Computer-aided Teaching Software of Three-dimensional Model of Sports Movement Based on Kinect Depth Data

Feng Li*[1] and Peng Li[2]

[1]Department of Public Sports, Xuhai College, China University of Mining & Technology, JiangSu, XuZhou 221008, China, jm829@cumt.edu.cn
[2]Department of Public Sports, Xuhai College, China University of Mining & Technology, JiangSu, XuZhou 221008, China, LP566@cumt.edu.cn

Corresponding author: Feng Li, jm829@cumt.edu.cn

Abstract. This paper is based on Kinect's three-dimensional model of sports movement computer-aided teaching software. The framework of this system mainly includes three parts: input, auxiliary sports training system and output. Through input, calculation, comparison and other methods, it provides users with the functions of motion posture analysis, so as to realize the human motion analysis based on computer vision. In the input part of the system, the image sequences of athletes and coaches are input into the auxiliary sports training system through Kinect image acquisition. The auxiliary sports training system first detects the input image sequence and constructs the contour map of human body, then calculates the joint angle track of athletes and coaches when they move their limbs according to the corresponding algorithm, and finally compares the calculated results and calculates the similarity.

Keywords: Kinect; Physical Training; 3D Model; Computer Assisted Instruction
DOI: https://doi.org/10.14733/cadaps.2021.S2.123-134

1 INTRODUCTION

With the rapid development of mass sports, the rapid development of science and technology, especially the development of electronic information industry represented by electronic products, has attracted more and more attention. This industry brings high-quality computing speed to people, at the same time, it brings cheap and acceptable products to people [1]. The embodiment of mass sports is the development of sports capture equipment. From the expensive video capture equipment at the beginning to the low-cost body sense equipment Kinect and xton. People have been thinking, why don't we make a sports teaching system of mass sports, using computer and sense of body equipment instead of professional coaches to provide us with professional guidance, but also provide us with cheap services [2]. It can replace human eyes, provide more accurate and 360 degrees evaluation than professional coaches, and provide comprehensive services to the public in different time dimensions, such as action review, fast and slow playback, etc [3].

Action reproduction generally refers to the use of captured human or animal action data to reproduce the action track on the screen. This definition includes two steps: one is the capture and storage of action data; the other is the reproduction of action data. Claes has studied the automatic recovery of the human body's three-dimensional motion posture from the synchronous video sequence [4]. Kuzuhara has started with the motion video information, extract the spatial motion track of human body, use the radial basis function to deform on the basis of the general model, so as to obtain the middle model of reconstruction [5]. Chen's main research content is human motion understanding technology based on video content [6]. Under the condition of monocular vision, Teng multimedia laboratory uses spatial feature points and head statistical model [7]. Zhang used the method of image target detection to complete the correction of different sports movements, thus improving the movement standard of athletes [8].

A high-quality sports teaching system must include three necessary conditions: 1) strong intelligence: good recognition effect for people of different body and different environment; powerful motion analysis function; high-efficiency operation speed. 2) Good interactivity: for people with different learning abilities, it can provide satisfactory characteristics such as easy to understand, learn and use. 3) Low price: in order to achieve the widespread popularity of mass sports, this teaching system must provide a price that can be accepted by the general public, so as to truly realize mass sports. This paper attempts to use Kinect, xtion and other cheap and powerful sense of body equipment and mass sports to create a new way of sports teaching, to promote the rapid development of mass sports.

2 THE OVERALL FRAMEWORK OF COMPUTER AIDED SYSTEM IN PHYSICAL EDUCATION

2.1 System Software Platform

The teaching system of physical education mainly includes four modules, which are motion data collection, motion data processing, motion data analysis and motion animation. Among them, motion data collection is completed by Kinect equipment, including the students' motion data collection and the teachers' standard motion data collection [8]. Collecting sports information of physical education teachers, then establishing standard sports database; collecting sports data of teachers is mainly for teaching and training of physical education. The motion data is processed by occluding the joint information [9]. Data analysis is to compare the students' sports data with the teachers' sports data, and put forward relevant training guidance suggestions according to the comparison results, so as to quickly improve the students' sports action level. The motion picture is reproduced, and the motion file is made from the captured motion data, which drives the virtual human body to reproduce the human motion [10].

