Three-Dimensional Animation Capture Driver Technology for Digital Media

Wanjie Dong
School of Arts and Media, Wuhan Vocational College of Software and Engineering, Wuhan 430000, China

***Abstract*—For the motion capture driving technology of three-dimensional animation, this study combines skeleton extraction methods and human motion pose data to construct the human skeleton of three-dimensional animated characters. Combining matching algorithms and action recognition techniques, the postures of the human three-dimensional model were tested and analyzed. The experimental results showed that the level-set central clustering method extracted shoulder joint position values of 0.26, 0.24, 0.28, and 0.21 in the four models, respectively. The error value was the smallest among the skeleton extraction algorithms, indicating that this skeleton extraction algorithm had high accuracy in extracting human skeleton information. In addition, the depth information of human joint points was compared using the parallax ranging method, and the highest error was 1.57%. This further demonstrated that the coordinate error of the three-dimensional joints was relatively accurate, which also proved the effectiveness of the binocular stereo vision system. The system had an accuracy of over 80% in recognizing joint rotation information and dynamic movements in the human three-dimensional model. Finally, the highest accuracy of inertial sensors in capturing human movements was 97%, indicating the superiority of digital media in capturing three-dimensional animation technology. This also provides a theoretical basis and technical reference for animation production and other aspects.**

***Keywords*—3D animation; computer vision; motion matching algorithm; human 3D skeletal model; motion capture technology**

I. INTRODUCTION

With the rapid development of digital media technology and virtual reality technology, animation and film production require increasingly precise character models [1]. However, the three-dimensional (3D) animated character models still requires professional production software and generation systems. Therefore, regarding specific character modeling and motion driving, computer vision and human motion simulation techniques are used to match human 3D skeleton model data, thereby achieving smooth animation effects [3]. Digital media technology mainly processes, stores, and transmits information through computers and digital devices. The application of 3D modeling and rendering functions in digital media in animation production makes it more accurate and realistic, thereby enriching the visual experience [5]. The innovation of digital media technology has also provided new technological platforms and experiential conditions for the cycle and cost of animation production. In addition, regarding the design of motion postures for animated characters, computer vision and computer graphics are used. Combining wearable devices to collect data on human body movements and postures, it has been applied in practical applications such as robot gait rehabilitation, motion analysis, and film and television animation [6]. Finally, based on techniques such as image processing and pattern recognition, human motion analysis is performed on the collected motion data to complete the animation driving of the computer interface. However, for the collection of human motion data, auxiliary tools of wearable devices cannot meet the requirements of 3D animation display and accurate joint position matching. Therefore, the study first combines the skeleton extraction algorithm to build the human 3D skeleton model. Skin technology is used to achieve topology matching of animation models. Secondly, a binocular vision camera system is used to recognize human motion fonts to precisely match human posture movements and joint positions. This research method effectively combines the human 3D skeleton model with posture motion matching, fully utilizing the joint depth information of the skeleton model, and providing accurate data matching for human motion trajectory and 3D animation simulation. Finally, the research combines computer vision and artificial intelligence technology to test and verify motion capture devices, aiming to prove the effectiveness of 3D animation capture driving technology and provide technical means and realistic 3D visual effects for the smooth movements and behavioral postures of character models, thereby promoting the high-quality development of film and television animation production.

The research is mainly divided into six sections. Section II elaborates on the current research results. Section III conducts algorithm analysis on the constructed 3D animated human model to promote skeleton extraction and matching of motion postures. Section IV is to verify and analyze the motion recognition and capture equipment. Results and discussion is given in Section V and finally, Section VI concludes the paper.

II. LITERATURE REVIEW

Due to the advancement of 3D animation and virtual animation technology, motion capture techniques for animation models have been extensively studied. In recent years, domestic and foreign scholars have conducted a lot of research on digital media technologies such as virtual acquisition methods and computer vision in 3D animation production. Jiao L et al. proposed a node encoding classification for graph learning and computer vision applications, focusing on the development of graphic structures and computer vision. The applications of visual tasks based on neural network methods were also summarized [8]. Wang Y et al. used Kinect fusion algorithm and function to evaluate the tracking confidence of virtual reality simulation technology.

Then a prototype system was established to evaluate the tracking skeleton of moving objects, thereby proving the good fusion performance [9]. Gao P proposed a multi-dimensional data model for video image motion recognition and motion capture based on a deep learning framework. It combined deep learning features and datasets to achieve high recognition accuracy for gesture actions [10]. For the 3D modeling of film and television animation, Xu L combined local binary fitting algorithm and convolutional neural network to construct a single perspective 3D face model. The results showed that it was feasible in film and television animation and human-computer interaction [11]. Wang X P et al. extended the corresponding relationships to functions using the balanced function map algorithm for 3D shape registration. Experimental analysis was conducted on the character animation dataset in function space, demonstrating the superiority of the algorithm [12]. For the application of computer vision and graphic vision, different algorithms have various effectiveness and feasibility in action recognition technology.

Regarding the human 3D skeleton model, researchers from different fields have achieved many results. Bhogal R K et al. used convolutional neural networks to search for optimal features for action recognition in multi-view skeletal 3D data. The long and short-term memory layering was used to achieve model accuracy, thereby proving the high accuracy of the model on the human dataset [13]. Setiawan F et al. used graph convolutional neural networks to simulate human skeleton for action recognition. The Laplace matrix was used to encode graph attributes, thereby achieving high recognition accuracy on human datasets [14]. Mao W S et al. used radio frequency identification technology and bicycle motion networks to label human posture data for 3D human posture tracking, which improved tracking accuracy [15]. Lin Y et al. used velocity threshold correction method to adjust joint data for human 3D posture detection. The camera was used to detect the depth value of posture data, thereby improving the accuracy of human 3D posture detection [16]. Ahad M A R et al. proposed a method for extracting motion posture features based on skeleton data for 3D skeleton joint position recognition. The high accuracy of its method was validated in the benchmark dataset [17].

