Hindawi
Advances in Multimedia
Volume 2023, Article ID 9764212, 1 page
https://doi.org/10.1155/2023/9764212

Retraction

Retracted: Generating 3D Virtual Human Animation Based on Facial Expression and Human Posture Captured by Dual Cameras

Advances in Multimedia

Received 15 August 2023; Accepted 15 August 2023; Published 16 August 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Wang, X. Chai, G. Han, and J. Wang, "Generating 3D Virtual Human Animation Based on Facial Expression and Human Posture Captured by Dual Cameras," *Advances in Multimedia*, vol. 2022, Article ID 4833436, 9 pages, 2022.

Hindawi
Advances in Multimedia
Volume 2022, Article ID 4833436, 9 pages
https://doi.org/10.1155/2022/4833436

Research Article

Generating 3D Virtual Human Animation Based on Facial Expression and Human Posture Captured by Dual Cameras

Junming Wang,[1] **Xiaojie Chai**,[2] **Guo Han**,[3] **and Jixia Wang**[2]

[1]*Yantai Nanshan University, College of Humanities, Yantai 265713, China*
[2]*Yantai Nanshan University, Office of Teaching Affaires, Yantai 265713, China*
[3]*Nanshan Tourism Group Co., Ltd., Yantai 265713, China*

Correspondence should be addressed to Junming Wang; wangjunming2@nanshan.edu.cn

Received 22 May 2022; Revised 26 June 2022; Accepted 6 July 2022; Published 9 August 2022

Academic Editor: Qiangyi Li

In order to generate 3D virtual human animation with smoother actions and richer and more vivid expressions, this article proposes a system based on dual cameras to synchronously capture human facial expressions and human posture, so as to better generate 3D virtual human animation. Firstly, the 2D features of the image are extracted through facial expression, and the benchmark is provided for the standard 3D coordinate values. At the same time, the 3D pose fusion is realized by using the pose measurement algorithm for human motion capture. Through the test, it is found that the two fusion algorithms can meet the requirements of the static accuracy of 0.5° and the dynamic accuracy of 0°, and the three-dimensional model of a virtual human body has a good effect in data-driven animation. The experiments show that the system can better capture facial expressions and human posture synchronously, and the frame rate in the experimental test can reach 20 fps, which can generate more natural and realistic 3D virtual human animation in real time.

1. Introduction

With the continuous development of intelligent technology and its wide application in different industries, computer virtual reality technology has gradually become a hot topic of current research, and humans, as one of the most important roles in virtual scene, have naturally become the focus of virtual technology research. At this stage, virtual human technology has been widely used in game production, multimedia, film and television production, video conference, and other fields [1]. Through three-dimensional virtual human recognition and embodiment, it can show the virtual human with facial emotion expression characteristics and further improve the real feeling under the virtual scene and the immersion of human-computer interaction [2]. To achieve this goal, we must synthesize realistic and easy to control 3D virtual human animation, scientifically and reasonably match and recognize facial expression information and head pose estimation information, and provide a more realistic virtual reality technology experience for public life.

2. Literature Review

At present, there are three basic methods in the field of face modeling at home and abroad. The first is to use a 3D scanner to scan 3D face data and obtain data points to build a face model. The advantage of this method is that the constructed face model is very realistic, but the disadvantage is that the equipment is expensive, the final generated model is too complex, the control is difficult, it is not conducive to expression synthesis, and it is easy to be affected by object material and highlight in the scanning process [3]. Second, the face model is reconstructed by using the front and side photos of the face, or multiple photos from multiple angles. This method is convenient to import materials, and the presented model is closer to the real face. However, before modeling, the corresponding relationship between multiple images must be established, and a lot of manual annotation work needs to be done, which is cumbersome and complex. Thirdly, a 3D scanner is used to scan multiple photos and establish a face database. When a single face image is

imported, the faces in the photos are matched according to the linear combination of face models in the database, and the matching results can be regarded as the face model to be constructed. The advantage of this method is that the process of importing materials is simple, only front photos are needed, and the synthesis effect is realistic enough. However, the establishment of a face model database is a very cumbersome process, which requires a lot of manpower and material resources [4].

In recent years, some researchers have proposed a new method to obtain the data needed to construct 3D virtual face through projector and camera. By adjusting the position of camera and projector, 3D data are obtained, and then, image matching is carried out to obtain a realistic 3D face model. In the aspect of feature modeling, it is proposed to use three cameras to shoot the face from different angles and construct the three-dimensional face model based on the two-dimensional image [5]. Firstly, the face range is determined by detecting the change of color brightness, and then, the face frame structure is obtained by using the active surface model. Finally, the corresponding face model is obtained by adjusting and texture matching combined with the processed image, but it needs postprocessing to get better results [6]. Zhou proposed an automatic 3D face reconstruction method based on a single photo. Firstly, the improved algorithm is used to extract the face range and feature points from the imported face photos, and then, the position of feature points is corrected by using color information. Finally, through texture matching, the specific face model is put into the general face model to realize expression change. In the aspect of statistical modeling, the three-dimensional information of feature points is reconstructed by stereo vision. Based on the radial basis function of multilayer and multiregion, the feature points and general face model are used for the reconstruction of specific face, the spline curve is used for the reconstruction of the face surface, and finally, the spherical mapping is used for texture restoration. This method is more natural and can get realistic effects. The research on Chinese head and face modeling is carried out. The three-dimensional point cloud data used are from the literature of Chinese face size statistics. The data are the average level of Chinese face and head size. Then, the modeling is based on the point cloud data, and the human model is derived through the three-dimensional modeling software. The advantage of this method is the high fidelity of the model, but the disadvantage is the large amount of data. It is easy to control after simplifying it [7].

