Research on Photorealistic Virtual Face Modeling

Xiangzhen He, Northwest Minzu University, China
Shengyin Zhu, Northwest Normal University, China*
Yihao Zhang, Northwest Minzu University, China
Yerong Hu, Northwest Minzu University, China
Dengyun Zhu, Northwest Minzu University, China
Xiaoyue Liu, Northwest Minzu University, China
Fucheng Wan, Northwest Minzu University, China

ABSTRACT

Most of the traditional face modeling methods adopt the parameter-based method to construct, but the face model constructed by this method is too smooth and ignores the detailed features of the face. To solve this problem, a virtual face modeling method based on 3D motion capture data is proposed in this paper. In order to improve the deformation method for the realistic modeling purpose, this paper divides the modeling process of personalized face into two parts: overall modification and local deformation. The overall deformation modifies the facial shape and the position of the facial features of Maya universal model. Based on the principle of radial basis function interpolation algorithm, a smooth interpolation function is constructed, and the new position coordinates of non-characteristic points are obtained by solving the linear equations, so that they are more in line with the physiological characteristics of human faces. The end result is a more realistic virtual face that you model.

KEYWORDS

3D Motion Capture Data, Deformation Model, Face Modeling, Personalized Face, Virtual Human

1. INTRODUCTION

With the rapid development of science and technology, how to reconstruct 3d facial expressions realistically has always been a research hotspot in the field of computer graphics and computer vision (Wang and Shen, 2018; Wang et al., 2020; Mohammad et al., 2018). Facial expression plays a crucial role in human face-to-face communication, which can directly express the internal emotion changes of the communicated objects (Sheng, 2017). Facial movements are abundant, and people are sensitive to facial expressions (Ge, 2014). In recent years, the deep integration of expression capture and interactive experience has brought new direction to human-computer real-time interactive experience (Ran and Chen, 2016). The single human-computer interaction mode can no longer meet the needs of human beings. The user-centered, highly personalized and omni-directional perception interaction method has become the development trend and demand. Efficient tracking of face data with limited hardware resources (Jon, 2020). In this situation, integrated, networked, intelligent and standardized new generation of interaction methods emerge at the right moment, emphasizing that

DOI: 10.4018/JITR.299949
*Corresponding Author

human-computer interaction can be as natural and convenient as human interaction. Human interaction includes language, expression, body movements and many other aspects, especially facial expressions. Face modeling is the key to modern visual effects in special-effects movies and computer games (Yang and Tong, 2018). Human face contains a large amount of feature information and is the most important organ for character recognition. The study of computer simulation of specific faces is of great significance (Lv and Li, 2018), which plays a crucial role in information communication. The authenticity of facial expression animation is one of the most critical evaluation indexes of facial expression simulation. In the process of traditional facial expression animation, the lifelike effect of its animation often needs a large amount of human intervention as the cost (Fang, 2010). With the continuous development of computer animation, in order to obtain realistic character animation, people use motion capture equipment to capture the movements of performers, and directly drive the movement of virtual characters through these motion data (Wei, 2016). Therefore, in order to meet the needs of human visual experience, the study of realistic human face becomes an urgent need.

1.1 Face Motion Capture Data

The facial motion capture data used in the experiment was obtained through the Vicon Cara system, which is a type of optical motion capture system. Compared with other motion capture systems, Vicon cara has the following advantages: high definition, four high-resolution high-speed cameras in the system work simultaneously, can capture the subtle movements of the face, custom built-in lighting, and ensure data integrity. Based on the principle of computer vision, Vicon Cara sets up multiple high-definition cameras in three-dimensional space to scan and track the feature points of the captured object from various angles to achieve the purpose of motion capture. In theory, at any point in the motion capture area, as long as the point can be captured by at least two cameras at the same time, we can determine the three-dimensional position of the point. When the camera configuration is high enough, the pixels are clear enough, and the capture speed is fast enough, we can quickly get the movement track of the point