In summary, researchers have conducted many model constructions and algorithm applications on animation production technology and human motion recognition methods. However, there is a lack of testing for the construction and simulation of internal skeletons in human 3D posture recognition. The research on the application and production effects of 3D animated characters is also limited, resulting in limited research on character simulation and motion posture in film and television animation. Therefore, the study utilizes motion-matching algorithms and skin animation algorithms to construct 3D skeleton models of animated characters. The binocular stereo vision system has high advantages in 3D animation motion capture technology.

III. 3D Animation Human Motion Capture System Construction

This section combines matching algorithms and skin techniques to connect skeleton motion and 3D data to analyze the motion trajectory of the skeleton model. A binocular stereo vision system is used to extract features from human joint points. A binocular local matching algorithm is combined to improve the 3D spatial information of human actions, thereby constructing a human 3D skeleton model and action capture system.

A. Motion Matching Algorithm for 3D Animation Model

3D animation character generation includes extracting skeletons, embedding skeletons, matching actions, and skin binding. The posture model of animated characters identifies joint positions and matches motion data for skeleton extraction to generate 3D animated characters. Therefore, the skeleton discrete embedding is used to identify the positions of key joints. The positions of other joint points are calculated based on the standard proportion relationship, thereby obtaining a complete character skeleton model [18]. However, skeleton motion control in 3D animation requires data that matches with the motion data. The skeleton extraction structure based on the same topology and the motion data structure of the Bio Vision Human Motion Capture (BVH) file are the same. In data matching, the BVH file is scaled to fit the same topology. The data scaling is shown in Eq. (1).

$$A_{dest} = A_{src} * \left(\frac{L_{dest}}{L_{src}} \right) \tag{1}$$

In Eq. (1), A_{dest} represents the motion data of the target skeleton. A_{src} is the motion data of the source animation data file. L_{dest} and L_{src} represent the length of the target joint skeleton and the joint skeleton length of the source skeleton, respectively. The ratio of two skeleton lengths can achieve data scaling. In addition, a hierarchical structure between skeleton joints is established to perform skeleton motion. The motion of the parent node affects the child nodes. Then the joint coordinate system completes the matrix transformation, as shown in Eq. (2).

$$P_a = P_0 E_{m-1}^{m}(t) E_{m-2}^{m-1}(t) \ldots E_{n-1}^{n}(t) \tag{2}$$

In Eq. (2), a represents the specified joint point. $E_{n-1}^{n}(t)$ is the transformation matrix. P_0 is the reference matrix, which is the initial posture. The transformation formula continuously converts the root node coordinates to the local coordinates of the target node, thereby completing the associated motion between skeletons. The skeleton motion data and 3D model are independent of each other. To achieve the 3D animation effect, the Linear Blending Skinning (LBS) algorithm is used to bind or deform skeleton and skin. The LBS algorithm labels human motion joints to calculate the vertex changes of the model. The vertex is related to the motion state of skeleton. The specific joint transition relationship is shown in Eq. (3).

$$v' = E_i \times A_i^{-1} v \tag{3}$$

In Eq. (3), v represents the coordinate of a vertex in the skeleton. v' is the vertex coordinate that has been converted through coordinate transformation. A_i^{-1} is the local coordinate system that converts the coordinate points in the global state to joint J_i. E_i is its vertex motion control matrix. A vertex is affected by the joint motion of multiple skeletons, so different vertices are weighted to calculate the motion transformation matrix of the vertex. The weight is shown in Eq. (4).

$$\sum_{i=n}^{n} w_i = 1 \tag{4}$$

In Eq. (4), w_i represents the weight of a vertex. The sum of the weights of its vertices affected by different skeletons is 1. To improve skin technology and animation effects, the motion trajectories of all joints in the skeleton model are calculated, as shown in Eq. (5).

$$v' = \sum_{i=n}^{n} w_i E_{ji} A_{ji}^{-1} v \tag{5}$$

In Eq. (5), v' represents the vertex coordinates converted by coordinates. E_i is its vertex motion control matrix. To make the skin effect smoother and more realistic, as well as avoid collapse, the vertex weight values are kept in the model vertex direction to achieve continuous smooth motion. For the skin deformation problem of 3D animation models, the LBS algorithm is used to calculate the proportion of skeletons and skin at joints, conforming to the same model structure. In addition, the structure of the 3D animation model generation system is divided using the skin technology, as shown in Fig. 1.

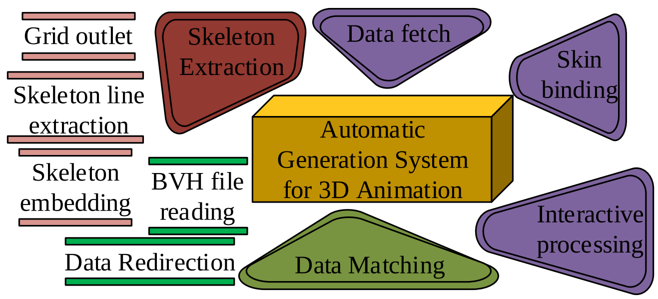


Fig. 1. Schematic diagram of the 3D animation model system structure.

