A survey of real face modeling methods

Xiaoyue Liu, Yugang Dai, Xiangzhen He*, Fucheng Wan
Northwest Minzu University, National languages Information Technology research institute, Lanzhou city, Gansu province, China

Abstract. The face model has always been a research challenge in computer graphics, which involves the coordination of multiple organs in faces. This article explained two kinds of face modeling method which is based on the data driven and based on parameter control, analyzed its content and background, summarized their advantages and disadvantages, and concluded muscle model which is based on the anatomy of the principle has higher veracity and easy to drive.

1 Introduction
With the continuous development of computer technology, multimedia technology in particular, how can let the computer can express themselves more human and have natural communication with people caught more and more attention from the researchers. Realistic 3 d face animation technology as an important way of human-computer interaction, can be applied to the humanized human-computer interaction, virtual host, web conferencing, video phone, remote education, remote medical diagnosis, movies, games, entertainment, and many other fields. The most obvious application of face animation is film production. In "terminator II", "true lies", "avatar" and other films production are all show the charm of human face modeling and animation technology. Animators are constantly looking for a more promising animation system, hoping to use the latest academic research to modify and extend the current animation system.

At present, there are many ways of human faces modeling and can be divided into two types, one is based on data-driven method and the other is based on parameter control. The method of parametric control is divided into two models based on physical and geometric models[1], which is the focus of the research on face animation. Physical animation includes pseudo-muscle animation and muscle animation. The pseudo muscle animation refers to a grid that has no muscles in the model, only the face skin grid, which only calculates the displacement of the epidermis. Muscle animation refers to add some grid to the model (or other variable shape) to represent the muscles, animation as the adjacent grid by calculating to get the displacement of the face skin to make all kinds of action. The geometric animation includes parameterized animation, interpolation animation, and spline distortion animation. The common feature of these three animations is that when you are animated, you don't have to think about the geometry of the model, just using some kind of constraint method to transform the model.

2 Based on data-driven approach
Method based on data driven largely through two steps, the first step is to model training, through the establishment of large-scale database of speech samples, sample mass is obtained by the method of statistical data of regular function, by computer simulation, and the second step is data synthesis, through the model has been, to test corpus is to synthesize a number of discrete points to carry on the synthesis, article eventually synthetic face model. In order to get a satisfactory facial model,
database model of quality and quantity determines the performance of the pros and cons, of course, the method of statistical also occupies a very important content, at present, with the method of three kinds, the first is the Bregler, put forward the method based on image sequence splicing; The second is based on key frame-shifting methods, such as Ezzat, by the Massachusetts institute of technology (MIT). The third is a method of splicing of faces by people such as Cosatto. Besides, the fundamental idea is through the database to select the most suitable for synthesis of face of the various components, and then through the computer technology will these artifacts splicing, synthesis, achieve a vivid facial model. In addition, Gawain and others have proposed a facial image synthesis method based on component subgraph, concrete is the human face when divided into 8 classes components or eight modules, but his model is only used in the synthesis of the static image, and there is no effect in the field of visual speech synthesis.

Method based on the data in order to achieve the high quality human face model, need to build a huge database resources, this is only for a specific database resources of experiment personnel, the model can only use a specific staff above, if you want to replace the corresponding experiment, so it is necessary to establish another database, thus, the model's ability to adapt and database of general building methods still need to improve.

3 Parameter-controlled methods
Based on parameter control method, according to the different method of human face model, can be subdivided into two kinds, one kind is the geometric network model, the other kind is physiological anatomy model[2].

(1) A method based on the geometric grid model

Began in the 1970s, study researchers faced synthesis. Parker and others through the establishment of a 3 d face geometric network model to research the human face synthesis, by adjusting the parameters to control model, the movement of some key points in the final in through image processing technology to realize the sports action softness, so as to realize the action of pronunciation facial organs.

Perception Science Laboratory of the university of California at Berkeley PSL (UCSC Perceptual Science Laboratory) in Parke and others on the basis of detailed research, a series of model is put forward, in the end, the success of the development of a virtual host, this is one of the earliest research results. Cohen and Massaro coming through to the people of different organs such as the teeth, tongue, jaw and so on carries on the detailed description, create very fine 3 d face model, thus help correct pronunciation hard-of-hearing people[3][4]. CSLU PSLPSL (Oregon) with Oregon university (Center for Spoken Language Understanding) research Center, developed can be used in the study of voice and facial synthesis toolkit, it includes many of the very practical tool[5].

