Video-Based Fast 3D Individual Facial Modeling

Mandun Zhang, Peng Lu,Xiangsheng Huang, Xiaoxu Zhou, Yangsheng Wang
Institute of Automation, Chinese Academy of Sciences, 100080, China
{mdzhang, plu, xshuang, xxzhou,wys}@mail.pattek.com.cn

Abstract

In this paper, we present a video-based fast 3D facial modeling algorithm. Different from traditional complex stereo vision procedure, our new method needs only one frontal image for fast 3D modeling without any camera calibration. The proposed method has three steps. Firstly we propose an improved Active Shape Models (ASM) method to extract frontal feature points. Then according to MPEG-4 protocol for 3D face structure, we appoint and deform feature points by Radial Basis functions (RBF) corresponding to the generic model. At last the texture mapping is carried out with regard to different directional projections. The experiments demonstrate that our new algorithm can fast photo- realistically render 3D face with very limited computation.

Key words: Facial Modeling, Active Shape Models, Radial Basis Functions

1. Introduction

From the viewpoint of both computer vision and computer graphics, fast 3D facial modeling and animation is one of most challenging and interesting research topics. And facial animation has been applied in human-computer interfaces, interactive games, multimedia titles, VR, and, as always, in a broad variety of animations production. Moreover it also becomes a hotspot especially in recent years because of online game burgeoning. One of the most noticeable trends is the boom toward “realistic” looking and feeling games. If the player can become a role and see a realistic copy of the self within a game system, the game will gain more attention from its player.

By now various 3D face modeling approaches, such as a Laser scanner [1,2,3], a stereoscopic camera [5], or an active light stripper [4], have been proposed. However this kind of facial 3D data only include shape information without texture information, and always easily contaminated by measuring noise. It’s difficult for them to be used in facial animation. The procedure of the facial modeling and animation can be summarized as follows: 1) a structural generic model selection, 2) a specific face information based on computer vision or based on image acquirement, 3) the generic model deformation, and 4) texture mapping.

It is computational expensive to apply these approaches based on computer vision. Therefore many research efforts have been made to generate realistic facial modeling from photos taken with an ordinary video camera. Reconstructing the true 3D shape and texture of a face from a single image is an ill-posed problem [13]. It’s a generic method to reconstruct an individual facial modeling from two orthogonal photos [6,7,8,9,10,11], but they are too complex and inconvenient for serious applications. When generating model from two arbitrary pictures [15], if skins are so smooth and free of blemishes, the corner point matching may fail. Creating realistic textured 3D facial models from multiple photographs of a human subject [12,14] is theoretically possible, but computationally too intensive.

In this paper, we advance a practical approach for reconstructing an individual facial modeling from a frontal view of the person. Firstly the 2D positions (x- and y-coordinates) of key feature points from frontal image are detected by an improved ASM [17] method. Then they are devoted to deform a generic model with corresponding feature points using RBF. The z-coordinates for the individual face points are derived from the general model. Then texture mapping is performed for realistic modeling. The rest of this paper is organized as follows. Section 2 describes our technique for extracting facial feature points from video. Section 3 describes how to choose a generic model. Section 4 describes deformation algorithm and texture mapping. The results are shown in Section 5 and are concluded in Section 6.

2. Extraction of frontal facial feature points

We propose a novel face alignment algorithm, in which local appearance models of key points are modeled using statistical learning. RealBoost can effectively select most significant Gabor features to build the likelihood model that ensures the ground truth position of each key point will more likely have a higher likelihood than its neighbors. Instead of using principle components analysis (PCA) and one pixel Gabor coefficients, we use Gabor features of key point and its neighbors, which provide rich information to model local structures of a face.

2.1. Statistical shape models

The ASM technique [17] relies upon object (e.g. face) structure being represented by a set of points. The points can represent the boundary, internal features, such as the eyes and mouth of a face. Given a set of training images for a given object, points are manually placed in the same location on the object in each image. By examining the statistics of the positions of the labeled points, a "point Distribution Model" is derived. The model gives the average positions of the points, and has a number of parameters that control the main variations found in the training set. The points from each image are represented as vector x and aligned to a common co-ordinate frame. A principle component analysis is applied to the aligned shape vector. The model can be used to generate new shape using equation

$$x = \bar{x} + Pb \tag{1}$$

Where $x = (x_0, y_0 \ldots x_{n-1}, y_{n-1})^T$, (x_k, y_k) is the k^{th} model point, $\bar{x}$ is the mean shape vector, P is the set of orthogonal models of shape variation and b is a set of shape parameters.