From Fig. 1, the modules of the 3D animation model system mainly include data reading, skeleton extraction, data matching, skin binding, and interaction processing. The skeleton extraction module includes mesh processing, skeleton line extraction, and skeleton embedding. Data matching is the process of reading animation data from a BVH file and redirecting it to an existing model skeleton. In addition, the skin binding module utilizes the LBS skin deformation algorithm to bind skin and skeletons. By calculating the weight relationship between skeletons and model vertices, the motion trajectory of model vertices can be calculated. The final interaction processing of the system implements a visual display window to facilitate data import and parameter settings.

B. Construction of Human 3D Skeleton and Motion Capture System

Based on 3D animation motion posture, the human 3D action skeleton is constructed to meet the animation posture needs. The 3D action skeleton requires obtaining joint localization and recognition information. The binocular stereo vision system extracts features from human joints and combines optimization algorithms to obtain 3D information of human movements, thereby constructing a 3D skeleton model [19]. To accurately obtain 3D information of human motion joints, binocular camera calibration and 3D coordinate solution are used to ensure that human posture movements are consistent with joint positions. The visual distance measurement of the binocular camera is obtained by the principle of triangulation. The depth information of the target point is shown in Eq. (6).

$$Z = \frac{J}{X_l - X_r} \times f \tag{6}$$

In Eq. (6), J represents the distance between the center-line of the left and right optical centers of the binocular camera. f is the camera focal length. The mapping abscissa on the left is X_l, and the imaging abscissa on the right is X_r. The disparity value of the left and right mapping points is shown in Eq. (7).

$$d = X_l - X_r \tag{7}$$

In Eq. (7), d represents the disparity value between the left and right image points. Therefore, the depth information conversion of the target point is shown in Eq. (8).

$$Z = \frac{J}{d} \times f \tag{8}$$

In Eq. (8), Z represents the depth value of the target point in the physical world. To accurately obtain the 3D coordinates of human motion joints, this study combines the least squares method and the inner and outer parameter matrices of binocular cameras. Then, combined with the binocular local matching algorithm, the joint depth of human motion skeletons is calculated, as shown in Fig. 2.

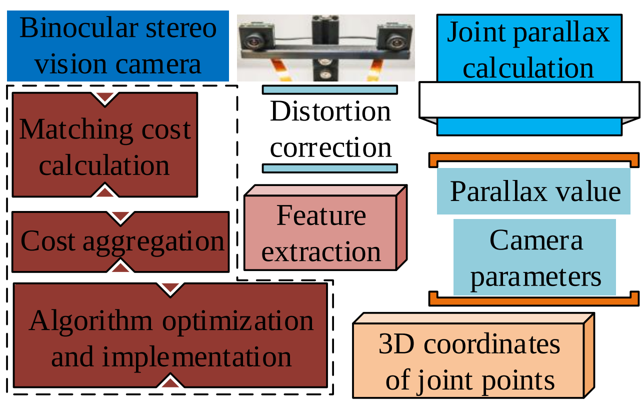


Fig. 2. Structure of binocular stereo matching algorithm.

From Fig. 2, feature constraints are an important step in matching algorithms, which can be used for edge extraction of left and right images, thereby reducing the matching range and obtaining gradient information of the image. In addition, joint disparity mainly includes feature constraints and matching cost calculation, cost aggregation, and algorithm optimization. The disparity value and camera parameters are combined to obtain the 3D coordinates of the joint points. Finally, by calculating the parallax value of human motion joints and accumulating the relevant state equations, the two-dimensional skeleton joints of human motion can be corrected. Then, based on the coordinate transformation formula, the 3D joint coordinates can be calculated to construct the 3D skeleton of human actions. Based on the constructed 3D skeleton, the model is used to transform the 3D skeleton into a human 3D action that is consistent with the action posture. The motion capture and model error analysis are performed. The structure of the motion capture system is shown in Fig. 3.

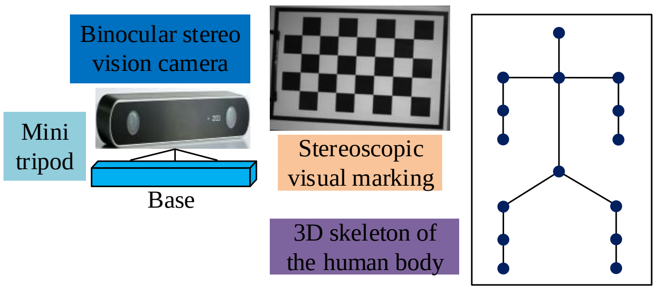


Fig. 3. System structure and model diagram of a 3D motion capture system.

From Fig. 3, the binocular stereo vision camera system used in the study includes a camera, a miniature tripod, and a base for easy portability and camera parameter adjustment. To improve the running speed of the system, platforms such as Windows and Linux are combined with high-performance processors to synchronously obtain images from left and right cameras. Finally, the images are transmitted to the computer through an interface for data processing. Due to the camera system saving data on the 3D skeleton model, the 3D coordinate information of joint points is input into the software to construct the human body skeleton, thereby obtaining action posture spatial information.