In coarticulation increase in model establishment, Cohen mainly through statistics of the clustering method to establish the basic lip static view model, Wang Zhiming also through this method to establish the dynamic visual model, so as to realize the model of human face synthesis based on parameter driving. In this way, the model is established and the parameters are simulated dynamically. Taiwan university Perng and Lin, also by the same method for analysis, they are 2 d grid technology for synthesis of face, using a variety of parameters, including a total of 200 vertices and more than 300 triangles, and so on. By adjusting the parameters, the fitting function, the lip form is synthesized, and the natural degree is also relatively good.

In the research of visual speech synthesis is relatively more, generally is conducted using statistical methods in the training, is advantageous to the synthesis of model parameters.

(2) A method based on the anatomical structure of anatomy

The corresponding approach to the geometric network model is the physical anatomical structure model method, also known as the physical model method. This method is mainly from the Angle of physiological anatomy, to look at the problem, on the basis of a variety of interdisciplinary knowledge, with the movement of human physiological organ as the prototype, through the computer technology synthesis has high identification degree facial model. The interdisciplinary includes anatomy,
histology and biomechanics, kinematics, and other aspects, human physiological organ contains the bones, muscles, skin, and so on.

In exploring based on physiological anatomy model, Waters and Terzopoulos also build their own model, mainly divided into six levels from low to high, control layer, layer, physical layer, muscular layer, image and geometry layer. At the high level, the control parameters of nature and semantics are provided, and in the low level, the realistic face simulation animation is implemented.

Magnenat and other people aimed at physiological anatomy model, put forward the abstract muscle action program that a concept, by simulating movement of human organs and muscles to simulate human face synthesis. Bars constructed a model that could be used to implement different changes in faces by adjusting parameters, which can be used to visualize speech synthesis. Fang Wei studied different parts of the movement used three methods of facial muscle deformation, linear local deformation, density of systolic and diastolic virtual models to realize the different faces, so as to achieve the effect of various needs. Zou north types will face subdivided into eight major parts, detailed analysis of the changes in each part of the organization, through this segmentation, face a more simple and faster computer synthesis. Xiang-yu zhang, lan-fang dong people use of front and side two photo synthetic facial expressions, but also realizes the visualization of English speech synthesis, but the Chinese visual speech synthesis system does not implement.

The anatomical structure model has its own merits, the most obvious advantage is that it can describe the human face in detail, and the effect of more realistic action. However, this model also has many shortcomings, the most obvious weakness is the characteristics of the model too much, too complex, and this effect is caused by the workload is bigger, the parameters of the complex model sometimes does not necessarily reflect good effect, but also relatively large amount of calculation[6].

Based on the above analysis, two models because of the advantages of convenient control, strong portability, therefore, in the animation design, game production, has a profound influence on many aspects, such as human-computer interaction, and a wide range of applications.

4 Muscle model
According to the analysis of facial muscle movement characteristics, muscle is divided into two types: the linear muscle that stretches, the plane muscle and the sphincter of the contraction. Linear muscle control insertion point to the contraction and expansion of the attachment point; The smooth muscle is a set of parallel muscles, with a set of attachment points on both sides of the muscle, distinguished from a single attachment point of the linear muscle. The sphincter's function is to shrink the ring muscle, which can be seen as a muscle that contracts to an imaginary center. The user can directly or indirectly set up the activity of the muscle to drive the motion of the facial model, resulting in various expressions. This model is widely used in the synthesis of human face animation due to its simplicity and the independence of face grid[7].

![Figure 4.1 Schematic diagram of linear muscle](image)

Figure 4.1 shows the linear muscle diagram, $V_1$ as a fixed endpoint for the muscle vector; $V_2$ is the other endpoint of the muscle vector; $P$ is any point in the scope; $P'$ prime is the position where $P$ is moved; $D$ is the distance from the fixed end point; $a_2$ Angle the scope; $\mu$ is the angle at which the
vector is from the main line. So the displacement expression for P to P' after the muscle contractions is:

\[
\overrightarrow{PP'} = K \cdot A \cdot h(D) \cdot \overrightarrow{V_{p}} \\
= \frac{K \cdot A \cdot h(D) \cdot V_{p}}{||V_{p}||}
\]

The k is constant

\[
A = \cos(\mu)
\]

\[
h(D) = \begin{cases} 
\cos(1 - D/R_s), & D \leq R_s \\
\cos((D - R_s)/(R_f - R_s)), & R_s \leq D \leq R_f 
\end{cases}
\]

Figure 4.2 Schematic diagram of the sphincter

Figure 4.2 shows the sphincter diagram, and the movement of the sphincter point p is:

\[
f_i = \alpha \theta \left( r_i \right) \frac{(o - \mathbf{p})}{||o - \mathbf{p}||}
\]