If the shape parameter b is chosen inside suitable limits (derived from training set), we can ensure that generated shapes are similar to those in the original training set.

2.2. Key points localization based on realBoosting learning in Gabor space

The ambiguity between the truth position and its neighbors requires that the local likelihood model should be able to correctly rank the likelihood of these ambiguous positions. We adopt RealBoost to learn the ranking prior likelihood models that not only characterize the local features of a ground truth position, but also preserve the likelihood ranking order between the ground truth position and its neighbors. In RealBoost learning procedure, Gabor features [18,19] of each key point and its neighbors are extracted to construct "weak" ranking evaluation function set.

RealBoost is adopted to solving the following three fundamental problems in one boosting procedure: (1) learning effective features from a large feature set, (2) constructing weak classifiers each of which is based on one of the selected features, and (3) boosting the weak classifiers into a stronger classifier.

2.3. Experiment results

Fig. 1 illustrates results of our feature extraction system. We manually labeled 700 images, each of size 200 × 200.On each image 79 key points are labeled. We select 450 images as the training set and the others as the test images.

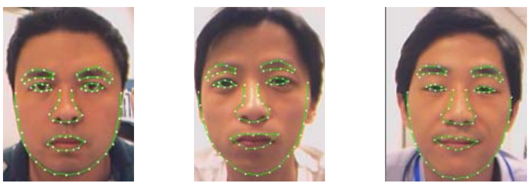
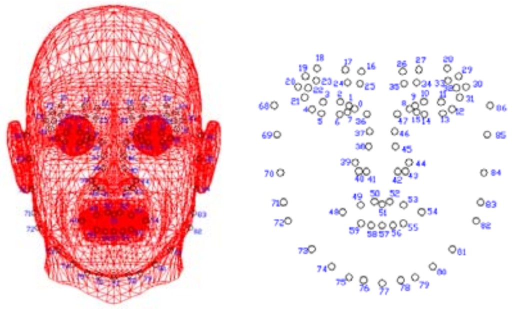


Fig. 1 Result of detecting facial feature points

3. Generic model choice

There is the same basic structure such as eyes, nose and mouth etc for different people. It's easy for us to identify a generic model with them. Everyone has different features that make one unlike others. A generic model should be a structural one for facial animation. The wireframe model is gettable from much software such as 3DMAX, Poser or Maya. It's usually chosen as a facial model.



(a) Feature points identifying (b) Feature points

Fig. 2 A generic model

We choose a general facial model derived from Poser 5.0 that is made of 8172 3D vertexes and 15980 triangles. It could reflect tiny facial features.

According to MPEG-4, there are 84 feature points defined on a neutral face that provides referenced space for defining facial animation parameter (FAP). They suffice for identifying proper shape of facial model. Feature points are divided into several groups such as lip, eyes and mouth etc. There are 79 feature points selected according to MPEG-4 and facial shape. Figure 2 shows their positions which correspond to .the frontal face detection result.

4. Deformation and texture mapping

4.1. Scattered data interpolation

The facial model adjustment is a problem about 3D space deformation in fact (In this paper, we only have 2D positions of control points and the z-coordinates for the individual face are direct from the general model). We identify limited control points on it and compute their displacements, then choose an interpolation function which accommodates displacements of control points. Locations of the other nodes are transformed upon the function. At last the final deformation result of the model is reached.

There are several choices for how to construct the interpolating function. We use a method based on radial basis functions (RBF) as approximation functions for their power to deal with irregular sets of data in multi-dimensional space in approximating high dimensional smooth surface [14,16]. We choose 2D coordinates of origin mesh nodes as embedding space to construct the interpolation function f (p) that suffices displacements of control points:

$$u_i = f(p_i) \quad (0 \le i \le N-1) \qquad (2)$$

Where u_i is displacements of all control points and N is the number of feature points. We exploit RBF volume morphing to directly drive 2D geometry deformation of face models.

$$f(p_i) = \Sigma c_i \phi(\| p_j - p_i \|) \quad (0 \le j、i \le N-1) \qquad (3)$$

where $\phi(\| p_j - p_i \|)$ are RBF and the coefficients c_i are the vector coefficients of control points.