C. *3D Animation Model Technology Driven by Action Posture*

The human posture and expression behavior unit data of 3D animation are synchronously captured by a dual camera system, which in turn generates the human animation model. The motion posture capture data is mainly controlled through key-frame interpolation and inverse kinematics to control model motion. Key-frame interpolation is generated through interpolation algorithms to generate intermediate transition frames to simulate real motion effects. To ensure that key-frame interpolation methods generate realistic and motion-compliant animations, quaternions are used to represent the rotation information of human joints. The interpolation algorithm is combined to compensate for missing frames in the rotation information. The Spherical Linear Interpolation (SLERP) method in quaternion interpolation can facilitate smooth interpolation of joint information. The ordinary linear interpolation function is shown in Eq. (9).

$$\vec{d} = \vec{d}_0 + t\left(\vec{d}_1 - \vec{d}_0\right) \tag{9}$$

In Eq. (9), $\vec{d}_0$ and $\vec{d}_1$ represent vectors in two directions, and the angle between them is φ. $\vec{d}$ represents the joint rotation information by taking two quaternions from the median vector of two directional vectors, as shown in Eq. (10).

$$\begin{cases} \vec{q}_0 = \left(x_0, y_0, z_0, w_0\right) \\ \vec{q}_1 = \left(x_1, y_1, z_1, w_1\right) \end{cases} \tag{10}$$

In Eq. (10), $\vec{q}_0$ and $\vec{q}_1$ are the direction vectors of two quaternions, respectively. The surface interpolation between two quaternions is shown in Eq. (11).

$$\begin{cases} \vec{q} = a(t)\vec{q}_0 + b(t)\vec{q}_1 \\ a(t) = \dfrac{\sin\left[(1-t)\varphi\right]}{\sin\varphi} \\ b(t) = \dfrac{\sin t\varphi}{\sin\varphi} \end{cases} \tag{11}$$

In Eq. (11), $(1-t)\varphi$ is the angle between $\vec{d}$ and $\vec{d}_1$. $t\varphi$ is the angle between $\vec{d}$ and $\vec{d}_0$. Therefore, the spherical interpolation is shown in Eq. (12).

$$SLERP\left(\vec{q}_0, \vec{q}_1, t\right) = \frac{\sin\left[(1-t)\varphi\right]\vec{q}_0 + \sin t\varphi\vec{q}_1}{\sin\varphi} \tag{12}$$

In Eq. (12), the angle between $\vec{d}$ and $\vec{d}_1$ is $(1-t)\varphi$. The angle between $\vec{d}$ and $\vec{d}_0$ is $t\varphi$. The dot product between two directional vectors is calculated to determine the angle between them, as shown in Eq. (13).

$$\cos\varphi = \vec{q}_0 \cdot \vec{q}_1 = x_0x_1 + y_0y_1 + z_0z_1 + w_0w_1 \tag{13}$$

In Eq. (13), the angle between the two directional vectors is φ. $(1-t)\varphi$ is the angle between $\vec{d}$ and $\vec{d}_1$. $t\varphi$ is the angle between $\vec{d}$ and $\vec{d}_0$. When the dot product result is negative, the interpolation will move the longest path around the sphere. When the angle between two directional vectors is too small, the denominator results tend to approach 0, and linear interpolation is used to replace the minimum angle. The spherical interpolation method is used to obtain key-frame sequences of uniform motion, but there are still some motion sequences that do not meet the laws of human motion. Therefore, it is necessary to combine the kinematic method to correct parameters and obtain more suitable control parameters for human motion laws. The human motion state is usually based on posture initialization, adding time and parameter changes in kinematics, including forward kinematics and inverse kinematics. In inverse kinematics, the intermediate joint points are calculated based on the position of the end node. The child nodes drive all parent nodes to achieve motion constraints layer by layer, specifically, as shown in Eq. (14).

$$R = f^{-1}(W) \tag{14}$$

In Eq. (14), R represents the joint rotation angle. W represents the end node position. The inverse kinematics analysis method reduces the complexity of node calculation, but it is only suitable for solving nodes with fewer degrees of freedom. In addition, the inverse kinematics numerical method can solve the nodes with larger degrees of freedom to obtain complex human postures. Finally, combined with the joint limit state of the human skeleton model, the rotation information of all nodes is continuously adjusted. The standard range of human joint angle motion is shown in Fig. 4.

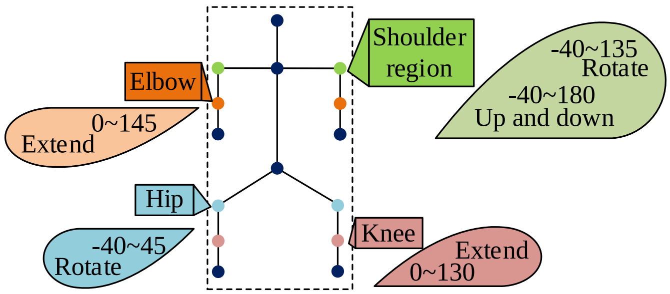


Fig. 4. Schematic diagram of the motion units of some joints in the human body.

From Fig. 4, the range of joint points through extreme motion is maximized when setting the action state of the 3D animation model. According to the constraint conditions, the motion constraint for the rotation angle of the joint point is shown in Eq. (15).

$$\phi_i = \begin{cases} \sum_{n=i-2}^{i} \frac{\phi_n}{n} & if\left(\phi_i < \alpha\right) \\ \phi_i & if\left(\alpha < \phi_i < \beta\right) \\ \sum_{n=i-2}^{i} \frac{\phi_n}{n} & if\left(\phi_i > \beta\right) \end{cases} \tag{15}$$

In Eq. (15), ϕ_i represents the rotation angle of a certain joint point, and its motion range is $\left[\alpha_{axis}, \beta_{axis}\right]$. To improve the motion and posture control of the 3D animated human body, a method combining key-frame interpolation and inverse kinematics is used to drive the 3D model, as shown in Fig. 5.

