International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.12, (2016), pp.61-70
http://dx.doi.org/10.14257/ijsip.2016.9.12.07

Research on the Physical Education Teaching System Based on 3D Human Motion Capture

Liu Zhier

Jilin Institute of Physical Education, China

Abstract

At present, people get less and less exercise. This virtually led to a decline in the quality of people's physical. The best way to solve this problem is to develop a set of easy to install, inexpensive sports teaching system, so that people can work in a better way to get sports guidance. In view of this, this paper studied physical education teaching system, and the system has the function of 3D human motion capture, 3D human motion reconstruction and three-dimensional (3D) human motion analysis. First of all, this paper uses the Kinect platform to achieve three-dimensional human motion capture, and it focuses on the principle of 3D motion capture and technological development of 3D motion capture technology based on Kinect; Then, the motion data storage and 3D motion reconstruction of OGRE graphics rendering engine are described in this paper; Finally, this paper makes a research on the method of 3D human motion comparison, and designs the method of dynamic time warping (DTW) based on the problems existing in the actual situation, at the same time, this paper carries on the verification experiment. The experimental results show that: the three-dimensional human motion capture based on Kinect platform can meet our requirements for human motion capture accuracy. By comparison, the OGRE (Object-Oriented Graphics Rendering Engine) graphics rendering engine is more suitable to complete the rendering of the character model in the motion reconstruction. Most of the action comparison analysis method can only carry on the qualitative analysis to the movement study quality, but cannot carry on the quantitative analysis. And DTW algorithm can be a good solution to the above problems, and the physical education teaching system has high application value.

Keywords: *Physical education teaching system, 3D human motion capture, dynamic time warping*

1. Introduction

A complete set of popular sports teaching system involves three-dimensional (3D) human motion capture, three-dimensional human body motion reconstruction and other aspects [1]. Although there have been some high-end sports capture and analysis system for professional athletes in the market, because these systems are extremely expensive, they cannot be directly applied to the sports teaching of ordinary people [2]. At present, the sports teaching system is still in the experimental stage, and it has the following two questions [3]: (1) the human body model in the process of motion reconstruction is rough and single, and the sports teaching system is more emphasis on realistic and personalized human body model; (2) the existing research mainly focuses on the motion capture and reconstruction, which is lack of effective analysis on the problem of the movement effect and the movement.

For a long time, China's competitive sports training science and technology level is relatively low, whose teaching methods are always subjective and experience based that coaches guide the technological actions of athletes with the naked eye and experience. The athletes can grasp the key of technique only through a number of repetitive exercises, which have seriously affected the further improvement of the level of sports in China. The

ISSN: 2005-4254 IJSIP

domestic and foreign sports experts, through long-term practice and research, have unanimously recognized that the introduction of digital technology of graphics and image technology in sports training can help the athletes to master the technical essentials of the actions as soon as possible in the training process. As a result, it reduces blind repetition, greatly improves training efficiency and reduces the possibility of doing harm to the athlete, so as to achieve the best training effect. Based on the above reasons, we develop a 3D human motion imitation for sports training and analysis of the system, provide powerful technical guarantee and favorable advanced technological tools for our country's athletes preparing for the Olympic Games, to ensure that the gold medal plan for the Olympic Games, and promote the development of sports industry in China. For this purpose, this paper studies the three-dimensional human motion capture, 3D human motion reconstruction and 3D human motion analysis of the teaching system of physical education.

1.1. Research Background

The work carried out around the human body movement can be traced back to the research work for human motion perception done by the psychologist Johansson in 1973. In his experiments, attaching to the bright spot at the joint point of people, and let people in a dark environment, so only the joint is visible. Experimental results show that for static light collection, the human visual perception system cannot obtain any meaningful information. While for the spot set sequence generated in the movement, one can identify the form of movement such as walking and running, and even determine the gender of the one who is moving. Therefore, the motion recognition problem is proposed.

