Journal of Zhejiang University-SCIENCE C (Computers & Electronics)
ISSN 1869-1951 (Print); ISSN 1869-196X (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

Statistical learning based facial animation*

Shibiao XU†, Guanghui MA, Weiliang MENG, Xiaopeng ZHANG†‡

(*National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China*)

†E-mail: aquathinker@gmail.com; xpzhang@nlpr.ia.ac.cn

Received Dec. 29, 2012; Revision accepted May 7, 2013; Crosschecked June 6, 2013

Abstract: To synthesize real-time and realistic facial animation, we present an effective algorithm which combines image- and geometry-based methods for facial animation simulation. Considering the numerous motion units in the expression coding system, we present a novel simplified motion unit based on the basic facial expression, and construct the corresponding basic action for a head model. As image features are difficult to obtain using the performance driven method, we develop an automatic image feature recognition method based on statistical learning, and an expression image semi-automatic labeling method with rotation invariant face detection, which can improve the accuracy and efficiency of expression feature identification and training. After facial animation redirection, each basic action weight needs to be computed and mapped automatically. We apply the blend shape method to construct and train the corresponding expression database according to each basic action, and adopt the least squares method to compute the corresponding control parameters for facial animation. Moreover, there is a pre-integration of diffuse light distribution and specular light distribution based on the physical method, to improve the plausibility and efficiency of facial rendering. Our work provides a simplification of the facial motion unit, an optimization of the statistical training process and recognition process for facial animation, solves the expression parameters, and simulates the subsurface scattering effect in real time. Experimental results indicate that our method is effective and efficient, and suitable for computer animation and interactive applications.

Key words: Facial animation, Motion unit, Statistical learning, Realistic rendering, Pre-integration
doi:10.1631/jzus.CIDE1307 **Document code:** A **CLC number:** TP391.9

1 Introduction

Facial animation is a continuous process involving coordination control for many facial parts. Face muscles are complicated in structure, leading to rich expressions. Moreover, facial animation is appreciated and evaluated by humans themselves, who are very sensitive to the subtle discrepancy between facial animation effects and the real actions on a face. In this respect, facial animation needs to satisfy more requirements compared to other types of animation; therefore, realistic facial animation simulation becomes a crucial problem in computer graphics, and arouses researchers' interest for computer animation and interactive applications. To achieve realistic facial animation, three requirements must be met: building a good head model, generating real-time natural expression motion, and rendering realistic skin effect for the head model. This paper deals with an effective facial animation simulation and realistic rendering method for computer animation and interactive applications.

At present, there are two main types of methods for facial animation simulation: image-based methods and geometry-based methods (Yao and Chen, 2008). Image-based methods include mainly the morphing method, performance driven method, and expression coding system. Using image-based methods, facial animation effects are calculated according to the corresponding image transformation, and realistic facial animation can be generated flexibly. However, the acquisition of image features and the expression redirection between different models are not very easy. Geometry-based methods include mainly the blend

‡ Corresponding author
* Project supported by the 2013 Annual Beijing Technological and Cultural Fusion for Demonstrated Base Construction and Industrial Nurture (No. Z131100000113007), and the National Natural Science Foundation of China (Nos. 61202324, 61271431, and 61271430)

shape method, parametric method, and muscle-based method. Geometry-based methods refer to direct processing on the surface of a 3D model, which enables fast computation; nevertheless, they require a high-precision head model, and the combination of multiple facial animations is difficult to control. For realistic rendering of facial animation, related research works are deficient. Traditional methods use texture synthesis and mapping technology to directly complete facial animation rendering, which leads to poor rendering effects (Blanz *et al*., 2003; Zhang and Chen, 2003; Zhang *et al*., 2006). The difficulty of facial animation rendering lies in realistic skin rendering. Skin has many subtle visual characteristics, and viewers are acutely sensitive to the appearance of skin. Simple rendering for the target model can generate only insubstantial, blunt, and dry appearance.

We provide a framework to overcome the aforementioned problems of facial animation based on statistical learning.

2 Related works

Image based expression animation: The expression coding system is an expression mapping technology, which provides a muscle motion database to produce all possible facial expressions. This system completes the generation of facial animation by encoding each facial expression, which can achieve a heightened movement effect. The Facial Action Coding System (FACS) is a widely used facial expression coding system, first proposed by Ekman *et al*. (2002). Considering that emotional expression is consistent globally, FACS divides a face model into 44 independent expression motion units, and gives the definition of seven basic expressions (Fig. 1). Happiness is embodied in crow's feet wrinkles, pushed up cheeks, and movement from the muscle that orbits the eye. Sadness is embodied in drooping upper eyelids, losing focus in eyes, and the slight pulling down of lip corners. Surprise is embodied in raised eyebrows, widened eyes, and an open mouth. Anger is embodied in furrowed eyebrows, glaring eyes, and a narrowing of the lips. Fear is embodied in raised and pulled together eyebrows, raised upper eyelids, tensed lower eyelids, and lips slightly stretched horizontally back towards the ears. Disgust is embodied in nose wrinkling and upper lip raised. Contempt is embodied in tightened lip corner and raised on only one side of the face. However, the complicated corresponding relationship in FACS is difficult to implement and extend, and thus it is not suitable for computer facial animation production or interactive expression animation generation.