In Fig. 5, the human 3D skeleton model is mapped after importing data. Based on the corresponding skeleton posture data, the structure of the human model is set up to improve the human 3D skeleton model. The key-frame interpolation method involves interpolating the 3D estimation data to ensure smooth, stable, and continuous model motion. Finally, the skeleton data is bound and refreshed at the sent frame rate to obtain 3D animation effects.

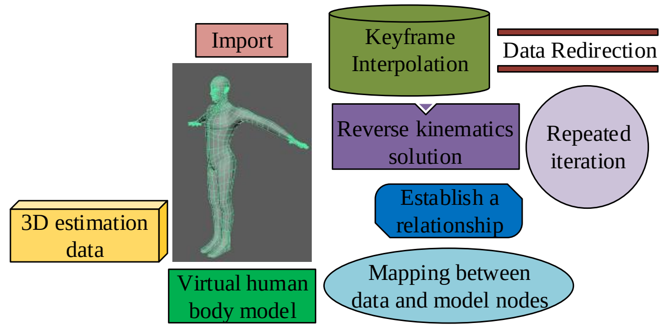


Fig. 5. The action and posture driving process of the 3D animation model.

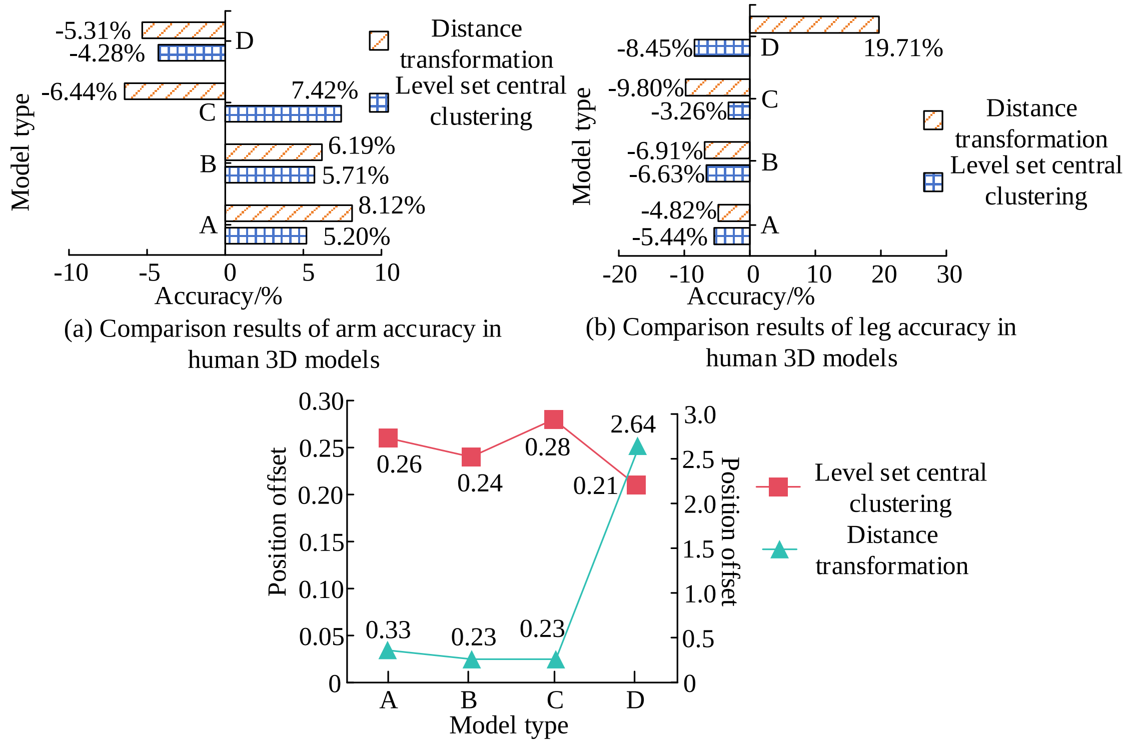
IV. Human Motion Capture and 3D Animation Driving Analysis

The 3D animation generation system and the motion capture system of the human skeleton model are interactively validated on the system platform to compare the motion posture and data errors of the 3D animation. Compared with other motion capture devices, it obtains smoother and more continuous data information in 3D animation information. The interactive 3D animation model is combined with the Microsoft Basic Class Library to establish a system interaction interface for importing BVH files and playback control. The hardware environment of the platform and the software system of the human 3D skeleton are shown in Table I.

In Table I, the interactive interface of the 3D animation model was used to open the BVH file and performed data processing on the 3D model. The motion posture and joint points of the human 3D skeleton were extracted through a camera. Combined with the software platform, image processing and 3D skeleton extraction were completed to construct 3D actions. 3D animation models were combined with skeleton extraction algorithms to conduct comparative experiments on four models. A-D was used to represent them. The number of vertices and polygons in model A was 2541 and 5078, respectively, while the number of information in models B, C, and D was the same, which was 13336 and 26668. The skeleton extraction algorithm adopted the level set central clustering method and distance transformation method. The accuracy of the arms and legs of the four models was compared, as well as the displacement of the shoulder joints, as shown in Fig. 6.