Almost at the same time, the theory of computer vision was put forward. It is believed that the study of computer vision is how to dynamically obtain the understanding of the corresponding scene from one or more 2D images. In this understanding system, the visual system is divided into three stages from the bottom to up, and the visual process is represented as a process of abstraction from the underlying features to the advanced features. The three stages are low level vision, middle level vision and high level vision, in which the low-level vision mainly handles the input images. While the main task of the middle vision is to restore the depth of the three-dimensional scene and 2.5-dimension information related to the scene. The task of high level vision is to restore the complete 3D information of the object on the basis of the original input image, the basic feature of the image and the 2.5-dimension image, and identify the object's position and orientation. After 20 years' technological development, computer vision technology has been rapidly developed in the theory and practice. In the meanwhile, human motion analysis has shown more and more extensive application prospects in the fields of advanced man-machine interaction, safety monitoring, medical diagnosis, etc. In consequence, people has generated strong interest in using computer vision techniques to solve human motion analysis, proceeded in-depth exploration, and achieved many important results.

1.2. Research Purpose and Significance

Human motion analysis not only has extremely important significance, but also has broad application prospects in intelligent monitoring, sports analysis, animation, human-computer interaction, video conference, medical diagnosis, virtual reality and so on. In consequence, it has attracted more and more researchers' interest. Its application is specifically embodied in the following aspects

(1) Smart surveillance system

One application is to identify the identity of a person in an access control occasion. Through the face recognition, gait analysis and so on, decide whether the one comes has the right to enter the safe area. Another application is more concerned about people's action in this scene rather than just identify people's identity, mainly from those occasions are

sensitive in requirements to the security, such as parking, supermarkets, vending machines, and traffic management, *etc.* When a suspicious behavior occurs in the scene, the surveillance system can send out warning timely for the security personnel, so as to avoid the occurrence of crimes.

(2) Virtual reality

The goal of virtual reality is to provide people with a virtual interactive world. In this virtual world, expecting to reproduce a user's gesture, we must firstly get the body posture in the real physical space, and then map it to the virtual space. This is the reason why human motion vision analysis is needed. Increasing the user's gestures, head movements and facial expressions and so on will provide users with a broader space for interaction. In addition, human motion analysis has a wide range of applications in virtual reality occasions, such as virtual game, video conference, character animation and so on.

(3) Advanced user interfaces

In the interaction of human and computer in the future, we hope that the machine can be like the human, taking the visual information as the effective complement to the voice and natural language to complete a more intelligent human-computer interaction. This requires a computer possesses, in addition to the traditional keyboard and mouse, but also independent perception of the external environment. It can extract the effective information in the environment such as detection of the existence of a human being, and further recognize the human posture and understand behavior, combined with facial expressions, body posture and gestures to proceed high-level human-computer interaction.

(4) Motion analysis

Motion analysis is mainly used in three aspects. One is the image search based on the content from the database of sports activities. The second is, in the dance, sports and other training, make use of visual methods to establish the geometry model of human body, through the joint motion analysis to guide and correct the action of trainer, which can achieve a very intuitive effect. The last one is the application in medical gait analysis. Current medical gait analysis is a research field that aims to provide diagnostic and therapeutic support.

(5) Model-based image coding

In the long-distance transmission of digital images, we are more concerned about the human face and body posture, which are also the contents appear frequently. As a result, we can proceed parameterized modeling of face and body. In image transmission, the sender only codes and transfers these parameters and the receiver restores the original face or body posture through the received parameters, so as to achieve the goal of greatly reducing the amount of data transmitted.

2. Research Contents

2.1. Literature Review

As is well known that motion capture is defined as the process of recording the movement of objects or human, and this technology has been widely used in many application fields. Q Ren studied on how to apply motion capture technology to enhance teaching quality for physical education [4]. Jia Yi Chow explores the possibility and potential of using motion capture technology to augment pedagogical practices in PE. Using examples from its implementation in sports science investigations and pilot work in a Singapore school, and discuss how such motion-tracking systems can be incorporated in schools, assisting practitioners in refining pedagogical practices. The implications of its wider use in a PE context will also be discussed [5]. FU Quan，HQ Zhao and WU

Zhuang-Zhi reviewed the application status of the motion capture technique in physical education and training, and points out the prospects of the motion capture technology in physical education and training [6]. Based on these researches, the Physical education teaching system based on 3D human motion capture is proposed.