The software platform of physical education teaching system is built under the environment of Visual Studio 2008, including real-time motion capture, motion analysis and motion standard. Drive Kinect to start motion capture, end motion capture and save motion data. The movement analysis includes the plane judgment of students' movements and the comparison of joint physical parameters. The standard animation is made by collecting the high-level motion data of the teacher, which contains all the information of each joint of the complete motion. The hardware of the teaching system of physical education mainly includes two parts: Xbox360 Kinect and PC. The Kinect is connected with PC through USB interface, and the image is collected through Kinect, then the human body is tracked. The PC analyzes the collected data in real time.

Kinect is a somatosensory peripheral product developed by Microsoft for Xbox 360 as shown in Figure 1. Through an infrared laser projection lens, a group of infrared laser dot matrix is projected onto the user. The other two CMOS cameras carry out 3D scanning on the user to obtain the information of X, y and Z coordinates. Kinect can obtain three kinds of information at one time, including color image of ordinary camera, depth data and sound signal reflected by 3D infrared ray. There are three lenses on the Kinect machine, the infrared transmitter on the left, the infrared CMOS camera on the right, and the ordinary RGB color camera in the middle. The left and right lenses form a 3D depth sensor [11]. Kinect mainly relies on the 3D depth sensor to detect the player's actions. RGB color camera is mainly used to identify the user's identity, using face or body features for recognition. In addition, it can also be matched with the motor tracking mode of the base to automatically rotate the lens position following the target to find the most suitable central focusing position of the screen. Kinect's most important core technology is 3D depth information processing technology, which mainly deals with the relevant information received by infrared transmitter and camera, so as to judge the distance of the target object. CMOS infrared sensor collects every point in the field of view of the camera according to the set distance reference plane. The sensor generates a depth of field image stream at the speed of 30 frames / s, which can reproduce the surrounding environment in real time [12].

2.2 Teaching System Architecture

Based on the course design planning in the research method, the course design and development of Kinect body feeling technology interactive teaching of physical movement are carried out, which is presented in the form of web page as a whole. This course takes university students as the main research object and uses problem-solving teaching method to assist teaching. The teaching system is mainly presented in the form of web pages, and the learning content adopts unified design user interface and production specifications, which will help users to be familiar with the operation of user interface. The design and architecture of user interface are shown in Figure 1. Learners will observe the operation of physical education teachers through terminal devices such as computers. There are not only online demonstrations, but also offline teaching materials. The learning materials include sports tasks, sports schools, sports example observation, sports operation library, etc.

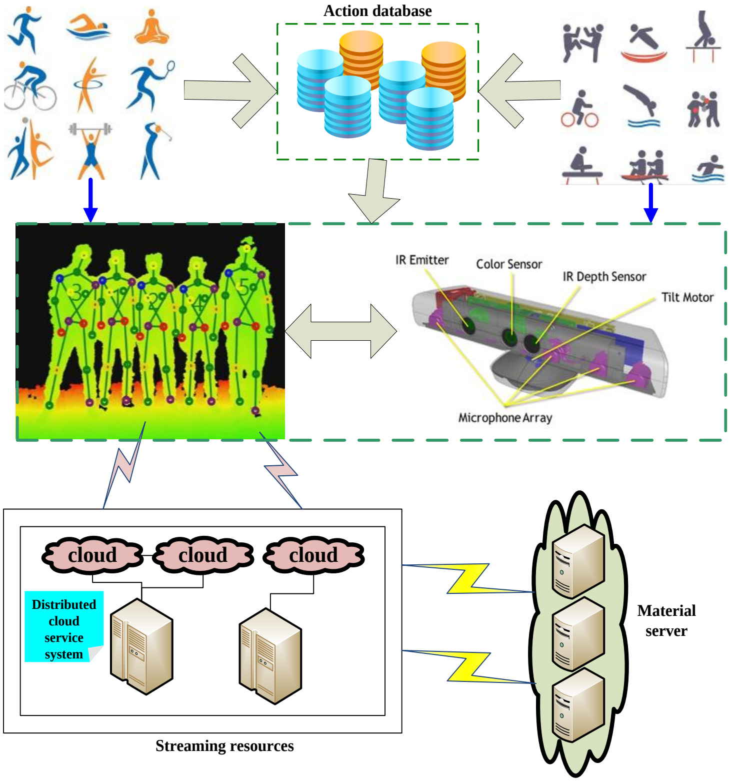

Figure 1: User interface design and architecture diagram.

