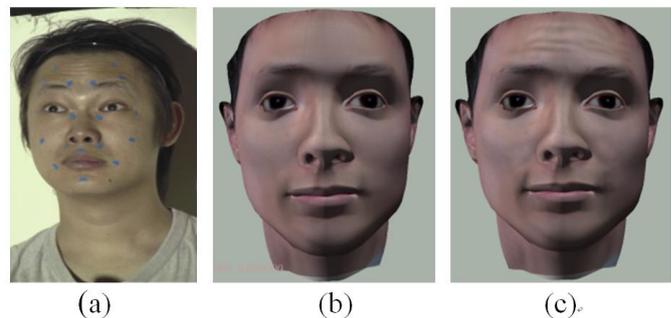E-mail: taiwc@mail.cptt.com.tw
E-mail: chiousc@mail.cptt.com.tw

These motion capture data can be used to drive 3D faces efficiently, but there are still subtle portions, such as wrinkles or creases, whose variations are much smaller than markers' sizes. As shown in Fig. 1(b), while synthesizing only approximate geometry without details, the fidelity will be insufficient.

To deal with such critical problems, several researches work on transferring feature motions to pre-computed prototypes. For instances, Z. Deng et al. [1] proposed using learnable weighted combinations of predefined key expressions. Nevertheless, the blended details may not be the same as those performed by the subject, and it will suffer blurring effects during blending.

Stereo triangulation, based on two or more image data, is the most typical approach to calculate 3D positions, but pixel correspondences of un-textured regions are usually ambiguous. Structured-light reconstruction, using a camera and a projector to acquire depth images, is another popular triangulation method. In 2004, L. Zhang et al. [2] proposed an impressive system for estimating dynamic face surfaces. They exploited space-time coherence for a more reliable pixel-correspondence. However, they had to use devices with higher resolution and capturing speed due to inherent properties of coded structure-light

systems.

On the other hand, shape-from-shading (SFS) and photometric stereo methods estimate shapes according to variations of image intensity. Most of the existing methods are under the assumption of the Lambertian reflectance model. Photometric stereo recovers the shape from images under different light directions but in a fixed view point. Using photometric stereo for dynamic objects will require expensive high-speed cameras and a light dome [3]. Most shape-from-shading techniques assume the scene with a single light source and use only intensity gradients within a single image for shape recovery. Hence, the troublesome problem of pixel correspondence can be avoided. Also, SFS is sensitive to subtle variations and is therefore adequate to facial details. Nevertheless, its sensitivity to noise may result in serious tremble.

In this paper, we combine the benefits of stereo triangulation and shape-from-shading. Stereo-based 3D motion tracking is employed to evaluate the rough geometry of an expressive face. A novel space-time constraint over shape-from-shading is proposed to estimate the detailed facial motion. Moreover, to efficiently deal with non-Lambertian reflection on a face, we choose Phong reflection model. With the proposed method, reliable detailed facial motion and approximate BRDF reflectance properties can simultaneously be estimated during optimization.

The goal of our work is to enhance the feature-driven animation with captured facial details. The proposed framework can be divided into two phases: acquisition and synthesis. The acquisition phase reconstructs the 3D detailed surface and estimates the reflectance parameters. The synthesis phase is about generation of facial animation. The flowchart is shown in Fig.2.
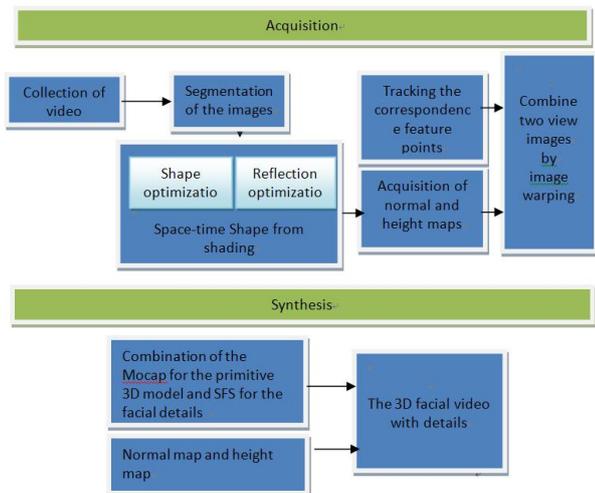

Fig. 2 The flowchart of the proposed system.