Among them, \( \alpha \) is an elastic parameter and is aggressive based on the actual situation:

\[
\theta(\pi) = \cos \left( \left( 1 - r_i \right) \frac{\pi}{2} \right), 0 \leq r_i \leq 1
\]

\[
\theta(\pi) = \cos \left( \frac{r_i - 1}{R - 1} \frac{\pi}{2} \right), 1 \leq r_i \leq R
\]

\( R \) is a threshold, more than that, that's zero; \( r_i \) does not mean the distance to O, but a relative weighted distance, defined as follows:

\[
r_i = \sqrt{y_i^2a^2 + x_i^2b^2}
\]

\( y_i \) and \( x_i \) are the ordinate and abscissa of the point \( p \).

Figure 4.3 Schematic diagram of plane muscle

As shown in figure 4.3, the flat surface muscle diagram, the \( W_i \) is the width of the flat muscle (the length of the non-moving side); The \( L_i \) is the length of the plane muscle; \( x_i \) is the one that affects the domain; \( f_{ij} \) is the displacement of the point \( x_i \) movement on the horizontal muscle, \( m_{j1}^{A1}, m_{j2}^{A2} \) to determine the two endpoints of the maximal muscles. \( m_{j1}^{A1}, m_{j2}^{A2} \) to determine the two endpoints of the
non-moving line of the muscle; $m_{j}^{AC}$, $m_{j}^{lc}$ to determine the two end points of the Central Line; $l_{j,i}$ is the distance between the line of the point. $f_{j,i}$ calculate as the following:

$$ f_{j,i} = \alpha \theta(\lambda_{j,i}) \frac{(m_{j}^{AC} - m_{j}^{lc})}{\|m_{j}^{AC} - m_{j}^{lc}\|} $$

$\alpha$ is the parameter, depending on the situation; The $\theta$ function is as follows:

$$ \theta(\lambda_{j,i}) = \cos \left( (1 - \frac{\lambda_{j,i}}{\lambda_{j,i}^{\eta}}) \cdot \frac{\pi}{2} \right), 0 \leq \lambda_{j,i} \leq 1, $$

$$ \theta(\lambda_{j,i}) = \cos \left( \frac{\lambda_{j,i}^{\eta} - 1}{\delta_{j}^{\eta} - 1} \cdot \frac{\pi}{2} \right), 1 \leq \lambda_{j,i} \leq \delta_{j} $$

The index $\eta$ is the parameter that needs to be taken by itself; $\lambda_{j,i}$ is defined as:

$$ \lambda_{j,i} = \frac{l_{j,i}}{\|m_{j}^{AC} - m_{j}^{lc}\|} $$

$\delta_{j}$ is defined as:

$$ \delta_{j} = \frac{L_{j}}{\|m_{j}^{AC} - m_{j}^{lc}\|} $$

5 Conclusion

There are many ways of modeling human faces, but there are two categories, one based on data-driven methods and one based on parameter control. Based on the data driven face modeling in order to achieve the high quality human face model, need to build a huge database resources, this is only for a specific database resources of experiment personnel, model can only use a specific staff. While face modeling based on parameter control, especially muscle modeling method can precisely describe the face, reflect the action effect of more vivid, the user can be set directly or indirectly, muscle activity to drive the movement of facial model, based on the can realize animation more real.

Acknowledgements

This work was supported by Northwest MinZu University the central university of basic scientific research business funds(No.31920170156,No.Yxm2017109),Gansu Province Education Science PlanningProject(No.GS[2016]GHB0203)

References

[1] Cosatto E, Graf H P, Potamianos G, et al. Audio-visual selection process for the synthesis of photo-realistic talking-head animations: US, US 6654018 B1[P]. 2003.
[2] Morishima S, Aizawa K, Harashima H. An intelligent facial image coding driven by speech and phoneme. In: International Conference on Acoustics, Speech, and Signal Processing, 1989. 3: 1795~1798.
[3] http://psl.ucsc.edu.
[4] Cohen M M, Beskow J, Massaro D W. Recent developments in facial animation: an inside view. In: Proceedings of AVSP'98, Workshop on Audio-Visual Speech Processing, Terrigal, Australia, 1998.
[5] http://cslu.cse.ogi.edu/toolkit/.
[6] Morris T G. Application of personality models and interaction with synthetic characters in a computing system: US, US6526395[P]. 2003.
[7] Z. Deng, U. Neumann, Expressive speech animation synthesis with phoneme-level controls, Computer Graphics Forum 27 (8) (2008) 2096–2113.