In the aspect of human pose estimation, someone designed OpenPose, which can detect multiple human poses in one image in real time, predict the joint point confidence and part of the affinity field vector in the multicamera scene, and match the 3D information from the 2D pose detection results, which has achieved good robustness; Purnama and Sari put forward the method of fitting dense 3D human body with skinned multiperson linear model (SMPL), constructs DensePose network to regress the human body surface, and maps the human body pixels of the image into 3D human body surface [8]; Li et al. proposed a real-time and stable attitude estimation method based on monocular camera, which has high real-time detection speed. However, in some scenes, the joint prediction is not accurate enough to solve the problem of limb occlusion [9]; He et al. propose to use Kinect camera and color camera to synchronously collect facial expression and body posture, use Kinect fusion to scan the character model, use robust ICP method to realize pose estimation, and reconstruct the face to obtain facial animation [10]. However, it is easy to introduce noise during Kinect acquisition, resulting in jitter. In terms of facial expression animation, a FaceWarehouse 3D expression database is constructed to regress and predict 3D feature points, track facial expressions in real time, and register and generate 3D faces. It can generate realistic facial expression animation combined with blendshape, which improves the accuracy of 3D expression animation. However, in the offline training stage, it is necessary to collect data for each face, and the generated face model is relatively rough [11].

3. Attitude Measurement Algorithm for Human Motion Capture

3.1. Design of Attitude Measurement Unit. Taking the anisotropic magnetoresistive sensor AMR, which is widely used at present, as an example, its resistance is composed of permalloy covered on silicon wafer, and the change of external magnetic field is judged according to the change of its resistance value. When the ferromagnetic material is in the applied magnetic field, the included angle between the magnetization direction and the current changes, resulting in the change of its resistance. The included angle changes between $-90^\circ \sim +90^\circ$. When the included angle is 0°, the resistance of the ferromagnetic material is the largest, and when the included angle is $\pm 90^\circ$, the magnetoresistance is the smallest. Generally, the magnetic sensor is in the linear region of about 45° to ensure the linear characteristics of magnetic intensity output [12]. The relationship is shown in Figure 1.

3.2. Magnetic Sensor Correction Error. When the magnetic sensor is used, it relies on the internal magnetic sensing element to convert the external magnetic intensity into electrical physical quantity. When the magnetic sensor works in the magnetic field, its internal sensitive elements will also be magnetized, resulting in its own magnetic field. This part of the magnetic field is different from the magnetic field in the external environment, which is called remanence [13]. This residual magnetism is also included in the measured value of the magnetic sensor. In addition, in the process of converting magnetic field into electrical signal, the zero point of analog circuit and a/D conversion is not zero, which will also cause error in the measured value. The residual magnetism and the offset of the circuit act on the magnetic sensor, which is equivalent to applying a fixed magnetic field on the sensitive axis. This part of the offset can be regarded as translation error.

Suppose that the ideal values of the magnetic field component are m_{x0}, m_{y0}, m_{z0} without the influence of zero bias, while the actually measured magnetic intensity

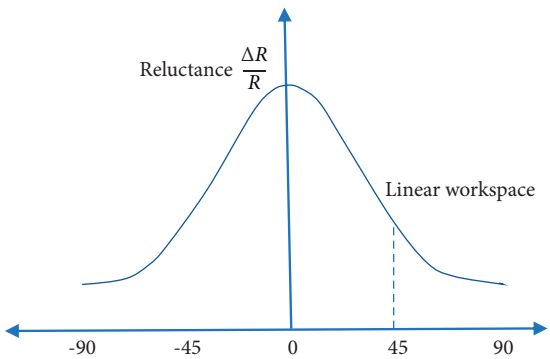

Figure 1: Relationship between magnetoresistance and angle between magnetic field and current.

components in the geomagnetic environment are m_x, m_y, m_z, and the three-axis zero bias error is u_x, u_y, u_z. The relationship is as follows:

$$\begin{cases} m_{x0} = m_x + u_x, \\ m_{y0} = m_y + u_y, \\ m_{z0} = m_z + u_z. \end{cases} \tag{1}$$

Which is expressed in the matrix form as follows:

$$\begin{bmatrix} m_{x0} \\ m_{y0} \\ m_{z0} \end{bmatrix} = \begin{bmatrix} m_x \\ m_y \\ m_z \end{bmatrix} + \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix}. \tag{2}$$

Let the ideal values of the magnetic component be m_{x0}, m_{y0}, m_{z0} without being affected by the scale error, while the measured values of the magnetic component in the measurement path are m_x, m_y, m_z, and the error of the triaxial scale coefficient is r_x, r_y, r_z. The relationship is as follows:

$$\begin{cases} m_{x0} = r_x m_x, \\ m_{y0} = r_y m_y, \\ m_{z0} = r_z m_z. \end{cases} \tag{3}$$

which is expressed in the matrix form as follows:

$$\begin{bmatrix} m_{x0} \\ m_{y0} \\ m_{z0} \end{bmatrix} = \begin{bmatrix} r_x & 0 & 0 \\ 0 & r_y & 0 \\ 0 & 0 & r_z \end{bmatrix} \begin{bmatrix} m_x \\ m_y \\ m_z \end{bmatrix}. \tag{4}$$

