MVA2000 IAPR Workshop on Machine Vision Applications, Nov. 28-30, 2000, The University of Tokyo, Japan

3—16 Realistic Virtual Face and Body Synthesis

Jie Yan   Hongjiang Zhang
Microsoft Research China

Abstract 3-D face and body model has been successfully explored in the computer graphics, computer vision and model-based image coding communities. This paper developed a virtual human face and body generation technique that synthesize virtual human face and body of arbitrary views. Realistic human Faces synthesis plays an important role in animating virtual human behaviors. Therefore, this paper will focus on the realistic virtual face synthesis methods. The virtual human body synthesis method will follow the same mechanism as face dose. As we know, faces are characterized by their shapes and their textures. By using a frontal and profile images of a specific subject, a deformation technique allows interactive alignment of features in the 3-D general graphic face model with the features of the pre-provided images of the specific subject. The deformation result is a 3-D face model of the specific human face. It precisely reflects the correspondence geometric features of the specific subject. After that, subdivision spline surface construction and multi-direction texture mapping techniques are presented which endow the photometric information to the specific 3-D geometric face model. Therefore, we could get the synthesized virtual faces with geometric and texture features of the specific subject.

1. Introduction

3-D human face and body synthesis is very useful today, this technique can be used in many areas such as virtual environment, television conference, game and special effects film and so on. Much progress has been made in the field of computer animation, but the task of animating the human face has always remained a major challenge. One of the chief difficulties of implementing realistic human face animation is how to reflect the differences between different specific human faces. Faces are characterized by their shapes and textures. Therefore, in order to enhance the realism of the synthesized virtual face, automatically facial model deformable technique and multi-direction texture mapping technique are proposed to endow both the geometric features and the texture features to the synthesized virtual images.

Different faces share with common facial features, for example, the face contour is resemble to an ellipse, the shape and the spatial relationship of facial organs such as brow, eye, nose and mouth are steady. However, one can not deny the existence of the individuality differences which may sometimes appear to be very distinct between two faces, even the face of "identical twins" differ in some aspects[1]. As far as a given human face is concerned, the size of eye, height of nose, shape of lip all portray the facial features of the face. We define that the general face is a typical human face without any distinctive features. The differences between the given face and the general face is described as the individuality features of the given face[2]. To personalize the general face model, our approach is extract distinct features points from images. Then apply these deformable parameters to the general face model. However, considering the precision of the automatic feature extraction, our deformation technique also employs interactive mechanism. Based on the multi-direction images pre-provided by the animator, we proposed a general method for manipulating polygonal mesh of human face model, which deform the current face model to a new face model by specifying the new 3-d positions of a set of specific vertexes(feature points). When move a vertex to match the specific feature point in the image, the deformation algorithm automatically calculate the new position of the relative vertexes around the moved vertex. Thus simulate the plastic-visco-elastic behavior of the facial skin. After the shape deformation, the face model was provided with the shape characters of the specific face. In order to endow the individual texture characters to the face model, multi-direction texture mapping technique is developed to further enhance the realism of the synthesized human face, an algorithm is designed to select texture information from the frontal view and profile view images of the specific subject's face.

2. Generation of Individual Geometric Face Model

Face model is an important part in human face synthesis techniques. The similarity and reality of synthesized image will depend on the model. We build a general 3-D mesh model to reflect facial structure, some important area, such as mouth, eye and brow are modeled independently. The whole general 3-D mesh model we use is consists of 1229 vertices and 2056 triangles. In order to reflect the smoothness of the real human face, we use polynomial patch to represent the mesh model. More detail is discussed in papers[3][4].

Comparing with the general human face model, any specific human face image is different from the general human face. We assume that we only have one frontal and one profile face images of the specific subjects.

Now it is necessary to adjust the general model to match the specific human face in accordance with the two input human face images. This process is parameter fitting from the general human face to the specific human face.The general face model is transformed to a specific face model using a deformation technique. Here global transformation and local transformation are introduced in order to translating the general neutral face model to the specific neutral face model[5]. Firstly, we accomplish global transformation to adapt the whole facial contour and facial features contours such as brow, eye, mouth and nose, so that the corresponding contours of the general neutral face will coincide with that of the given face. Global transformation can only finish the modification of outline of whole face and organ contour. To attain the purpose of adjusting the general human model automatically when one or several vertices are moved, a local deformable face model needs to be adopted.

Surface interpolation is a simple method to implement face model deformation[6]. The position of the vertex of the mesh is multiple by a transform matrix which is specified by the original and new positions of the feature points. The main drawback of the surface interpolation method is a single feature point's movement will affect the whole model's deformation thus make it impossible to achieve local adjustment. In order to edit the human face model more flexible and controllable, we develop a new deformation algorithm to drag the vertex to the position where it should be. As a result of modify the face model, we describe the main idea of this algorithm using a single moved vertex, it can be easily extended to the condition of numerous moved vertexes.

