


Research Article

Group Animation Motion Capture Method Based on Virtual Reality Technology

Songlin Zhang¹ and Wei Xun^{1,2} 

¹*Yibin Vocational and Technical College, Yibin 644003, China*

²*Silliman University, Dumaguete, Philippines*

Correspondence should be addressed to Wei Xun; xunwei@su.edu.ph

Received 30 March 2022; Accepted 12 May 2022; Published 8 June 2022

Academic Editor: Qiangyi Li

Copyright © 2022 Songlin Zhang and Wei Xun. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the group animation motion capture effect and enhance the design technology of group animation, this paper studies the method of group animation motion capture combined with virtual reality technology. This paper constructs a crowd motion capture system based on virtual reality technology and describes the attributes by organizing crowd profiles of different levels and structures. Moreover, this paper conducts collaborative modeling of multiple attributes and constructs a model based on structured interactive attributes. The strength of the model is that it takes into account scene scale, chaos, and crowding properties to effectively characterize the potential interactions of group movement patterns. Combined with the simulation research, it can be seen that the group animation motion capture method based on virtual reality technology proposed in this paper can play an important role in group animation motion capture and animation design.

1. Introduction

Technically speaking, there are four main steps in 3D animation production: geometric modeling, material adjustment, motion trajectory setting, and coloring technology. At present, the main problem of geometric modeling is the reproduction of the model in the virtual world. The main problem of material finishing is how to make virtual material feel like material in reality [1]. The main problem with motion trajectories is the movement of characters, especially the limb movements of humans and animals. In addition, the main problem of shading and rendering technology is how to make corresponding static or dynamic pictures according to the setting of the scene. In the process of 3D animation production, the most difficult problem is the movement trajectory, the adjustment of character movement, and the coordination of limbs. Although we often see domineering scenes in ordinary animations, for a good animation work, the movement of the characters is the object of attention [2]. Like film and television works, even though the special effects of the scenes are well done, the

acting skills of the actors still come first. If there is no good acting and lack of realism, then this must not be a favorite movie. In the production of 3D animation, the movement trajectory of the character is equivalent to the acting skills of the actors in the film and television performance. The set movement of the character is usually adjusted by the animator, but for a long animation, it is very difficult to just rely on the animator to manually adjust the movement of the character. Therefore, the setting and adjustment of the character's motion trajectory have become the most difficult bottleneck in the process of making 3D animation [3].

The animation system of traditional film and television performances is based on the graphics processing principle of computer technology, through many video capture devices, the movement of things is recorded in the form of image recording, and then the image information obtained by the video capture device is captured by computer graphics processing technology. To be processed, image capture technology mainly includes acoustic capture, optical capture, mechanical capture, and electromagnetic capture. Optical motion capture system is the technology used in

most performance animations. The general process of its work is as follows. Write and then draw the sub-shot script, design the atmosphere map and a different character modeling, and make the model required for the animation according to the requirements of the script. Usually in order to facilitate later video editing and animation production, performers need to wear green clothes and then install light-emitting points on their key body parts, such as body joints, wrists, and elbows. The settings of such light-emitting points are conducive to the recognition of the visual system. Then, by means of camera cooperation, the video of the performer is captured from all angles, and then according to the mark of each frame in the image obtained by the camera, the movement of the performer is recorded, so that many points of movement according to time can be obtained. Then, according to the three-dimensional technology, the motion trajectories of these points are made into the movements of the skeleton model, so that the movements of the performer can be transformed into the movements of these points.

This paper studies the group animation motion capture method combined with virtual reality technology, constructs a group motion capture system based on virtual reality technology, and improves the effect of subsequent group animation motion capture.

2. Related Work

Since optical motion capture technology is more sensitive to light, it is easy to be interfered by external light in the actual application process, which leads to the phenomenon that the overall motion capture effect is reduced. However, when the motion capture work is performed in the studio, due to the large difference between the scenery environment in the studio and the actual environment, and the lack of overall literature interaction, it makes the performance of the overall observed object more difficult, and it is necessary to rely on the director's guidance. As well as the imagination of the actors, this makes the entire motion capture more difficult, and relatively more equipment is required in the process of optical motion capture. If the character cannot quickly enter the state, it will make the time for motion capture longer, resulting in serious resource consumption [4]. With the continuous improvement of the level of optical motion capture technology, in order to better shoot outdoors and reduce the impact of external light factors on the optical motion capture technology, corresponding researchers have gradually improved the tracker installed on the observed object. The light source receiving method is changed to the method of actively emitting light source, so the influence of external optical factors on the optical motion capture technology can be effectively avoided [5].

The use of virtual shooting technology in the process of optical motion capture can enable the director to more intuitively reflect the operation status of the current scene, so as to help the director to better guide the live action actors, so that the corresponding observed actors can express their performance. With the expressions and actions that the animated objects in the scene need to perform, the quality of the overall animation production is improved [6]. In the process

of using optical motion capture technology for motion capture, virtual shooting technology is used to simulate different scenes in the film and television animation, so that the corresponding shooting team can more intuitively feel the production results of the corresponding film and television animation, and can quickly analyze the results. If there are deficiencies in the beginning process, make the director communicate with the actors and the postproduction team in time, adjust the corresponding actions and performances, reduce the phenomenon that needs to be changed after the film and television animation production is completed, reduce the cost of film and television animation production, and ensure the film and television animation production cost [7].