TABLE I. Basic Information of System Hardware Environment and Software Platform

Hardware Environment for 3D Animation Models		**Software platform for human 3D skeleton models**	
CPU	Intel(R)Core(TM)i5-3230M CPU @2.60GHz	Windows10 system	MATLAB platform
Memory	8G	Binocular camera calibration	Visual Sudio
Graphics card	NVIDIA GeForce GT 750M	Nvidia GeForce GTX1080	16G DDR4 2333MHz
Processing models and analysis results		Image processing and skeleton information extraction	



(a) Comparison results of arm accuracy in human 3D models

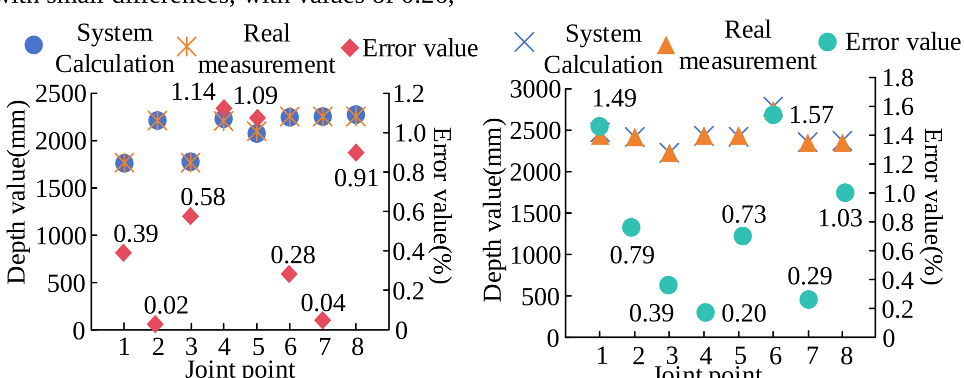
(b) Comparison results of leg accuracy in human 3D models

(c) The displacement results of shoulder joint positions in human 3D models

Fig. 6. Comparison of two skeleton extraction algorithms in human 3D models.

In Fig. 6 (a), the arm accuracy of the four models in the level set central clustering method was 5.20%, 5.71%, 7.42%, and -4.28%, respectively. The values in the distance transformation method were 8.12%, 6.19%, -6.44%, and -5.31%. The overall value of the level set central clustering method was low, with relatively high accuracy. In Fig. 6 (b), the leg accuracy of the human 3D model in the level set central clustering method was -5.44%, -6.63%, -3.26%, and -8.45%, respectively. The values of the distance transformation method were -4.82%, -6.91%, -9.80%, and 19.71%, with an overall difference greater than the former. Fig. 6 (c) shows the shoulder joint displacement in a human model. The level set central clustering method was relatively balanced with small differences, with values of 0.26, 0.24, 0.28, and 0.21, respectively. Therefore, it indicated that the level-set central clustering method had higher accuracy in extracting human skeletons. Afterwards, combining skin binding promoted superior smoothing effects in 3D animation, thereby extracting motion data.

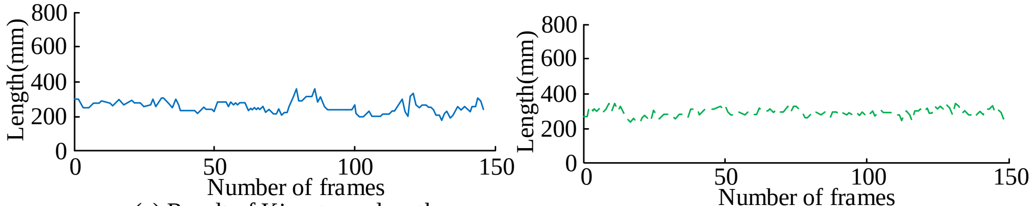
Based on the software platform of the binocular camera system and the constructed human 3D skeleton, the error analysis of joint depth values for human motion posture is carried out to achieve motion capture. The error comparison between the parallax ranging method and the real measurement is conducted using the human skeleton and its joint point model. The results are shown in Fig. 7.



(a) The joint depth value and its error value of the first set of actions

(b) The joint depth value and its error value of the second set of actions

Fig. 7. Results of joint depth values and error values for two groups of actions.



(a) Result of Kinect arm length

(b) The results of Openpose weight optimization

Fig. 8. Comparison of results of action capture methods.