2.2. Three Dimensional Human Motion Capture Based on Kinect Platform

Three dimensional human motion captures is a very important part of this project, and the motion capture data directly determines the effect of the subsequent 3D motion reconstruction. Therefore, this paper will focus on the three-dimensional motion capture technology. For a given video image, how to recover the corresponding 3D human pose information is an essential and fundamental problem of human motion reconstruction, which is also an important step for the generation of 3D human motion. We use contour-based method to reconstruct 3D human body posture. Specifically, first of all, we use motion segmentation method to extract the human body's 2D contour information in the image and use a generic parameterized virtual human model to customize the human body model. Then extract corresponding camera information from the motion, and in this perspective use personalized human body models to proceed action driven and 2D projection of the data in the moving 3D motion database, so as to obtain 2D body contour database. Finally, according to the contour similarity criteria, proceed contour matching in the 2D motion contour, and find the contour data that has the highest contour similarity with the contour data to be reconstructed. The 3D pose information the contour corresponds to is the initial reconstruction results. This is the basic idea of human posture reconstruction in this paper.

2.2.1. Principle of 3D Motion Capture: June 2010, the Microsoft Corp released the XBOX360 game console peripherals, and named Kinect. Kinect completely changed the single mode of original game operation, and it paid more attention to the concept of human-computer interaction. And it had a real-time dynamic capture, speech recognition, community interaction, image recognition and other functions. It can get three kinds of different signals at one time. And they are color video information, sound signal and 3D depth information [5]. If we want to use Kinect for three-dimensional motion capture, we must put the focus on how to get the human body's three-dimensional motion information. Firstly, Kinect uses Coding Light technology to obtain the depth information of the human body. And then, it carries out different marks on each part of the human body through the algorithm of random decision forest. Finally, it extracts the skeleton information of the human body from different marks, which can save the human body's movement data. The 3D motion capture recognition based on depth information can be summarized into the following points:

1) Depth data is processed in the rendering pipeline. The data are matched by the decision tree algorithm, and then give an inference of the human body segmentation region.

2) Once all the segmented regions are identified by the labeled data, the sensor can identify the distribution of the joint points of the human body.

3) Calculation of 3D view point sensor joint is inferred from the top, front and left.

4) Based on the proposed joint point and 3D view, the sensor begins to track the body's skeletal and body movements.

2.2.2. Technological Development of 3D Motion Captured Technology Based on Kinect: To carry out the development of related applications based on Kinect, we must have to understand the Kinect SDK (Software Development Kit) [6]. It is a Microsoft Corp for Kinect developers who provides software tools development kit and the development package set Kinect driver and related development routines in one. Skeleton tracking

technology is one of the core technologies of Kinect, and the most important of the skeleton tracking method based on Kinect SDK is to call the Skeleton Class, Skeleton Frame Class, Skeleton Stream Class and Collection Joint. The human skeleton tracked by Kinect consists of 20 joint points, and they formed 19 bones together. Skeletal tracking determines whether the bone is traced. SDK Kinect provides a Skeleton Tracking State enumeration type to check the correlation value. There are three possible states of skeletal tracking: tracked, inferred and not tracked. The bone smoothing process is mainly for the setting of the skeletal smooth parameters, which requires extensive experience and constant debugging, and the bone smoothing parameters mainly include the correction value, the jitter radius, the maximum jitter radius, the prediction of the frame size and the smooth value. Through the reasonable setting of the above parameters, the program can be a larger bone shaking place to implement the difference of treatment, so that the bone does not appear larger jitter.