Use three cameras to capture the changes of 102 landmark points to achieve voice expression animation [8]. Literature [9] uses two camcorders, combined frame receivers to transmit data to the computer, and uses direct linear transformation to process the captured point information into available information. Literature [10] uses marked points of different colors to capture faces at different angles, after obtaining data, recognizes and trains templates according to different color points, and uses image processing technology to calculate 3D point information. Literature [11] proposed a method to obtain three-dimensional information with only one camera, using two mirrors to reflect the faces covered with fluorescent markers. When shooting, irradiate the face with a violet light, the reflection effect of the fluorescent marker points is obvious, and the image has a high contrast, which is convenient for tracking. Then, use the principle of space geometry to calculate the three-dimensional information of the point, and obtain the initial captured data. Literature [3] smears colored pigments on the face, the blue marker points are used to track the overall expression changes, and the colored stripes in other regions are used to track the changes of detailed wrinkles. Literature [12] proposed to achieve real face geometry reconstruction by projecting different types of textures.

Literature [13] uses somatosensory peripherals as their data capture tools, which can simultaneously capture two-dimensional images and obtain three-dimensional position information, which has the advantages of high speed and synchronization, but the noise and error of the three-dimensional information it acquires are relatively large, requiring software methods for denoising and correction. Literature [14] uses the trained facial expression model and uses the matching of the model to limit the influence of noise. The advantage of this is that it has good real-time performance, but the expression changes of characters are somewhat monotonous, lacking variability and degrees of freedom. Literature [15] uses structured light to illuminate a face without marked points, calculates the depth difference between two image sequences, and matches the depth with the face template to drive the model change. This method has higher requirements on lighting and is more cumbersome to use in general. Literature [16] uses five synchronous cameras to shoot faces without markers, supplemented by two-dimensional grid tracking to establish the connection between

frames to achieve facial expression animation reproduction. It consumes more time.

3. Group Motion Capture Algorithm Based on Virtual Reality Technology

The proposed method first utilizes a nontracking method to obtain group motion information of dense scenes. The ideal situation for analyzing group behavior is to keep track of all individual goals. However, there are still a lot of occlusions in dense scenes, making it difficult to track all individual targets. Therefore, we use particle flow to approximate crowd trajectories to overcome the difficulty of tracking animated pedestrians. This method regards dense pedestrians as a particle country and uses the advective motion of particles to propagate in the optical flow field to approximate the motion of pedestrians and capture the continuous group motion. Its effect is much better than the motion information generated by the pure optical flow representation.

A video of a group scene is given, and a series of video frames are first divided into video blocks of size $T \times W \times H$ that can be represented:

$$\{(U_w^t, V_h^t) | w \in [1, W], h \in [1, H], t \in [1, T-1]\}. \quad (1)$$

The particles are evenly placed in the optical flow field according to a certain step size, and the average optical flow in the space-time cube representing the surrounding fixed area is obtained by calculation. Each particle moves with the average optical flow of its covering points, and the corresponding particle trajectory is accumulated by the fourth-order Runge-Kutta numerical calculation algorithm. The particle flow forms a trajectory as follows:

$$\begin{aligned} X_w^{t+1} &= X_w^t + U_w^t, \\ Y_h^{t+1} &= Y_h^t + V_h^t. \end{aligned} \quad (2)$$

Particles follow the fluid motion to generate trajectories guided by the average neighbors. The particle trajectory $P_{t_1}^{t_T}(t)$ includes T-tuples in the optical flow field as follows:

$$P_{t_1}^{t_T}(t) = (s_i, v_i, t_i)_{t_i=t_1}^{t_T}. \quad (3)$$

Among them, s and v represent the position vector of particle i obtained from the optical flow field at (X_w^{t+1}, Y_h^{t+1}) and the velocity vector obtained by particle i at time r and displacement (U_w^t, V_h^t) .

This clustering method can obtain reliable clustering patterns determined by particle density (unlike the K-means method), and it is robust to different trajectory types. In particular, c samples $x_i, i = 1, \dots, c$ are given in the T-dimensional space Λ [17]:

$$\hat{f}(x) = \frac{c_{K,h}}{c} \sum_{i=1}^c K\left(\frac{d^2(x, x_i)}{h}\right). \quad (4)$$

The algorithm randomly selects a point x_j and drifts to $m_h(x)$ with the highest probability density at the current scale by computing the mean vector x_{j+1} , as follows:

$$\begin{aligned} m_h(x) &= c'_{K,h} \frac{\nabla f(x)}{f(x)} \\ &= \frac{\sum_{i=1}^c x_i Q(x_i)}{\sum_{i=1}^c Q(x_i)} - x\lambda. \end{aligned} \quad (5)$$

Among them, $c'_{K,h}$ is the inference coefficient obtained using the gradient descent algorithm. The next point x_{j+1} is drifted from x_j as follows:

$$x_{j+1} = x_j + m_h(x_j), \quad j = 1, \dots, c. \quad (6)$$

The graph structure is as follows: in order to express the graph structure of the group, we use the spectral information to reflect the structural properties of the graph, because the Laplacian spectrum achieves good results in recognition and classification problems. We assume that there are N graphs in T-frame video clips containing m trajectories. For each graph $G_k = (V_k, E_k, W_k)$, the Laplacian matrix can be defined by F as [18]

$$A_k = \begin{cases} d_j, & \text{if } i \neq j \\ \sum_{j=1}^m d_{ij}, & \text{-otherwise} \end{cases}. \quad (7)$$

Among them, $i, j = 1, 2, \dots, m, k = 1, 2, \dots, N, d_{ij} = W_k$ represents the distance between different vertices. The eigenvalue G_k of the Laplace matrix can be obtained by the singular value decomposition (SVD) method.

$$A_k = U\Delta U^T. \quad (8)$$

Among them, $\Delta = \text{diag}\{\tau_1, \tau_2, \dots, \tau_m\}$, $\tau_1 \geq \tau_2 \geq \dots \geq \tau_m \geq 0$. We select the largest 3 eigenvalues as graph structure representation features, which can discriminatively distinguish diverse graph structure patterns.

