FEATURE-POINT DRIVEN 3D EXPRESSION EDITING

Chii-Yuan Chuang, I-Chen Lin, Yung-Sheng Lo and Chao-Chih Lin
Department of Computer Science, National Chiao Tung University, Taiwan

Keywords: Facial expression, facial animation, graphical interfaces, surface reconstruction.

Abstract: Producing a life-like 3D facial expression is usually a labor-intensive process. In movie and game industries, motion capture and 3D scanning techniques, acquiring motion data from real persons, are used to speed up the production. However, acquiring dynamic and subtle details, such as wrinkles, on a face are still difficult or expensive. In this paper, we propose a feature-point-driven approach to synthesize novel expressions with details. Our work can be divided into two main parts: acquisition of 3D facial details and expression synthesis. 3D facial details are estimated from sample images by a shape-from-shading technique. While employing relation between specific feature points and facial surfaces in prototype images, our system provides an intuitive editing tool to synthesize 3D geometry and corresponding 2D textures or 3D detailed normals of novel expressions. Besides expression editing, the proposed method can also be extended to enhance existing motion capture data with facial details.

1 INTRODUCTION

Nowadays, 3D characters or avatars have been popularly used in various kinds of media; however, generating realistic facial expression is still a labor-intensive work for animators. Human faces are the most expressive parts of our appearance, and any subtle difference may have dramatically different meanings.

Recently, motion capture (mocap) techniques are popularly utilized to speed up the production of 3D facial animation. However, there are still subtle portions, such as wrinkles or creases, whose variations are smaller than markers. These subtle portions are difficult to be acquired by mocap techniques. Our goal is to enhance the feature-driven face system with facial details.

Figure 1 shows the difference among a real face, synthetic faces with or without the facial details. With the same motion of lifting eyebrow, the detailed face is more expressive.

To efficiently generate 3D expressions with details, two challenges have to be tackled. The first challenge is the estimation of facial surface details. In our research, we make use of stereo triangulation for the approximate geometry of facial expression. To estimate the facial details, we propose using a shape-from-shading (SFS) method since SFS techniques can avoid unreliable point matching.

The second challenge is expression synthesis.

Figure 1:(a) a synthetic neutral face, (b) a synthetic face with lifting eyebrows but without details; (c) a synthetic face with lifting eyebrows and estimated details; (d) a real face with lifting eyebrows.

Chuang C., Lin I., Lo Y. and Lin C. (2007).
FEATURE-POINT DRIVEN 3D EXPRESSION EDITING.
In *Proceedings of the Second International Conference on Computer Graphics Theory and Applications - AS/IE*, pages 165-170
DOI: 10.5220/0002074201650170

Since it's infeasible to acquire every expression by the capture technique, we propose using optimal weighted blending. Given a set of feature point positions adjusted by users, our system utilizes an optimization approach to select appropriate prototype expressions from the data pool and calculates the best weights for blending. The surface details are represented in terms of normal difference maps, which can be efficiently rendered by pixel shaders.

The major contribution of our system is that we propose an intuitive and inexpensive framework for acquisition and editing detailed expressions. Sample expressions are evaluated by feature-point-driven face reconstruction and shape-from-shading techniques. The novel expressions can be synthesized by optimal projection in expression space. Moreover, novel expressions can also be retargeted to models of other subjects.

1.2 System Overview

Our system can be divided into two parts: offline processing and online novel expression synthesis. Figure.2 shows an overview of our system.

Offline processing is about the preprocessing of our prototype images. This offline part will describe in Section 3. The online synthesis will be described in section 4.

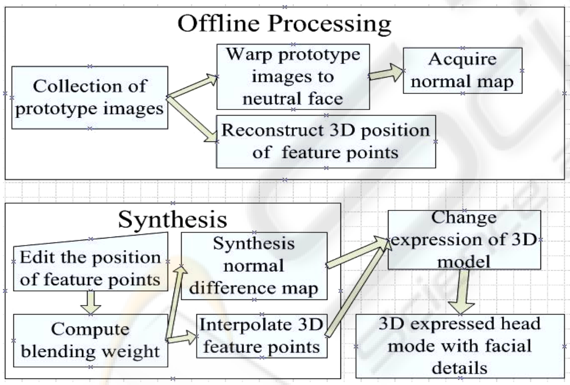

Figure 2: The system overview.

