Development significance and structure analysis of sports video image analysis system based on VR

Xianming Chen*
Wuhan Railway Vocational College of Technology, Wuhan, 430205, China

*Corresponding author: aisyfsb706@163.com

Abstract. Modern science and technology is highly differentiated and concentrated, and has also entered a new stage of multidisciplinary comprehensive utilization. Especially, VR technology has played a great role in competitive sports training. Intelligent video analysis system is a key point of modern video development. With the development of computer technology and the gradual progress of video technology, intelligent video analysis system has broad development space. In this paper, a key algorithm SISA (Sports Image Segmentation Algorithm) is designed. The algorithm improves and optimizes the semantic segmentation algorithm by using hole convolution, pyramid pooling and trainable conditional random field structure, and uses key node detection algorithm to enhance the segmentation effect of small objects. The common methods to realize virtual human animation are manual marking, motion capture and automatic recognition. In order to reduce the workload of manual marking, avoid the high cost of motion capture and make up for the precision error of automatic recognition, it can be practically applied to the guidance of physical training.

Keywords: VR; Sports video image analysis system: Key node detection; key frame

1. Introduction
With the promotion of social development, sports competition has been favored by people, and sports training has also achieved rapid development under the influence of the update and progress of science and technology. It can be said that sports competition achievements are closely related to the development of science and technology [1]. Sports need not only similarities but also similarities. In actual scenes, it is usually necessary to record athletes' movement process and convert it into 3D model for calculating energy conversion efficiency. Because of the development of sports, the audience of various teaching and auxiliary software is increasing, but because of various limitations, it is impossible to calculate and analyze the energy conversion during sports.

It can be predicted that with the deepening of the research on the sports training assistant analysis system, its functions will be constantly improved. The application of sports training assistant analysis system in various sports training can improve the high-tech content of each training, and promote coaches and athletes to change from the traditional sports training mode to the more scientific and vivid
digital and simulated sports analysis mode [2-3].

Intelligent video analysis technology mainly adopts back-end PC server processing mode and front-end DSP processing mode. DSP is a high-speed digital signal processor, which performs real-time processing at the front end of video acquisition [4]. VR technology is used for simulation training to realize the human motion measurement method of sports training methods from traditional human eye observation to high-precision motion capture and analysis; Sports video image analysis system plays an extremely important role in the field of sports. It can accurately and real-time measure the movement curves of various parts and the relative rotation angles of joints during athletes' movements, which helps coaches correct athletes' wrong movements, prevent sports injuries and improve sports performance.

2. Present situation of sports video image analysis system
In recent years, with the rapid development of science and technology and the wide application of high-speed and high-definition cameras, the function, quality, test speed and test accuracy of sports video image analysis system have been greatly improved. The measurement of sports video image analysis system has developed to be marked by the synchronization of three-dimensional high-speed images, three-dimensional dynamic force measurement and multi-channel electromyography measurement, which makes the kinematics and biomechanics of human body build a clear mechanical image in three-dimensional space [5]. Data smoothing, barycenter calculation and velocity parameter form discussed in the past have been completed by standardized software.

The research on automatic recognition of joint points of human motion images on films and videotapes has achieved initial success and has been applied to practice, but this technology relying on gray recognition can only be realized by pasting obvious markers on joint points of human body, and the field test, especially when joint points are blocked, can only rely on manual interpretation of joint points [6]. In addition, the standard manikin is not completely suitable for all kinds of sports, especially for some sports with special requirements for body shape. Therefore, the individualized model of geometric distribution of human body mass is also an urgent problem to be solved.