2. MATERIALS AND METHODS

2.1 Data Acquisition

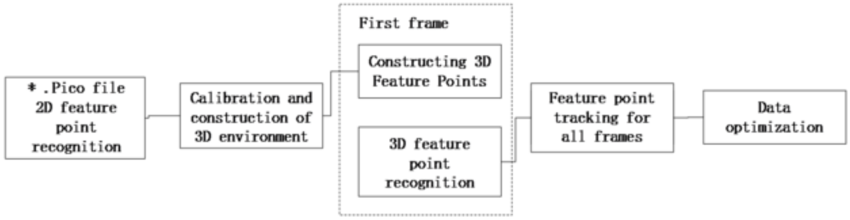
First before experiment officially facial data need to record the calibration files, for real-time monitoring of feature points, by observing the calibration plate movement, detect facial capture the quality of the environment, and by adjusting the focal length of the camera and the parameters setting, so as to ensure the clarity of video data and achieve better capture effect. In order to achieve good data detection, we usually need to spend a lot of time in video calibration to ensure the quality of video data collection, which greatly reduces the workload of subsequent data processing. After the calibration file collection is completed, facial expression and movement data collection is carried out on the actors. The video data collected by Vicon Cara system is first stored in Logger. In case of network, the video data file should be downloaded from Logger and exported in the format of * .pico and * .xcp, and. Pico is the original video data file. We will output the calibration video in * .xcp format and the expression video in * .pico format.

2.2 Data Processing

Since the video data captured by Vicon Caralive in this experiment are only two-dimensional images in four channels and cannot be directly used for research, data processing, including three-dimensional construction and data optimization, is required. The data processing flow chart is shown in Figure 1.

(1) Identify the two-dimensional feature points in the * .pico file. According to the feature points detected by the classifier, the image processing technology is used to extract the boundaries of each feature point, and finally an ellipse is fitted to each of the extracted feature points.

Figure 1. Data processing flowchart



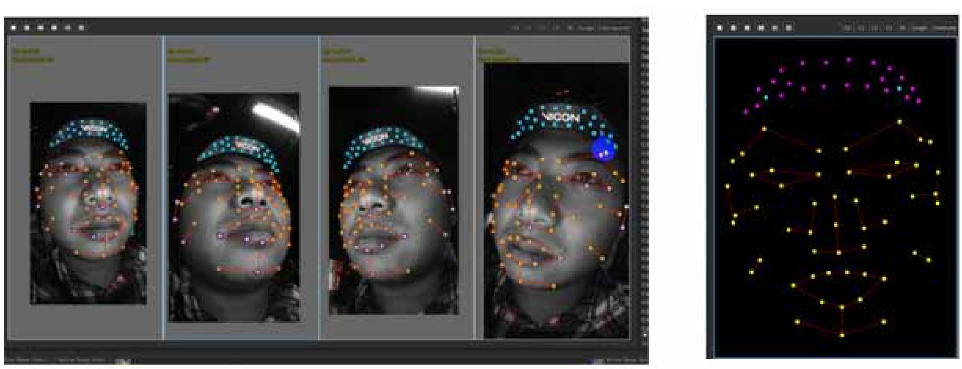
(2) Calibration and construction of three-dimensional environment. Import the * .xcp file, use the calibration edge algorithm to initialize the data points of the Brim board, establish the 3D data point set of the Brim board in the 3D environment, and then collect all the points on the Brim board and initialize the automatically generated point set After correlation, the initial estimates of the position and orientation of the two-dimensional data points of the four channels in three-dimensional space are finally obtained.

(3) Construct 3D feature points and 3D feature point recognition. Obvious feature points are selected as seed points at various important parts of a human face, thereby identifying points around it. Starting at the specified seed point, repeatedly review the newly matched 2D data points in the two camera views until no more matches are found. In each iteration process, the longitudinal constraint method is used to find all clear candidate matches, and three-dimensional feature points are generated at specified positions.

(4) Track the feature points of all frames. The three-dimensional position of the marker point in the next frame is estimated from the first frame, and the corresponding two-dimensional coordinates are calculated in each channel view. Then, according to the position and shape of the feature points in each view, the best matching point of the next frame is detected, and the position of the three-dimensional marker point is recalculated.