2 RELATED WORK

Facial animation can be roughly divided into two groups according to their basic structures. The first one is image-based facial animation, and the other is model-based facial animation. Image-based approaches employ one or several real facial images to synthesize novel images. These kinds of systems can reach a photorealistic quality but are difficult to relight and to alter view directions.

In 2002, Ezzat et al., developed a multidimensional morphable model (MMM) to synthesize unrecorded mouth configuration from the set of mouth prototypes.

In contrast to image-based facial animation, 3D model-based animation is more versatile. However, it needs more control parameters for modelling, animating, or rendering.

In Sifakis' research (Sifakis et al., 2005) they proposed an anatomical face model controlled by muscle activations and kinematic bone degrees of freedom. Their system can automatically compute control values from sparse captured maker input.

Based on a large set of 3D scanned face examples, (Blanz and Vetter, 1999) built a morphable head model. By the linear combination of prototypes in MMM space, new faces can be modelled. In 2003, Blanz et al., further transferred facial expressions by computing the difference between two scans of the same person in a vector space of faces.

In contrast to static scans, (Zhang et al., 2004) proposed a structured light approach to capture the dynamic variation of a face. Their system utilized two projectors and six cameras for the structured light-based depth estimation. Besides, they presented a keyframe interpolation technique to synthesize in-between video frames and a controllable face model.

A geometry-driven approach proposed by Zhang et al. in 2006, synthesizes facial expression through the relation between positions of specific feature points and expressions. They utilized an expression vector space and a new 2D expression can be synthesized by solving an optimization. We adapt their approach and extract expression details in terms of normal maps.

In addition to facial details due to different expressions, (Golovinskiy et al., 2006) developed a statistical model for static facial details. They acquired high-resolution face geometry across a wide range of ages, genders, and races. They further use the texture analysis/synthesis framework to enhance details on a static face.

In addition to facial animation, our research is also related to extract the surface variation from images. Weyrich et al., in 2006 proposed a practical skin reflectance model whose parameters can be robustly estimated from measurements. They utilized the photometric stereo to reconstruct the 3D face geometry model but the experiment devices are expensive. (Fang et al., 2004) adapted Horn's (Horn, 1990) approach and simply utilized Lambertian reflection model to extract the normal map from a single image. Their approach spends less time and doesn't need expensive equipment. If readers are interested in SFS techniques please refer to a detailed survey by Zhang et al., in 1999.

3 ACQUISITION OF EXPRESSION DETAILS

In our research, we synthesize novel expression from several sets of prototype samples. In order to acquire 3D detailed face structure, we use stereo triangulation for conspicuous markers in two views. While morphing a generic model according to these 3D positions, we can acquire an approximate geometry. However, stereo triangulation is unreliable for detailed variation, since the point correspondences are difficult to find. Therefore, we utilize a modified shape-from-shading (SFS) technique proposed by Fang et al., in 2004 and Horn in 2000, to deal with the facial details.

3.1 Normal Recovery

First, we simply assume that the color of the skin is uniform. The intensity variations one result only from the variations of the angle between surface normals and incident light directions. Based on these assumptions, the normal can be efficiently extracted from a single image under a single light source as follow.

As shown in Figure. 3, let L be the unit vector of the light source direction. To evaluate the surface normal N_{xy} of a pixel in the image, first, we have to estimate a projection vector P_{xy} as shown in Eq. (1).

$$P_{xy} = \nabla I_{xy} - (\nabla I_{xy} \cdot L)L \qquad (1)$$

where $\nabla I_{xy} = (\frac{\partial I_{xy}}{\partial x}, \frac{\partial I_{xy}}{\partial y}, 0)$ is the image gradient and P_{xy} is the projection of vector ∇I_{xy} to the plane perpendicular to L.

We assume that the darkest intensity value, $I_{\min}$, implies the intensity of ambient light in the scene and the brightest value, $I_{\max}$, indicates the intensity when a pixel faces the light source.