First of all, Kinect's development environment needs to be solved in system construction. The development tools of this system must be built in visualization and java development environment. The networking method of human-computer interaction system and user interface system is to use socket for networking transmission and wireless network address for networking. The human-computer interaction system is mainly based on the notebook computer platform, and transmits information to the user interface system through Kinect's bone detection results for information transmission operation. The system can display the azimuth coordinate position map in real time through the coordinate information returned from the user interface in this system to track the location of its user interface system in real time. The user interface system is mainly used for information transmission and reception on the user interface platform, and transmission networking through wireless control chip. The network transmission process of this system can be transmitted to the user interface system through the requirements of human-computer interaction system, and the information received by the system can be controlled by the server in wireless transmission.

The main function of the system is the main control system of human-computer interaction with Kinect body sense device and mobile device, and the human-computer interaction system will receive the coordinate information returned by the user interface to control the location of mobile device in real time. The human-computer interaction system mainly uses Kinect to analyze the body operation and transmit information to the user interface in the way of network packet transmission. The system controls the mobile device. In order to avoid the conflict caused by the user's similar posture, the posture swing is classified based on the basic operation. Because the system mainly cooperates with the way of sports search theme and user interaction to achieve the teaching of the system, the system uses the way of database to record the time spent by the user in the process of sports learning and store the user information into the database for analysis. In order to facilitate the user to record information, the system will confirm the identity with the lens in the human-computer interaction system device before the start of sports. Through the method of face identity confirmation, the user can no longer fill in complicated data.

3 RESEARCH ON KINECT BASED 3D MODEL CONSTRUCTION ALGORITHM OF SPORTS ACTION

3.1 Kinect 3D Sensor Image Acquisition and Attitude Estimation

The actual image acquisition distance of Kinect 3D sensor is about 1-4 meters, the horizontal field angle is 57 °, the vertical field angle is 43 °, and the frame number of image acquisition is 30 frames / second. Color image data can be generated by colorimagesteam function, while depth image data can be generated by depth imagesteam function. The auxiliary sports training system built in this paper uses Kinect three-dimensional sensor coding to achieve image acquisition and the extraction of main human joint data.

In order to get the pose of sports object, it is necessary to detect the object first, so as to facilitate the segmentation of the object. At present, the methods of target motion detection include background subtraction, frame difference and so on. In this paper, the frame difference method is used to extract and segment the target.

The experiment first obtains the image $I_{k-1}(x,y)$ of the previous frame; then obtains the differential image $D_k(x,y)$ by subtracting the gray value of the adjacent frame image, and takes the threshold t. In this paper, the threshold value t = 20, the difference image is thresholded; finally, the image is denoised by morphological processing method. In this paper, Canny edge detection algorithm is used to calculate the contour edge, and horizontal line scanning is used to further process the image, so as to finally get the coordinates of human joints. Specifically, the method of human contour edge feature extraction adopted in this paper mainly includes the following steps:

1) Gaussian filter is used to smooth the image and suppress the noise. This step can be realized by two methods: one is the method of weighting two one-dimensional Gaussian kernels twice, the other is the method of convolution of two-dimensional Gaussian kernels once.

One dimensional Gaussian kernel function:

$$G = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{x^2}{2\sigma^2}} \tag{1}$$

Two-dimensional Gaussian kernel function:

$$G = \frac{1}{\sqrt{2\pi\sigma} * \sigma} e^{\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

2) The partial derivative matrix in X and Y direction of image is calculated by the first-order partial derivative finite difference method, and the magnitude of gray value gradient of image is characterized at the same time. The expression of image gradient is:

$$T_1 = \begin{vmatrix} -1 & -1 \\ 1 & 1 \end{vmatrix}, T_2 = \begin{vmatrix} 1 & -1 \\ 1 & -1 \end{vmatrix} \tag{3}$$

$$\xi_1(m,n) = g(m,n) * T_1(x,y) \tag{4}$$

$$\xi_2(m,n) = g(m,n) * T_2(x,y) \tag{5}$$

$$\xi(m,n) = \sqrt{\xi_1^2(m,n) + \xi_2^2(m,n)} \tag{6}$$

$$\delta_\xi = \tan^{-1} \frac{\xi_2(m,n)}{\xi_1(m,n)} \tag{7}$$

3.2 3D Motion Reconstruction Based on Human Body Model

(1) 3D motion reproduction based on Kinect SDK and Ogre

Kinect SDK is a development kit corresponding to Kinect sensor, which contains various functional interfaces for users to use. Liu Zheng uses the development kit to extract the action data captured by Kinect, then uses the function nuitransformsmooth of Kinect SDK to smooth the data, and finally converts it to the quaternion action data format required by ogre for matching. The highlight of his method is that the Kinect SDK is directly used as the driver to improve the operability of the program. At the same time, the action does not need to be matched in advance. It is very efficient and the effect has been improved. Figure 2 is a set of examples of Liu's action reproduction.Among them, we can find some problems, such as (a) hand shaking in the action posture, (b) shoulder collapse in the action posture, and (c) serious distortion of the model skin eyebrows. In this paper, some improvements have been made to solve these problems.