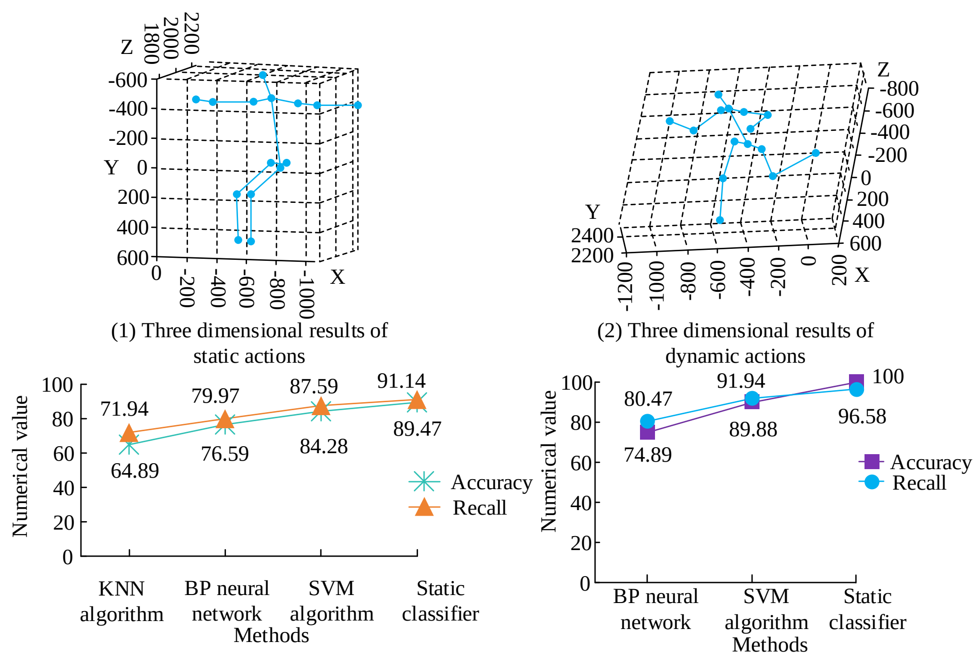
The action depth of the human 3D skeleton model is validated. The system calculation and actual measurement values are compared. The highest error was 1.14%, while the lowest was 0.02%. From Fig. 7 (b), the depth information error value for another set of actions was the lowest at 0.20% and the highest at 1.57%. Therefore, the 3D coordinate error of human joints was relatively accurate, which also proved the effectiveness of the binocular stereo vision camera system. Afterwards, joint movements of different human models are compared. Different binocular camera systems are used to capture 3D movements. The measurement length of the motion frame rate is shown in Fig. 8.

In Fig. 8 (a), the length of the limbs captured by the Kinect device remained basically unchanged, which was between 200mm and 400mm. In Fig. 8 (b), the weight optimization of Openpose multi-camera had a lower length fluctuation in the number of frames compared with Kinect devices. Therefore, it indicated the accuracy and superiority of the binocular stereo vision camera system in capturing motion. Finally, regarding the driving system of 3D animation, to simulate real-time human body movements and human-computer interaction movements, the image data of human body movements and postures is captured, as shown in Table II.

According to Table II, the binocular camera system used a Logitech C525 camera with a resolution of 1280×720. The maximum acquisition frame rate was 30fps. The Unity development platform has flexibility and convenience in constructing 3D animation models and their driver programs. It is feasible to solve joint rotation information of human body posture. Therefore, this study selects 3D human postures with static movements to compare the performance of different methods. The results are shown in Fig. 9.

TABLE II. CONFIGURATION OF BINOCULAR STEREOSCOPIC CAMERA SYSTEM PLATFORM

Driving system	Dual camera system
CPU parameters	Intel(R)Core(TM)i7-8700K CPU
Memory	NVIDIA GeForce GTX 1080
Operating system	Ubuntu 16.04
Development platform	Unity/Visual Studio2017



(a) Test results of static actions

(b) Test results of dynamic actions

Fig. 9. Action test results of human 3D model.

Fig. 9 (a) displays the static actions of the human 3D model in Fig. 1. Its accuracy was above 70%, and the static classifier was more accurate in recognizing human movements. Fig. 9 (b) shows the recognition accuracy of dynamic movements in Fig. 2. The results were all above 80%, with a recall rate of over 74%, proving that the action testing of human 3D models was superior. Afterwards, the motion capture is performed, as shown in Fig. 10.

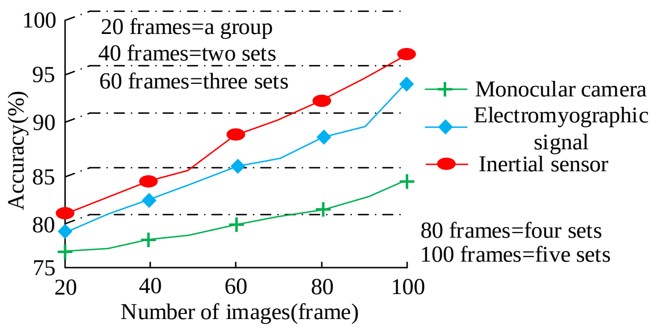


Fig. 10. Test results of different methods for human motion recognition.

In Fig. 10, the accuracy of motion capture improved with the increase in the number of motion images. The monocular camera system method had the lowest accuracy in recognizing human movements, with a maximum value of 84%. The highest accuracy of the action feature extraction method using electromyographic signals was 94%. The action capture accuracy using inertial sensors was as high as 97%. According to the motion capture method, the driving technology of 3D animation is continuously improving, and the motion capture of human 3D models is more accurate, thereby satisfying smooth animation effects.

V. Results and Discussion

As one of the key technologies in 3D animation production, computer vision and human motion simulation technology are important research directions for human 3D models. In the construction of a 3D skeleton model, the skin technology and matching algorithm were used to analyze the trajectory of the 3D data and motion features of the human skeleton. Among them, for the displacement test of the shoulder joints, the difference between the level set center clustering method was small, with specific results of 0.26, 0.24, 0.28, and 0.21. Afterwards, the binocular stereo vision system combined with the binocular matching algorithm to calculate the joint depth of the human action skeleton. The highest error of the skeleton model was 1.57%, which met the accuracy requirements of the 3D skeletal model of human movements. The highest motion capture accuracy for obtaining posture features of the binocular vision camera combined with the 3D skeleton model was 97%.

Based on the above results, it indicates that this study effectively improves the smoothness of 3D animation using matching algorithms and computer vision technologies, while enriching the visual effects and motion smoothness in animation character production. However, the motion recognition and posture data of human joints in this research system are still not complete enough, which affects the detailed effect of animated character models. At the same time, the motion capture driving technology lacks specific parameter moduli for joint points and posture that target the human 3D motion characteristics. In the future, the development and design of motion driving systems for 3D animation still need to continue exploring computer vision and motion capture driving technology to achieve innovative design in film and television animation and game production.