This paper is organized as follows. Related work is introduced in Section 2. In Section 3, we propose combining motion capture and shape-from-shading for facial detail acquisition. In Section 4, various issues such as initialization, spatial-temporal constraints, and progressive refinement are also presented. Section 5 explains our experiment and results. At last, we discuss the advantages and disadvantages of the proposed work and make conclusion.

2 Related Work

The proposed method applies shape-from-shading based on graphics reflection model and space-time constraints to estimate detailed facial motion. We introduce recent related literatures in this section.

To estimate 3D positions and motions, stereo triangulation is the main stream in computer vision for decades. However, reliable pixel correspondence is still a troublesome issue. Readers may refer to the review paper by T.S. Huang and A.N. Netravali [4] for details.

For precise tracking, conspicuous markers can alleviate the correspondence problem. B. Guenter et al. [5] proposed a dynamic face digitization system. This research used 182 special markers, and they also took into account the spatial and temporal consistency for reliable tracking. For time-vary facial details, they simply recorded dynamic texture and avoided surface recovery. I.-C. Lin et al. [6,7] proposed capturing dense facial markers with a single video camera and mirrors in near real time. Even though 300 markers were tracked, wrinkles or dimples were still difficult to be reconstructed.

Recently, G. Vogiatzis et al. [8] proposed a novel multiview stereo reconstruction method. This method used the visual hull as the initial volumetric shape and optimized the results by the graphic-cut. Their system can provide a more reliable 3D surface but are also not adequate for details.

Based on bio-mechanical hypotheses, anatomical models can also generate wrinkles and creases [9]. In 2005, E. Sifakis, et al. [10] proposed an automatic approach to determine muscle parameters according to sparse markers; therefore, facial details can also be synthesized by anatomical models. However, it cannot generate the actual details of the performer.

On the other hand, photometric stereo and shape-from-shading are two approaches to estimate subtle surface normals from variations of image intensity. They are capable of dealing with surface details. Images for photometric stereo are usually captured in a fixed view point but under various lighting directions. According to the

Lambertian reflection model, given accurately aligned pixels, photometric stereo methods can estimate surface normals by a simple least-square method.

Empirically, to acquire reliable normals, we will need more than eight images of different light directions. To capture moving objects at 30 frames per second (fps), we have to employ a 240-fps high speed camera with synchronized directional lights. In the work of T. Weyrich et al. [11], they produced a face-scanning dome composed of 16 digital cameras, 150 light sources and a commercial 3D face-scanning system. Their system can accurately acquire the reflectance as well as facial normals.

M. seitz et al. [12] proposed an example-based method. They introduced orientation-consistency to reconstruct surfaces by comparing images of reference objects with a known shape. Constraints, such as distant lighting and reference objects with an identical material, will limit the applications of this method.

By contrast, shape-from-shading can avoid errors due to inaccurate pixel correspondence, and this low-cost approach may require only one image for shape recovery. In 2004, H. Fang et al. [13] adapted B.K. Horn's [14] approach and utilized Lambertian reflection model to extract the normal map from a single texture image. This kind of approach doesn't need expensive devices. However, it is error-prone due to the simple Lambertian assumption, sensitivity to input noise, etc. Usually, manual adjustments are required for post processing.

SFS is also difficult for real data due to its intrinsic ill-condition. In interactive modeling, G. Zeng at al. [15] proposed a semi-automatic solution for continuous surface. Users assigned surface normals on specific feature points and the system then refined the surface variations to the whole face. This method applied a fast marching method (FMM) to speed up computation and solved its ambiguity by human assistance.

T. Yu et al. [16] proposed an optimization method to obtain the shape and reflectance parameters on a static model with Phong reflection model. This method initiated the reflectance parameters from different scales and further refined the estimation with multi-view information. However, they did not yet apply to real persons. To stabilize the iterative SFS algorithm, constraints such as smoothness, intensity gradients will be used to obtain the reliable result.

B. Bickel et al. [17] further proposed an impressive multi-scale capturing system for facial motion. They first used conventional motion capture for large-scale motion. For middle-scale motions, they painted a specific color on each wrinkle of a subject and estimated parameters of the "valley-shape" wrinkle model from video. Our multi-level framework is similar to this work. Nevertheless, we don't restrict the shape of wrinkles or Lambertian reflection. Our proposed approach can directly estimate facial details, e.g. wrinkles or dimples, without additional paint.