The structure framework of this reproduction system is shown in Figure 3. The system designed by Liu Zhengcun uses Kinect SDK as the driver, while my system uses openni + nite as the driver, which is one of the main differences in our work. Other differences include the algorithm of bone mapping and action data smoothing.

(2) 3D motion reproduction based on openNI, nite and Ogre

OpenNI (open natural interaction) is a multilingual, cross platform framework developed by some industry-leading non-profit organizations, and defines openNI API for natural interaction. One of the members of the organization is primesense, which provides Kinect's core chips, so it is very conducive to the development of related applications in this project.

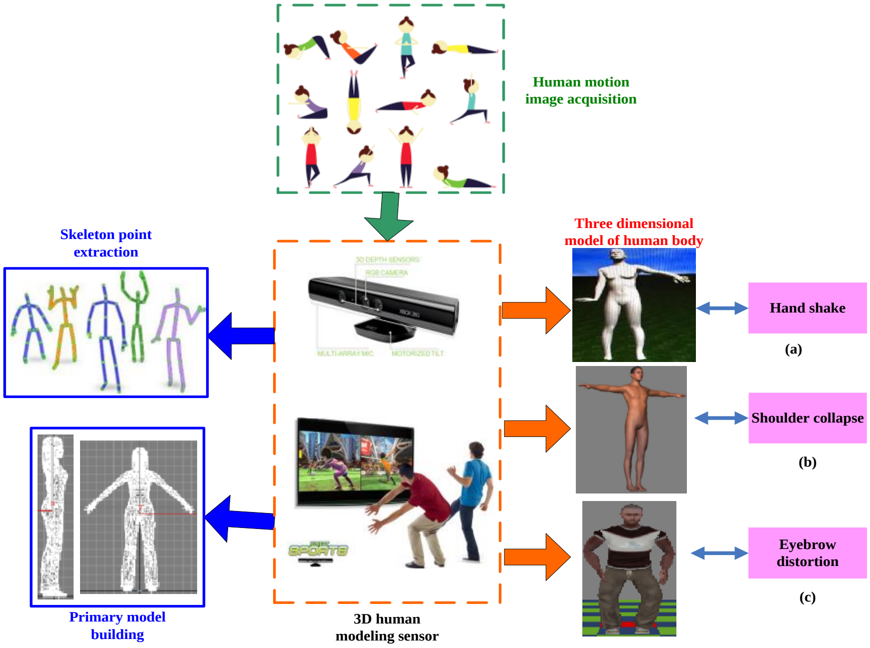

Figure 2: Reproduction effect chart based on Kinect SDK.

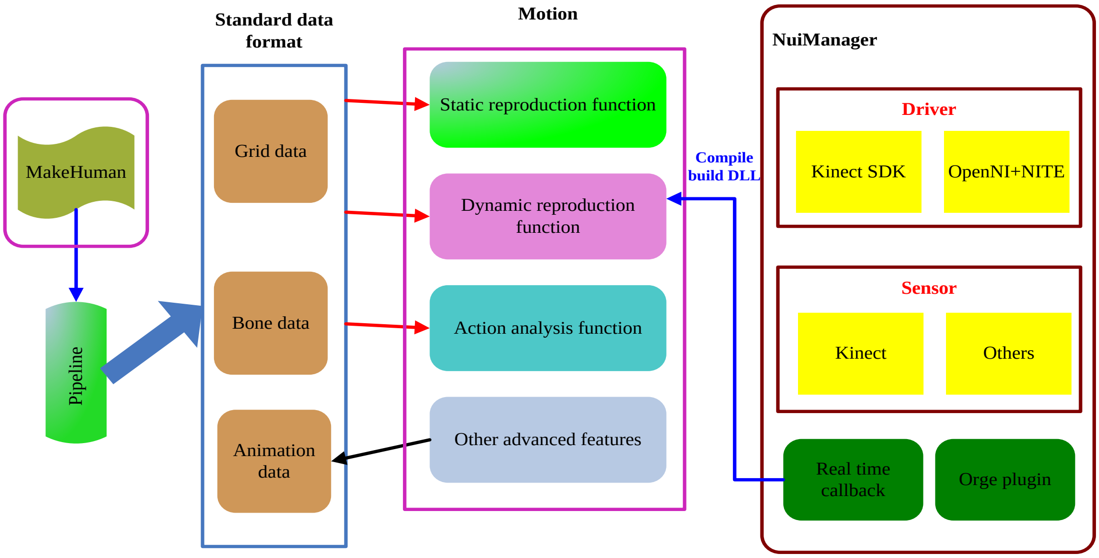

Figure 3: Structure diagram of action reproduction system.