Performance driven facial animation: The main idea is to manually set a number of feature points on the face of actors, capture the motion vectors of these feature points for various facial expressions, and then use these motion vectors to drive the corresponding feature points of a head model generating facial animation. Lance (1990) was first to present the performance driven facial animation techniques, by using a static facial texture image to capture 2D feature point motion. Guenter *et al*. (2006) extended this method by establishing a facial animation generation system, which can obtain facial geometric data through a 3D scanner. At the same time, they used multiple cameras to capture 182 feature points on the face and achieved desirable results. Since then, several performance driven facial animation techniques have been proposed (Blanz *et al*., 2003; Zhang and Chen, 2003; Zhang *et al*., 2006). Note that the positions of facial feature points in the images must be calibrated prior to starting all of the previously mentioned methods. However, labeling on face is presumptuous, and sometimes unpractical. In addition, landmarks restrict the scope of the geometry information captured.

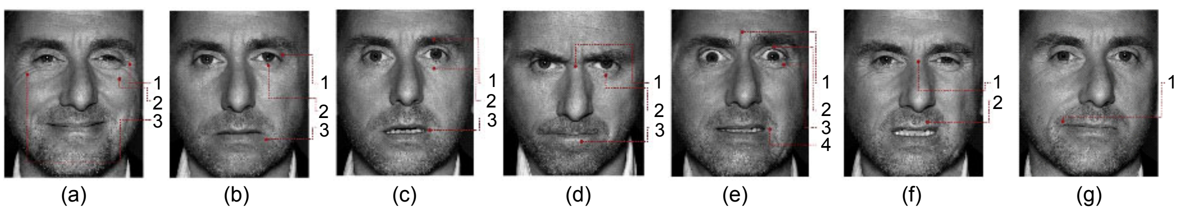

Fig. 1 Seven basic emotional expressions given in the Facial Action Coding System (FACS) (Ekman *et al*., 2002)
(a) Happiness; (b) Sadness; (c) Surprise; (d) Anger; (e) Fear; (f) Disgust; (g) Contempt

Face feature detection: To avoid using the labeling method directly from the video sequences to recover expression movement, in this paper we adopt the facial feature point identification technology, which can track characteristic movement automatically. In recent years, researchers have proposed many detection methods for facial feature points, among which the active shape model (ASM) (Cootes *et al*., 1995) and active appearance model (AAM) (Cootes *et al*., 2001) are the most popular. The ASM method applies sample shape parameterized to construct the shape model of a target object, and then principal component analysis (PCA) to establish the motion model to describe the shape of the control points. It requires a set of parameters to control the position changes of shape control points to make them converge to the current object shape. This method can be applied for direct identification and positioning of face shape and feature points. However, the ASM method requires a higher degree of dependence on the initial conditions, and considers only the shape information of the face, which leads to low accuracy. Cootes *et al*. (2001) extended the ASM method to the AAM method. The differences between these two methods are that the AAM method considers not only local feature information but also global shape and texture information to establish a face model. The AAM method establishes the hybrid model through statistical analysis for the face shape features and texture features, and thus it is intuitive and easy to apply. However, statistical analysis of the ASM and AAM methods highly depends on numerous training samples. As a result, the manual feature point labeling process is very tedious and results in low accuracy.

Geometry-based facial animation: The blend shape method (Parke, 1972) is an intuitive geometry-based method for producing facial animation. This method defines an interpolation function in unit time interval, and can generate smooth movement through the specified key-frames. In fact, the model interpolated can be regarded as a sum of basic models weighed, and each basic model weight can be specified artificially or determined by a certain algorithm (Buck *et al*., 2006). Note that different interpolation functions can produce different results. The linear interpolation function (Deng *et al*., 2006) has been used widely. Being simple and intuitive, the blend shape method has been widely used in most 3D animation software and graphics engines, such as Maya and Unity3D. However, traditional methods manually specify weights for basic models. It is thus difficult to achieve the rich natural combination of many facial expressions. An automatic weight mapping method is required for facial animation.