Assume the current single vertex v has a space position $v=[x,y,z]$, it is moved to a new position $v'=[x',y',z']$, when this movement occur, the total vertexes of the face mesh are separated into two sets called inner vertexes set and outer vertexes set. The inner vertexes set is defined as a set consist of all the vertexes influenced by this movement. Accordingly, the outer vertexes set contains those vertexes not influenced by this movement. The inner set can be specified by the facial muscle's distribution and alignment. However, it is a complex task for involving anatomic knowledge which destine to become a barrier to real time interactive process. To solve this problem, a replaceable simple method is presented which considers the vertexes within a certain span to the moved vertex v are belong to the inner vertexes set, the span n can be adjusted arbitrarily during the edit process. **Fig.1** shows the inner vertexes of v, the span is 2.

Human facial skin is an elastic surface, to simulate the facial deformation, springs between neighboring vertexes is a reasonable hypothesis. If p_0 is a vertex in inner vertexes set, there exist m vertexes $p_1, p_2, \ldots\ldots p_m$ around p_0, see **Fig.1**. Assume the original meshes is a balance object, the movement of v disturbed the balance, the vertexes in the inner vertexes set are driven to the new place to obtain another balance. Accordingly, we can write down an equation about p_0 to describe the new balance.

$$\sum_{i=1}^{m} \frac{p_i'-p_0'}{\|p_i'-p_0'\|}\left(\|p_i'-p_0'\|-\|p_i-p_0\|\right)k=0$$

Note that if $p_i=v$ then $p_i'=v'$ is known, otherwise it is an unknown variable. all of the vertexes in the inner vertexes set has such an equation. Solving this non-linear equations group we can obtain the new three-space positions of the inner vertexes. Unfortunately the equation is not robust when the grade is too high.

A relative simple method to perform the deformation is assume the face skin has elasticity property, we may approximately think that the sum of distance between p_0 and its ambient vertexes tend to be minimize.

$$\partial \sum_{i=1}^{m}\left(\|p_i'-p_0'\|\right)^2 \Big/ \partial p_0'=0$$

It is clear that the equation group is a linear equation and can be easily solved now. We adopt both the two deformation algorithms to the edit process, it can be randomly selected by the animator.

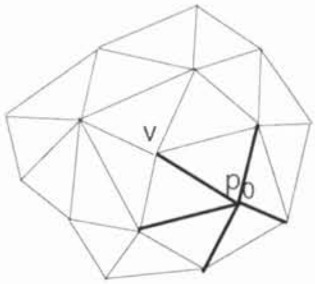


Fig.1 The inner vertexes of the moved vertex v

Using this deformation scheme, we may even generate dramatic effects of the specific subject. See **Fig.2**.

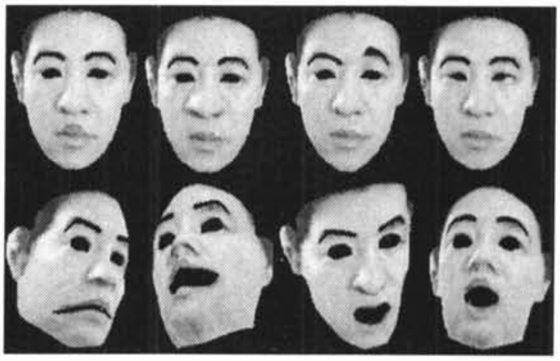
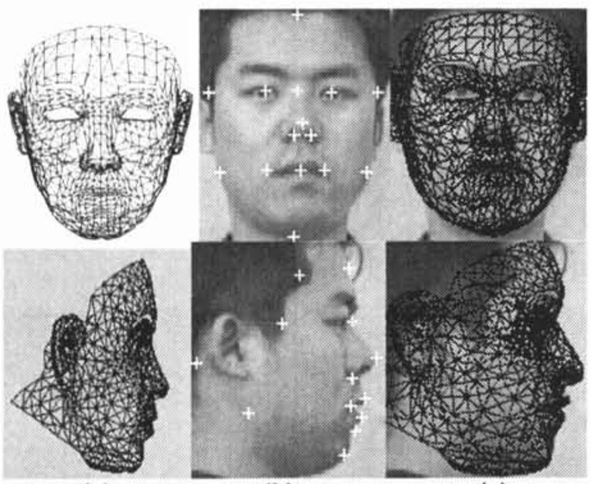


Fig.2 Some dramatic effects of the specific subject

We adopt an auto-fitting method. The feature points set can be extract from the specific subject's frontal view and profile view images, Then we can accomplish modification from the general human face to the specific human face in virtue of comparison and mapping between two groups of feature points set.