Assuming that the measurement axes are orthogonal, the three-axis magnetic field values are m_{x0}, m_{y0}, m_{z0}, while the magnetic field components measured in the application process are m_x, m_y, m_z [14]. The following formula can be obtained through coordinate transformation (see Figure 2 for coordinates):

$$\begin{cases} m_{x0} = m_x \cos\beta \cos\gamma, \\ m_{y0} = m_y \cos\alpha + m_x \cos\beta \sin\gamma, \\ m_{z0} = m_z + m_y \sin\alpha + m_x \sin\beta. \end{cases} \tag{5}$$

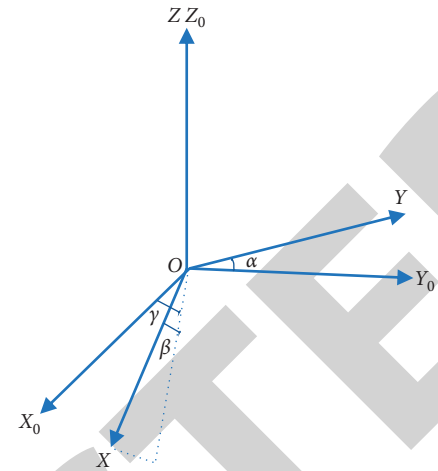

Figure 2: Schematic diagram of nonorthogonal triaxial.

Since the nonorthogonality of the three measuring axes is relatively small in the manufacturing process, and α, β, γ are close to 0,

$$\begin{aligned} &\cos\alpha \approx \cos\beta, \\ &\sin\alpha \approx \alpha, \\ &\sin\beta \approx \beta, \\ &\sin\gamma \approx \gamma. \end{aligned} \tag{6}$$

This leads to

$$\begin{cases} m_{x0} = m_x, \\ m_{y0} = m_y + m_x\gamma, \\ m_{z0} = m_z + m_y\alpha + m_x\beta. \end{cases} \tag{7}$$

which is expressed in the matrix form as follows:

$$\begin{bmatrix} m_{x0} \\ m_{y0} \\ m_{z0} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \gamma & 1 & 0 \\ \beta & \alpha & 1 \end{bmatrix} \begin{bmatrix} m_x \\ m_y \\ m_z \end{bmatrix}. \tag{8}$$

When the magnetic sensor measures in the geomagnetic field, the measured value will shift due to the influence of hard magnetic materials such as magnetized metals in the surrounding environment. Once the hard magnetic material is magnetized, its magnetic field intensity is relatively stable, and the interference projected onto the three measurement axes remains basically unchanged [15]. This part of the interference acts on the magnetic sensor and can be equivalent to the translation error of the magnetic field component. Let the ideal values of the magnetic component be m_{x0}, m_{y0}, m_{z0} without the influence of hard magnetic interference, the triaxial components actually measured under hard magnetic interference are m_x, m_y, m_z, and the components along the triaxial direction of hard magnetic interference are h_x, h_y, h_z. The relationship is as follows:

$$\begin{cases} m_{x0} = m_x + h_x, \\ m_{y0} = m_y + h_y, \\ m_{z0} = m_z + h_z. \end{cases} \tag{9}$$

which is expressed in the matrix form as follows:

$$\begin{bmatrix} m_{x0} \\ m_{y0} \\ m_{z0} \end{bmatrix} = \begin{bmatrix} m_x \\ m_y \\ m_z \end{bmatrix} + \begin{bmatrix} h_x \\ h_y \\ h_z \end{bmatrix}. \tag{10}$$

In addition to hard magnetic materials, there is often a kind of soft magnetic material in the external environment. After being magnetized, the magnetic field of soft magnetic materials is not as stable as that of hard magnetic materials and is easily changed by the magnetic field change of the surrounding environment [16]. Under the geomagnetic field, it is generally considered that the magnetic field generated by soft magnetic materials is directly proportional to the magnetic intensity received by itself [17]. This proportional coefficient is called magnetic susceptibility. Suppose that the ideal value of the magnetic component is m_{x0}, m_{y0}, m_{z0} without the influence of soft magnetic interference, the triaxial component actually measured under interference is m_x, m_y, m_z, and the factor of soft magnetic interference is s_x, s_y, s_z. The relationship is as follows:

$$\begin{cases} m_{x0} = s_x m_x, \\ m_{y0} = s_y m_y, \\ m_{z0} = s_z m_z. \end{cases} \tag{11}$$

Therefore, combining the influence of the above five errors, the error model is established as follows:

$$\begin{bmatrix} m_{x0} \\ m_{y0} \\ m_{z0} \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix} \begin{bmatrix} r_x & 0 & 0 \\ 0 & r_y & 0 \\ 0 & 0 & r_z \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ \gamma & 1 & 0 \\ \beta & \alpha & 1 \end{bmatrix} \begin{bmatrix} m_x \\ m_y \\ m_z \end{bmatrix} + \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} + \begin{bmatrix} h_x \\ h_y \\ h_z \end{bmatrix}. \tag{12}$$

For the above errors, the scale coefficient errors r_x, r_y, r_z and the triaxial soft magnetic errors s_x, s_y, s_z have the same form, which are proportional to the magnetic field component. The three-axis zero bias error u_x, u_y, u_z and the three-axis hard magnetic error h_x, h_y, h_z also have the same form. Although all error factors have corresponding meanings in the previous analysis, the purpose of magnetic sensor error compensation is to obtain the ideal value of geomagnetic field intensity through the measured value of triaxial magnetometer under interference [18]. Therefore, the above parameters can be regarded as unknowns. The error model of the magnetic sensor is established by combining the above error factors as follows:

$$\begin{bmatrix} m_{x0} \\ m_{y0} \\ m_{z0} \end{bmatrix} = \begin{bmatrix} k_x & 0 & 0 \\ 0 & k_y & 0 \\ 0 & 0 & k_z \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ \gamma & 1 & 0 \\ \beta & \alpha & 1 \end{bmatrix} \begin{bmatrix} m_x \\ m_y \\ m_z \end{bmatrix} + \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix}. \tag{13}$$

where

$$\begin{bmatrix} k_x & 0 & 0 \\ 0 & k_y & 0 \\ 0 & 0 & k_z \end{bmatrix} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & s_z \end{bmatrix} = \begin{bmatrix} r_x & 0 & 0 \\ 0 & r_y & 0 \\ 0 & 0 & r_z \end{bmatrix}, \quad \begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} = \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} + \begin{bmatrix} h_x \\ h_y \\ h_z \end{bmatrix}. \tag{14}$$