Skin rendering: At present, realistic skin rendering technology contains mainly three methods: texture space based sub-surface scattering simulation (d'Eon *et al*., 2007), screen space based sub-surface scattering simulation (Jimenez *et al*., 2009), and pre-integrated sub-surface scattering simulation (Penner and Borshukov, 2011). The texture space based method renders the irradiance information of model vertexes to a 2D texture, and approximates the skin diffuse profile through a multi-step Gaussian convolution of the texture. This method can simulate very realistic and accurate scattering effects, even better than many off-line rendering methods, so it is usually regarded as a measurement for realistic skin rendering effects. However, this method needs to compute multiple textures for repetitive convolution operations according to different illumination conditions and animation effects, which requires many hardware computing resources. The screen space based method simplifies rendering for the previously mentioned questions, and calculates the convolution on the rendering result of a head model directly through the deferred shading technology, which can dramatically improve the rendering efficiency. This approach, however, assumes that the eyes and the mouth are closed, and thus there are some restrictions on facial animation. The pre-integrated method is expected to comprehensively consider the calculation complexity of skin rendering. This method renders the diffuse quantity of lighting to a 2D texture with an offline method according to the curvature of the head model, and completes skin rendering in different parts through texture sampling technology, which can avoid the constraints of different illumination conditions and facial animation effects. Furthermore, the popular Phong illumination model is used to deal with specular effects for fixed light. But the Phong model fails to capture increased specular at grazing angles and is not physically plausible. Therefore, a more accurate physically based surface reflectance model is required for realistic skin rendering.

3 Real-time simulation of facial animation

3.1 Simplification of the expression motion unit

Considering the complexity of the FACS system, too many motion units are unbefitting for computer facial animation production and interactive application facial animation generation. The seven basic expressions include the combinations of motion units from eyes, eyebrows, and lips, and we can produce different expressions through changing the related attributes of these motion units. In this study we simplify the expression motion units as seven basic motion units based on the above three main areas of the head model (Fig. 2). The head model is generated through 3D laser scanning for the human face. The scale of the model determines the plausibility of the head model, and also affects the calculation for animation. We adopt the head model on the medium scale as a sample, with 6043 mesh vertexes and 6410 facets. We employ a blend shape method to complete realistic facial animation production. For different facial animations, the weight of each basic action is computed and mapped automatically.

3.2 Statistical training based labeling

In the AAM model 'landmarks' is a basic and simple concept; it is not easy, however, to obtain nice landmarks (Chai *et al.*, 2005). In general, this labeling process needs to manually place hundreds of landmarks on each sample image and keep the corresponding sequence, which is quite difficult.

In this study we propose a statistical training based semi-automatic labeling method for expression images to conquer the problem (Fig. 3). First, we set up an initial and sparse model. The face model of Carnegie Mellon University (Matthews and Baker, 2004), which is the training result from CMU PIE image sets, can be taken as the initial training set

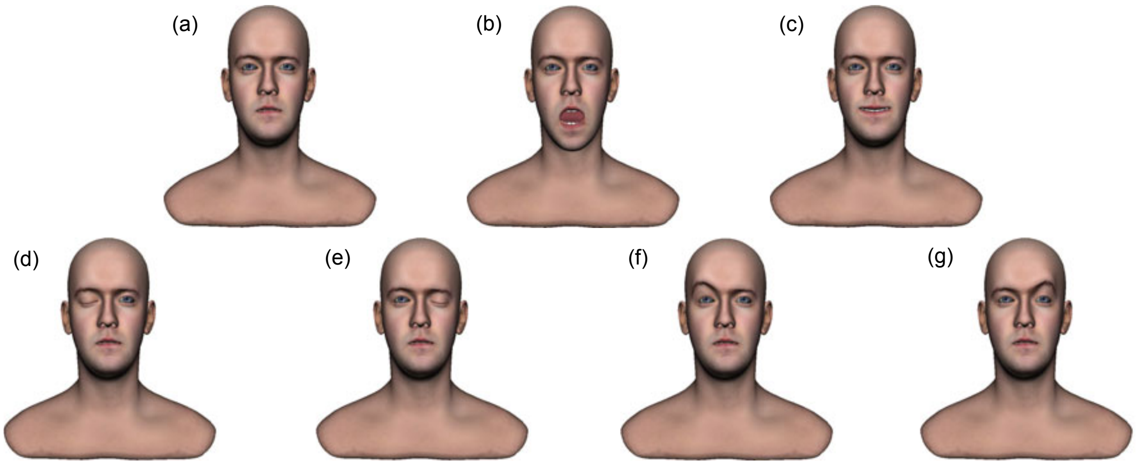

Fig. 2 Seven basic actions of head mode defined in this study

(a) Normal; (b) Mouth opened; (c) Smile; (d) Right eye closed; (e) Left eye closed; (f) Right eyebrow raised; (g) Left eyebrow raised