The group attribute is as follows: the group attribute is used to express the characteristics of the group, including the orientation distribution and the velocity distribution. In each trajectory graph $G_k = (V_k, E_k, W_k)$, for each node $G_k = (V_k, E_k, W_k)$, we calculate the basic velocity $S(\cdot)$ and orientation $O(\cdot)$ feature channel as follows:

$$\begin{aligned} S(v_k) &= \sqrt{(x_T - x_1)^2 + (y_T - y_1)^2} \\ O(v_k) &= \arctan\left(\frac{x_T - x_1}{y_T - y_1}\right). \end{aligned} \quad (9)$$

Among them, $s_i = (x_i, y_i)$, $i \in [1, T]$ is the position of the graph node trajectory v . Attributes can be viewed as quantitative measures to represent attributes of group behavior. We use a histogram containing n bins ($n=8$) of uniform orientation and velocity to define the group properties as follows:

$$\begin{aligned} H_{\text{ori}} &= \{O_i\}_{i=1}^n \\ H_{\text{spd}} &= \{S_i\}_{i=1}^n. \end{aligned} \quad (10)$$

Movement dynamics are combined with the internal attributes of the group, and the movement information outside

the group also needs to be considered to describe the group. For each trajectory graph G_j , we choose the highest three $S(v_k)$ as the trajectory graph velocity and consider the average node position as the position of the graph to record the dynamic motion.

These features record the motion information of the structure and trajectory graphs, and effectively express typical group behavior patterns. Based on this, all features can be expressed as a unified 24-dimensional vector (concatenated by 3+8+8+3+2-dimensional vectors), which describes the group-level structure and apparent patterns. Next, a bag-of-words model is constructed to quantify the trajectory graph patterns.

The dictionary construction of trajectory graphs is inspired by visual words to represent local patterns of images, and trajectory graphs represent group behavior patterns for specific video sequences, which can be applied to group recognition tasks. The concatenated feature vectors are clustered using the K-means method to build a dictionary of trajectory graph words. The trajectory graph word bag model BoTG represents the group behavior pattern through a histogram vector h_j , as shown below [19]:

$$\begin{aligned} h_j &= (f_1, f_2, \dots, f_i, \dots, f_d)^T, \\ f_i &= \frac{n_{ij}}{n_j}. \end{aligned} \quad (11)$$

Among them, d is the number of words selected as a dictionary. n_{ij} is the frequency with which the i -th trajectory map appears in the j -th video clip. The vector h_j contains the pattern distribution information of a particular group scene, which is normalized by the number of all images in the j -th video segment. Therefore, each T-frame video can be represented as a trajectory graph word packet BoTG. According to this, BoTG captures informative group information by maintaining the symbiotic pattern. We extracted BoTG from swarm video clips to train an SVM model to recognize different event types. BoTG can serve as an efficient behavioral representation of global groups with symbiotic structural coherence.

We assume that there are N types of attributes to describe each group video, and there are M group videos in total. First, a graph structure is constructed separately for all attributes to measure the video relationship. The graph is defined using W_i , as the similarity matrix, where W is denoted as the similarity corresponding to the n -th attribute, the i -th and the j -th video among all N attributes. For the n -th class of attributes ($n = 1, \dots, 7$), each subgraph can be constructed as follows:

$$\begin{aligned} W_{n,ij} &= \exp\left(-\frac{d(x_{n,i}, x_{n,j})}{\sigma_n}\right), \\ W_{n,ij} &= \exp\left(-(x_{n,i} - x_{n,j})^T M(x_{n,i} - x_{n,j})\right), \\ W_{n,ij} &= \exp\left(-\|A_n(x_{n,i} - x_{n,j})\|^2\right). \end{aligned} \quad (12)$$

A feature x_n can be mapped to $A_n x_n$ in the original feature space. This mapping A_n can be a linear or nonlinear mapping. Inspired by deep autoencoders, we expect that both the group data relationship and the underlying manifold structure of the data distribution can be better learned by nonlinear transformations in deep models. Therefore, we next add the properties of deep embedding to the graph sorting framework.

Overall, graph-based ranking methods can be formalized as the following regularization framework:

$$\operatorname{argmin} \mathfrak{F}(f) = R(f, x, W) + \mu(f, Y). \quad (13)$$

In particular, the label Y , the attribute vector x_n (for the n -th class attribute of the video), and the similarity matrix W_n for the n -th class attribute are set, and our goal is to find the relevance ranking score f . The specific form is as follows [20]:

$$\begin{aligned} \mathfrak{F}(f) &= \sum_{n=1}^N \alpha_n \left(\sum_{i,j} W_{n,ij} \left| \frac{f_i}{\sqrt{S_{n,ii}}} - \frac{f_j}{\sqrt{S_{n,jj}}} \right| + \mu_n \sum_i |f_i - Y_i|^2 \right) \\ f &= \operatorname{argmin}_f \mathfrak{F}(f), \quad \text{s.t. } \alpha_n \geq 0, \sum_{n=1}^N \alpha_n = 1. \end{aligned} \quad (14)$$