To stabilize our optimization result, we apply the concept of space-time constraints. Space-time constraints have proved to be a powerful approach in character animation, especially in motion editing. Such approaches will gradually adjust motion capture data to fit constraints from users' assignments or kinematic properties [18,19]. In 2004, L. Zhang et al. [2] proposed a structured light approach to capture the dynamic variation on faces. They used space-time coherence to match corresponding pixels; they also assumed short-term linearity in space-time domain for computational efficiency. H. Fang et al. [20] proposed a "RotoTexture" synthesis technique. They utilized spatial and temporal smooth constraints to reduce the visual noise of texture mapping in video.

Unlike most SFS under Lambertian assumption, we apply more general reflectance models. In computer graphics, Bidirectional Reflectance Distribution Function model (BRDF) is widely used to represent the reflectance model of human faces. Assume that human skin are composed of the oil layer, epidermis and dermis, facial skin reflectance can be approximated by a specular component at the oil-air interface and a diffuse reflectance component due to subsurface scattering.

Jensen et al. [21] introduced a Bidirectional Surface Scattering Reflectance Distribution Function model (BSSRDF) that combined dipole diffusion approximation and single scattering computation. Based on Jensen's approach, Donner et al. [22] presented a new efficient technique with multiple dipoles to account for diffusion in thin slabs.

In our experiment, we have also tried Jensen's BSSRDF model in optimization of shape-from-shading. However, due to its larger degrees of freedom and complex correlations, the computation was inefficient and it can barely improve the result compared to Phong model. Therefore, we employ Phong reflection model as the analytic shading model.

3 Acquisition of Facial Motion

As above-mentioned, stereo reconstruction can estimate accurate depths with conspicuous features; shape-from-shading can avoid the pixel-correspondence problem in textureless regions.

We propose combining the benefits of both approaches. Two video cameras are utilized for sparse 3D marker

tracking. With captured feature motion, primitive 3D face geometry can be evaluated. Then, these captured image sequences are further utilized for facial details by shape-from-shading.

Since stereo triangulation, motion tracking and automatic false corrections have been addressed in various literatures, please refer to the reference articles for details [4–7].

3.1 Primitive Facial Surfaces from Motion Capture

In order to evaluate the variations of primitive 3D face surfaces, we adapt a model-based approach. Assume that the expression at the initial frame is neutral. First, 3D positions of markers at the first frame are estimated by stereo triangulation. We then characterize a generic face model by feature fitting, which is similar to the deformation method mentioned later.

For each of the following frames, we track the markers and deform the characterized model according to markers' 3D positions. The deformation method we used is radial-basis-function-based (RBF-based) data scattering.

Given a set of corresponding pairs p_i, q_i between the neutral face and an expressional face, where p_i is the 3D position of marker i on the neutral face, and q_i is the position on an expressional face. Given the displacement of each marker $u_i = q_i - p_i$, we use scattering function $S(p)$ to estimate the displacement of a non-feature point. The scattering function is

$$S(p) = \sum_i c_i \phi(||p - p_i||) + Ap + B \tag{1}$$

where $\phi(r) = e^{-r/32}$ is a radial basis function, and c_i are weighted coefficients, and A, B are affine terms. Coefficient c_i, A and B can easily be solved by linear equations with constraints: $u_i = S(p_i), \sum_i c_i = 0, \sum_i c_i x_i = 0, \sum_i c_i y_i = 0$ and $\sum_i c_i z_i = 0$.
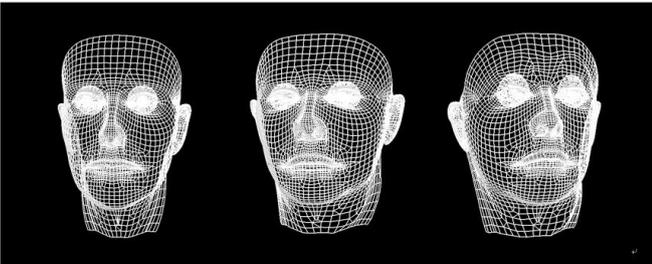

Fig. 3 Deformation of a generic model for personalization and facial expression.

3.2 Detailed Facial Motion by Shape-from-shading

For subtle motions, e.g. wrinkles or creases, we propose recovering time-varying detailed surface V from shading information, spatial and temporal coherence. To solve the intrinsic ill-condition problem of shape-from-shading, we adopt an optimization method that minimizes the difference between captured image I_t and synthesized image Syn_t . The objective function O becomes

$$O = \sum_{t=1}^{Num_T} (Syn_t - I_t)^2 \tag{2}$$

, where Num_T is the frame amount. For simplification of introduction, our explanation in the following subsections is for a single image sequence. It can then be extended to two-view sequences. Besides, for computational consistency, all of the expressional images are automatically warped to the neutral face before any process.