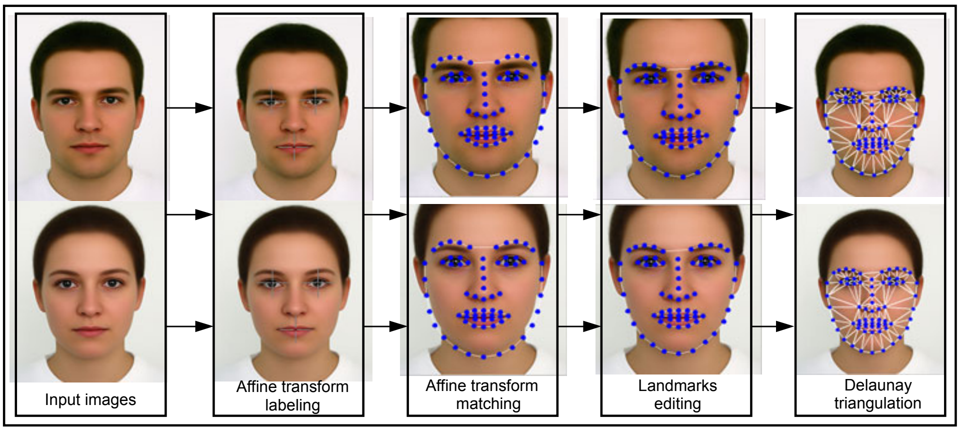

Fig. 3 Overview of semi-automatic labeling (sample images are from Little *et al*. (2012))

model. Second, the standard shape of the initial model is placed on the input image using affine transform technology. Third, we adjust the position of each point interactively until each point is placed on the corresponding position, which means that the process of shape labeling is optimized to the process of shape editing. Finally, the training sets labeled are used to complete the process of shape feature modeling, Delaunay triangulation, and texture feature modeling.

Based on the above optimization, the image labeling process and the model building process are fused, which is a process from coarse to accurate model construction. What is more, this semi-automatic method requires only a little user correction by applying the initial model learned previously, which can dramatically improve the efficiency of model training.

3.3 Statistical learning based identification

The traditional AdaBoost face detection method is dependent on the integral diagram, the cascade detector, and the Haar features, suitable only for the face detection close to the upright direction (Viola and Jones, 2001). While the face largely rotates, the detection performance will fall down sharply, and the accuracy and speed for expression feature point detection will be influenced dramatically. This study includes a rotation invariant face detection method to extend the above face detection method. Fig. 4 shows the detection results. According to the coherent characteristic of face movement in the video, we divide the rotation space, as shown in Fig. 5a, where each color represents a different detection priority. Then we compute the current rotation angle of the face based on previous detection results, which means that the current face is detected within the adjacent orientation of previous detection results. At the first detection stage, the upright orientation is given the top priority, its adjacent parts the secondary priority, and so on. According to the face orientation in the detected image, each part of the divided image has a different priority. In the next frame, the orientation with the top priority is used to detect the face. If the face in the top priority orientation is not detected, we detect the secondary priority orientation, and so forth. In general, this method can significantly improve the detection speed, which is roughly two times faster than AdaBoost face detection. Fig. 5b shows the running time of our face detector according to the different priorities, and the user can choose different detection priorities, which represent different detection ranges.

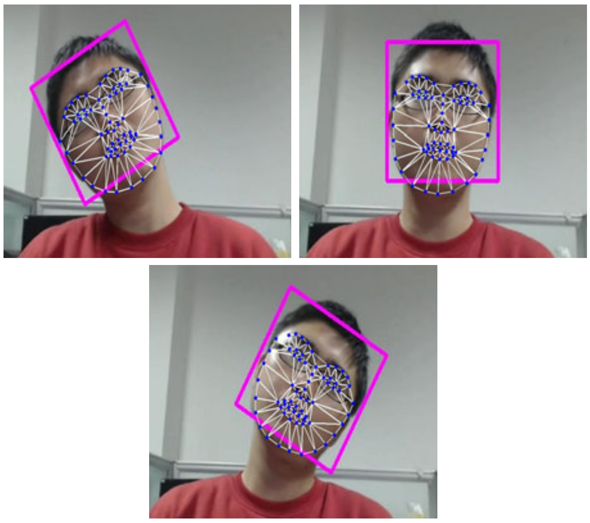

Fig. 4 The results obtained using a rotation invariant face feature detection method