Among them, $\alpha = [\alpha_1, \dots, \alpha_N]$ represents the weights of different graphs, and D_n is a diagonal matrix whose $D_{n,ij} = \sum_j W_{n,ij}$. In order to better integrate different properties and improve computational efficiency, α can be regarded as a variable, and formula (5) should be able to be optimized simultaneously.

$$\begin{aligned} &\operatorname{argmin}_{f, \alpha\{A_1, A_2, \dots, A_N\}} \mathfrak{F}(f, \alpha\{A_1, A_2, \dots, A_N\}), \\ &= \sum \alpha_n^\lambda \left(\sum_{i,j} W_{n,ij} \left| \frac{f_i}{\sqrt{S_{n,ii}}} - \frac{f_j}{\sqrt{S_{n,jj}}} \right| + \mu_n \sum_i |f_i - Y_i|^2 \right), \\ &\text{s.t. } \sum_{n=1}^N \alpha_n = 1, W_{n,ij} = \exp\left(-\|A_n(x_{n,i} - x_{n,j})\|^2\right). \end{aligned} \quad (15)$$

The system parameters can be derived through the structure of the deep network and the multilevel nonlinear transformation abstraction, and the depth transformation metric A_n can be derived when f and α_n are fixed. The deep structure of the stacked autoencoder (SDAE) can autonomously abstract higher-level semantic information through a series of nonlinear reconstruction transformations. In particular, each layer of the stacked autoencoder (SDAE) is a hidden layer representation generated by data training, which is equivalent to a higher-level abstract response.

A linear transformation function and a continuous nonlinear transformation function both transform x through h into r neurons.

$$\begin{aligned} z &= h_\theta(x), \\ &= s(Wx + b). \end{aligned} \quad (16)$$

The representation of the manifold structure present in the data is introduced by such a decoder reconstruction form to describe the manifold distance relationship between the data. The total reconstruction error is thus defined as follows:

$$\operatorname{argmin}_\theta \sum 1/\Delta \|g_\theta(h_\theta(x_i)) - x_j\|_2^2, \quad x_j \in \Delta. \quad (17)$$

Afterward, x can be updated by implicitly representing $h(x)$, resulting in the $(t+1)$ -th iteration. The hidden layer can embed semantic attributes through each iteration as a depth layer, which can well explain the relationship between the input data. Furthermore, the transformation metric matrix can gradually approach the inherent manifold structure of the population data, thus promoting the fusion of diverse and heterogeneous properties of population patterns.

The maximization (M) step is to optimize f and α .

It is worth noting that by the time A is fixed, the main function is a convex function with respect to \mathfrak{F} and α^λ . Formula (19) for f and α_n can likewise be solved by iteratively minimizing $\mathfrak{F}(f, \alpha_n)$. When $W_{nij} = \exp(-\|A_n(x_{ni} - x_{nj})\|^2)$ is fixed, the partial derivative of L can be obtained as follows:

$$\begin{aligned} \frac{\partial \mathfrak{F}(f, \alpha_n)}{\partial f} &= 2 \sum_{n=1}^N \alpha_n^\lambda (L_n f + \mu_n (f - Y)), \\ \frac{\partial \mathfrak{F}(f, \alpha_n)}{\partial f} &= \lambda \alpha_n^{\lambda-1} (f^T L_n f + \mu_n (f - Y)^2). \end{aligned} \quad (18)$$

The specific details of the iterative algorithm process and the corresponding convergence proof can be referred to. When f is fixed, we can derive formula (18) to obtain α_n , which has the constraint $\sum_{n=1}^N \alpha_n = 1$, as follows:

$$\alpha_n = \frac{(1/f^T L_n f + \mu_n |f - Y|^2)^{1/\lambda-1}}{\sum_{n=1}^N (1/f^T L_n f + \mu_n |f - Y|^2)^{1/\lambda-1}}. \quad (19)$$

If α_n is fixed, f can be solved as follows:

$$\begin{aligned} f &= \left(I + \frac{\sum_{n=1}^N \alpha_n^\lambda L_n}{\sum_{n=1}^N \alpha_n^\lambda \mu_n} \right) Y, \\ Y_i &= \begin{cases} 1, & \text{If the video I will query the video} \\ 0, & \text{otherwise} \end{cases}. \end{aligned} \quad (20)$$

It is worth mentioning that we ended up adopting 7 kinds of attributes, so we constructed 7 separate graphs.

The entire iterative optimization process can be found in Algorithm formula (19). At the same time, our method can

easily extend to a wider variety of properties and incorporate more semantic structural information.

4. Group Animation Motion Capture Method Based on Virtual Reality Technology

The design of the group animation capture system is shown in Figure 1. The key technology of motion capture is marker tracking and three-dimensional reconstruction of spatial coordinates. In addition, in computer vision, it is necessary to use the position information of the viewpoint and the orientation information of the viewpoint to calculate the three-dimensional spatial structure from the two-dimensional image information. This uses various parameters of the camera. The relationship between the three-dimensional geometric position of a point on the surface of the space object and its corresponding point in the image is determined by the geometric model of the camera's imaging. These geometric model parameters are camera parameters, which can only be obtained through camera calibration calculation. Therefore, camera calibration is also one of the key technologies of motion capture.

Video capture cards offer two methods of storing data into memory. In this paper, the method of double buffering with better performance is selected, and two buffers are opened in the memory; one is used to capture images. Another block is used to analyze the image data that has been acquired. In this way, the collected image data can also be analyzed and processed while being captured. After completing the acquisition of one frame of image, switch the two buffers to the entire video capture module, as shown in Figure 2. A video capture thread is opened for each capture card, which is responsible for its image capture and analysis. In addition, a calculation thread is opened, which uses the image coordinates of the analyzed marker points to calculate its spatial position.