(5) Data optimization is the most critical step in data processing. The tracked data is missing and incomplete due to human factors or system software, hardware, environment and other factors; there are problems such as interference points and noise points. Therefore, we need to optimize the data.
   - Missing mark points in the acquisition process are completed. The missing point complement is completed by nonlinear interpolation technique, and the optimal missing point position is predicted based on the known position of the adjacent frame mark point and the moving relation between the two frames.
   - Remove redundant noise points. The software comes with the ability to delete all unnamed points, using the topological connections we made in the first step to remove all interference points and redundant noise points.
   - Topologically connect the landmarks. The construction of topological relations can clearly display the distribution of data points on the face, and can directly see the restored motion of feature points, so that the identification and inspection of mark points can be carried out frame by frame. Topological connection of faces in 3d environment is shown in Figure 2.

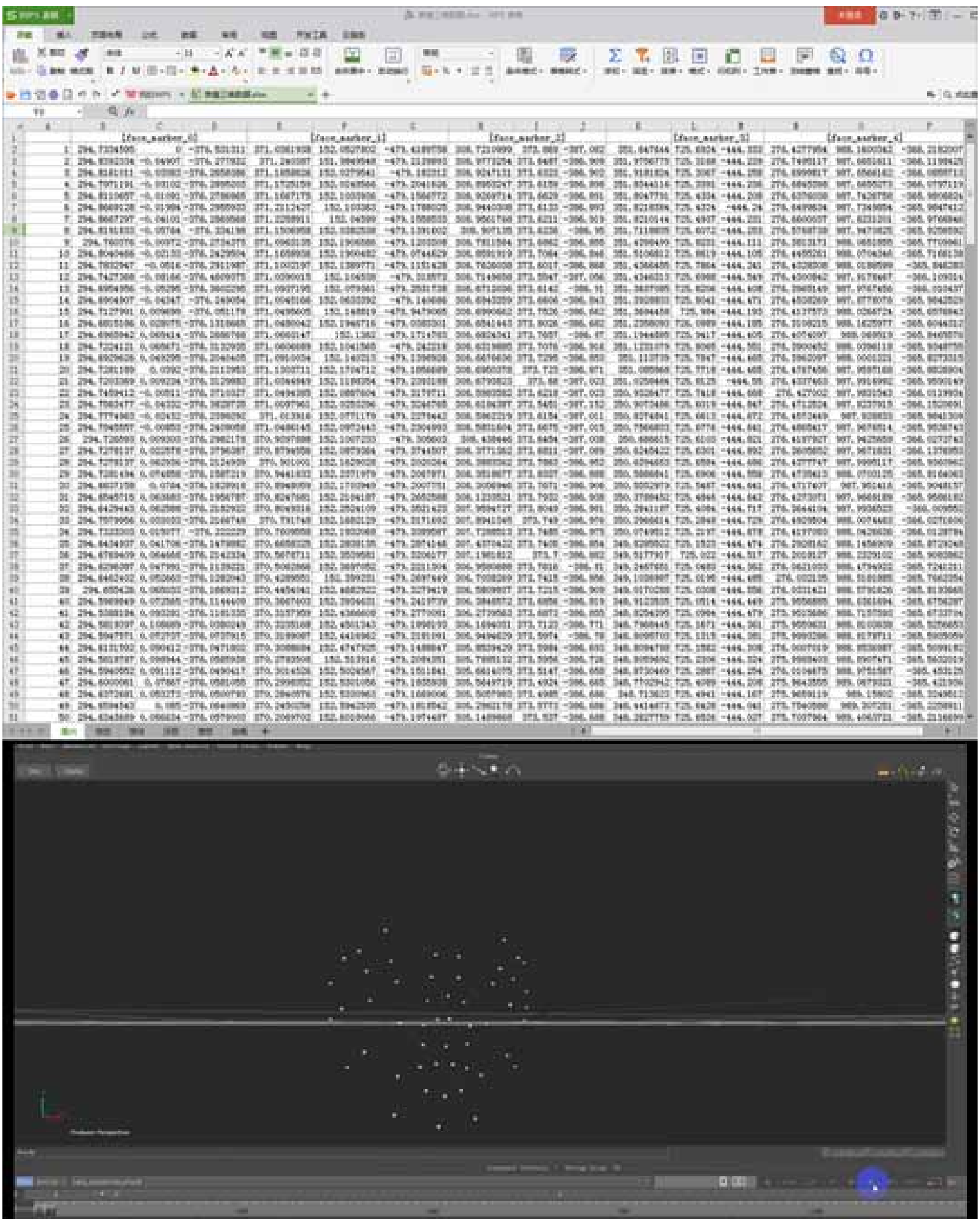
Figure 2. Face topology connection in a three-dimensional environment



- Standardization of 3D space coordinates. Manually select three clear mark points with little motion change as the stability points, and standardize the THREE-DIMENSIONAL space coordinates. Since the stability points are not affected by motion changes, the feature points on the current frame can be mapped into the three-dimensional space for each frame.