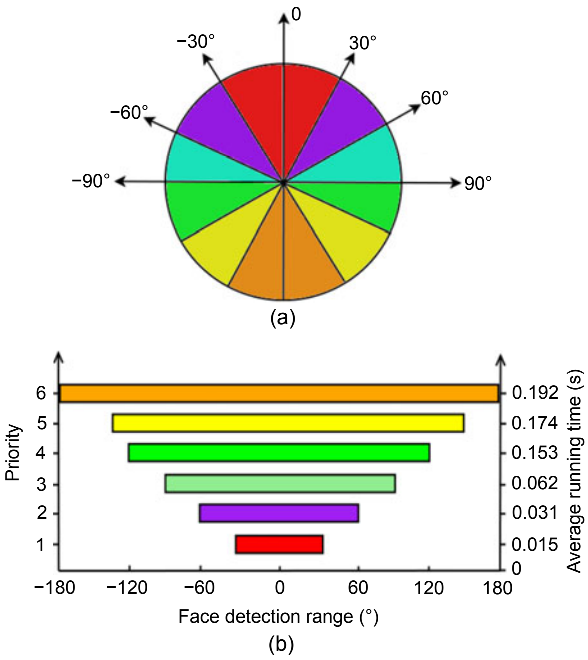

Fig. 5 Analysis and evaluation of our method
(a) Priority of the orientation; (b) Running time evaluation

Through PCA for the shape feature of the training set, we find that not all feature points are significantly changed for different expression images, and that the changes concentrate mainly in the eyes, eyebrows, and lips, indicating that the expression motion unit simplified in this study is reasonable. Considering the automatic computing and mapping of various basic action weights, we construct the basic expression database corresponding to the basic action of the head model, and train them respectively (Fig. 6). Therefore, in the process of facial animation simulation, different facial animations can be expressed as a linear combination of each basic action for a head model:

$$\boldsymbol{M}' = \boldsymbol{M}_0' + \sum_{i=1}^{5} b_i \boldsymbol{M}_i', \tag{1}$$

where $\boldsymbol{M}_i'$ is the basic action for the head model and b_i is the basic action weight. b_i can be solved through the least squares method, $\min\left\| \boldsymbol{S}' - \left(\boldsymbol{S}_0' + \sum_{i=1}^{5} b_i \boldsymbol{S}_i' \right) \right\|$, where $\boldsymbol{S}'$ is the AAM accuracy fitting result of the shape feature, and $\boldsymbol{S}_i'$ is the average shape feature for the trained basic expression corresponding to the basic action of a head model. The weight can be directly applied to Eq. (1) after being normalized to complete the simulation of facial animation.

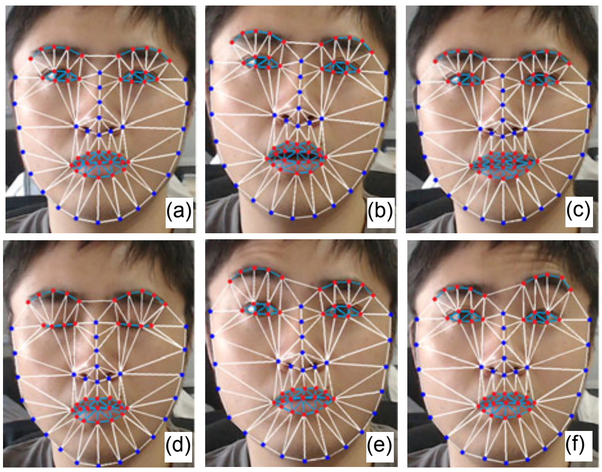

Fig. 6 Training of significant regional shape features
(a) Neutral shape; (b) Mouth opened; (c) Smile; (d) Eyes closed; (e) Right eyebrow raised; (f) Left eyebrow raised

4 Real-time realistic facial rendering

The traditional Lambert diffuse lighting model (Blinn, 1977) is an empirical model, used mainly to simulate the lighting phenomenon of a rough surface. Basically, this model hypothesizes that the surface is an ideal diffuse reflector, and the diffuse distribution for lighting depends only on the cosine value of the angle between incident light and the normal surface, making it insufficient to simulate the skin surface. To improve the plausibility and efficiency of facial animation rendering, we adopt the physically based diffuse model. Considering the sub-surface scattering of the skin surface, the diffuse distribution of lighting can be expressed as an integral expression according to the curvature value of a model surface and incident light:

$$D(\theta, r) = \frac{\int_{-\pi}^{\pi} \cos(\theta + x) \cdot R(2r\sin(x/2))\mathrm{d}x}{\int_{-\pi}^{\pi} R(2r\sin(x/2))\mathrm{d}x}, \tag{2}$$

where D is the diffuse distribution function of lighting, θ is the angle of incident light, r is the curvature value of the model surface, and R is the diffuse profile function. Because Eq. (2) is very complicated, we pre-integrate the diffuse distribution into a 2D texture, and complete skin rendering in different parts through texture sampling technology (Fig. 7a).