We compute equation (3) and get the vector coefficient of every control point. Displacements of other non-feature points may be computed in the form

$$u = \Sigma c_i \phi(\| p - p_i \|) \ (0 \le i \le N-1) \qquad (4)$$

In this paper, we have chosen to use $\phi(\| p - p_i \|) = e^{-\| p - pi \| / 64}$.

4.2. Texture generation

Texture mapping makes the synthetic face more photorealistic. The frontal face shape can be automatically extracted as described in the previous section. The parts out of the contour lines are filled with nearly similar color. Then, it's merged into a size 256× 256 image with an ear (Ears texture coordinates can be derived from it). With the average template operation, the face contour becomes enough blurring. The result of blended texture is shown in Figure 3.

Fig. 3 The texture map

4.3. Texture fitting

The size of frontal face image is w1×h1. In our implementation w1= 86.x – 68.x where 86.x represent the x coordinate of point 86 (Figure 2 (b)), and h1 =YSCALE×(18.y – 71.y) where YSCALE is a scale factor and its value is 1.1 here. The total size of the image is w × h where their values are all 256. The maximum x value (xmax) is 86.x and the minimum x value (xmin) is 68.x. The maximum y value ymax = YSCALE × 18.y and the minimum y value (ymin) is 77.y (In this paper, we take point 77 as the origin point). The maximum and the minimum z coordinates are represented with zmax and zmin respectively. The 3D coordinates of every point are x, y and z. We mark texture coordinates of every point with xtex and ytex. The left and down corner of frontal face image is (xt, yt) where xt and yt are 50 in this paper. According to linear interpolation, if one point is projected into the front, the left or the right view, the texture coordinates of every point are:

$$\begin{cases} \begin{cases} xtex = xt/\mathrm{w} + (x - xmin)/(xmax - xmin) \times \mathrm{w1/w} & (5) \\ xtex = \mathrm{xt/w} \times (z - zmin)/(zmax - zmin) & (6) \\ xtex = (\mathrm{w1+xt})/\mathrm{w} + (z - zmax)/(zmin - zmax) \times \mathrm{w1/w} & (7) \end{cases} \\ ytex = yt/h + (y - ymin)/(ymax - ymin) & (8) \end{cases}$$

For example, if it's projected into the frontal view, its texture coordinates are derived from (5) and (8).

As for forehead and neck texture, we assign some texture near the frontal face image. They can be derived from the similar method as above.

5. Experiment results

The implementation of the above described individual face generation system is written in VC++ and OpenGL. A detailed individualized head is shown in Fig. 4 where input image is shown in Figure 1. It has proper shape and texture in several views.

Fig. 4 Snapshots of reconstructed heads

Other examples from video are shown in Figure 5. The texture image is stored in BMP format and has sized about 200K depending on the quality of pictures.

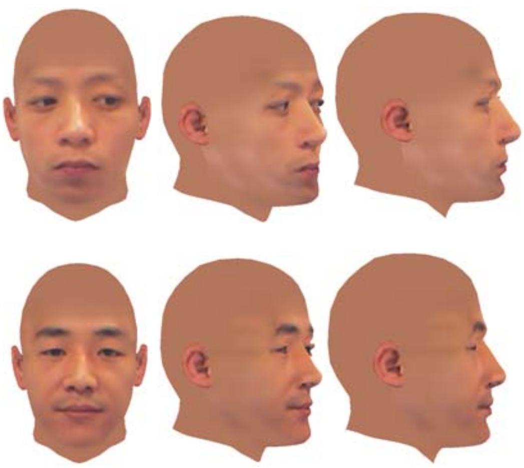


Fig. 5. Examples of reconstructed heads from video

6. Conclusion and future work

In this paper, we introduce a video-based approach to fast reconstruct face modeling. The efficient and robust face modeling method shows the process of modifying a generic head for shape modifying and producing texture image by extracting frontal facial image with an improved ASM method. It's a simple, quick and economic method on which facial animation can be based.