2.3. Personalized 3D Motion Reconstruction Based on OGRE Rendering Engine

Three dimensional reconstructions are a technique which is suitable for the computer representation and processing of three-dimensional objects in space. It is also the basis of processing, operating and analyzing the properties of the computer in the computer environment. And it is also the key technology to establish the objective world (virtual reality) in the computer.

2.3.1. Action Data Storage for Skeletal Representation: The main storage formats of motion data mainly include the rotation matrix, Euler angle and Quaternion. Each format has its own advantages and disadvantages. At the same time, they can be converted to each other. The rotation matrix usually describes the orientation of a coordinate system by means of the base vector of the coordinate system, and it uses another coordinate system to describe these basic vectors, and then we can get a 3 * 3 matrix. By this matrix, we can get the current position. The basic idea of Euler angle is to represent an angular displacement by the sequence of three rotations about the vertical axis in the three dimensional space. Although the selection of the three axes is not fixed, people usually use the Descartes coordinate system and rotate it in the most usual order. The Quaternion contains some advantages of the rotation matrix and Euler angles, but it is relatively difficult to learn. Its invention is a further extension of the complex field.

2.3.2. 3D Motion Reconstruction under OGRE Graphics Rendering Engine: Object-Oriented Graphics Rendering Engine (OGRE) is an object oriented graphics rendering engine. And it is a scene oriented 3D graphics rendering engine, which is developed by the C++ language. It can help developers to use hardware accelerated 3D graphics system quickly and directly to develop the relevant application. Taking into account the needs of the individual human body, we select MakeHuman to obtain personalized human model. After obtaining the 3D human motion data and skeleton model, the model can be used to reproduce the movement data. The 3D motion reconstruction flow chart below the OGRE rendering engine is shown in the Figure 1:

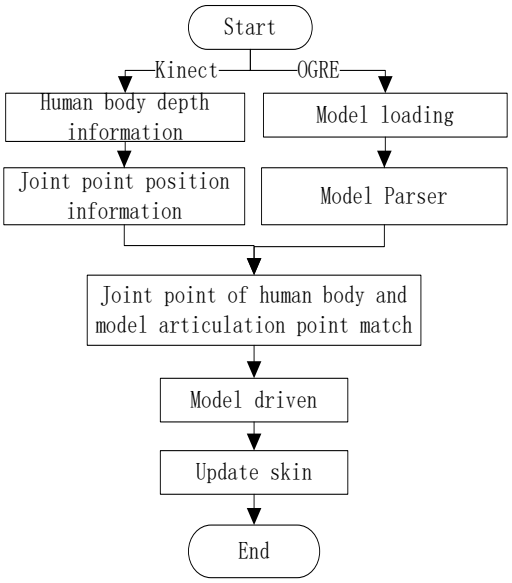

Figure 1. Motion Reconstruction Flow Chart under the OGRE Rendering Engine

2.4. Research on the Method of Three Dimensional Human Motion Comparison for the Physical Education Teaching System

Dynamic time warping (DTW) algorithm needs to use the concept of several distance measures. Before introducing the DTW algorithm, this paper first introduces the concept of several distance measures. They are Euclidean distance, Chebyshev distance and absolute distance. In the DTW algorithm, we usually use Euclidean distance measure when calculating the distance between the two vectors, so, we only introduce the calculation formula of Euclidean distance measure. Its formula is as follows:

$$D(X,Y)=\frac{1}{K}\sqrt{\sum_{i=1}^{k}(x_i-y_i)^2} \tag{1}$$

DTW algorithm can be a good solution to the problem of different length of time sequence. It can be used to construct the Euclidean distance matrix, and use the regular function to find the shortest path, which can calculate the shortest distance between two different length sequences. The principle of data correlation based on Euclidean distance is: the method is based on the Euclidean distance of the data points of the two sequences in a strict chronological order (it does not take into account the consistency e waveform; the time series is not equal to the case).