The sum of the modulus of the triaxial magnetic component of the geomagnetic field is a constant (related to the geographical location). However, due to the influence of noise in the measurement process and the instrument error of the sensor itself, the modulus of the actually obtained triaxial component often changes in a large range [19].

Under static conditions, the root mean square errors of roll angle, pitch angle, and heading angle obtained by the above two compensation fusion algorithms are shown in Table 1.

Under low dynamic conditions, the comparison and error between the heading angle solution based on gradient descent compensation fusion and the theoretical value are shown in Figure 3.

Taking the heading angle as an example, the comparison between the calculation results of the attitude fusion heading angle based on complementary filtering and PI adjustment and the theoretical value and the error angle under low dynamic conditions are shown in Figure 4.

Taking the heading angle as an example, the comparison between the calculation results of the attitude fusion heading angle based on gradient descent and the theoretical value and the error angle under high dynamic conditions are shown in Figure 5.

Taking the heading angle as an example, the comparison between the calculation results and theoretical values of the compensated fusion heading angle based on complementary filtering and PI adjustment under high dynamic conditions and the error angle are shown in Figure 6.

Under high dynamic conditions, the root mean square of roll angle, pitch angle, and heading angle calculated by the above two compensation fusion algorithms are shown in Table 2.

To sum up, according to the above three groups of the experimental results, the solution effect of the two attitude fusion algorithms is good, with a static accuracy of less than 0.5° and a dynamic accuracy of less than 2°. In contrast, the fusion algorithm based on complementary filtering and PI regulation is simple to realize and the solution accuracy is more stable [20].

4. Dual Cameras Synchronously Capture Facial Expressions and Human Posture

4.1. Camera Synchronization. In the process of human posture acquisition, the camera is far away from the human face, and the ability to deal with subtle facial expressions is poor. In order to solve this problem, two monocular cameras

Table 1: Comparison of attitude solution errors based on gradient descent, complementary filtering, and PI adjustment under static state.

Error	Gradient descent method	Cross filtering and PI regulation
RMSE[0]	0.2005	0.0313
RMSE[ρ]	0.1862	0.0263
RMSE[ψ]	0.2205	0.0657

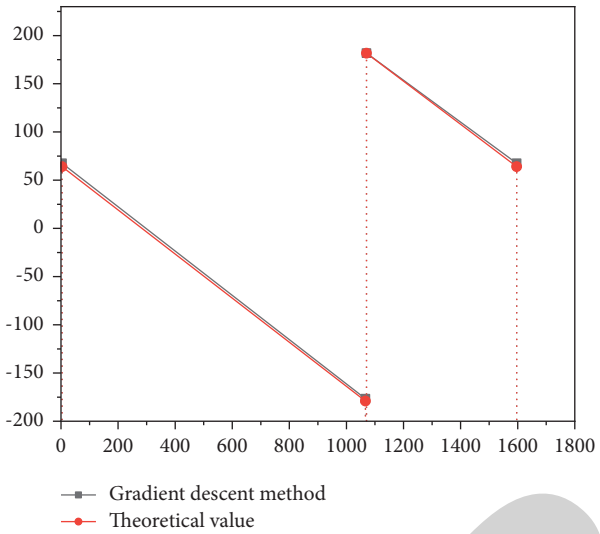

Figure 3: Comparison between calculation results and theoretical values of heading angle based on gradient descent under low dynamic.

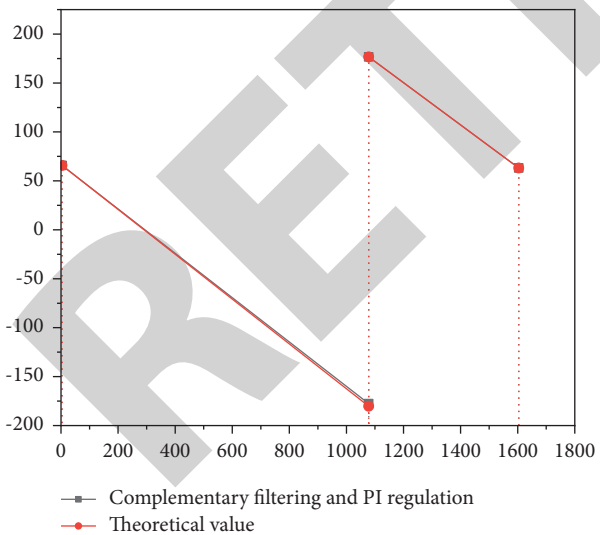

Figure 4: Comparison of course angle calculation results and theoretical values based on complementary filtering and PI adjustment under low dynamic.

are used to collect face and body posture data, respectively. However, the simultaneous acquisition of two cameras will cause the problem of time and space synchronization, resulting in errors in the fusion and solution of subsequent facial data and pose estimation data. Therefore, it is