After the data processing is completed, the three-dimensional data is exported in the format of * .fbx. * .fbx is a public domain binary file format for storing three-dimensional motion capture data. The format is converted using motion builder and converted into * .htr format data. Organize and build. * xls format expression database. The facial expression database is shown in Figure 3.

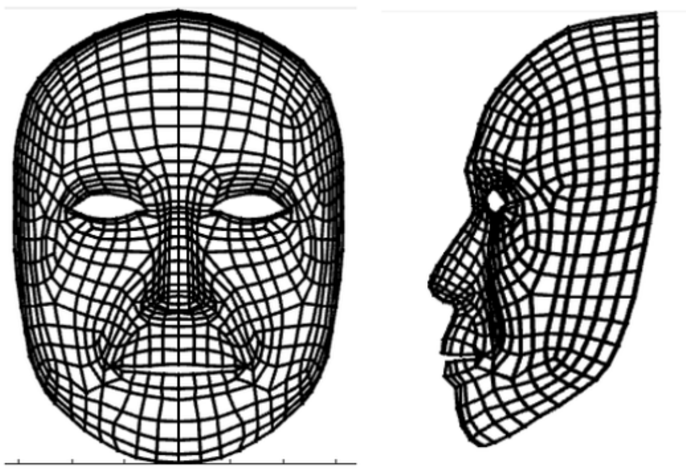
Figure 3. Facial expression database



3. RESULTS AND DISCUSSION

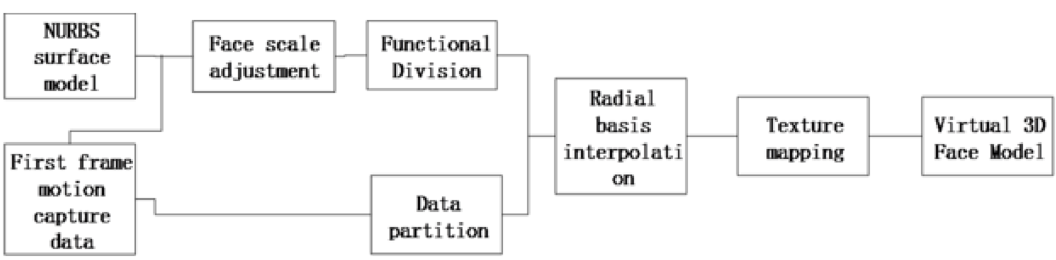
We can distinguish between different people because each individual's facial features are different, and there are hundreds of millions of ways to represent each of the world's hundreds of millions of people. The NURBS universal face model is obtained through Maya. Although the structure is perfect and the lines are smooth, the lack of face features reduces the visual sense of reality. In order to make the face model more realistic, the Vicon Cara face capture system used in this experiment can accurately capture the detailed features of the face, modify the general face model based on the 3d data of operation and capture, and finally build the personalized face model. See Figure 4.

Figure 4. NURBS face surface mesh model



In order to achieve a realistic personalized face, In this paper, the deformation process is divided into two parts: global deformation and local deformation. The overall deformation is to modify the face shape and the positions of the five features of the NURBS face model, but the overall deformation is only similar. Enhance the effect, partition the face model, and deform the RBF of each local area to achieve both form and spirit. The face modeling technology roadmap is shown in Figure 5.