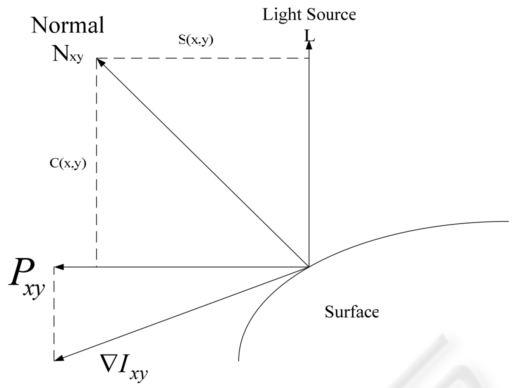

Figure 3: Normal recovery.

Then the cosine of angle between the surface normal and the incident light direction can be evaluated as follows:

$$C(x,y) = (I_{xy} - I_{\min})/(I_{\max} - I_{\min})$$

Therefore, the sine value S(x,y) between the surface normal and the incident light direction can also be calculated.

The normal can be estimated through eq.(2)

$$N_{xy} = C(x,y)L + S(x,y)P_{xy} / \|P_{xy}\| \qquad (2)$$

We demonstrate the estimated normals in Figure 4

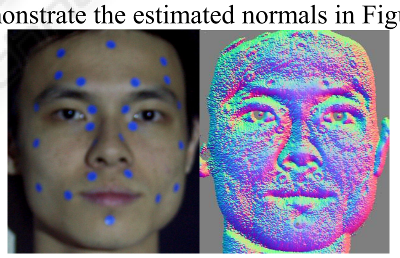

Figure 4: The left figure is the acquired image and the right one is the recovered normal array illustrated in R,G,B channels.

3.2 Normal Difference Map

When we applied the shape-from-shading (SFS) technique based on the uniform-skin-color assumption, some defective normals occur. Color variations on human skin, acnes, scars etc. may also make the image gradients changes dramatically. Instead of applying normal maps directly, we propose using a normal difference map to alleviate defects.

The normal difference map can be calculated by subtraction of the normal map of expressed face to the normal map of the neutral face.

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

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.

We can add the facial details to 3D model by modifying the original surface normal according to the normal difference. With the normal difference maps, the defects of uniform-skin-color assumption are alleviated.

Due to the error caused by pixel alignment, input noise, and digitization, etc, our normal difference maps still have some unavoidable estimation errors. We utilize an adaptive Gaussian filter to reduce the noise problem.

4 FEATURE-POINT-DRIVEN SYNTHESIS OF NOVEL EXPRESSIONS

In the previous section, we described how to acquire facial details from an image. However, it is infeasible to acquire all possible facial expressions. Thus, we develop an approach to synthesize novel expressions from prototypes.

We assume that expressions are highly related to the movements of specific feature points. Here, we assume that similar expressions will have similar movements of feature points.

This assumption is similar to the work proposed by Zhang et al., in 2006. They use an optimization approach which focused on synthesizing 2D textures, but we further synthesize 3D geometry and normals and also apply additional constraints on prototype images.

4.1 Calculation of Blending Weights

We take the concept of vector space interpolation to deal with these two targets. We regard an expression as a combination of 3D geometry and appearance (2D textures or normal difference maps). If we establish a proper vector space to represent the 3D geometry and appearance of expressions, we could approximate novel expressions by interpolating several prototypes with appropriate blending weights. The interpolation can be represented as Figure 5 and the calculation of blending weights is described below.

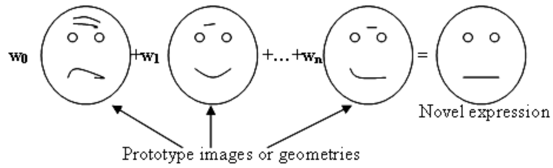

Figure 5: w_i represents the blending weight. A novel expression is synthesized by interpolating prototypes.