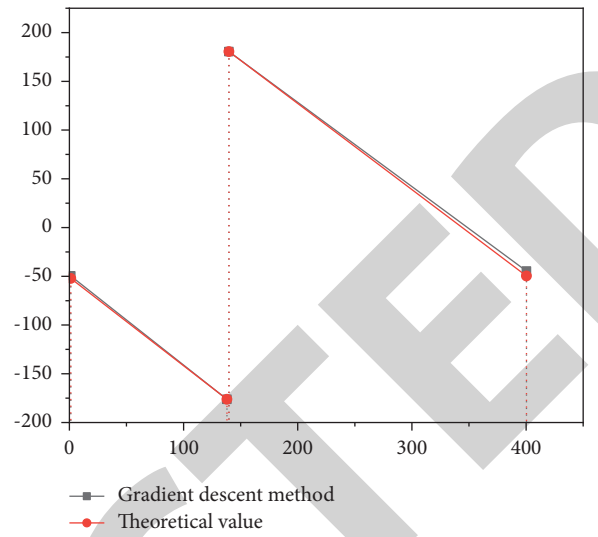

Figure 5: Comparison of course angle calculation results and theoretical values based on gradient descent under high dynamic.

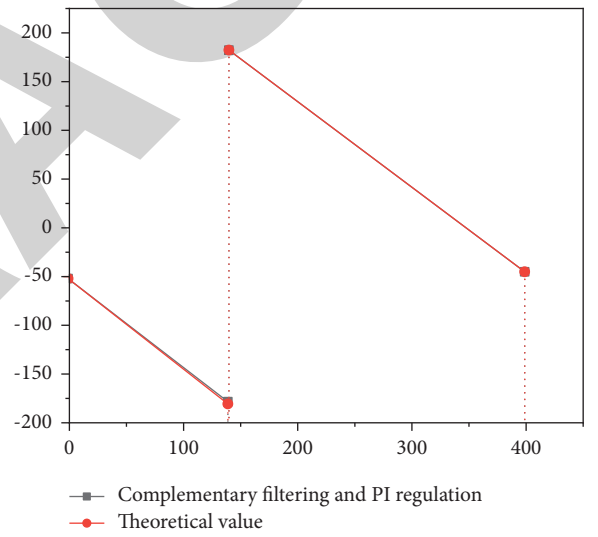

Figure 6: Comparison of course angle calculation results and theoretical values based on complementary filtering and PI adjustment under high dynamic.

Table 2: Attitude calculation error based on gradient descent, complementary filtering, and PI adjustment under high dynamic.

Error	Gradient descent method	Cross filtering and PI regulation
RMSE[0]	1.4387	1.1331
RMSE[ρ]	1.4728	1.1219
RMSE[ψ]	1.6572	1.2515

necessary to synchronize the camera in time and space. In order to ensure that the spatial coordinates of the character data collected by the two cameras are consistent. In order to keep time synchronization, the TCP network timestamp method is used to synchronously fuse facial and body information [21].

The purpose of camera calibration is to correct the camera distortion parameters and correct the image. According to the calibration method, two cameras are used to take pictures of different directions of the same chessboard picture, and their internal parameters, external parameters, and distortion coefficients are obtained. In order to solve the problem of camera time synchronization, this article proposes a synchronization method using TCP network to transmit timestamp, which integrates expression and attitude data synchronously. TCP is a transport layer protocol that provides reliable and orderly data transmission for connection, but there may be packet loss, packet sticking, and so on in the transmission process. In order to solve the similar problems in the sending and receiving process of the program, each data packet is always encapsulated into a fixed length at the sending end according to the transmission data type [22]. According to the processing frame rate, call the sleep method at the sending end to ensure the TCP sending rate, and set the same animation refresh rate at the receiving end to ensure the reception of both parties. After adopting this method, since the transmission after synchronization can reach 20 FPS, even if a small part of data is lost, it can still achieve a more real effect. Figure 7 shows the TCP transmission process.

The facial image, using the TCP network, transmits the data structure including timestamp to the attitude estimation program. Secondly, after encapsulating the data obtained from the pose estimation part, match the timestamp corresponding to the facial expression data, and integrate it into new expression and pose structure data in order. Finally, the fused data are transmitted to the UE4 engine to control the model to generate animation. Because the frame rate is low when the monocular camera collects images to estimate 3D posture, in order to make better use of the data obtained by posture estimation and reduce frame loss, this article uses smoothing and prediction methods to process the facial expression behavior unit data to ensure that the facial expression animation is more smooth, and the face and body behavior in subsequent frames can be synchronized through prediction [23].

4.2. Capture Facial Expressions in Real Time. In order to get a clear and real facial expression animation from the monocular camera, this article extracts 2D feature points from the image, uses 2D feature point regression to obtain the parameters of facial action coding system (FACS) facial action unit (AU), and drives the model to obtain facial expression animation according to the combination of behavior unit parameters and expression baseline.