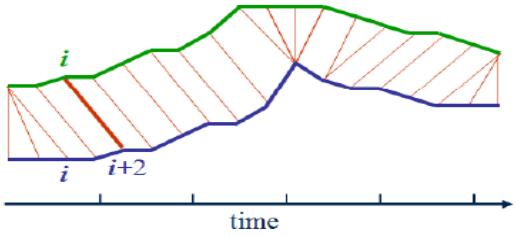

Figure 2. Principle Diagram of Data Comparison based on DTW Algorithm

Figure 2, shows the data comparison principle based on DTW algorithm. It can be seen from the graph, the data comparison method based on DTW algorithm takes into account the similarity of waveform, and it is not strictly in accordance with the time point for the comparison of the two sequences. It can be no one corresponding situation.

The essence of DTW is to use dynamic programming techniques to transform a difficult global optimization problem into a number of local optimization problems, and achieve the optimal decision through a step by step decision. It tries to find a regular function, $m = \phi(in)$, which maps the time axis N of the test vector i to the time axis j of the reference template M, and make the function to meet the Formula 2:

$$D = \min_{\varphi(i_n)} \sum_{i_n=1}^{N} d(T(i_n), R(\varphi(i_n))) \tag{2}$$

Among them, D is the matching path between the two vectors in the optimal time warping. Because the distance of the two vectors is calculated by DTW which to find the optimal matching path, the regular function of the minimum cumulative distance of the two vectors matching is obtained. This ensures the maximum similarity between them.

The essence of dynamic time warping can be summed up in the following two steps: (1) Constructing a distance matrix: suppose there is a M dimension standard reference template from X= {x(1), x(2)...x(m)} and a N dimension template to be tested Y= {y(1), y(2)...x(N)}. And N≠M, each component of X and Y is a number or a multidimensional vector. But if it is a vector, each component must have the same dimension. Because N≠M, if the traditional Euclidean distance comparison method is used to calculate the similarity between the template X and the template Y, the result will be a large deviation. Therefore, we use the DTW algorithm to calculate the distance between each component of X and the Y of each component, and form an N *M matrix. (2) Finding an optimal path in the distance matrix formed in the first step. The path should be derived from the distance matrix on the left, stop at the bottom right of the distance matrix, and it should make the minimum accumulation distance of the whole path. The optimal path is the nonlinear relationship between two time scales, which can be obtained by dynamic programming method. In the operation process, we need to use the local decision function and distance matrix to form another matrix, which is plus distance matrix. In order to find the best path, we need to keep in mind that each time the local decision is selected, which is the local path. And then, starting from the bottom right corner of the local decision function, we can find the actual minimum cost path.

3. Experimental Results Based On DTW Comparison Analysis Method

Figure 3, is the interface diagram of the action analysis and comparison, which mainly include the following main functions of the action file recording, the action file playback, and the action learning quality evaluation. At the same time, in order to facilitate the visual comparison, we added the color image real-time display, and the depth image real-time display function in the interface.

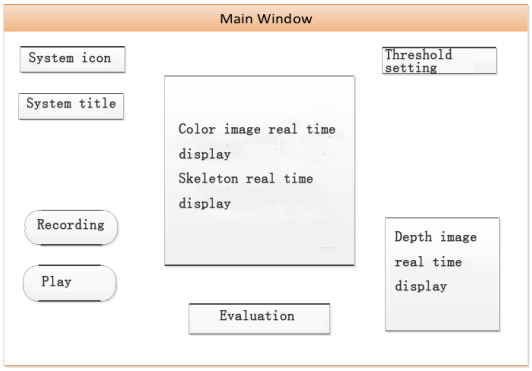

Figure 3. System Interface Structure Diagram

This paper made a simple set of movements, and the recording process will be color images and depth of the image of real-time feedback. In the recording process, there will be color images and depth images of real-time feedback. It should be noted that the action recording needs to be carried out in an environment of open and less interference, while it should allow the Kinect to capture the joint point of the body at the same time, to improve the accuracy of motion alignment. Figure 4 and Figure 5, are two groups of actions to compare the results of the experiment.