In order to decrease the degrees of freedom (DOF) of the objective function, without loss of generality, we represent the 3D data in terms of height maps. These heights can easily be transformed to normals from gradients of neighbor z values.

Hence, only the z (height) values of aligned pixels (vertices) are evaluated. The shape parameters V are therefore defined as:

$V = (V_1, V_2, ...V_t, ...V_{Num_T})$, where
$V_t = (z_{t1}, z_{t2}, ...z_{tp}, ...z_{tNum_P})$ and Num_P is the amount of vertices.

We choose Phong model as the analytic reflection model, and other BRDF model can also be applied. Phong model is widely used in the computer graphics and differentiation of parameters is relatively straightforward. Given a light source L with direction N_L and the surface normal N_{tp}, the reflection intensity on vertex p related to normal can be written as:

$$Syn_{tp} = I_d \cdot k_d(N_L \cdot N_{tp}) + I_s \cdot k_s(e \cdot r)^\alpha \tag{3}$$

where k_d and k_s are the diffuse and specular coefficients and α is the Phong exponent term. The vector e denotes the eye direction and r is the reflection vector with respect to N_L and N_{tp}.

Assume that the reflectance parameter of a subject's face $R = (k_d, k_s, \alpha)$ is uniform in a region. The Eq(2) becomes

$$O(V, R) = \sum_{t=1}^{Num_T} \sum_{p=1}^{Num_P} (Syn_{tp} - I_{tp})^2 \tag{4}$$

In other words, our goal is to find the best surface sequence V^* and reflectance parameter R^* that will minimize the objective function.

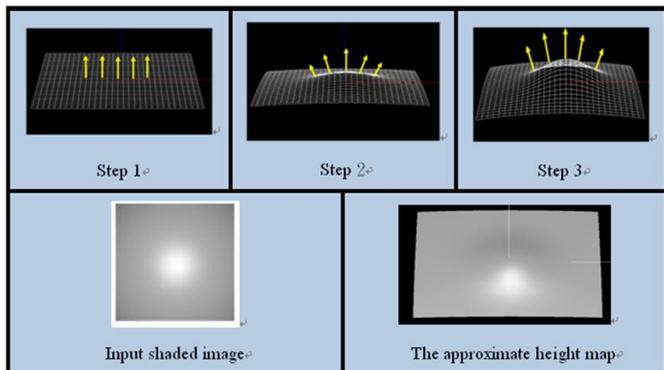$$< V^*, R^* >= \arg \min O(V, R) \qquad (5)$$

Fig. 4 A conceptual diagram of progressive surface estimation with a height map.

4 Space-time Shape-from-shading

In our experiments, we found that the estimated normals or surfaces are highly sensitive to noise, while applying direct shape-from-shading methods [13,14].

While using optimization methods on Eq(5), we used the primitive surfaces from motion capture as initial values, and we apply modified conjugate-gradient methods to approach the variables to real surfaces. Nevertheless, they were easily trapped into local-minimums in early stages and undesired trembling effects occurred.

For more reliable detailed motions, we propose using a multi-stage optimization scheme and space-time constraints.

4.1 Multi-stage Optimization

Our objective function has two sets of parameters, the shape parameter V and reflectance parameter E. R is a time-invariant parameter set, but V represents only local geometry. If we simultaneously estimate these two parameters into a single stage, we need a proper scale between these two kinds of parameters for balanced influences. In order to avoid bias, we optimize these two sets of parameters separately.

Moreover, from our experiment, we found that directly apply all BRDF parameters in optimization, the specular components will dominate the objective function, and therefore, the process will be trapped into local minimum. On the contrary, the reflection of human skin is mainly contributed by diffuse components.