Extracting effective feature points from the image is the core step of facial expression analysis. The obtained feature points can be used as the basic data of expression parameters. In this article, the constrained local neural domain model is used to detect facial feature points and track the face. The model construction stage is divided into shape model and patch model construction process. When capturing facial expressions, first build CLNF model, generate objective function, and get facial feature points; Then, the appearance features are extracted by directional gradient histogram, and the dimension is reduced by principal component analysis. Finally, support vector machine is used to detect the presence of facial behavior unit Au, and support vector regression is used to estimate Au intensity, so as to obtain the classification and regression effect of expression. Because the average frame rate of facial expression acquisition process is 30 FPS, which is better than the effect of pose estimation, this article uses the Holt two-parameter exponential smoothing method to smooth and predict facial expression parameters and match the data obtained from pose estimation [24]. It is expected to infer the expression state of the next moment by using the prior information of the current character's expression state. Through this method, the facial expression data can be smoother, the trend prediction in a certain range can be obtained, and the detection accuracy of facial expression can be improved. The following equation is the smoothing equation:

$$\begin{cases} y_t = \alpha x_t + (1-\alpha)\left(y_{t-1} + b_{t-1}\right), \\ b_t = \beta\left(y_t - y_{t-1}\right) + (1-\beta) b_{t-1}. \end{cases} \tag{15}$$

The following formula is the prediction model:

$$y_{t+A} = y_t + hb_t. \tag{16}$$

where α is the smoothing parameter, β is the trend smoothing parameter, y_t is the smoothing value at time t, x_t is the actual value, and b_t is the trend value at time t. After smoothing by the Holt algorithm, the sudden change of expression can be avoided and a smoother and natural animation effect can be obtained. The prediction model can reasonably predict the expression state in a range. In order to verify the prediction effect, the average absolute percentage error is used to evaluate the overall prediction result of the continuous time series. The following equation is the calculation formula:

$$\text{MAPE} = \frac{1}{T}\left(\sum_{t=1}^{T} \frac{\left|a_t - \widehat{x}_t\right|}{x_t}\right) \times 100\%. \tag{17}$$

The smaller the error, the better the prediction effect and close to the real value. Otherwise, the effect is poor. Evaluate the 9 values in the expression behavior unit, select 25 data within 1 s, and predict the data of the next 5 frames. The MAPE results are all within 10%, indicating that the prediction results can reflect a certain change trend. Because the camera close to the face can collect the head information more accurately, in order to represent the human head position information, this article uses an efficient N-point projection algorithm to estimate the head pose in the process of facial expression acquisition. Based on the 3D coordinate value of the standard head and the internal parameters of the camera, the pose estimation information of the human head can be obtained. By matching and fusing the head coordinate information, expression behavior unit and pose estimation information, reasonable character pose, and expression driving data can be generated.

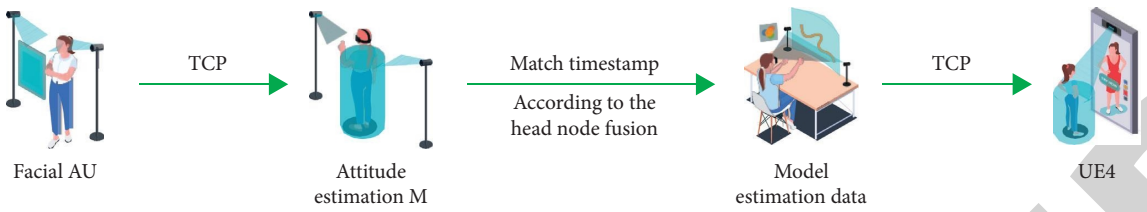

Figure 7: TCP transmission process.

4.3. Real-Time Human Posture Estimation. In order to get clear and real body animation of virtual characters from monocular camera, this article uses the camera to capture human posture, obtain 3D spatial coordinates and rotation angle data of human key points, and drive the body motion of 3D animated characters. The occlusion robust pose map (ORPM) method proposed by Mehta and others is used to realize human pose estimation. The ORPM method is a method to estimate two-dimensional and three-dimensional pose at the same time. For image I obtained by an RGB camera, calculate the human pose in the image, and output the position and rotation angle of each bone point. It can output complete pose estimation data even under strong local occlusion. The collected human body is divided into trunk, neck, head, limb stem, and other human key points. The predicted heat map h and partial affinity field were used to realize the correlation of 2D joint points. Figure 8 shows the network structure of ORPM. Firstly, the ResNet-50 network is used to extract the features of the input image, and then, these features are sent to the two-dimensional attitude estimation network to obtain the heat map h and partial affinity field of each key point. Then, this part of information is sent to the three-dimensional attitude estimation model together with the feature information obtained from the basic network to obtain the three-dimensional attitude estimation results. This part of the results not only includes the heat map of each key point, but also contains the redundant information of pose map, which is used to correct the final estimation result at the end. By adding redundant information, the 3D pose model has good occlusion robustness.

In the process of pose estimation, if there is obvious occlusion of the human body in the image, the trunk and neck nodes are defined as the main nodes to judge the body pose. If the end node has occlusion, the position of its parent node is taken to estimate the node. If the trunk center node is blocked, the basic posture is maintained according to the position of the main node. This can ensure that no matter whether there is occlusion or not, each group of data output contains complete joint points, so as to prevent unnatural animation due to missing part of joint data. In order to prevent sudden change of data jitter, the data shall be smoothed. Specifically, after obtaining the node data, add a smoothing item, which is represented in the following equation:

$$Q = \left\| q_{i-1} - 2q_{i-1} + q \right\|^2, \tag{18}$$

where Q represents the motion of the current frame.

Different human posture can be represented by the position and rotation angle of the model skeleton in space. The bone point data are represented according to the following equation :

$$M_i = (x_i, y_i, z_i, \theta_i, \psi_i, \phi_i), \tag{19}$$

where (x_i, y_i, z_i) is displacement information and $(\theta_i, \psi_i, \phi_i)$ is rotation information. In this way, the human posture estimation data driving the 3D character limb animation is obtained.