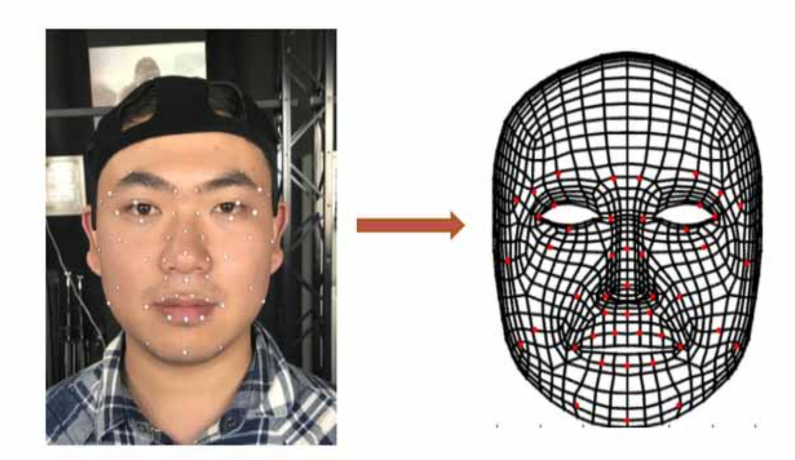
Figure 5. Face Modeling Technology Roadmap



3.1 General Face Model Overall Transformation

The model established by NURBS includes 790 polygons and 775 mesh vertices. According to the method of 3D motion capture experiments to mark feature points, 52 feature points are defined on a standard face model. The deformation is based on the neutral expression, and the corresponding feature points in the motion capture experiment are labeled in the NURBS face model. Feature point marks are shown in Figure 6.

Figure 6. NURBS model face marking



The essence of the overall deformation is to stretch and compress all feature points on the NURBS model in the *X*, *Y* and *Z* directions. The degree of change in each direction R_x, R_y, R_z is calculated by equation (3-1):

$$\begin{cases} R_x = \dfrac{H_1}{H_0} \\ R_y = \dfrac{W_1}{W_0} \\ R_z = \dfrac{L_1}{L_0} \end{cases} \tag{3-1}$$

In the above formula, H_1 is the horizontal length in the *X* direction of the eyebrow peak point and the outer corner point of the neutral expression facial capture data, and H_0 is the horizontal length in the *X* direction of the eyebrow peak point and the outer corner point of the NURBS facial model; W_1 is the neutral expression facial capture. Data The horizontal length of the left and right outer corner points in the *Y* direction. W_0 is the horizontal length of the left and right outer corner points of the NURBS facial model in the *Y* direction. L_1 is the eyebrow peak point of the neutral expression facial capture data. The horizontal length of the mouth corner point in the *Z* direction,

L_0 is the horizontal length of the eyebrow peak point and the mouth corner point of the NURBS facial model.

After obtaining the changes in each direction, the new positions of the feature points of the NURBS face model are calculated according to formula (3-2).

$$\begin{cases} x_1 = x_0 * R_x \\ y_1 = y_0 * R_y \\ z_1 = z_0 * R_z \end{cases} \tag{3-2}$$

In the above formula, x_0, y_0, z_0 are the lengths of the original NURBS facial model feature points in the *X*, *Y*, and *Z* directions, and x_1, y_1, z_1 are the changed NURBS facial model feature points in the *X*, *Y*, and *Z* directions, respectively length.

The location of feature points is calculated in order to carry out overall deformation of universal NURBS face, which can make face features close to real faces. However, due to the fact that other parts cannot be driven to be deformed except for the location of feature points, local deformation needs to be carried out by using radial basis function.

3.2 Local Deformation Based on Radial Basis Functions

The local deformation adopts the radial basis interpolation method. The basis function contains only a single variable. The basis function of a point (x,y) is represented by $h_k\left(x,y\right)=h\left(d_k\right)$, where d_k is the horizontal distance from the point (x,y) to the k-th feature point. Generally, the accuracy of the polynomial expression of the above basis function is low, and it needs to be optimized to obtain a high-precision polynomial RBF formula (3-3):

$$F\left(x,y\right)=\sum_{k=1}^{n}a_k h_k\left(x,y\right)+\sum_{k=1}^{m}b_k q_k\left(x,y\right) \tag{3-3}$$

Among them, $q_k\left(x,y\right)$ is a polynomial basis whose power is less than m . ak , bk is a coefficient that satisfies the equations:

$$\sum_{j=1}^{n}a_j h_j\left(x_i,y_i\right)+\sum_{j=1}^{m}b_j q_j\left(x_i,y_i\right)=f_i \;;\quad i=1,2,\cdots,n \tag{3-4}$$

$$\sum_{i=1}^{n}a_i h_j\left(x_i,y_i\right)=0 \;;\quad j=1,2,\cdots,m \tag{3-5}$$

Equation (3-4) includes *n* equations that meet the requirements of RBF interpolation. Equation (3-5) contains the number of equations as *m*, ensuring the high accuracy of the polynomials, and solving the pending coefficients for the two equations simultaneously. The following is the specific operation of local deformation.