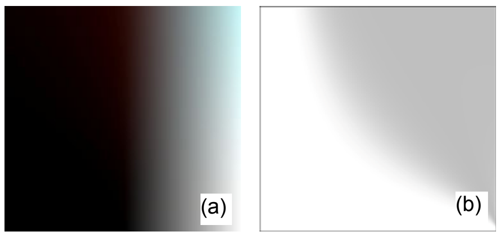

Fig. 7 Pre-integrated diffuse (a) and specular (b) distributions of lighting
In (a), the horizontal axis represents the change of $\cos\theta$ (θ is the angle of incident light), and the vertical axis represents the change of surface curvature; In (b), the horizontal axis represents the change of $\cos\alpha$ (α is the angle between the normal and half vector), and the vertical axis represents the change of surface roughness

The traditional Phong specular lighting model (Blinn, 1977) is an empirical model, used mainly to simulate the lighting phenomenon of a smooth surface. This model hypothesizes that the surface is an ideal specular reflector, and that the specular distribution for lighting depends only on the cosine value of the angle between the reflected light and the view. Therefore, it is not enough to simulate the skin surface. To improve the quality of facial animation rendering, we adopt the physically based bidirectional reflectance distribution function (BRDF) specular model. The specular distribution of lighting can be expressed as a Beckmann distribution:

$$S(\alpha)=\frac{\exp(-\tan^2\alpha/m^2)}{\pi m^2\cos^4\alpha},\quad \alpha=\arccos(\boldsymbol{N}\cdot\boldsymbol{H}),\quad (3)$$

where S is the specular distribution of lighting, $\boldsymbol{H}$ is the half vector between the incident light and the view, $\boldsymbol{N}$ is the normal vector, m is the roughness of the model surface, and α is the angle between the normal and the half vector. To simplify the computation, we pre-integrate the specular distribution into a 2D texture, and complete skin rendering in different parts through texture sampling technology (Fig. 7b).

We obtain a better rendering result of skin than the traditional method under the same conditions by combining the above two physically based distribution functions of lighting. As shown in Fig. 8a, the contrast effects are clearly visible in the regions around the ears, nose, and eyes, where the values of curvature change largely. The traditional lighting model leads to an asymmetric distribution of lighting, and does not have a translucent effect, often resulting in an insubstantial, blunt, and dry appearance. Our method can capture the increased specular at grazing angles and is physically plausible with respect to sub-surface scattering effects, which makes the skin surface more translucent and soft.

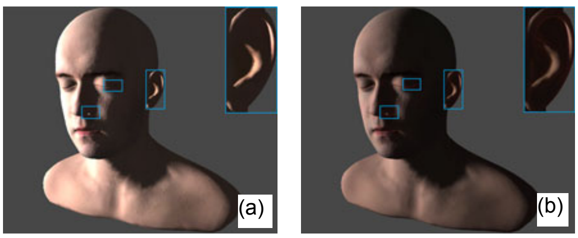

Fig. 8 Rendering results of skin using the traditional lighting model (a) and our proposed model (b)

5 Results and comparisons

To effectively synthesize facial animation for computer production and interactive applications, we present a real-time and realistic framework for facial animation simulation and rendering. All the simulations are performed on an AMD PC with an NVIDIA GeForce GTX460 graphics card, and implemented using OpenGL and GLSL at about 60 frames/s.

5.1 Facial animation for interactive applications

We have implemented the simulation and realistic rendering method for facial animation in graphics engine Unity3D. Fig. 9 shows some screen shots from the facial animation video. Changes in facial expressions can be captured through a web camera, and then used to drive the head model to automatically complete facial animation generation, which satisfies the requirements of both real-time simulation and realistic rendering. Experimental results indicate that our method is effective and efficient.

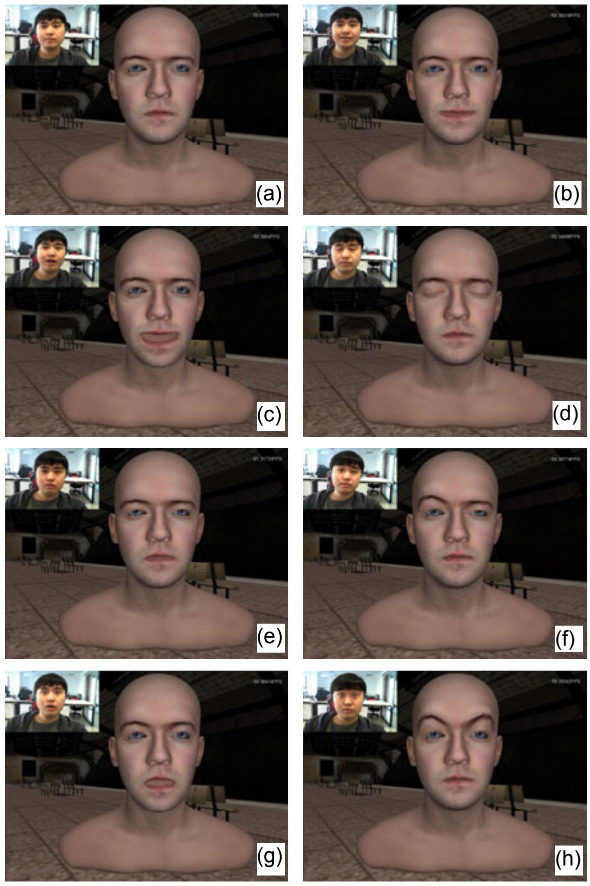

Fig. 9 Facial animation simulation and rendering

(a) Neutral expression; (b) Smile; (c) Laugh; (d) Blink; (e) Sadness; (f) Contempt; (g) Surprise; (h) Fear