4.4. 3D Virtual Human Animation Driver. When generating virtual human body animation, due to the clear and small number of bone points, the corresponding animation can be generated by controlling the relative position of bone points. When generating facial expression animation, a large number of facial points and facial muscles need to be driven to obtain realistic animation effects. If the driving data jitter or deviation is large, the generated expression animation will be funny and exaggerated. Therefore, this article uses the bone animation method to drive the body posture of the model and uses the blendshape model method to drive the facial expression, so as to obtain the overall transition natural and real whole-body animation effect. Based on the obtained posture data, this article drives the virtual character to generate animation in the UE4 engine to simulate human posture. According to the captured body bone points, the mapping relationship between the animation character joint points and the captured data nodes is established in the UE4 character blueprint to determine the position of the joint points. Initially, the coordinate system of the collected body feature points is different from that of the UE4 engine, and each bone also has an independent local coordinate system. If the coordinate system is not converted, the direction of the bone points will be wrong. Therefore, it is necessary to unify the coordinate system before inputting data. In this article, the three-dimensional model, the local coordinate system of feature points, and the world coordinate system of feature points are unified to UE4 world coordinates.

Due to the inherent correlation between the nodes of the human skeleton, it is easy to produce uncoordinated actions by directly specifying the translation and rotation values of each joint. In this article, the physics-based human body simulation methods, forward dynamics, and reverse dynamics are used to obtain a smoother animation effect. Because the height ratio of different collectors is different, giving the position of the end node directly may lead to the penetration of the model and the separation of adjacent

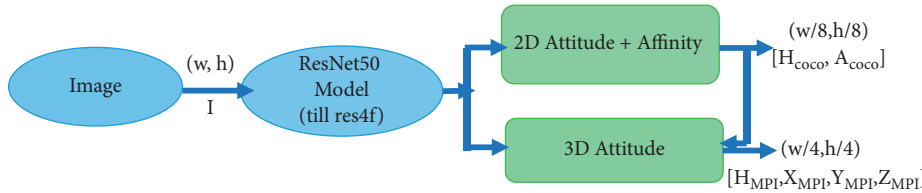

Figure 8: ORPM network structure.

Table 3: Animation frame rate generated by different methods.

Method	Facial expression (Fs)	Attitude estimation (FPS)	Integrated animation (FPS)
Paper method	30	20	20
Traditional method	20	5	—

joints. Therefore, this article uses animated IK to deal with the end nodes of the body of the model (such as wrist and ankle), even if the parent node is pulled by the child node. When the position of the end joint is determined, the information of the middle bone point is calculated according to the position and rotation information of the end bone point. The deflection angles of the corresponding two joints in the bone structure can be determined by the Jacobian matrix. Because the intermediate bone calculated by IK animation has multiple directions, it is necessary to adjust the node target position parameters in its animation blueprint for different bones to ensure the positive direction of the bone. The main support nodes of the body (such as hip and shoulder) use animation FK; that is, the parent node of the bone is used to drive the movement of the child node to determine the accurate position of the core bone relative to the UE4 world coordinate system. The character blueprint transmits the bone data to the corresponding animation blueprint bone node and refreshes the animation according to the sent frame rate to get the body animation driving effect of the model. Table 3 shows the frame rate of animation generated by different methods.

In the process of expression animation driving, first, bind the character's facial bones in 3 ds max, use the Maya design model blendshape expression controller, and import the UE4 engine to generate the corresponding morph targets. Then, the mapping relationship between the controller and the behavior unit parameter Au obtained in the expression acquisition process is established. Finally, use the TCP network to transmit Au data to the UE4 engine, and drive the corresponding expression controller according to the expression weight. Refresh the facial expression of 3D characters according to the transmission frame rate to get the real-time facial animation effect.

5. Conclusion

This article presents a method of using dual cameras to synchronously detect human posture and expression and drive 3D animation. In order to test the effectiveness of the method, the experiments were carried out using two Logitech cameras with 720p resolution, Intel i7-8700k processor, and NVIDIA GeForce GTX 1080 graphics card. First, make the performer perform daily behaviors, including body movements and a variety of facial expressions. After the camera is synchronized, the performer's actions are collected in real time to obtain facial expression and pose estimation data. Secondly, the driving data are processed according to the method proposed in this article, and the three-dimensional model bound with facial expression controller and body bone points is established. Finally, use the TCP network to transmit data and drive the model in the UE4 engine to produce animation effect. In the laboratory environment, the facial expression and body posture are captured at the same time, and the animation is generated. The processing frame rate can reach 20 FPS to generate the animation acceptable to the human naked eye. In order to verify whether the acquisition effect of the two cameras is better than the animation effect obtained by collecting face or posture alone, this article compares and analyzes the animation effect obtained by collecting part of human motion.

Considering that the camera close to the face can more accurately collect facial expression and synchronous posture, this article uses two cameras to collect facial expression and body posture, respectively. Firstly, the Zhang Zhengyou calibration method is used to synchronize the camera in space, and the TCP network timestamp is used to synchronize the camera in time. Then, using dual cameras to collect facial expressions and human posture, respectively, the obtained data of facial behavior units and body bone points are matched and fused, and the world coordinate system is unified. Then, the model including facial expression controller and bound body bones is established, and the mapping relationship between data to controller and bone points is created in UE4. Finally, TCP is used to transmit data and drive the model to generate animation in UE4. In the laboratory scene, human posture detection, expression detection, and animation driving are realized. The results show that this method can generate animation effects in line with the law of human motion and can achieve the frame rate acceptable to users. In general, the method proposed in this article can capture the human facial expression and body posture in real time, and drive the three-dimensional model to generate animation.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Yantai Nanshan University.

References

[1] A. Yiannakides, A. Aristidou, and Y. Chrysanthou, "Real-time 3D human pose and motion reconstruction from monocular RGB videos," *Computer Animations and Virtual Worlds*, vol. 30, pp. 3-4, Article ID e1887, 2019.