3. VR technology and sports system simulation

3.1 VR technology
Essentially, VR technology is an advanced computer user interface technology, which provides users with various intuitive and natural real-time interactive means such as vision, hearing, touch, taste, etc., so as to maximize the convenience of man-machine interactive operation, without the need for tedious keyboard input, so as to improve the working efficiency of the whole system. There are three basic characteristics [7]: immersion, Interaction and Imagination, which is commonly called "3I". The relationship is shown in Figure 1.
3.2 Sports system simulation

Sports system simulation point to a all round and integrative cross-experimental discipline, which uses the theory and method of modern system simulation, intimately combines sports practice, and absorbs and draws lessons from the study outcome of other subject to conduct comprehensive and systematic research on sports problems and serve sports practice. It reproduces the teaching experience of physical education teachers, the training intention of coaches, the organization scheme of managers and the training process of athletes through computer simulation technology, so as to explain, analyze, predict, organize and evaluate the sports system. Nowadays, the study hot spots of system simulation mainly include object-oriented simulation method, qualitative simulation, distributed interactive simulation, visual simulation, multimedia simulation, VR-based simulation and intelligent simulation.

4. Structure analysis of sports video image analysis system based on VR

4.1 Overall structure design of the system

In sports training, sports video image analysis system is based on computer platform, which can collect sports videos in different operating systems and scientifically process video images. The video analysis system can be installed with a variety of systems, namely Windows system and Linux system, and can load sports video process management and control information programs through embedded Linux kernel structure.

Send the root file system of embedded Linux system based on CAN, and comprehensively store the motion video frame sequence in relevant directories. The overall structure design of the system mainly includes three modules, namely, user control module, video information acquisition module and output module. In the output module, the GUI man-machine interaction system is used as the carrier to analyze the video and image information of sports training in detail, and download the image processing program in time in the development board, so as to promote the video analysis of sports training and the system software drive [8]. Based on this, the general structure model of sports video image analysis system in sports training is designed, as shown in Figure 2.

![System overall structure model](image)

Fig.2 System overall structure model

4.2 Design and Implementation of SISA Neural Network

SISA (Sports Image Segmentation Algorithm) applies two deep learning models to the semantic extraction of characters and clubs, semantic segmentation model and key node detection model in sports video. Semantic segmentation model aims to extract the characters and related sports equipment in each frame of sports video [9]. The key node detection model aims to double check the semantic segmentation model, and ensure the extraction effect of small semantics such as clubs.

Only the rough boundary of silhouette can be obtained by using deep neural network and larger receptive field. At this time, the relatively small semantics such as golf clubs will be covered by the rough boundary of background semantics. In order to make the boundary of golf club clearer, a fully connected conditional random field layer is added after the output of neural network to recalculate the correlation between pixels in order to optimize the silhouette boundary.
The conditional random field model is composed of two-dimensional energy functions, and its energy equation can be obtained for a label value:

\[ E(x) = \sum \varphi_u(x_i) + \sum_{i,j} \varphi_p(x_i, x_j) \]  

(1)

The unary potential function \( \varphi_u(x_i) = -\log P(x_i) \), where \( P(x_i) \) represents the probability that the pixel point \( i \) is the label \( x \), that is, the classification result of the current pixel point.

After a picture passes through convolutional neural network, every pixel has been preliminarily classified, so the input of one-dimensional potential function in conditional random field is the output of neural network. The binary potential function \( \varphi_p(x_i, x_j) \) represents the probability that the labels of pixel \( i,j \) are \( x_i \) and \( x_j \), that is, it represents the possibility that \( i,j \) exists at the same time, so \( \varphi_p(x_i, x_j) \) can be characterized as:

\[
\varphi_p(x_i, x_j) = \mu(x_i, x_j) \left[ \omega_1 \exp \left( -\frac{\|P_i - P_j\|^2}{2\sigma^2} - \frac{\|I_i - I_j\|^2}{2\sigma^2} \right) + \omega_2 \exp \left( -\frac{\|P_i - P_j\|^2}{2\sigma^2} \right) \right]
\]  

(2)

Among them,

\[
\mu(x_i, x_j) = \begin{cases} 0, & x_i \neq x_j \\ 1, & Otherwise \end{cases}
\]  

(3)

This shows that the correlation between two nodes will be recalculated only when the labels of the two nodes are different. Two Gaussian kernels in binary potential function represent different feature spaces. The first Gaussian kernel is based on both the position of pixels and the RGB color of pixels. The idea of this Gaussian kernel function is that pixels with similar spatial distance have a higher probability of being the same object. The second Gaussian kernel is only based on the position information of \( i,j \) pixels. The first Gaussian kernel function will make pixels with similar position and color information have consistent labels; The second Gaussian kernel function can make the pixels with similar positions have consistent labels, thus making the extracted semantics smoother and removing some isolated small areas in the segmentation results.