To calculate the blending weights, we utilize an expression vector space. Each expression is represented as $E_i = (G_i, N_i, T_i)$ where E_i, G_i represents an expression, and geometry respectively. N_i is the surface normal and T_i is the face texture. Let

$$\left\{\left(\sum_{i=0}^{m} w_i G_i, \sum_{i=0}^{m} w_i N, \sum_{i=0}^{m} w_i T\right) \middle| \sum_{i=0}^{m} w_i = 1, \text{ and } w_0, w_1, ..., w_m \geq 0\right\}$$

$H(E_0, E_1, ,...,E_m)$ be the space of all possible convex combinations of these examples.
$H(E_0, E_1,...,E_m)$=

We can represent novel expressions as follows:

$$E_{(new)} = [G_{(new)}, N_{(new)}, T_{(new)}]$$

$$\text{where } G_{(new)} = \sum_{i=0}^{m} w_i G_i, T_{(new)} = \sum_{i=0}^{m} w_i T_i,$$

$$\text{and } N_{(new)} = \sum_{i=0}^{m} w_i N_i$$

As shown above, the blending weight w_i is related to normal, texture, and geometry. Therefore, we can calculate the weights from one of the three components. In our approach, geometric relation between prototypes is employed to get the blending weights. Let G_i^S denote the feature point set of prototype expressions in our data pool and G^N denote the set of new positions of feature points which is assigned by user. By projecting G^N into the convex hull of $G_0^S G_1^S ... G_m^S$, the weights can be found. Thus, the estimation of blending weight can be written as an optimization problem:

$$\min\left\{\left(\mathrm{G}^{\mathrm{N}} - \sum_{i=0}^{m} w_i G_i^S\right)^T \left(\mathrm{G}^{\mathrm{N}} - \sum_{i=0}^{m} w_i G_i^S\right)\right\},$$

subjects to: $\sum_{i=0}^{m} w_i = 1, w_i \geq 0,$ *for i = 0,1,...,m.*

The objective function of the optimization problem above can be rewritten as:

$$W^T g^T g W - 2G^{N^T} g W + G^{N^T} G^N$$

$$\text{where } g = (G_0^S, G_1^S, ..., G_m^S), W = (w_0, w_1, ..., w_m)$$

This optimization is a positive semi-definite quadratic programming problem with linear constraints. We used the active set strategy to solve this optimization problem.

4.2 Synthesis of Textures and Normals

After calculating the blending weights, we synthesize the texture or normals by interpolating the prototypes. For convenience of interpolation, we align the pixels of prototypes by the warping method proposed by Beier and Neelyin 1992. After alignment of pixels, the synthesis can be performed as weighted summation of corresponding pixels in all images.

Since there are only 15~25 prototypes in our database, if simply interpolating the prototypes, the variations of novel expression will be few. Thus, we divide a prototype image into 8 sub-regions. We synthesize each sub-region individually. By dividing sub-regions, we can compose more expressions with a smaller data pool.

To avoid image discontinuities along the sub-region boundaries, we also make use of gradual blending between the sub-region boundaries.

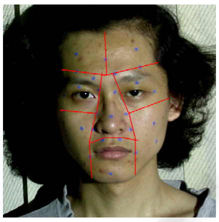

Figure 6: We divide a face into 8 sub-regions to increase the probable novel expression.

5 EXPERIMENT AND RESULTS

In our system, we use 15~25 prototype images for synthesizing 2D novel expressions. In order to acquire feature details from our synthesized images, our prototype images are taken under an illumination-controlled environment.

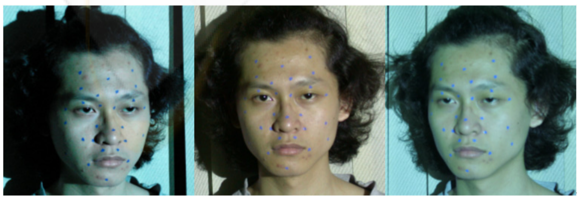

Figure 7: A set of prototype image consist of 3 different views. Left and right ones are used to recover the 3D geometry of feature points. The central view is use to acquire normal maps.

Only a single light source is applied in this environment. Three cameras are used to take pictures from different views of our model. Figure 7 shows a set of our prototype images.

Figure 8: The synthesized facial expressions.

6 CONCLUSION AND FUTURE WORK

In this paper, a 3D detailed facial expression synthesis system is presented. The system consists of three functions — 3D expression editing, 2D texture synthesizing, and acquisition of normal difference maps. By manipulating the positions of feature points, users could modify the expression of a 3D head model. Then, by deriving the relationship between modified geometry and geometry of our prototypes, the corresponding normal difference maps could be synthesized. We demonstrated that a corresponding normal difference map could be evaluated by a low cost method.