For 52 feature points, the three-dimensional feature points collected by the Vicon Cara face capture system are labeled p_i, the corresponding feature points on the NURBS face model are recorded as $p_i^{(0)}$, and the spatial displacement of this point is expressed as (3-6):

$$u_i = p_i - p_i^{(0)} \tag{3-6}$$

Definition $f\left(p_i\right)$ satisfies $f\left(p_i\right) = u_i$, displacement of feature points p_i on NURBS face model is solved from $f\left(p_i\right)$, and the following form is established for RBF function. As shown in Formula (3-7):

$$u = \sum_{k=1}^{n} c_i \varnothing\left(p - p_i\right) + M_p + t \tag{3-7}$$

In the formula, *n* is the total number of feature points, c_i is the weight coefficient of the feature point p_i, $\varnothing$ is a radial symmetric function, *M* and *t* are affine components, where *M* is a 3×3 matrix and t is a 3×1 vector of. The subscripts corresponding to the three directions of p_i, u_i, c_i, and t_i are *x*, *y* and *z*. $M = \left(M_j^i\right)$, *i,j* ⊨ 0,1,2. The determination of the coefficients c_i and the affine components *M* and *t* depends on the constraints of (3-8).

$$\begin{cases} u_i = f\left(p_i\right) \\ \sum_{i=1}^{n} c_i = 0 \\ \sum_{i=1}^{n} c_i \bullet p_i^T = 0 \end{cases} \tag{3-8}$$

The purpose of the latter two equations is to reduce the adverse effects of the RBF affine component. A Gaussian function is used to set the radial symmetry function $\varnothing$. As shown in Formula (3-9):

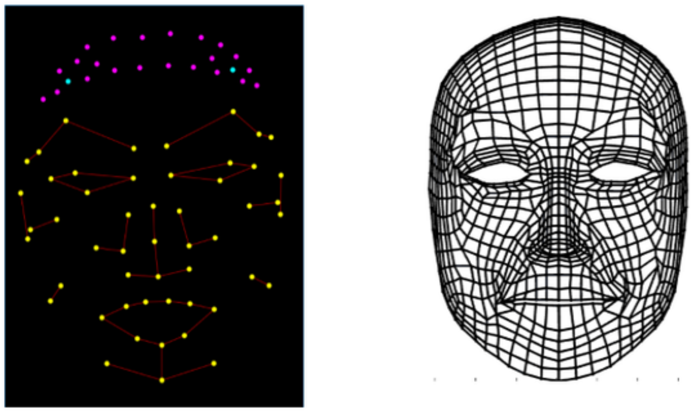
$$\varnothing(r) = e^{-\alpha r^2}, \alpha > 0 \tag{3-9}$$

In this paper, α is set to 1/64.

After solving the parameter values of formula (3-9), all feature points of the NURBS face model are deformed, and a partially modified personalized virtual face model is established as shown in Figure 7.

In this experiment, a general NURBS surface model is established, and then the overall deformation and RBF local deformation of the NURBS model are processed based on the 3D captured data, and finally the virtual face model of specific face is established.