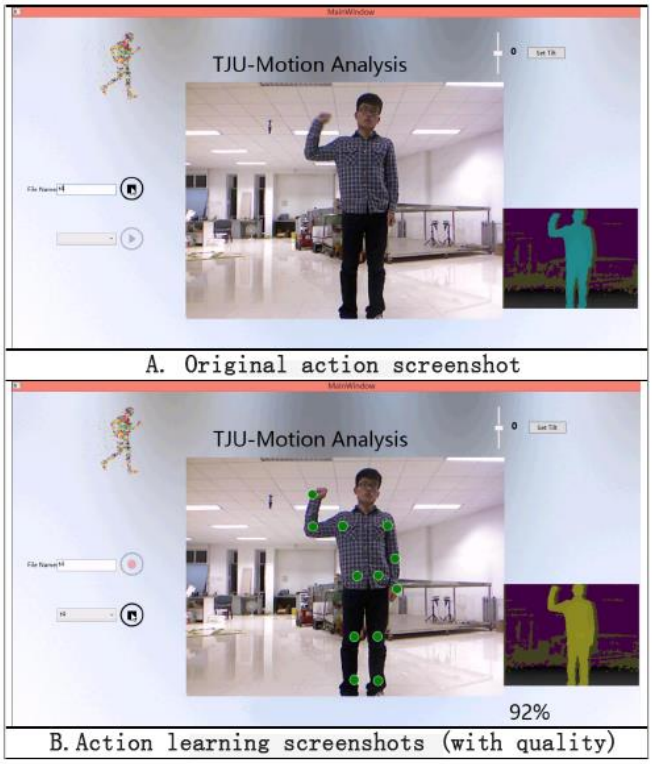

Figure 4. Action Comparison Test Results (1)

Figure 4, shows a good action learning quality result of comparative analysis. we can see from the following two aspects of the quality of action learning: (1) Color video window shows that the vast majority are green joints; (2) The lower right corner of the screen gives the similar degree of the two groups of movements, and their similarity is 92%, which indicates that the two groups are basically matched.

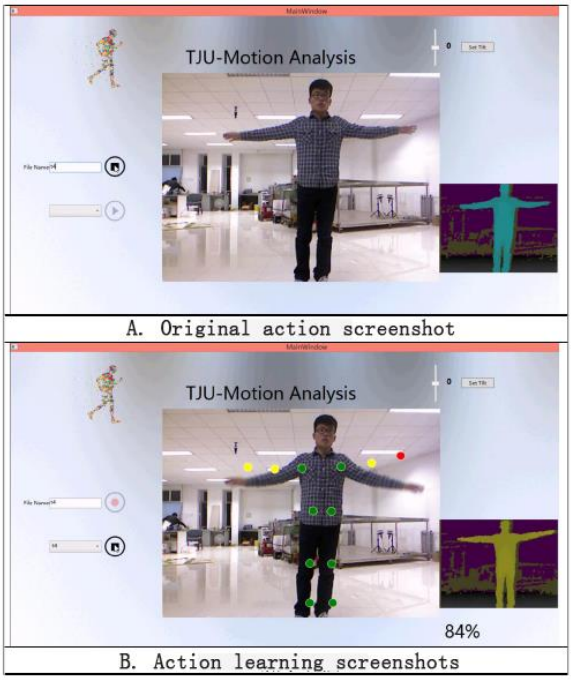

Figure 5. Action Comparison Test Results (2)

Figure 5, shows the comparative analysis result of the poor quality. We can see from the chart that in the action learning process, the deviation of hand joint action of the left arm is larger; the deviation of the left arm elbow is slight; the deviation of the right arm elbow is slight; the deviation of the right hand joint is slight. The action learning of other body joints is basically in place. We can find out the deficiency of the action and correct them by using different colors according to the different degree of joint learning.