Motion capture is a technology that captures and records human movement. The skeleton captured by this technique contains some typical joint points. As shown in Figure 3, the skeleton contains 21 joint points, and the root node contains 6-dimensional data, which are the three-dimensional translation and three-dimensional rotation information. Motion capture data can be represented by a matrix, in which each row represents a frame and each column represents a dimension.

In the early stage of designing the system, the requirements of the system functions are firstly analyzed. This section will focus on expounding the function and specific implementation process of each module of the animation synthesis prototype system based on 3D motion capture data key frames. The functional modules of the prototype system are shown in Figure 4.

This prototype system is mainly composed of four main modules: Bvh loading, player, key frame visualization, and animation synthesis. The system frame structure diagram is shown in Figure 5.

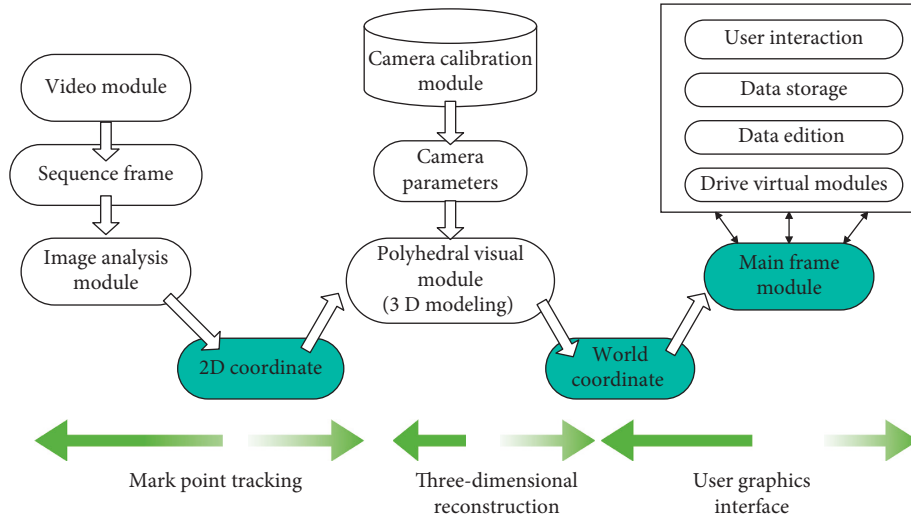


FIGURE 1: Group animation motion capture model.

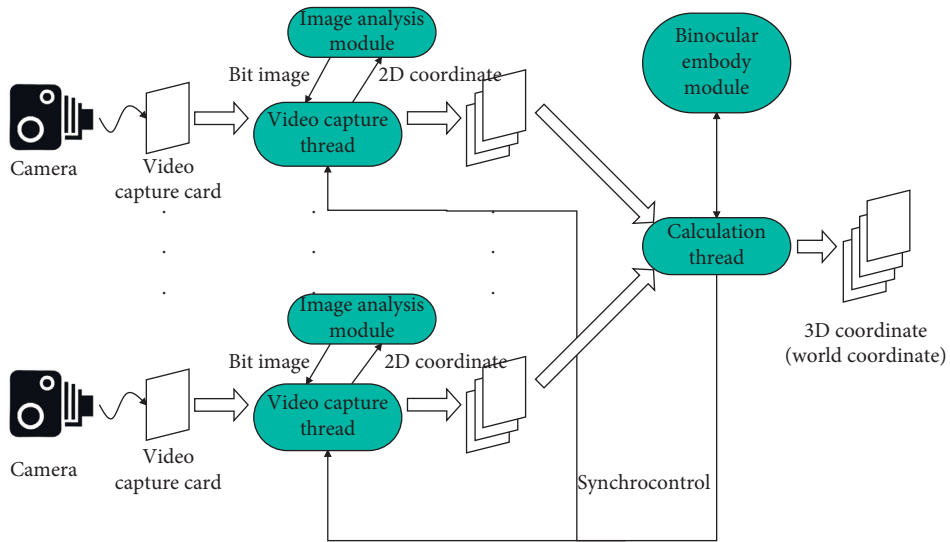


FIGURE 2: Video capture module.

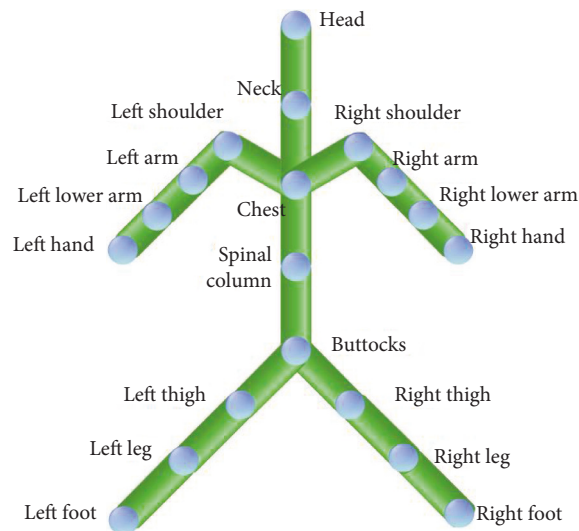


FIGURE 3: An example of a human frame for motion capture.