[2] M. Nowak and R. Sitnik, "High-detail animation of human body shape and pose from high-resolution 4d scans using iterative closest point and shape maps," *Applied Sciences*, vol. 10, no. 21, p. 7535, 2020.

[3] K. Aberman, R. Wu, D. Lischinski, B. Chen, and D. Cohen-Or, "Learning character-agnostic motion for motion retargeting in 2d," *ACM Transactions on Graphics*, vol. 38, no. 4, pp. 1–75, 2019.

[4] J. Sun, M. Wang, X. Zhao, and D. Zhang, "Multi-view pose generator based on deep learning for monocular 3d human pose estimation," *Symmetry*, vol. 12, no. 7, p. 1116, 2020.

[5] C. R. D. Souza, A. Gaidon, Y. Cabon, N. Murray, and A. M. López, "Generating human action videos by coupling 3d game engines and probabilistic graphical models," *International Journal of Computer Vision*, vol. 128, no. 5, pp. 1505–1536, 2020.

[6] J. C. Núñez, R. Cabido, J. F. Vélez, A. S. Montemayor, and J. J. Pantrigo, "Multiview 3d human pose estimation using improved least-squares and lstm networks," *Neurocomputing*, vol. 323, pp. 335–343, 2019.

[7] L. Zhou, Y. Chen, J. Wang, and H. Lu, "Progressive bi-c3d pose grammar for human pose estimation," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 07, pp. 13033–13040, 2020.

[8] J. Purnama and R. F. Sari, "Unobtrusive academic emotion recognition based on facial expression using rgb-d camera using adaptive-network-based fuzzy inference system (anfis)," *International Journal of Software Science and Computational Intelligence*, vol. 11, no. 1, pp. 1–15, 2019.

[9] J. Li, Y. Mi, G. Li, and Z. Ju, "CNN-based facial expression recognition from annotated RGB-D images for human-robot interaction," *International Journal of Humanoid Robotics*, vol. 16, no. 04, Article ID 1941002, 2019.

[10] L. He, D. Jiang, and H. Sahli, "Automatic depression analysis using dynamic facial appearance descriptor and dirichlet process Fisher encoding," *IEEE Transactions on Multimedia*, vol. 21, no. 6, pp. 1476–1486, 2019.

[11] T. SapińskiSapiński, D. KamińskaKamińska, A. PelikantPelikant, and G. Anbarjafari, "Emotion recognition from skeletal movements," *Entropy*, vol. 21, no. 7, p. 646, 2019.

[12] Y. Wang, Y. Li, Y. Song, and X. Rong, "Facial expression recognition based on auxiliary models," *Algorithms*, vol. 12, no. 11, p. 227, 2019.

[13] K. Mohan, A. Seal, O. Krejcar, and A. Yazidi, "Fer-net: facial expression recognition using deep neural net," *Neural Computing & Applications*, vol. 33, no. 15, pp. 9125–9136, 2021.

[14] N. Taubert, M. Stettler, L. Sting, R. Siebert, S. Spadacenta, and M. A. Giese, "Cross-species differences in the perception of dynamic facial expressions," *Journal of Vision*, vol. 19, no. 10, p. 155, 2019.

[15] D. Sen, S. Datta, and R. Balasubramanian, "Facial emotion classification using concatenated geometric and textural features," *Multimedia Tools and Applications*, vol. 78, no. 8, pp. 10287–10323, 2019.

[16] A. Zinkernagel, R. W. Alexandrowicz, T. Lischetzke, and M. Schmitt, "The blenderface method: video-based measurement of raw movement data during facial expressions of emotion using open-source software," *Behavior Research Methods*, vol. 51, no. 2, pp. 747–768, 2019.

[17] S. P. Yadav, "Emotion recognition model based on facial expressions," *Multimedia Tools and Applications*, vol. 80, no. 6, pp. 1–23, 2021.

[18] R. J. N. Stopyn, T. Hadjistavropoulos, and J. Loucks, "An eye tracking investigation of pain decoding based on older and younger adults' facial expressions," *Journal of Nonverbal Behavior*, vol. 45, no. 1, pp. 31–52, 2021.

[19] J. Tang, W. Zhu, and Y. Bi, "A computer vision-based navigation and localization method for station-moving aircraft transport platform with dual cameras," *Sensors*, vol. 20, no. 1, p. 279, 2020.

[20] Q. Zhang, "Relay vibration protection simulation experimental platform based on signal reconstruction of MATLAB software," *Nonlinear Engineering*, vol. 10, no. 1, pp. 461–468, 2021.

[21] R. Huang, S. Zhang, W. Zhang, and X. Yang, "Progress of zinc oxide-based nanocomposites in the textile industry," *IET Collaborative Intelligent Manufacturing*, vol. 3, no. 3, pp. 281–289, 2021.

[22] X. Liu, J. Liu, J. Chen, and F. Zhong, "Degradation of benzene, toluene, and xylene with high gaseous hourly space velocity by double dielectric barrier discharge combined with Mn3O4/activated carbon fibers," *Journal of Physics D: Applied Physics*, vol. 55, no. 12, Article ID 125206, 2022.

[23] P. Ajay, B. Nagaraj, B. M. Pillai, J. Suthakorn, and M. Bradha, "Intelligent ecofriendly transport management system based on iot in urban areas," *Environment, Development and Sustainability*, no. 3, pp. 1–8, 2022.

[24] D. Kumar, A. Sharma, R. Kumar, and N. Sharma, "Restoration of the Network for Next Generation (5G) Optical Communication Network," in *Proceedings of the 2019 International Conference on Signal Processing and Communication (ICSC)*, March 2019.