Hence, we introduce an intermediate stage, diffuse-shape optimization, where the shape V and diffuse parameter $R_{diffuse}(kd)$ are iteratively optimized. After this stage, the BRDF-shape optimization is performed for all diffuse and specular parameters. The flow chart of the multi-stage optimization is shown in Fig.5.
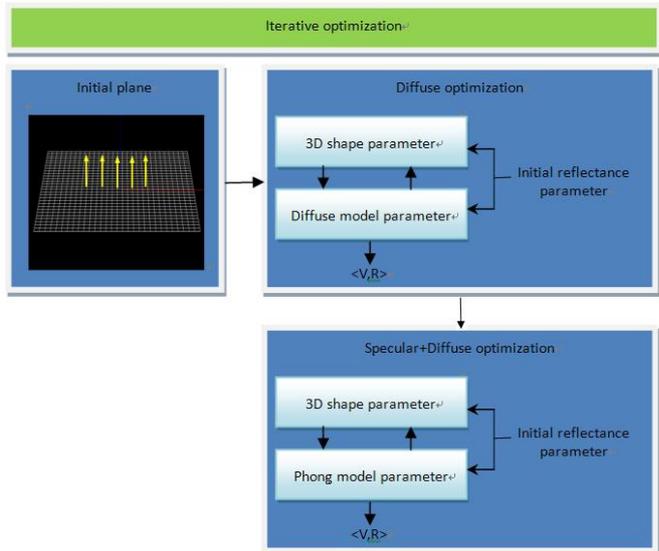

Fig. 5 The flow chart of multi-stage optimization.

4.2 Space-time Constraints

Since our faces are mostly continuous surfaces, the motion of a vertex p has high spatial coherence with its neighbors. Therefore, we use spatial coherence constraints to alleviate the inherent noise sensitivity in shape-from-shading.

$$CS_{tp} = k_{CS}(z_{tp} - \sum_j \frac{1}{w_j} z_{tj})^2, \text{for } j \in Neighbor(p) \qquad (6)$$

where $Neighbor(p)$ denotes the 8-neighbor pixel set, w_j is an adaptive weight and k_{CS} is the weight for spatial constraints.

When adopting only spatial constraints in optimization, we found that there're still flickers caused by input noise. According to biomechanics properties of facial muscles and tissues, a human facial surface should "gradually" transit between expressions. Hence, the temporal coherence can further improve the optimization.
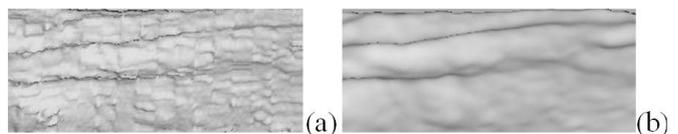

Fig. 6 The optimized shapes (a) without spatial coherence constraints (b) with spatial coherence constraints.

To include temporal coherence, most research tracked the pixel flow and added smooth constraints on successive pixels. However, it will produce new errors resulted from inaccurate pixel correspondence occlusion. To avoid these conditions, we directly apply our temporal constraints to each pixel in the height map sequence. The global head motion will later be compensated by head tracking data.

The temporal constraint is as follows:

$$CT_{tp} = k_{CT}(z_{tp} - \sum_i \frac{1}{w} z_{(t+i)p})^2, \text{where i} = [-3, 3] \quad (7)$$

Therefore, our objective function becomes

$$O = \sum_{t=1}^{Num_T} \sum_{p=1}^{Num_P} [(Syn_{tp} - I_{tp})^2 + CS_{tp} + CT_{tp}] \quad (8)$$

Applying optimization to the whole image sequences is extremely time-consuming. For computational efficiency, we only apply our optimization on a small window of frames at a time. When sweeping the windows from the start to the end and combining the individual results, a pseudo-optimal surface sequence can be acquired. This approach, significantly decreasing degrees of freedom, and dramatically reduces the processing time.

4.3 Post Processing

Due to the error caused by pixel alignment, input noise, and digitization, etc, there're still unavoidable estimation errors. Users may apply an adaptive bilateral filter to reduce these noises and retain the details.

The last stage of acquisition is two-view data fusing. In traditional computer vision, data of one view should be fused with the best correlated ones. Nevertheless, it will again encounter the troublesome corresponding problem at textureless regions. Or two-view optimization should be applied.

Instead, we apply the view-dependent concept from image-based rendering. Given two synchronized height maps and feature points, we apply the view morphing technique [23] to warp the two side views into the front view. Then, we weighted-blend the two view data according to the view angles. Data that are more close to their original view axis will get a higher ratio.

5 Experiments and Results

This section describes and discusses our experiment. Then, we show the synthesized results where primitive and detailed motions are included.