VI. Conclusion

To address the driving technology for 3D animation capture, the motion matching algorithm and human 3D skeleton model were used for data analysis of the 3D animation capture system. Firstly, the generation process of 3D animated characters was used to analyze the motion posture of the skeleton model, and then to match and partition its motion data. The LBS algorithm was used to set up and interact with the skeleton and skin at the joints of the model. Secondly, the human 3D skeleton model was constructed. Combined with a dual camera system to synchronously capture human movements, the depth information of 3D joints was obtained. According to the skeleton extraction algorithm, the accuracy of the arm joints for the four models was 5.20%, 5.71%, 7.42%, and -4.28%, respectively. The depth information verification of the human 3D skeleton movement showed that the lowest error values were 0.02% and 0.20%, respectively, indicating that the binocular stereo-vision camera system had high accuracy in joint recognition of the human 3D skeleton model. Finally, the motion capture system was validated and analyzed based on the captured data of the joint motion posture. The accuracy of human motion recognition was high, all above 80%. Therefore, the capture system platform for 3D animation satisfies the smoothness effect of 3D animation. However, the system's motion recognition animation for human joints is still not complete enough. The impact of additional effects on animation lacks quantitative analysis. Therefore, further research and improvement should be conducted on the development and application of the 3D animation capture driver system.

References

[1] Bo X U. Video,Internet and Metaverse:The Media Transitions of Interaction in Theatre. Journal of Literature and Art Studies, 2022, 12(8):855-861.

[2] Mokayed, H., Quan, T. Z., Alkhaled, L., & Sivakumar, V. Real-time human detection and counting system using deep learning computer vision techniques. Artificial Intelligence and Applications. 2023, 1(4): 221-229.

[3] Hfliger A, Kurabayashi S. Dynamic Motion Matching: Design and Implementation of a Context-Aware Animation System for Games. International Journal of Semantic Computing, 2022,16(2):189-212.

[4] Zhang J Q, Xu X, Shen Z M, Huang Z H, Zhao Y, Cao Y P, Wan P, Wang N. Write-An-Animation: High-level Text-based Animation Editing with Character-Scene Interaction. Computer Graphics Forum: Journal of the European Association for Computer Graphics, 2021,40(7):217-228.

[5] Zechao Li. Intelligent media computing technology and application for media convergence. CAAI Transactions on Intelligence Technology,2022,7(3):329-330.

[6] Ying X. The relation between body surface angle and apparel ease distribution under the motion state. International journal of clothing science and technology, 2023,35(2):293-311.

[7] Yiqiao Lin, Xueyan Jiao, Lei Zhao. Detection of 3D Human Posture Based on Improved Mediapipe. Journal of Computer and Communications,2023,11(2):102-121.

[8] Jiao L, Chen J, Liu F, Yang S, You C, Liu X, Li L, Hou B. Graph Representation Learning Meets Computer Vision: A Survey. IEEE Transactions on Artificial Intelligence, 2023,4(1):2-22.

[9] Wang Y, Chang F, Wu Y, Hu Z, Li L, Li P, Lang P, Yao S. Multi-Kinects fusion for full-body tracking in virtual reality-aided assembly simulation. International Journal of Distributed Sensor Networks, 2022, 18(5):625-636.

[10] Gao P, Zhao D, Chen X. Multi-dimensional data modelling of video image action recognition and motion capture in deep learning framework.IET Image Processing, 2020, 14(7):1257-1264.

[11] Xu L. Fast Modelling Algorithm for Realistic Three-Dimensional Human Face for Film and Television Animation. Complexity, 2021, 2021(2):1-10.

[12] Wang X P, Lei H, Liu Y, Sang N. Balanced Functional Maps for Three-Dimensional Non-Rigid Shape Registration. Journal of Electronic Science and Technology,2021,19(4):369-378.

[13] Bhogal R K, Devendran V. Action Recognition for Multiview Skeleton 3D Data Using NTURGB+D Dataset. Computer Systems Science and Engineering, 2023,47(12):2759-2772.

[14] Setiawan F, Yahya B N, Chun S J, Lee S L. Sequential inter-hop graph convolution neural network (SIhGCN) for skeleton-based human action recognition. Expert Systems with Application, 2022,195(6.):1-10.

[15] Mao W S. RFID-based 3D human pose tracking: A subject generalization approach. Digital Communications and Networks, 2022,8(3):278-288.

[16] Lin Y, Jiao X, Zhao L. Detection of 3D Human Posture Based on Improved Mediapipe. Computers and Communications, 2023, 11(2):102-121.

[17] Ahad M A R, Ahmed M, Antar A D, Makihara Y, Yagi Y. Action recognition using Kinematics Posture Feature on 3D skeleton joint locations. Pattern Recognition Letters, 2021,145(5):216-224.

[18] Zhu N, Zhao G, Zhang X, Jin Z. Falling motion detection algorithm based on deep learning.IET image processing, 2022,16(11):2845-2853.

[19] Sun J M, Han S Q, Shen Z C, Wu J P. Binocular Human Pose and Distance Identification Based on Double Convolutional Chain. Acta Armamentarii, 2022, 43(11):2846-2854.

[20] Wei H, Meng L. A binocular reconstruction based on perspective projection constraints and its application on robot eye-hand coordination.IET Computer Vision, 2022,16(4):333-349.