OpenNI API contains many interfaces, which can realize the application of natural interaction. The goal of openNI is to implement a standard API to build a bridge between visual / audio sensors and visual / audio aware middleware. Figure 4 shows the system framework diagram of openni, in which the bottom layer is the visual and audio sensors, and the top layer is some applications. OpenNI is the bridge connecting the top layer and the bottom layer in the middle, providing application components such as real-time data recording, analysis and processing, such as gesture tracking of video data. OpenNI breaks the dependence between sensor and middleware, and it does not need to spend a lot of energy to solve the compatibility problem between multiple middleware to write application program with its API, which realizes cross platform performance. Developers can use the most original data to directly implement the algorithm at the upper level of middleware, without considering that the original data is generated from the specific sensor. This way provides the space for the sensor manufacturers to make full use of. No matter what the upper application is, the manufacturers only need to make their own sensor parts. In the same way, it allows developers of natural interaction to directly use the standardized data transmitted by sensors to indirectly operate the original three-dimensional data.

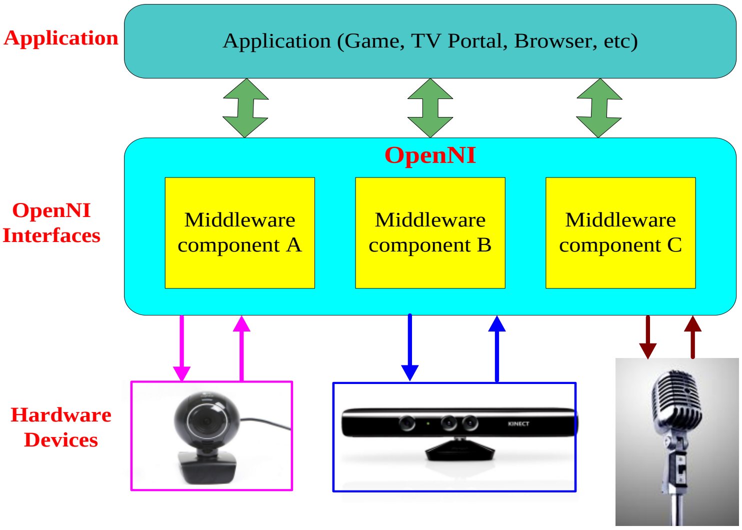

Figure 4: OpenNI framework.

Nite is a middleware in the middle layer of the above architecture level, which provides analysis algorithm. At the beginning, openNI framework didn't realize anything. It only provides an architecture. For example, we want to build a human body, but the system provides us with a skeleton to put there, and the rest needs to be implemented by ourselves. If we want to carry out some algorithm analysis, we must install Middleware that can carry out analysis. At present, the most commonly used one is nite published by primesense, which includes the currently important analysis algorithms, but it is not open-source, so there is no source code for us to analyze. Users can only develop and apply by downloading library files, and cannot understand the principle accurately. If you have other middleware in the future, you can also install other middleware to use.

On the basis of the above structure, firstly, the system smoothing algorithm is improved. By using the set smoothing (float T) interface function in nite, the smoothing degree can be determined by the control parameter T. the value range of T is 0 ~ 1. The closer to 1, the better the smoothing effect, the less near 0 smoothing.

Then, in order to alleviate the problem of bone collapse, this paper adopts a new bone mapping method, from the original makehuman body model to Kinect SDK body model to the makehuman body model to openNI body model as shown in Figure 5. Among them, makehuman bone model we selected the second life bone model in. DAE, which contains 26 joints; Kinect SDK body model contains 20 joints, openNI is even less, only 15 joints. We use the corresponding way of the main joints to map the skeleton, and debug the experiment, and finally determine the mapping way of the skeleton.