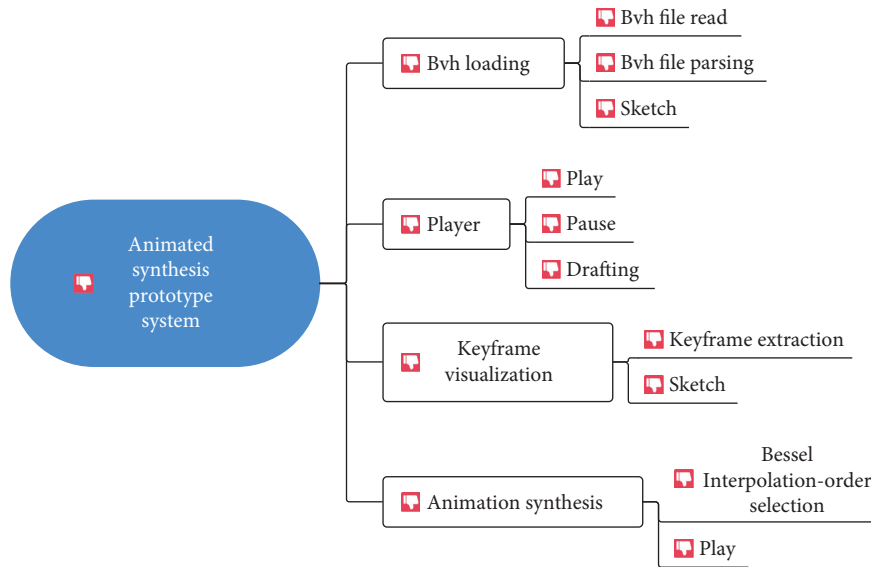


FIGURE 4: Functional structure diagram of animation synthesis prototype system.

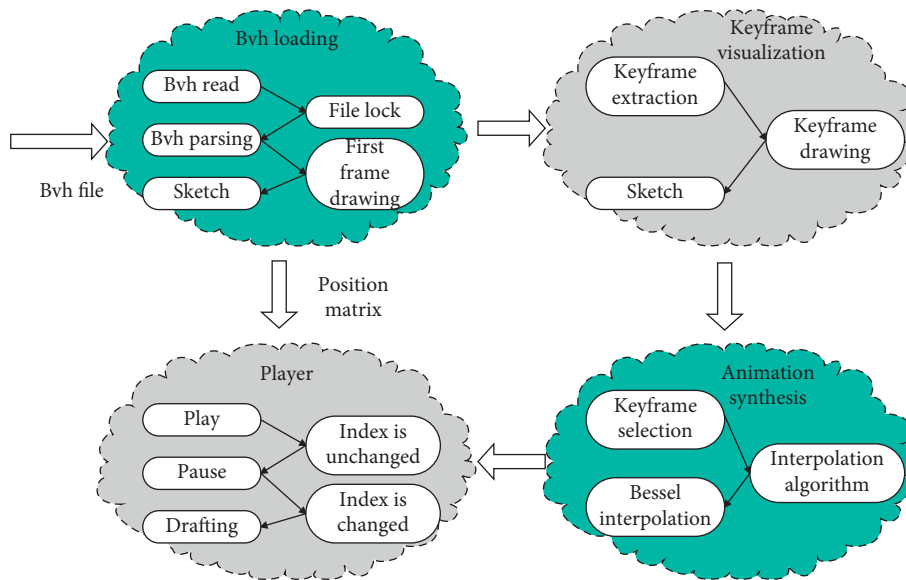


FIGURE 5: Functional framework of animation synthesis prototype system.

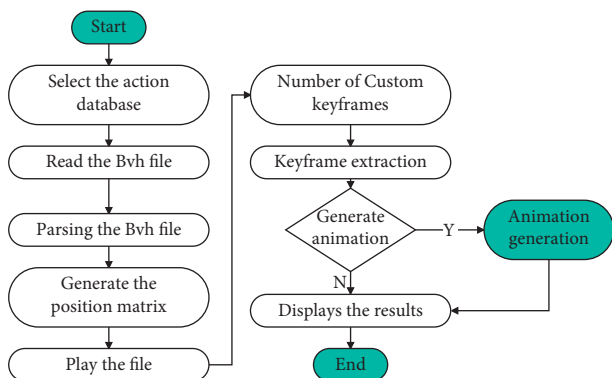


FIGURE 6: Flowchart of animation synthesis prototype system.

In order to verify whether the animation effect is good, the animation is synthesized on the basis of the extracted key frames, and the synthesized animation effect is displayed. The specific operation process is shown in Figure 6.

Figure 7 shows the group animation image designed by the group animation capture method proposed in this paper.

The effect of the group animation motion capture method based on virtual reality technology proposed in this paper is evaluated, and the motion capture effect and animation design effect are counted. The statistical test results are shown in Tables 1 and 2.

It can be seen from the above research that the group animation motion capture method based on virtual reality technology proposed in this paper can play an important