Figure 7. Personalized virtual face mesh model



4 CONCLUSION OUTLOOK

Face modeling technology involves image processing, advanced mathematics, computer graphics, face anatomy and other fields, and has a broad application prospect in face recognition, virtual teaching, film and television animation and other industries. As a subject with simultaneous development of research and application value, the updating and progress of facial modeling technology has far-reaching significance. Since the 1870s, many researchers have devoted their vision to this field and made numerous research achievements. Nevertheless, it is still a challenging task to build realistic face models. Based on the analysis and summary of previous studies, this paper makes a deep discussion on face modeling, aiming to build a realistic face model based on 3d motion capture system. The facial motion capture system is used to obtain accurate personalized facial data, and the geometric relationship between the general model and the 3D data is established. First, the general model is deformed as a whole, and the coordinates of the facial shape and facial features are adjusted to achieve "shape". To further improve the effect, the face model is partitioned and a smooth interpolation is constructed according to the principle of the radial basis function interpolation algorithm The function obtains the new position coordinates of the non-feature points by solving the linear equations, which makes it more consistent with the physiological characteristics of the human face, and achieves "likeness". Facial mesh is only equivalent to the skeletal part of anatomy, and the recognition and authenticity of the face model are mainly realized by texture mapping. Since the face model is similar to a cylinder, the method of cylindrical texture mapping is adopted to establish the corresponding relationship between 2d image data points and personalized 3D face grid model data points, and a 3d face model with sense of reality and recognition is established in OpenGL development environment.

The personalized face modeling in this paper is only the preliminary basis of facial animation. The key technologies of face grid modeling are deeply studied in this paper, and the facial expression database is established by using the THREE-DIMENSIONAL motion capture system to collect facial data. All these work are preparing for the later study of facial expression animation. In the following research work, the implementation of virtual facial expression animation system will become an important subject.

Face modeling is applied to face recognition, remote teaching and video keying. From the perspective of the development of face modeling, its value is mainly widely used in research and film and television animation, and many achievements have been made. However, due to its complexity, it is not widely used in daily life, such as face recognition, behavior analysis, tracking and monitoring medium engineering applications, and it is more inclined to simple and direct TWO-DIMENSIONAL technology. Through the continuous optimization of 3D modeling technology, the model has an intuitive and real visual effect. Based on this, it is also a very valuable research direction to study the application technology of automation to promote the engineering application.

In addition, due to time constraints and personal level defects, there are many deficiencies in the study of this paper, algorithm technology and research scheme still need to be further improved, and the implementation effect has not reached the most ideal state. In the process of face modeling, My thinking is not rigorous enough, and many factors are ignored. In the future work, face modeling technology will be constantly improved and improved.

DECLARATION

We declare that there is no conflict of interest in this paper.

FUNDING AGENCY

The publisher has waived the Open Access Processing fee for this article.

5 ACKNOWLEDGMENT

The research projects of the social science and humanity on Young Fund of the ministry of Education (No.17YJCZH057).

Supported by the Fundamental Research Funds for the Central Universities. (No:31920190059).

REFERENCES

Alsmirat, M. A., Al-Alem, F., Al-Ayyoud, M., Jararweh, Y., & Gupta, B. (2018). Impact of digital fingerprint image quality on the fingerprint recognition accuracy. *Multimedia Tools and Applications*.

Fang, X. (2010). *Research on facial motion capture data processing and expression animation reconstruction*. Dalian University of Technology.

Ge, X. (2014). *Research on facial expression animation based on motion capture data*. Hebei University of Technology.

Goenetxea, Unzueta, Dornaika, & Otaegui. (2020). Efficient deformable 3D face model tracking with limited hardware resources. *Multimedia Tools and Applications*.

Lu & Li. (2018). Overview of Realistic 3D Face Modeling Technology. *Software Guide, 17*(1), 1-3+7.

Nan, S. (2017). *3d Facial Expression Modeling and Recognition*. Beijing University of Technology.

Ran, X., & Hong, C. (2016). Design and Implementation of Face Real-time Interactive Experience based on facial Expression Capture Technology. *Computer Programming Skills and Maintenance*, (01), 79–80.

Shan, W., Xukun, S., & Qinping, Z. (2018). Article. *Xitong Fangzhen Xuebao*, *30*(7), 2423–2444.

Wang, H., Li, Z., Yang, L., & Gupta, B. B. (2020). Choi Chang. Visual saliency guided complex image retrieval. *Pattern Recognition Letters*, *130*, 64–72. doi:10.1016/j.patrec.2018.08.010

Wei, R. (2016). *Research on motion analysis technology based on human motion capture data*. Beijing Jiaotong University.

Yang, J., Li-jing, T., & Fu, X. (2018). Facial Modeling animation System based on Feature Extraction. *Digital Technology and Application*, *36*(9), 34–36. doi:10.1364/CLEO_QELS.2018.FW3F.4