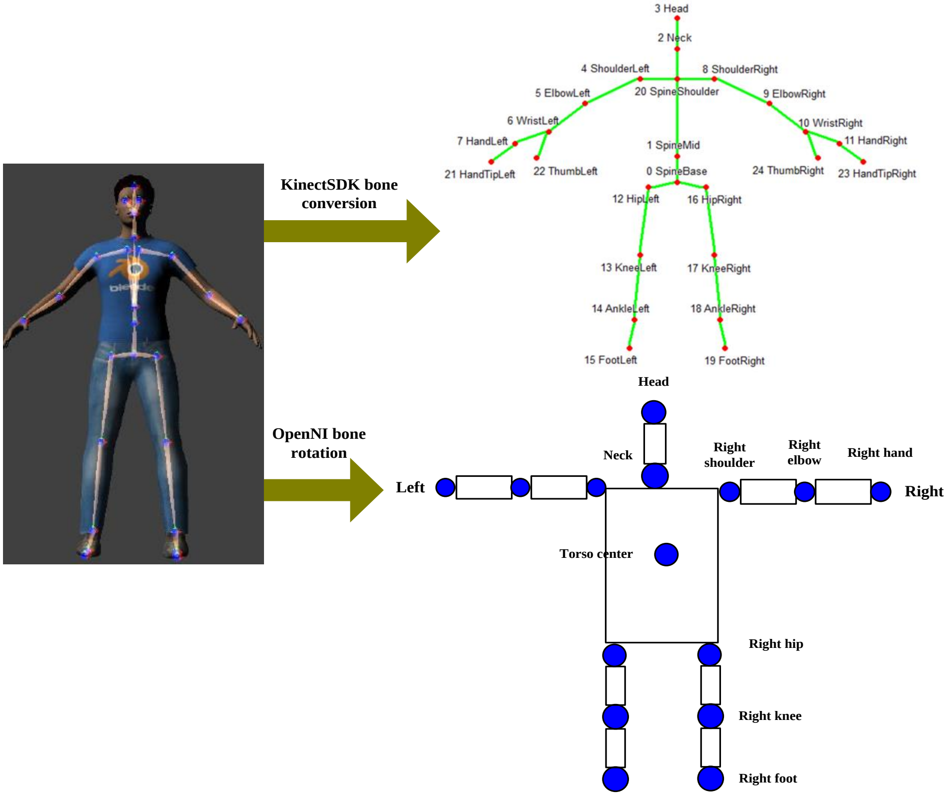

Figure 5: Makehuman bone to kinectSDK bone transformation and openNI bone transformation.

After the above skeleton mapping, the subject also debugs the initial attitude adjustment and other links. Because of the better compatibility of our openNI system framework, we don't need to develop the package link through Kinect SDK, so that makehuman and ogre can better fit Kinect, and make the rendering effect better, which is reflected in the model is more realistic, which is the reason why we choose this system, and solve the problem of model distortion. We can create a variety of personalized and realistic models for reproduction, and observe the reproduction effect in 360 degrees.

4 EXPERIMENT AND ANALYSIS

In the motion capture phase, Kinect provides the position and rotation information of 20 nodes, such as the human head, shoulders, hips and knees. At the same time, after being saved as BVH file, the file provides 63 data channels, which are 3 spatial position channels of root node and 60 rotation information channels of node. In the multi frame action sequence, 63 channels form a series of curve family. Except the absolute position of the root node has no direct comparability, other channels can be used as reference data of action similarity. The matching of all nodes (or some nodes, depending on the application needs) is the similarity calculation of human skeleton, which represents the similarity degree of two groups of actions.

They represent two groups of actions with large differences (intuitively different); and a curve family formed from all channels in an action, the former has 790 frame data (about 26 seconds). The larger fluctuation range is the limb part, while the curve near zero is the node with no offset from the parent node, such as the second spine node relative to the first spine node (there are three nodes to describe the spine).

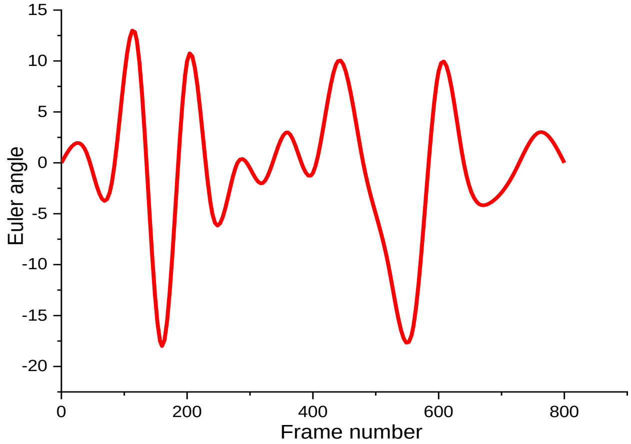

Figure 6: Single channel raw data.