After the global transformation, we classify the local deformation into two levels, one is the coarse level deformation and the other is the fine level deformation, both the two deformations follow the same deformation mechanism above. In the coarse level deformation, a set of vertices are allowed to be moved consequently, the set are defined by the animator without restriction. Generally, it may consist of an organ or a part of an organ, then the movement is propagated in the relative area. The coarse level is often used in the organ location and contour coincidence.

In the fine level deformation, a single vertex of the mesh is moved to a new position, it allows the animator adjust the facial meshes vertex by vertex, the 3-D coordinate of ambient vertices of the moved vertex are calculated by the deformation algorithm. Using the above deformation method, at first, the 3-d face meshes are scaled according to the image's size, then adjust the facial contour as well as the center positions of the organs using coarse level deformation. The fine level deformation is used to perform local adjustment. The two level's deformation is used until the model exactly matches the face images of the specific subject. **Fig. 3** shows the matching processes of the model and the frontal and profile image of a specific human. After the above global transformation and local transformation, a 3-d meshes of human face with individuality features of the given face is generated. Now that the individual human face model is obtained, we may further adopt multi-direction texture mapping technique (describe in the next section) to obtain realistic synthesized results.



(a) (b) (c)

Fig. 3 Modification from the general face model to the given person's face model (a) general face model (b) feature points extracted (c) given Person's Face

3. Multi-Direction Texture Mapping

Now that the individual human face model is obtained, we may further adopt spline surface construction technique and multi-direction texture mapping technique to obtain realistic synthesized virtual face.

At first, We consider construction of a smooth interpolation space surface of human faces. A radial basis function interpolation surface over space mesh is generated by the subdivision method. The technique provides advantages of generating generally nice shapes. The generating of subdivision spline surface can be considered as a polishing procedure similar to mesh refinement. After the mesh refinement, the subdivided facial meshes are represented with Bézier patches.

To enhance the realism, texture mapping of reflective data acquired from images of faces, provides a valuable technique for further enhancing the realism of the synthetic human face.

At present, the limitation of the traditional mapping technique is that the texture will distort greatly in some parts if only adopt texture information from a single image. For example, this texture mapping technique will cause greatly distortion of ear when using the frontal view image's information of the specific subject. To conquer this drawback, we employ a new multi-direction texture mapping technique, here at least two specific subject's images(including one frontal view and one profile views) are used in our texture mapping procedure. In order to decide which image should be chosen for texture mapping to current Bézier patch, we classify the Bézier patches according to the normal of each patch and adopt the texture information from the image of the corresponding view, and then, a multi-direction texture mapping technique is used to endow the texture information to the specific face model.

This procedure is that: For each Bézier surface patch of face surface, corresponding texture image is determined by mapping the boundary curve of the Bézier patch to face image. The face image which be chosen depends on the whole direction of the Bézier patch. When the angle of direction vector and Y-Z plane is less than 30 degree, frontal face image is used to be map on. Otherwise profile images are used. Let

$$p(u,v)=\sum_{i=0}^{2}\sum_{j=0}^{2} b_{ij}B_{i,2}(u)B_{j,2}(v)$$

be the bi-quadratic Bézier patch. The tangent plane can be represented as the span of a pair vectors.

$$\text{Along u: } r_u=\frac{\partial p(u,v)}{\partial u}=\sum_{i=0}^{2}\sum_{j=0}^{2} b_{ij}B_{j,2}(v)\frac{\partial B_{i,2}(u)}{\partial u}.$$

$$\text{Along v: } r_v=\frac{\partial p(u,v)}{\partial v}=\sum_{i=0}^{2}\sum_{j=0}^{2} b_{ij}B_{i,2}(u)\frac{\partial B_{j,2}(v)}{\partial u}.$$

The direction of Bézier patch can be estimated by average value of each point in the patch. That can be computed by the formula:

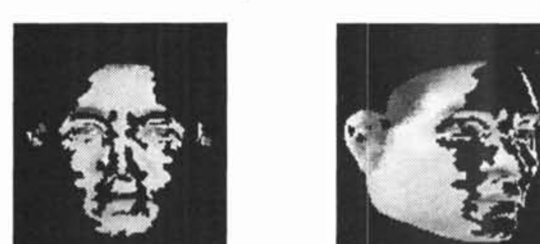
$$N=\int_0^1\int_0^1\frac{r_u\times r_v}{|r_u\times r_v|}dudv.$$

In the real computing procedure, we use the following formula to obtain the estimate value of each pitch. The complexity of the algorithm is $O(m^2)$.

$$f_N(u,v)=\frac{r_u\times r_v}{|r_u\times r_v|}$$

$$N\approx\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{1}{m^2}f_N\left(\frac{2i-1}{2m},\frac{2j-1}{2m}\right)$$