The purpose of the key node detection network is to serve the semantic segmentation network and solve the problem that the semantic segmentation network is not effective enough to extract the semantics of small objects. Sometimes, the semantic extraction of clubs is unclear, which will be separated by the background, and can't form a complete closed communication area with the characters. At this time, the problem is that the denoising module will clean up the broken clubs as noise. In order to ensure the accuracy of club extraction and avoid the problem of de-noising errors in the de-noising module, this paper uses the key node detection model with higher accuracy to double guarantee the club semantic extraction. The residual block structure is shown in Figure 3.
The VR-based sports system simulation can build virtual training equipment and scenes in the light of the characteristics of given sports events and training demand. The stereo image produce by computer technology turns a dark training room into a competition scene. Athletes can truly experience the feeling of driving a sled to slide down the runway on this device. Laboratory equipment can detect athletes' physical condition at any time, perform technical analysis and functional diagnosis, predict various possible problems in the competition, formulate the best strategy and design the best route, so as to achieve the goal of improving athletes' performance. By constructing virtual competition venues, practice scenes and competition scenes, player can train in a specific room, which is of special significance for athletes in recovery stage after injury, as well as for weather, venue conflicts and inconvenient use of away venues.

For the sake of make the animation of virtual human more natural and realistic, many technologies have been well-developed in computer animation, among which the conventional motion control methods are parametric key frame technology and process animation technology. In addition, simulation technology based on physics and motion capture technology and motion synthesis technology which are popular in commercial products are commonly used [10].

1) Parameterized key frame technique

Key frame technology is the earliest way to control the motion of virtual human, and its concept originated from the early cartoon production of Walt Disney Company. When computer technology has not yet entered this field, a lot of tedious intermediate frame work is done manually by assistant animators. Because the key frame interpolation does not consider the physical attributes of human body and the relationship between parameters, the motion obtained by interpolation is not necessarily reasonable, and usually requires careful adjustment by animators. However, because the key frame technology is easy to use, it is still the most commonly used animation generation method.

2) Interpolation algorithm

Interpolation algorithm plays an important role in generating the final animation by key frame technology. The simple interpolation method is linear interpolation method. Animation using this method often changes rapidly, which is mainly due to the discontinuous speed of moving objects. When using key frame technology to realize motion control of virtual human, the following two problems must be solved: position interpolation and orientation interpolation. The key frame position interpolation can be realized by spline interpolation and velocity curve interpolation. Orientation
interpolation can be realized by quaternion interpolation method.

(3) Quaternion interpolation

In computer graphics, when rotating around an axis that does not coincide with the coordinate axis, it can be obtained by combining translation and coordinate axis rotation. In order to get the required compound matrix, the traditional method is to first get the rotation matrix of the axis moved from the selected rotation axis to another coordinate axis, then get the rotation matrix of the axis according to the specified rotation angle, and finally get the inverse matrix sequence to change the rotation axis back to its original position. The calculation is complicated. A more effective way to obtain the rotation of a given axis is to use quaternion representation of rotation transformation, which requires less storage space than the matrix, and it is easier to write out the quaternion process of transformation sequence, which is especially important for motion interpolation animation which needs complex motion sequence and two positions of a given object.

4.4 Interactive design

The interaction of sports video image analysis system can be divided into model training module interaction design and prediction module interaction design. Among the three main modules in the sports video image analysis system, the data processing module is the bridge between the training module and the prediction module, so the interaction design between these two modules has been completed by the transfer of the data processing module.