5.1 Experiments

In our system, we used two video streams to track markers' motions and estimate time-varying height maps as well. To acquire more accurate facial details, our experiments are performed under an illumination-controlled environment. A calibrated spotlight was set as the single light source. Video streams were captured by two synchronized high-definition video cameras (HDV with resolution 1280*720 pixel and 30 frames-per-second). We pasted 25 to 30 markers on subjects' faces and avoided placing markers on regions with wrinkles or creases. Two subjects, one male and one female, were required to perform various facial expressions.

Since we aim at enhancing subtle variations for motion capture, for efficiency and reliability, users can assign regions for detail estimation. Fig.7 shows the captured neutral image and the user-assigned regions by a brushing interface. In our experiments, we preferred areas with more wrinkles and creases, such as the forehead, glabella, left and right cheeks.

Fig. 7 Two views of the neutral face and the user-assigned regions for detail estimation.

5.2 Results of Motion Acquisition

In our research, we utilize a novel approach to solve the ill-condition of detailed motion tracking. This method can optimize the geometry and reflectance parameters by minimizing the objective function. The cost values gradually decreased and the optimization was stopped when the improvement was less than the threshold.

Fig.8 shows the progress of optimization stages. In the first stage, we just optimized the diffuse term to get a more accurate initial shape. After the second stage, specular terms are included. The estimated surfaces gradually approached the input image sequence. Another result of shape recovery is shown in Fig.9. We also compare our space-time SFS with a direct lambertian-based method [13]. Our estimated surface is more accurate and noises are suppressed.

We used a conjugate gradient method for minimization.

The performance of our optimization is around 120 to 170 seconds per frame on a Pentium IV 3.2GHz with 1GB memory.
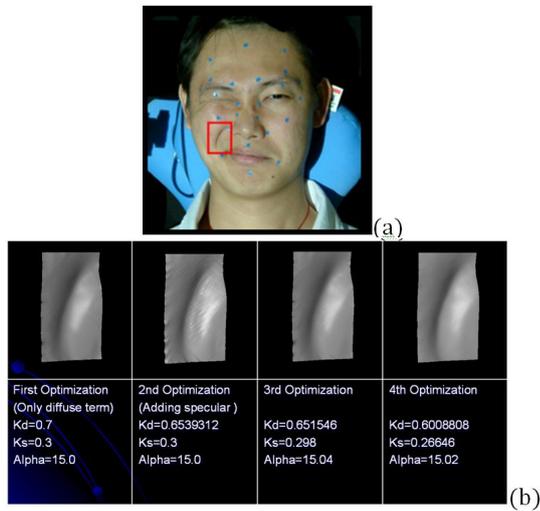

Fig. 8 A captured wrinkle image and the progressive refined surfaces in optimization stages.
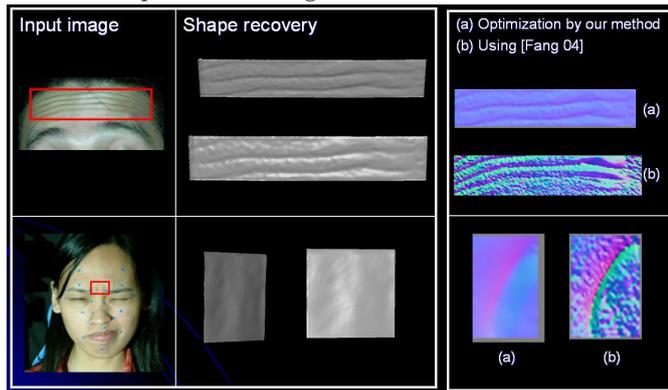

Fig. 9 Results of the proposed space-time SFS method and the Lambertian-based SFS method.

5.3 Synthesis of Facial Details

As mentioned in subsection 3.1, we drive a characterized neutral face according to motion capture data. Our 3D head model has 6078 vertices and 6315 polygons. We separate the face into segments including the forehead, nose, upper mouth, lower mouth, cheeks, jaw, etc. For deformation, we apply RBF-based data scattering to each segment and gradually blend at the boundaries.