Extract a channel in the curve family as shown in Figure 6, which we call raw data or standard data 1. In order to test the effectiveness of the algorithm, we choose to add quantitative noise and segmented stretch and compression (in time dimension) to the standard data. On the basis of the original data, we take out a piece of data and add noise instead of global noise. Because in general, Kinect makes more noise in joint estimation due to the occlusion of the target joint in a certain period of time, while the global data often does not have much noise.

The research group recorded a group of coaches' actions and students' actions. After removing the influence of initial value, edge, mutation and micro change data, the two-dimensional line image (taking the line width as 5) was obtained. Select the distratio of 0.9 to match and get the key point image. Then, according to the corresponding segmentation method, the global and local comparison results are obtained. According to the equal time interval segmentation method, we get the key point segmentation image. Then the key point images are analyzed to get the key point distribution of each group of images, as shown in Table 1 and Figure 7. Then the two groups of action segments are cross matched. For example, the first segment can correspond to one to five segments of the second group of actions, and the segment matching results are shown in Table 2 and Figure 8.

Parameter type	Section Ⅰ	Section Ⅱ	Section Ⅲ	Section Ⅳ	Section Ⅴ
Section of the first and second drawings	56-119	119-182	182-245	245-308	308-371
Key points of coach action	70	58	106	114	82
Key points of student action	43	61	91	93	82

Table 1: Key point allocation of equal time interval segments.

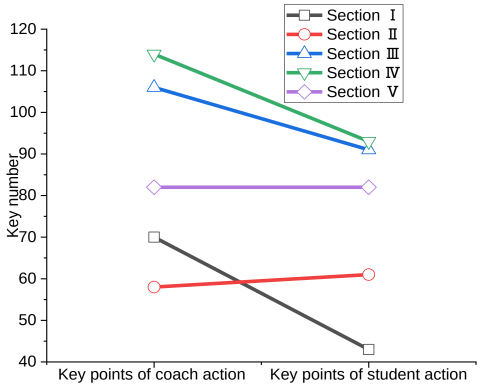

Figure 7: Key point data matching of segments.

	Section Ⅰ	Section Ⅱ	Section Ⅲ	Section Ⅳ	Section Ⅴ
Section Ⅰ	10	1	4	5	2
Section Ⅱ	7	11	4	6	2
Section Ⅲ	3	3	24	10	3
Section Ⅳ	1	13	9	15	4
Section Ⅴ	6	3	6	9	1

Table 2: Key point data matching of segments at equal time intervals.

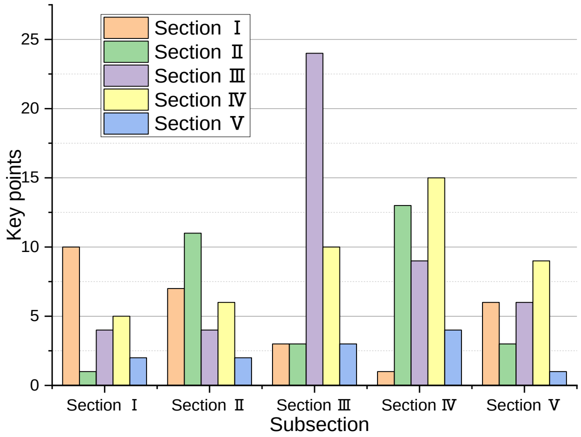

Figure 8: Key point data matching of segments.

From this, we can summarize some performance indicators of motion similarity measurement. The results are as follows. Global comparison index:

a) The overall complexity of the action: 426 key points are found from the first key point image, and 365 key points are found from the second key point image; we can get that the first image is more complex than the second image, and the action is more, that is, the coach's action is more complex and changeable than the student's action.

b) The overall matching proportion of the movements: the overall matching proportion is 137 / 365 = 37.53%, that is, the overall matching degree of the movements of the trainees and the coaches is 37.53%.

c) Zone action complexity: by comparing the key points in each zone of the same image, we can know which action in the same action sequence is more complex; by comparing the key points in the corresponding zone of the two groups of actions, we can know the block complexity of the actions of students and coaches.

d) Partition action matching proportion: divide the key points of each partition by the total key points of each partition, and we can get the matching degree of each partition (including mismatch).

e) Action advance and delay: by comparing the key points of different zones in the two groups of actions, for example, the first two zones of the students have less actions, and the last three zones have more actions, while the time of each zone is the same, so the first two zones are slow, and the last three zones are fast.