The prediction flow of the prediction module is a process of predicting and analyzing the motion video by using the model generated by the training module. Prediction module is mainly a module that interacts with users. In the process of prediction module, the functions of data preprocessing and data forwarding need to be completed through interaction with data processing module. ModelTrain is responsible for data distribution and function scheduling, the TminProces subclass is responsible for training the semantic segmentation model and the key node detection model, and the DataProcess subclass mainly preprocesses the data in the training process of the two models. The model training module ModelTrain class attribute definition is shown in Table 1.

| Attribute name   | Type                   | Access authority | Explain                                           |
|------------------|------------------------|------------------|---------------------------------------------------|
| id               | string                 | private          | Universal identifier                              |
| trainType        | int                    | private          | Training Type: Segmentation/Key Point Detection   |
| segData          | vector < Picture >     | private          | Semantic segmentation data                        |
| keyNodeData      | vector < Picture >     | private          | Key node detection data                           |
| tp               | TrainProcess           | private          | The ModelTrain sub-module                         |
| dp               | DataPreprocess         | private          | is used for different training tasks and data processing |

SDL diagram of model training module is shown in figure 4.

The model training module is mainly divided into two main stages: data processing stage and model training stage, namely DataProces and TrainProces. The main function of the data processing stage is to request training data from the database and perform corresponding preprocessing on the data according to the training target. In the training stage, the main functions are to initialize the deep neural network of SISA algorithm, adjust the optimization parameters and train the model. In the training stage, the work mainly centers on the construction of SISA algorithm’s deep neural network. It is difficult to show the process of optimizing and modifying the network according to the needs.
5. Conclusion
To effectively guarantee the scientificity and perfection of sports training guidance, measures must be taken to improve the efficiency and quality of guidance. Therefore, it is imperative to design and develop a new sports video image analysis system in sports training. Virtual human animation, editing of joint motion curve of virtual human and video synthesis technology are studied, and human posture is reconstructed and continuous human motion animation is generated. These works are the intersection of computer vision and computer graphics, which involves the research of computer vision, pattern recognition, artificial intelligence, cognitive psychology, graphics and other fields. Applying VR technology to simulate and train competitive sports. It can not only improve the effect and scientific and technological content of sports training rapidly, but also expand the application range of system simulation and promote the rapid development of VR technology. Among them, informationization and intelligence are a direction for the future development of large stadiums and gymnasiums. Intelligent video analysis has important application value in sports training.

References
[1] X Huang, Du L. Fire Detection and Recognition Optimization Based on VR Video Image. IEEE Access, no. 99, pp. 1-1, 2020.
[2] Zhou M. Feature Extraction of Human Motion Video Based on VR Technology. IEEE Access, no. 99, pp. 1-1, 2020.
[3] Zhang Xiaona. Application and Analysis of VR Technology in Digital Video Special Effects. China cable TV, vol. 000, no. 007, pp. 750-752, 2019.
[4] Ozcinar C, Cabrera J, Smolic A. Visual Attention-Aware Omnidirectional Video Streaming Using Optimal Tiles for VR. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, no. 1, pp. 1-1, 2019.
[5] Okolo S. Can VR Improve the Health of Older Sports Fans?. IEEE Potentials, vol. 38, no. 3, pp. 17-19, 2019.
[6] Oosterom M, Meershoek P, Welling M M, et al. Extending the Hybrid Surgical Guidance Concept with Freehand Fluorescence Tomography. IEEE Transactions on Medical Imaging, no. 99, pp. 1-1,
2019.

[7] Lu Y, An S. Research on sports video detection technology motion 3D reconstruction based on hidden Markov model. Cluster Computing, vol. 23, no. 3, pp. 1899-1909, 2020.

[8] Zhang Z, Cen Y, Zhang F, et al. Cumulus cloud modeling from images based on VAE-GAN. VR & Intelligent Hardware, vol. 3, no. 2, pp. 171-181, 2021.

[9] Wang X, Du J, Zou R, et al. Key node identification of wireless sensor networks based on cascade failure. Modern Physics Letters B, no. 1, pp. 2050394, 2020.

[10] Wood G, D J Wright, D Harris, et al. Testing the construct validity of a soccer-specific VR simulator using novice, academy, and professional soccer players. VR, no. 1, pp. 3, 2020.