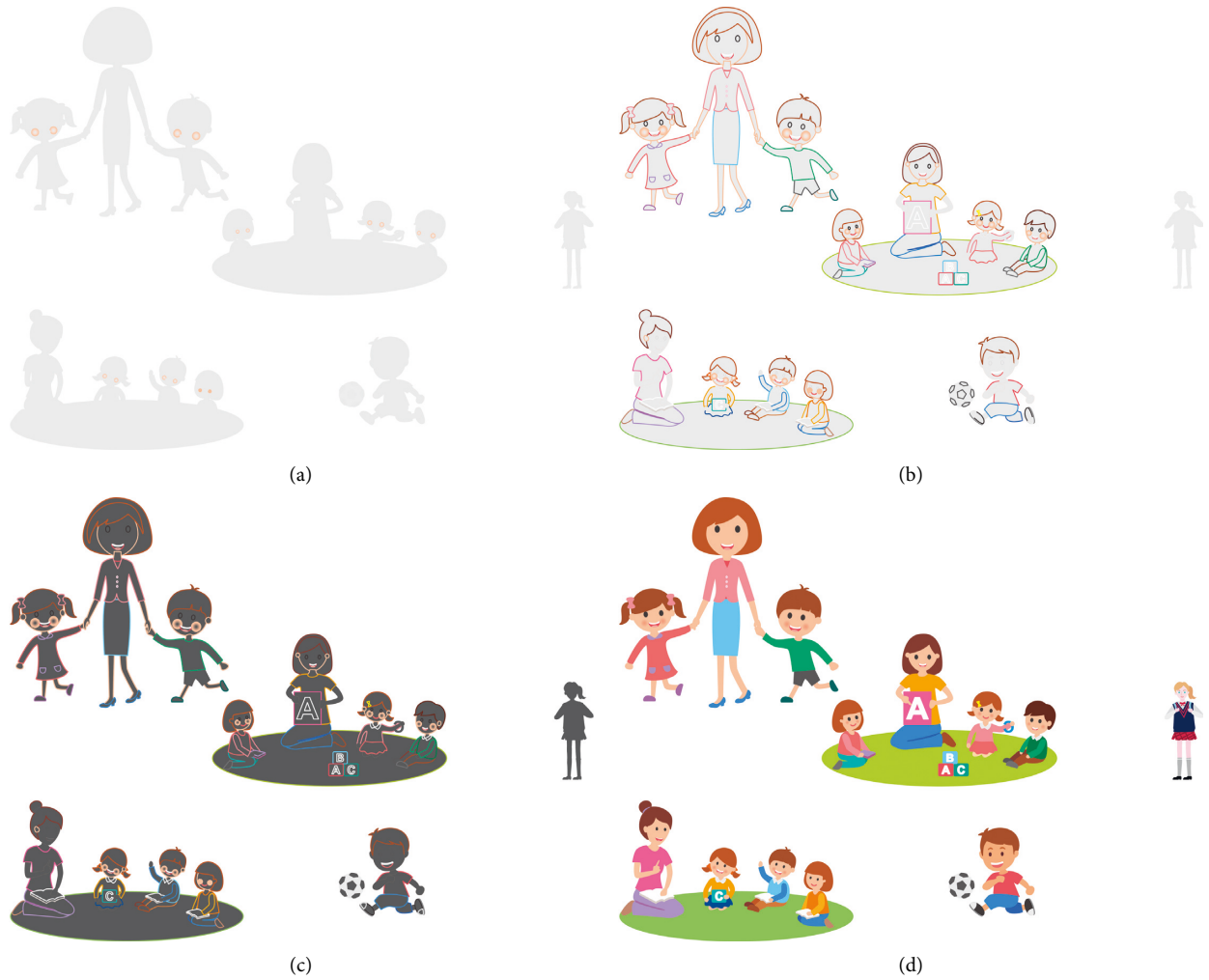


FIGURE 7: The group animation image designed by the group animation capture method: (a) shadow capture, (b) human feature capture, (c) preliminary background color treatment, and (d) color rendering.

TABLE 1: Motion capture effect of group animation motion capture method based on virtual reality technology.

Number	Motion capture	Number	Motion capture	Number	Motion capture	Number	Motion capture
1	81.56	17	81.30	33	81.45	49	79.32
2	82.22	18	79.49	34	90.27	50	79.94
3	87.45	19	85.81	35	89.95	51	90.85
4	79.85	20	84.39	36	89.44	52	85.53
5	80.80	21	86.78	37	80.52	53	85.04
6	88.24	22	90.43	38	84.62	54	81.22
7	81.61	23	86.35	39	80.71	55	83.69
8	81.80	24	83.78	40	87.56	56	82.42
9	83.74	25	84.98	41	88.08	57	85.36
10	84.32	26	80.19	42	86.68	58	86.24
11	85.43	27	84.91	43	89.28	59	89.44
12	86.56	28	85.99	44	89.44	60	81.50
13	82.29	29	84.14	45	79.90	61	85.55
14	84.38	30	86.39	46	80.96	62	81.11
15	81.80	31	83.04	47	88.91	63	89.16
16	90.23	32	80.75	48	85.32	64	89.99

TABLE 2: Animation design effect of group animation motion capture method based on virtual reality technology.

Number	Animation design	Number	Animation design	Number	Animation design	Number	Animation design
1	77.10	17	77.67	33	76.17	49	80.89
2	81.08	18	84.58	34	83.45	50	80.98
3	78.32	19	83.12	35	77.64	51	81.62
4	82.08	20	79.70	36	76.58	52	76.96
5	83.27	21	79.02	37	86.58	53	80.75
6	86.37	22	78.94	38	78.85	54	85.88
7	84.23	23	82.86	39	85.35	55	83.26
8	78.76	24	86.40	40	77.26	56	79.68
9	76.34	25	77.06	41	82.33	57	79.50
10	84.14	26	83.74	42	83.08	58	83.34
11	83.76	27	82.03	43	81.87	59	85.02
12	83.22	28	82.10	44	82.68	60	81.18
13	77.29	29	82.26	45	85.79	61	82.44
14	77.88	30	86.29	46	81.33	62	77.92
15	85.05	31	85.07	47	85.50	63	80.00
16	85.19	32	84.44	48	81.45	64	80.81

role in group animation motion capture and animation design.