To enhancing details with estimated height maps, the target polygons are first subdivided and per-pixel normal mapping is then applied. Fig.10 shows the subdivision results according to the masked regions and wrinkles. Fig.11 shows the facial animation with facial details. The synthesized results can be improved with displacement mapping and realistic face rendering.
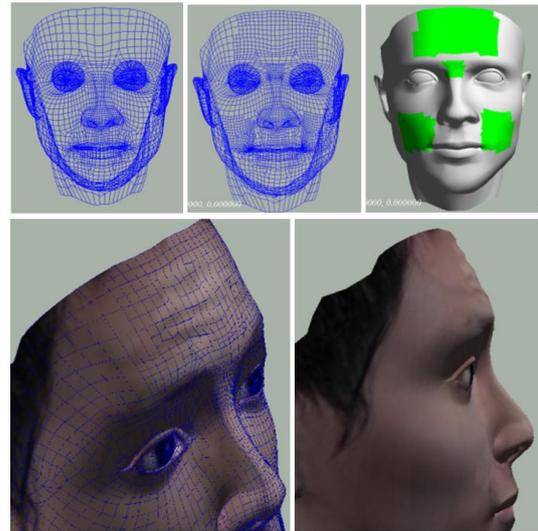

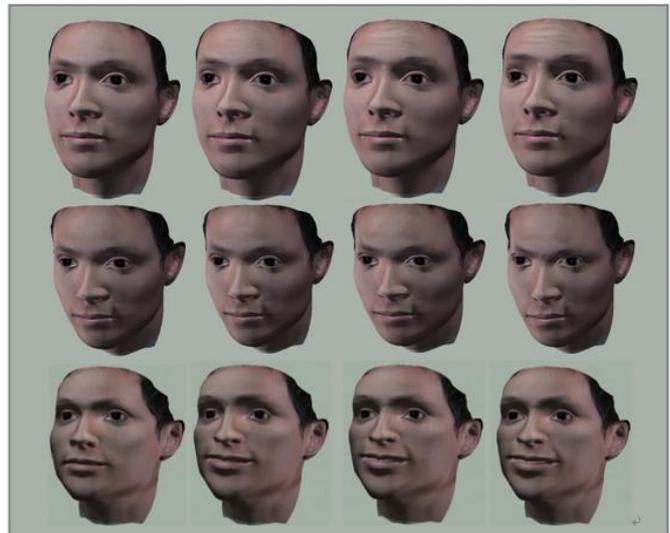**Fig. 10** Subdivision on marked regions and wrinkle synthesis.

Fig. 11 Retargetting the detailed motions to a personalized model.

6 Discussion and Conclusion

This paper is aimed at reliably estimating 3D facial details with inexpensive devices. We combine the advantages of stereo triangulation and shape-from-shading for detailed motion tracking. While including both spatial and temporal coherence as constraints, we can improve the reliability.

To optimize a large set of parameters, we propose an iterative scheme with multiple stages, where diffuse components, diffuse-specular components and time-varying surfaces are progressively improved. Our results demonstrate the effectiveness of the proposed methods.

Compared with related work, we don't assume the shape of wrinkles, and therefore, our method is more general for various details. Our method doesn't require additional projectors or painting on subjects' wrinkles. It can even apply to existing motion capture video if the illumination is controlled.

Several issues can be improved in the future. At present, we assume the skin reflection properties are uniform and alleviate this problem by compensating color differences and filtering. Individual reflection parameters for each pixel will be included.

As mentioned above, when we directly integrate BSS-RDF model in our shape optimization, the computation is inefficient and it can barely improve those with BRDF model. A more dedicated experiment or a simplified model should be conducted. Besides, other BRDF models, such as Cook-Torrance model, example-based models, can be utilized for better approximation.

Currently, our shape-from-shading is applied to regions with rich undulation, and therefore, the surface resolution is inconsistent. We plan to extend the proposed method to the whole face for more faithful capture and animation.

References

1. Deng, Z., Chiang, P.-Y., Fox, P., Neumann, U.: Animating Blendshape Faces by Cross-Mapping Motion Capture Data, Proceedings of ACM Symp. Interactive 3D Graphics and Games '06, pp. 43-48.
2. Zhang, L., Snavely, N., Curless, B., Seitz, S.M.: Spacetime Faces: High Resolution Capture for Modeling and Animation, Proceedings of ACM SIGGRAPH '04, pp. 548-558 (2004)
3. Wenger, A., Gardner, A., Tchou, C., Unger, J., Hawkins, T., Debevec, P.: Performance Relighting and Reflectance Transformation with Time-Multiplexed Illumination, ACM Trans. Graphics, vol.24, no.3, pp.756-764 (2005)
4. Huang, T.S., Netravali, A.N.: Motion and Structure from Feature Correspondences: A Review, Proceedings of IEEE, vol. 82, no. 2, pp. 252-268 (1994)
5. Guenter, B., Grimm, C., Wood, D., Malvar, H., Pighin, F.: Making Faces, Proceedings of ACM SIGGRAPH '98 , pp. 55-66 (1998)
6. Lin, I.-C., Yeh, J.-S., Ouhyoung, M.: Extracting 3D Facial Animation Parameters from Multiview Video Clips, IEEE Computer Graphics and Applications, vol.22, no. 6, pp.72-80 (2002)
7. Lin, I.-C., Ouhyoung, M.: Mirror Mocap: Automatic and Efficient Capture of Dense 3D Facial Motion Parameters