5.2 Comparison with other methods

There have been many methods for facial animation. To obtain plausible expression movement, traditional facial animation methods usually require too many calculations, and are not universally applicable. There is, however, less research on realistic rendering of facial animation, and the use of texture synthesis and mapping technology leads to poor rendering effects (Fig. 10). Zhang and Chen (2003) divided the face model according to various functional areas, and used the corresponding main control point to directly drive the face model. However, this kind of method is generally at low level control, and not convenient to operate in an actual application. Zhang and Chen (2003) introduced a top level muscle model, which can produce desirable facial animation effects; however, to create a simple rendering, more calculations are required (Fig. 10a). Lee *et al*. (1995) adopted the muscle model method to drive the head model, and restored and synthesized the texture image collected, which improved the rendering effect (Fig. 10b). Pighin *et al*. (2005) considered the changes of the head model and texture image at the same time; the texture image was calculated using the interpolation method according to the different expressions of the head model. Because numerous texture images need to be generated, this method requires many calculations (Fig. 10c). Weise *et al*. (2011) used depth information to reconstruct the head model, and then executed the matching process between the reconstructed model and the animation model, resulting in plausible facial animation. However, this method depends on a specific hardware equipment (Kinect sensor), and cannot be generalized. Also, Weise *et al*. (2011) did not consider the realistic rendering of facial animation (Fig. 10d).

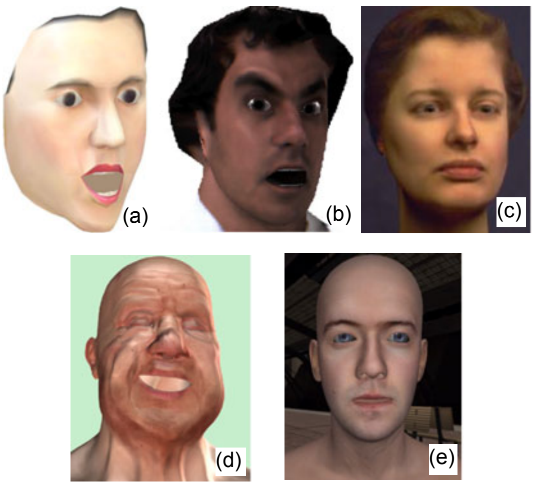

Fig. 10 Comparison with methods in the literature
(a) Zhang and Chen (2003); (b) Lee *et al*. (1995); (c) Pighin *et al*. (2005); (d) Weise *et al*. (2011); (e) Our method

As the above methods do not consider realistic rendering of facial animation, even if they have achieved real expression movement, the facial appearance will be insubstantial and blunt. Thus, these methods cannot be applied directly to computer animation and interactive applications. In this study, we combine the image- and geometry-based methods, calculate the sub-surface scattering characteristics for skin rendering based on the physical method, and generate slightly translucent and soft skin effects, making the facial animation more realistic and immersive (Fig. 10e).

6 Conclusions

We present a framework for real-time and realistic facial animation simulation and rendering based on statistical learning. By extending the basic AAM model, we develop a series of techniques, including the simplification of the expression motion unit, statistical training based semi-automatic labeling for expression images, rotation invariant face detection, a new algorithm for real-time facial animation, and a new algorithm for real-time realistic facial rendering. The head model is computed and mapped automatically according to the weights of the basic expressions, which dramatically improves the accuracy and efficiency of facial animation simulation. We consider the sub-surface scattering effect of skin, and pre-integrate the diffuse distribution and specular distribution of lighting based on the physical method, which greatly improves the plausibility and efficiency of facial animation rendering.

Our facial animation simulation and rendering framework is simple and efficient, and effective for computer production and interactive applications. There are still some limitations, however. The number of motion units used in this study is on a medium scale; more motion unit should be trained according to the application requirements. The method considers only face deformation, and ignores the rigid motion of the head and eyes, which is important for increasing the plausibility of facial animation. In the future, we will seek solutions to these limitations.