5. Conclusion

The development of motion capture technology itself also has certain limitations. For example, the price of the optical system is too expensive. Motion capture also needs to capture special light spots on the performer. The subsequent adjustment and data modification work is a particularly large workload. Therefore, the motion capture technology itself is constantly developing and improving. In optical motion capture, it is these blips that make postanimation processing data tricky. Now, some scholars propose to use video processing instead of directly capturing light spots to identify the movement trajectory of the performer, thereby simplifying the process of motion capture and reducing people's workload. This paper studies the group animation motion capture method combined with virtual reality technology and builds a group motion capture system based on virtual reality technology. The simulation results show that the group animation motion capture method based on virtual reality technology proposed in this paper can play an important role in group animation motion capture and animation design.

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 no competing interests.

Acknowledgments

This work was supported by Vocational Education Reform and Innovation Project of "Science, Innovation and Education" of the Ministry of Education (Grant HBKC217128), by Industry-University-Research Innovation Fund for

Chinese Universities, Ministry of Education (Grant 2021ALA02024), by University-Industry Collaborative Education Program of the Ministry of Education of China (Grant 201702028006), and by Team and Science Project Funds of Yibin Vocational and Technical College (Grants ybzysc20bk05, ybzy21cxt-d-06, and ZRKY21ZDXM-03).

References

- [1] E. Castano, "Art films foster theory of mind," *Humanities and Social Sciences Communications*, vol. 8, no. 1, pp. 119–210, 2021.
- [2] L. Jayyusi, "Hollywood's transnational imaginaries: colonial agency and vision from Indiana JonestoWorld War Z," *Continuum*, vol. 32, no. 3, pp. 355–369, 2018.
- [3] S. Kumar, K. De, and P. P. Roy, "Movie recommendation system using sentiment analysis from microblogging data," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 4, pp. 915–923, 2020.
- [4] E. Taiebi Javid, M. Nazari, and M. R. Ghaeli, "Social media and e-commerce: a scientometrics analysis," *International Journal of Data and Network Science*, vol. 3, no. 3, pp. 269–290, 2019.
- [5] B. Zou, M. Nurudeen, C. Zhu, Z. Zhang, R. Zhao, and L. Wang, "A neuro-fuzzy crime prediction model based on video analysis," *Chinese Journal of Electronics*, vol. 27, no. 5, pp. 968–975, 2018.
- [6] S. S. Sundar, "Rise of machine agency: a framework for studying the psychology of human-AI interaction (hai)," *Journal of Computer-Mediated Communication*, vol. 25, no. 1, pp. 74–88, 2020.
- [7] S.-C. Chang, "Market size matters? An approach to illustrate the market preference of Hong Kong-mainland China co-production cinema," *Journal of International Communication*, vol. 26, no. 1, pp. 125–149, 2020.
- [8] A. Kaplan and M. Haenlein, "Rulers of the world, unite! the challenges and opportunities of artificial intelligence," *Business Horizons*, vol. 63, no. 1, pp. 37–50, 2020.
- [9] B. Kuklick, "Fascism comes to America," *International Journal for History, Culture and Modernity*, vol. 6, no. 1, pp. 1–18, 2018.
- [10] A. Abbasi, H. Chen, and A. Salem, "Sentiment analysis in multiple languages," *ACM Transactions on Information Systems*, vol. 26, no. 3, pp. 1–34, 2008.

- [11] R. Piryani, V. Gupta, and V. K. Singh, "Movie Prism: a novel system for aspect level sentiment profiling of movies," *Journal of Intelligent and Fuzzy Systems*, vol. 32, no. 5, pp. 3297–3311, 2017.
- [12] L. Pang, "Mediating the ethics of technology: hollywood and movie piracy," *Culture, Theory and Critique*, vol. 45, no. 1, pp. 19–32, 2004.
- [13] J.-H. Shon, Y.-G. Kim, and S.-J. Yim, "Classifying movies based on audience perceptions: MTI framework and box office performance," *The Journal of Media Economics*, vol. 27, no. 2, pp. 79–106, 2014.
- [14] P. Bosc, D. Dubois, and H. Prade, "Fuzzy functional dependencies and redundancy elimination," *Journal of the American Society for Information Science*, vol. 49, no. 3, pp. 217–235, 1998.
- [15] S. Agrawal, R. K. Singh, and Q. Murtaza, "Prioritizing critical success factors for reverse logistics implementation using fuzzy-TOPSIS methodology," *Journal of Industrial Engineering International*, vol. 12, no. 1, pp. 15–27, 2016.
- [16] C. Porcel, A. Tejada-Lorente, M. A. Martínez, and E. Herrera-Viedma, "A hybrid recommender system for the selective dissemination of research resources in a technology transfer office," *Information Sciences*, vol. 184, no. 1, pp. 1–19, 2012.
- [17] T. Jojola, "Absurd reality: hollywood goes to the Indians," *Film and History: An Interdisciplinary Journal of Film and Television Studies*, vol. 23, no. 1, pp. 7–16, 1993.
- [18] R. Carrasco, M. G. Ordaz, and F. J. M. López, "Science fiction and bodies of the future: alternative gender realities in Hollywood Cinema," *Journal of Futures Studies*, vol. 20, no. 2, pp. 67–80, 2015.
- [19] R. D. Cohen, "The delinquents: censorship and youth culture in recent U. S. History," *History of Education Quarterly*, vol. 37, no. 3, pp. 251–270, 1997.
- [20] R. A. M. Noor, Z. Ahmad, M. M. Don, and M. H. Uzir, "Modelling and control of different types of polymerization processes using neural networks technique: a review," *Canadian Journal of Chemical Engineering*, vol. 88, no. 6, pp. 1065–1084, 2010.