Our contributions include: 1) a low-cost and efficient method to acquire and synthesize facial details, and 2) a framework for editing and synthesizing 3D detailed expression.

The current system has several feature work that can be improved. First, our normal difference map is acquired by a simplified method. Other shape-from-shading methods (Horn, 1990), (Wenger et al., 2005) can acquire a more accurate result. Second, the positions of feature points are decided empirically. Further analysis could increase the fidelity. Third, the expression produced could be extended for facial animation.

REFERENCES

Beier, T., and Neely, S. "Feature-based Image Metamorphosis", Proc. ACM SIGGRAPH'92, Pages 35-42, 1992.

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

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

Breger, C., Covell, M., and Slaney, M. "Video rewrite: Driven Visual Speech with Audio", Proc. ACM SIGGRAPH'97, Page 353-360, 1997.

Deng, Z., Chiang, P-Y., Fox, P., and Neumann, U. "Animating Blendshape Face by Cross-Mapping Motion Capture Data", Proceedings of the 2006 Symposium on Interactive 3D Graphics and Games, Pages 43-48, 2006.

Ezzat, T., Geiger, G., and Poggio, T. "Trainable Videorealistic Speech Animation", Proc. ACMSIGGRAPH'02, Pages 388-398, 2002.

Fang, H., and Hart, J. C. "Textureshop: Texture Synthesis as a Photograph Editing Tool", ACM Trans. on Graphics., Volume 23, Issue 3 (August 2004), Pages 354-359, 2004.

Fletcher, R. "Practical Methods of Optimization", Vol. 1, John Wiley & Sons, 1980.

Forsyth. Ponce "Computer Vision: a Modern Approach", Prentice Hall, 2003.

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.

Guenter, B. K., Grimm, C., Wood, D., Malvar, H. S., and Pighin, F. H., "Making Faces", Proc.ACM SIGGRAPH'98, Pages 55-66, 1998.

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

Laurence, B., Gergo, K., Nadia M. T., Prem K., "Simulation of Skin Aging and Wrinkles with Cosmetics Insight", Computer Animation and Simulation, Pagers 15-27, 2000

Lin, I.-C., Yeh, J.-S., and Ouhyoung, M. "Extracting 3D Facial Animation Parameters from Multiview Video Clips", IEEE Computer Graphics and Applications, Vol. 22(6), Pages 72-80, 2002.

Liu, Z., Shan, Y., and Zhang, Z. "Expressive Expression Mapping with Ratio Images", Proc. ACM SIGGRAPH'01, Pages 271-276, 2001.

Nielson, G. M. "Scattered Data Modeling" IEEE Computer Graphics and Applications, Vol.13, Issue 1, Pages 60-70, 1993.

Pighin, F., Hecker, J., Lischinski, D., Szeliski, R., and Salesin, D. H. "Synthesizing Realistic Facial Expressions from Photographs", Proc. ACM SIGGRAPH'98, Pages 75-84, 1998.

Sifakis, E., Neverov, I. and Fedkiw, R., "Automatic Determination of Facial Muscle Activations from Sparse Motion Capture Marker Data", ACM Trans. on Graphics, Vol. 24, Issue 3, Pages 417-425, 2005.

Tu, P-H., Lin, I-C., Yeh, J-S., Liang, R-H., and Ouhyoung, M. "Surface Detail Capturing for Realistic Facial Animation", Journal of Computer Science and Technology, Vol. 19(5), Pages 618-625, 2004.

Waters, K. "A Muscle Model for Animating Three-Dimensional Facial Expression", Proc. ACMSIGGRAPH'87, Pages 17-24, 1987.

Wenger, A., Gardner, A., Tchou, C., Unger, J., Hawkins, T., and Debevec, P. "Performance Relighting and Reflectance Transformation with Time-Multiplexed Illumination", ACM Trans. on Graphics, Vol. 24, Issue 3, Pages 756-764, 2005.

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.

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

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.

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.

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.

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