5 CONCLUSIONS

On the basis of the above research, combined with the advantages of Kinect collector in image acquisition, this paper constructs a general sports auxiliary training system, and designs some functions of the system in detail. In the input part of the system, the image sequences of athletes and coaches are input into the auxiliary sports training system through Kinect image acquisition. The auxiliary sports training system first detects the input image sequence and constructs the human body contour map, then calculates the joint angle trajectory of the athlete and the coach when the body moves according to the corresponding algorithm, finally compares the calculated results and calculates the similarity, and the resulting joint angle trajectory attitude similarity results are displayed to the system users through the output module. Through the above research, it can be seen that in the process of extracting the human sports target, using the target image feature extraction constructed in this paper has certain feasibility, can extract the key joint information of the moving human body, and can compare the difference between the standard action and the motion action with the aid of the auxiliary system.

Feng Li, https://orcid.org/0000-0002-6008-1952
Peng Li, https://orcid.org/0000-0002-2044-1167

REFERENCES

[1] Mentiplay, B.-F.; Hasanki, K.; Perraton, L.-G: Three-dimensional Assessment of Squats and Drop Jumps using the Microsoft Xbox One Kinect: Reliability and Validity, Journal of Sports Sciences, 36(9), 2018, 2202-2209. https://doi.org/10.1080/02640414.2018.1445439

[2] Carlos, Z., Chelsey, L., Pretesh, P.: The Use of Microsoft Kinect for Human Movement Analysis, International Journal of Computer Science in Sport, 54, 2015, 120-127. https://doi.org/10.5923/j.sports.20150504.02

[3] Zhang, H.; Takahashi, T.; Kageyama, Y.: Emotion Discrimination of Amusement Based on Three-Dimensional Data of Body Movements, International journal of the Society of Material Engineering for Resources, 23(2), 2018, 189-194. https://doi.org/10.5188/ijsmer.23.189

[4] Claes, J.; Buys, R.; Woods, C.: PATHway, I.: Design and Rationale for the Investigation of the Feasibility, Clinical Effectiveness and Cost-effectiveness of a Technology-enabled Cardiac Rehabilitation Platform, BMJ Open, 7(6), 2017, 1-12. https://doi.org/10.1136/bmjopen-2017-016781

[5] Kuzuhara, Y.; Kawamura, K.; Yoshitoshi, R.: A Preliminarily Study for Predicting Body Weight and Milk Properties in Lactating Holstein Cows using a Three-dimensional Camera System,

Computers & Electronics in Agriculture, 111, 2015, 186-193. https://doi.org/10.1016/j.compag.2014.12.020
[6] Chen, X.; Henrickson, K.; Wang, Y.: Kinect-Based Pedestrian Detection for Crowded Scenes, Computer Aided Civil & Infrastructure Engineering, 31(3), 2016, 229-240. https://doi.org/10.1111/mice.12163
[7] Teng, C.-H.; Chuo, K.-Y.; Hsieh, C.-Y.: Reconstructing Three-dimensional Models of Objects using a Kinect Sensor, Visual Computer, 34(11), 2018, 1507-1523. https://doi.org/10.1007/s00371-017-1425-2
[8] Zhang, Y.; Chen, C.; Wu, Q.: A Kinect-Based Approach for 3D Pavement Surface Reconstruction and Cracking Recognition, IEEE Transactions on Intelligent Transportation Systems, 19(12), 2018, 3935-3946. https://doi.org/10.1109/TITS.2018.2791476
[9] Nahangi, M.; Czerniawski, T.; Haas, C.-T.: Pipe Radius Estimation using Kinect Range Cameras, Automation in Construction, 99(3), 2019, 197-205. https://doi.org/10.1016/j.autcon.2018.12.015
[10] Precharattana, M.: Development of Computer-Assisted Instruction Lesson on Immune System Organs and Immune System Diseases, Applied Mechanics & Materials, 879, 2018, 276-283.https://doi.org/10.4028/www.scientific.net/AMM.879.276
[11] Patrizi, A.; Pennestri, E.; Valentini, P.-P.: Comparison between Low-cost Marker-less and High-end Marker-based Motion Capture Systems for the Computer-aided Assessment of Working Ergonomics, Ergonomics, 59(1), 2016, 155-162. https://doi.org/10.1080/00140139.2015.1057238
[12] Belles, C.; Pla, F.: A Kinect-Based System for 3D Reconstruction of Sewer Manholes, Computer-Aided Civil and Infrastructure Engineering, 30(11), 2015, 906-917. https://doi.org/10.1111/mice.12107