4. Conclusion

4.1. Conclusion

Research results show that: in the design of physical education teaching system, the three-dimensional human motion capture based on Kinect platform can meet our requirements for human motion capture accuracy. By comparison, the OGRE graphics rendering engine is more suitable to complete the rendering of the character model in the motion reconstruction. At the same time, most of the action comparison analysis method can only carry on the qualitative analysis to the movement study quality, but cannot carry on the quantitative analysis [7]. DTW algorithm can be a good solution to the above problems, and it can solve the problem that the length of time series of the coach action is not equal to the length of the student in the course of action comparison, so that it can give an objective evaluation of the quality of the action learning of the students. And it can accurately calculate the global similarity of the two series of unequal length sequences. All of these indicate that the physical education teaching system has high application value.

4.2. Expectations

In the work of this paper, we have primarily achieved the realization of human motion capture. But it should be said that this work is far from perfect, a lot of research ideas are not deep enough, and it still needs development in the system function and key technology. Generally speaking, the key technology research in the later period is embodied in the following aspects:

The first one is motion segmentation. The so-called motion segmentation is to detect the area where the human locates from the image sequence, and separates it from the background by binarization. The motion segmentation is a typical difficult problem based on human motion reconstruction of the contour line of the foundation. At the same time, it is also a problem in the research field in computer visual sense for body contour extraction. Human contour extraction of some simple movements, such as walking, running, jogging and so on has been able to achieve some good results [8]. However, it is very difficult to extract the contour of the human body in the dynamic complex background, and so far it has not achieved very good results. The reason is that the image capturing in dynamic environment is affected by various factors, such as changes in the weather, light change between objects, environment and conditions change, the chaos disturbance and so on, which bring difficulty for the accurate and effective human contour extraction. Dynamic and complex background motion segmentation will be an unavoidable problem. In consequence, a fast and accurate motion segmentation algorithm in a dynamic scene is a good technical guarantee for the general performance of the motion reconstruction system.

The second one is the simulation of human motion reconstruction. Taking the motion database truly captured as the foundation, at the same time, the reconstruction process strictly following the spatio-temporal model based on the motion database, our human motion reconstruction technology in a certain extent ensures the reality of the results. Even so, the process of contour matching and deformation based on contour similarity may affect the motion correlation and the physical reality of motion data. In this case, even if the visual effect tends to be consistent, it is hard to say that the reconstructed motion is consistent with the real results. Therefore, the simulation of the movement of the reconstruction results of human motion based on kinematics and dynamics simulation is a very meaningful follow-up study. Additionally, in any case, the development of computer vision technology will increasingly become an important technology for the benefit of people's lives, which is beyond doubt.

References

[1] J. Bernard, N. Wilhelm and B. Krüger, “MotionExplorer: exploratory search in human motion capture data based on hierarchical aggregation”, IEEE Transactions on Visualization & Computer Graphics, vol. 19, no. 12, **(2013)**.
[2] J. Selmer, “MotionFlow: Visual Abstraction and Aggregation of Sequential Patterns in Human Motion Tracking Data”, Biochemical & Biophysical Research Communications, vol. 1, no. 22, **(2016)**.
[3] X. Li, W. Hu and C. Shen, “A survey of appearance models in visual object tracking”, Acm Transactions on Intelligent Systems & Technology, vol. 4, no. 4, **(2013)**.
[4] Q. Ren, “Applying Motion Capture Technology to Enhance Teaching Quality for Physical Education”, Advances in Information Sciences and Service Sciences, vol. 5, no. 3, **(2013)**.
[5] J. Y. Chow, C. W. K. Tan, M. C. Y. Lee and C. Button, “Possibilities and implications of using a motion-tracking system in physical education”, European Physical Education Review, vol. 20, no. 4, **(2014)**.
[6] F. Quan, H. Q. Zhao and W. ZhuangZhi, “Application of Motion Capture Technology in Physical Simulation”, Journal of Shanxi Datong University, **(2013)**.
[7] S. Dubuisson and C. Gonzales, “A survey of datasets for visual tracking”, Machine Vision & Applications, vol. 27, no. 1, **(2015)**.
[8] Q. S. Yan, “Survey of Visual Tracking Algorithm”, Computer Science, **(2013)**