The emphasis in this paper has been on the quality of the results. Possible ways to improve it include:

- choosing more generic models to reduce deformation error
- developing a method for estimating the depth value of feature points
- improving the merging image to make it better
- improving ear modeling and adding hair modeling

7. Acknowledgment

We are grateful to people taken photos. This project is supported by Chinese National Research Foundation (No.2003AA114020).

References

1. Y.C. Lee, D. Terzopoulos, and K. Waters, "Realistic Face Modeling for Animation", Siggraph proceedings, 1995, pp.55-62.
2. Y.C. Lee, D. Terzopoulos, and K. Waters, "Constructing Physics-based Facial Models of Individuals", In Proceedings of Graphics Interface, 1993, pp. 1-8.
3. Yu Zhang, Edmond C. Prakashb, and Eric Sunga, "Constructing a Realistic Face Model of an Individual for Expression Animation", Animation International Journal of Information Technology Vol. 8, No. 2, September 2002, pp.10-25.
4. C. Beumier, and M. Acheroy, "3D Facial Surface Acquisition by Structured Light", In International Workshop on Synthetic-Natural Hybrid Coding and Three Dimensional Imaging, Santorini, Greece, 15-17 Sep, 1999, pp.103-106.
5. O.Faugeras, L.Quan, and P.Sturn, "Self-calibration of a 1D Projective Camera and its Application to the Self-calibration of a 2D Projective Camera", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.22, No.10, 2000, pp. 1179 – 1185.
6. Wen Gao, Jie Yan, Shigong Shan, and Xilin Chen, "Individual 3D Face Synthesis Based on Orthogonal Photos and Speech-Driven Facial Animation", 7th IEEE International Conference on Image Processing (ICIP 2000), Vancouver, British Columbia, Canada, 10-13 September, 2000.
7. Jie Yan, and Hongjiang Zhang. "Realistic Virtual Face and Body Synthesis", MVA2000 (International Workshop on Machine Vision Applications), Tokyo, Japan, November 28-30, 2000.
8. W. Lee, Kalra P., and N. Magenat Thalmann , "Model Based Face Reconstruction for Animation", Proc. Multimedia Modeling (MMM) '97, Singapore, 1997, pp. 323-338.
9. W. Lee, and N. Magnenat-Thalmann, "Fast Head Modeling for Animation", Journal Image and Vision Computing, Volume 18, Number 4, Elsevier, March, 2000, pp.355-364.
10. W. Lee, J. Gu, and N. Magnenat-Thalmann, "Generating Animatable 3D Virtual Humans from Photographs", Eurographics, Paris, 2000.
11. Horace H.S. Ip, and Lijin Yin, "Constructing a 3D Individual Head Model from two Orthogonal Views", the Visual Computer, Springer-Verlag, 1996, pp.254-266.
12. Chia-Ming Cheng, and Shang-Hong Lai. "An Integrated Approach to 3D Face Model Reconstruction from Video", *Proceedings of the IEEE ICCV Workshop on Recognition, Analysis, Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'01)*, July 2001.
13. V. Blanz, and T. Vetter, "A Morphable Model for the Synthesis of 3D Faces", In Proceedings of ACM SIGGRAPH 99, AddisonWesley, New York, A. Rockwood, Ed., Annual Conference Series, pp.187-194.
14. F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. H. Salesin, "Synthesizing Realistic Facial Expressions from Photographs", Siggraph proceedings, 1998, pp. 75-84.
15. Zicheng Liu, Zhengyou Zhang, Chuck Jacobs, and Michael Cohen. "Rapid Modeling of Animated Faces from Video", Technical Report, MSR-TR_2000-11.
16. D. Ruprecht, R. Nagel, and H. Müller: "Spatial Free-Form Deformation with Scattered Data Interpolation Methods", Computers & Graphics 19(1), 1995, pp.63-71.
17. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. "Active shape models: Their training and application".CVGIP: Image Understanding, 61:38–59,1995.
18. J. Friedman, T. Hastie, and R. Tibshirani."Additive logistic regression: a statistical view of boosting". The Annals of Statistics, 28(2):337–374, April 2000.
19. L. Wiskott, J. Fellous, N. Kruger, and C. V. malsburg. "Face recognition by elastic bunch graph matching". IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7):775–779, 1997.