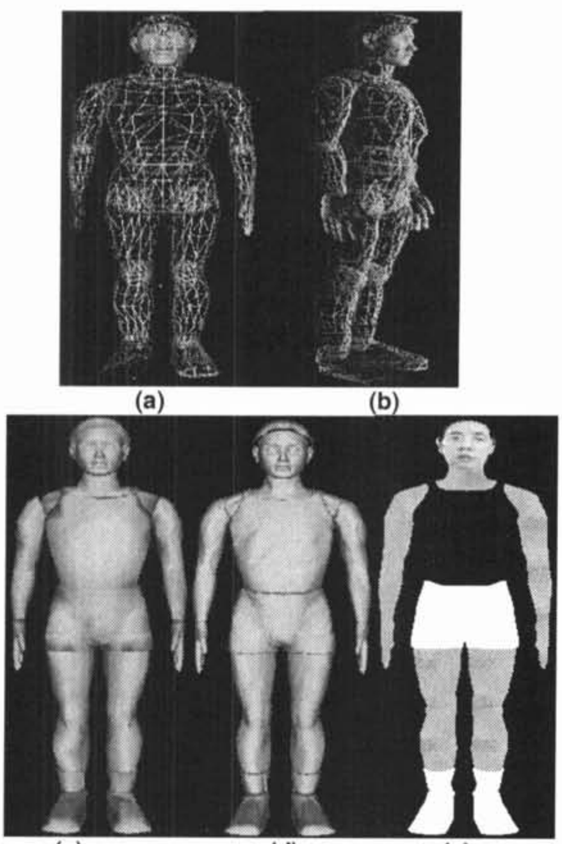
According to the direction of each patch, we select texture information from frontal and profile view images of the individual human face, see **Fig. 4**.



(a) (b)

Fig. 4 Texture mapping result according to each patch's direction (a) patches selected from frontal view image (b) patches selected from profile view images.

Fig. 5 Given person's synthesized virtual faces



(a) (b)

(c) (d) (e)

Fig. 6 3-D human body synthesized results (a)Body mesh(frontal) (b)Body mesh(profile) (c)Human body after phong processing (d)Human body after subdivision spline surface polishing (e) Human body after multi-direction texture mapping

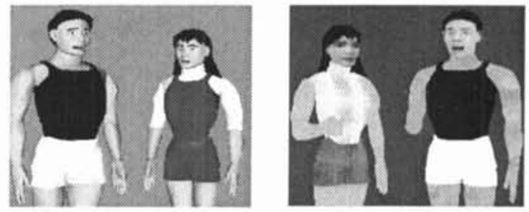


Fig. 7 The Synthesized virtual human before and after multi-direction texture mapping

4. Synthesized Results of Virtual Face and body

Using deformable human face model and multi-direction texture mapping techniques presented above, only by providing the frontal and profile images of a given person, we may expect to endow the individual geometric and texture features of the specific subject to the graphic model. Then we can synthesize the realistic individual virtual faces at arbitrary views. The center image of **Fig. 5** is a real frontal image of a specific subject. The ambient images around the real image are the synthesized virtual images at various views. The surface subdivision spline surface representation and multi-direction texture mapping techniques also can be used in generating realistic virtual human body. **Fig. 6** shows an experimental result. **Fig. 7** is the animation result of two talking virtual speakers in virtual environment.

5. Conclusion

In this paper, we present a feasible approach to generate virtual human face and body. In order to generate realistic virtual views of one subject, we use deformable human face model and multi-direction texture mapping techniques. Only by providing the frontal and profile images of a given person, we may expect to endow the individual geometric and texture features of the specific subject to the graphic model, To do this, we introduced techniques such as 3-D deformable human face and body model, spline surface construction as well as multi-direction texture mapping to synthesize realistic virtual human face and body at various views. The experimental results show that our approach is effective.

References

[1] Ashok Samal and Pkasana A. Iyengar, Automation Recognition And Analysis of Human Face And Facial Expressions: A Survey, Pattern Recognition, 1992, Vol. 25, No.1, pp 65-77.

[2] Kazuhito Murakami, Facial Caricaturing Based on Visual Illusion A Mechanism to Evaluate Caricature in PICASSO System, .IEICE TRANS.1NF.&SYST.April 1993, Vol. E76-D, No.4: pp 470-477.

[3] K. Aizawa, H. Harashima, and T. Saito. Model-based analysis synthesis image coding (MBASIC) system for a person's face. Signal Processing: Image Communication, 1:139-152,1989.

[4] D. Beymer and T. Poggio, Face Recognition from One Example View. A.I. Memo No. 1536, Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 1995.

[5] M. Patel, P. J.Willis. " FACES:Facial Animation, Construction and Editing System " .Proceedings Eurographics Vienna,Austria .September 1991,pp 33-45.

[6] B.C Yin and W. Gao, Radial Basis Function Interpolation on Space Mesh, ACM SIGGRAPH97, Virtual Proc. pp150.