from Video, The Visual Computer, vol.12, no.6, pp. 355-372 (2005)
8. Vogiatzis, G., Torr, P.H.S., Cipolla, R.: Multi-view Stereo via Volumetric Graph-cuts, Proceedings of IEEE Computer Vision and Pattern Recognition '05, vol.2, pp. 391-398 (2005)
9. Magnenat-Thalmann, N., Kalra, P., Luc Leveque, J., Bazin, R., Batisse, D., Querleux, B, IEEE Trans. Information Technology in Biomedicine, vol.6, no.4, pp.317 - 323 (2002)
10. Sifakis, E., Neverov, I., Fedkiw, R.: Automatic determination of facial muscle activations from sparse motion capture marker data, ACM Trans. Graphics, vol.24, no.3, pp.417-425 (2005)
11. Weyrich, T., Matusik, W., Pfister, H., Lee, J., Ngan, A., Jensen, W., Gross, M.: Analysis of Human Faces using a Measurement-Based Skin Reflectance Model, ACM Trans. Graphics, vol.25, no. 3, pp.1013-1024 (2006)
12. Seitz, M., Hertzmann, A.: Example-Based Photometric Stereo: Shape Reconstruction with General, Varying BRDFs, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27 no. 8 , pp. 1254-1264 (2005)
13. Fang, H., Hart, J.C.: Textureshop: Texture Synthesis as a Photograph Editing Tool, ACM Trans. Graphics, vol.23, no. 3, pp.354-359 (2004)
14. Horn, B.K.: Height and Gradient from Shading, Intl. Journal of Computer Vision, vol.5, no.1, pp.37-75 (1990)
15. Zeng, G., Matsushita, Y., Quan, L., Shum, H.Y.: Interactive Shape from Shading, Proceedings of IEEE Conf. Computer Vision and Pattern Recognition, vol.1, pp.343-350 (2005)
16. Yu, T., Xu, N., Ahuja, N.: Recovering Shape and Reflectance Model of Non-lambertian Objects from Multiple Views, Proceedings of IEEE Computer Vision and Pattern Recognition '04, vol.2, pp.226-233 (2004)
17. Bickel, B., Botsch, M., Angst, R., Matusik, W., Otaduy, M., Pfister, H., Gross, M.: Multi-Scale Capture of Facial Geometry and Motion, ACM Trans. Graphics, vol.26, no. 3, pp. 33.1-33.10 (2007).
18. Witkin, A., Kass, M.: Spacetime constraints, Proceedings of ACM SIGGRAPH '88, pp.159-168 (1988)
19. Gleicher M.: Retargetting Motion to New Characters, Proceedings of ACM SIGGRAPH '98, pp.33-42 (1998)
20. Fang, H., Hart, J.C.,: RotoTexture: Automated Tools for Texturing Raw Video, IEEE Trans. Visualization and Computer Graphics, vol. 12, pp.1580-1589 (2006)
21. Jensen, H.W., Marschner, S., Levoy, M., Hanrahan, P.: A Practical Model for Subsurface Light Transport, Proceedings of ACM SIGGRAPH '01, pp.511-518 (2001)
22. Donner, C., Jensen, H.W.: Light Diffusion in Multi-Layered Translucent Materials, ACM Trans. Graphics, vol.24, no.3, pp.1032-1039 (2005)
23. Seitz, S.M., Dyer, C.R.: View Morphing, Proceedings of ACM SIGGRAPH '96, pp.21-30 (1996)
24. Lin, I.-C., Yeh, J.-S., Ouhyoung, M., "Realistic 3D Facial Animation Parameters from Mirror-Reflected Multi-View Video", Proc. Computer Animation 2001, pp. 2-11 (2001)
25. Zhu, L. and Lee, W.-S.: Facial Animation with Motion Capture Based on Surface Blending, Proceedings of International Conference on Computer Graphics Theory and Applications (GRAPP'07), pp. 63-70 (2007)