References

Blanz, V., Basso, C., Poggio, T., Vetter, T., 2003. Reanimating faces in images and video. *Comput. Graph. For.*, **22**(3): 641-650. [doi:10.1111/1467-8659.t01-1-00712]

Blinn, J.F., 1977. Models of light reflection for computer synthesized pictures. *ACM SIGGRAPH Comput. Graph.*, **11**(2):192-198. [doi:10.1145/965141.563893]

Buck, I., Finkelstein, A., Jacobs, C., Klein, A., Salesin, D.H., Seims, J., Szeliski, R., Toyama, K., 2006. Performance-Driven Hand-Drawn Animation. Proc. 1st Int. Symp. on Non-photorealistic Animation and Rendering, p.101-108. [doi:10.1145/340916.340929]

Chai, X.J., Shan, S.G., Gao, W., Chen, X.L., 2005. Example-based learning for automatic face alignment. *J. Softw.*, **16**(5):718-726 (in Chinese).

Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active shape models—their training and application. *Comput. Vis. Image Understand.*, **61**(1):38-59. [doi:10.1006/cviu.1995.1004]

Cootes, T.F., Edwards, G.J., Taylor, C.J., 2001. Active Appearance Models. *IEEE Trans. Pattern Anal. Mach. Intell.*, **23**(6):681-685. [doi:10.1109/34.927467]

Deng, Z.G., Chiang, P.Y., Fox, P., Neumann, U., 2006. Animating Blend Shape Faces by Cross-Mapping Motion Capture Data. Proc. Symp. on Interactive 3D Graphics and Games, p.43-48. [doi:10.1145/1111411.1111419]

d'Eon, E., Luebke, D., Enderton, E., 2007. Efficient Rendering of Human Skin. Proc. 18th Eurographics Conf. on Rendering Techniques, p.147-157. [doi:10.2312/EGWR/EGSR07/147-157]

Ekman, P., Friesen, W.V., Hager, J.C., 2002. Facial Action Coding System: the Manual. A Human Face. Research Nexus, Salt Lake City, p.1-197.

Guenter, B., Grimm, C., Wood, D., Malvar, H., Pighin, F., 2006. Making Faces. ACM SIGGRAPH, p.55-66. [doi:10.1145/1185657.1185858]

Jimenez, J., Sundstedt, V., Gutierrez, D., 2009. Screen-space perceptual rendering of human skin. *ACM Trans. Appl. Percept.*, **6**(4), Article 23. [doi:10.1145/1609967.1609970]

Lance, W., 1990. Performance-Driven Facial Animation. ACM SIGGRAPH, p.235-242. [doi:10.1145/97879.97906]

Lee, Y.C., Terzopoulos, D., Waters, K., 1995. Realistic Modeling for Facial Animation. Proc. 22nd Annual Conf. on Computer Graphics and Interactive Techniques, p.55-62. [doi:10.1145/218380.218407]

Little, A.C., Hancock, P.J.B., DeBruine, L.M., Jones, B.C., 2012. Adaptation to antifaces and the perception of correct famous identity in an average face. *Front. Psychol.*, **3**:19. [doi:10.3389/fpsyg.2012.00019]

Matthews, I., Baker, S., 2004. Active appearance models revisited. *Int. J. Comput. Vis.*, **60**(2):135-164. [doi:10.1023/B:VISI.0000029666.37597.d3]

Parke, F.I., 1972. Computer Generated Animation of Faces. Proc. ACM Annual Conf., p.451-457. [doi:10.1145/800193.569955]

Penner, E., Borshukov, G., 2011. Pre-integrated Skin Shading. *In*: Engel, W. (Ed.), GPU Pro 2: Advanced Rendering Techniques. A K Peters/CRC Press, p.41-54. [doi:10.1201/b11325-1]

Pighin, F., Hecker, J., Lischinski, D., Szeliski, R., Salesin, D.H., 2005. Synthesizing Realistic Facial Expressions from Photographs. ACM SIGGRAPH, p.75-84. [doi:10.1145/1198555.1198589]

Viola, P., Jones, M., 2001. Rapid Object Detection Using a Boosted Cascade of Simple Features. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, p.511-518. [doi:10.1109/cvpr.2001.990517]

Weise, T., Bouaziz, S., Li, H., Pauly, M., 2011. Realtime performance-based facial animation. *ACM Trans. Graph.*, **30**(4), Article 77, p.1-10. [doi:10.1145/2010324.1964972]

Yao, J.F., Chen, Q., 2008. Survey on computer facial expression animation technology. *Appl. Res. Comput.*, **25**(11): 3233-3237. [doi:10.3969/j.issn.1001-3695.2008.11.009]

Zhang, Q.S., Chen, G.L., 2003. Realistic 3D human facial animation. *J. Softw.*, **14**(3):643-650 (in Chinese).

Zhang, Q.S., Liu, Z.C., Guo, B.N., Demetri, T., Shum, H.Y., 2006. Geometry-driven photorealistic facial expression synthesis. *IEEE Trans. Visual. Comput. Graph.*, **12**(1):48-60. [doi:10.1109/TVCG.2006.9]