

Research Article

Digital Twins by Physical Education Teaching Practice in Visual Sensing Training System

Xinran Liu and Ji Jiang 

Sports Faculty Department, Liaoning University, Shenyang, Liaoning, China

Correspondence should be addressed to Ji Jiang; jiangji@lnu.edu.cn

Received 18 October 2021; Accepted 24 November 2021; Published 10 January 2022

Academic Editor: Zhihan Lv

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

The paper expects to improve the efficiency and intelligence of somatosensory recognition technology in the application of physical education teaching practice. Firstly, the combination of induction recognition technology and the Internet is used. Secondly, through the Kinect sensor, bone data are acquired. Finally, the hidden Markov model (HMM) is used to simulate the experimental data. On the simulation results, a gait recognition algorithm is proposed. The gait recognition algorithm is used to identify the motion behaviour, and the results are displayed in the Web (World Wide Web) end built by the cloud server. Meantime, in view of the existing problems in the practice of physical education, combined with the establishment and operation of the Digital Twins (DTs) system, the camera source recognition architecture is carried out since the twin network and the two network branches share weights. This paper analyses these problems since the application of somatosensory recognition technology and puts forward the improvement methods. For the single problem of equipment in physical education, this paper puts forward the monitoring and identification function of the cloud server. It is to transmit data through Hypertext Transfer Protocol (HTTP) and locate and collect data through a monitoring terminal. For the lack of comprehensiveness and balance of sports plans, this paper proposes a scientific training plan and process customization based on Body Mass Index (BMI), analyses real-time data in the cloud, and makes scientific customization plans according to different students' physical conditions. Moreover, 25 participants are invited to carry out the exercise detection and analysis experiment, and the joint monitoring of their daily movements is tested. This process has completed the design of a feasible and accurate platform for information collection and processing, which is convenient for managers and educators to comprehensively and scientifically master and manage the physical level and training of college students. The proposed method improves the recognition rate of the camera source to some extent and has important exploration significance in the field of action recognition.

1. Introduction

In recent years, people's demand for data application in the process of sports and training has been increasing. In the practice of physical education and training, somatosensory recognition technology based on visual sensing systems has attracted more and more attention and research [1]. After 2010, college students' physical health problems were more prominent, showing a downward trend year by year, and the development of college physical education has urgent reform requirements [2]. The state and government attach great importance to the physical fitness of college students and the development of physical education in colleges and

universities. Accordingly, with the application of big data and the rise of visual technology, the demand for more intelligent and scientific technology is also more urgent [3]. Based on the motion somatosensory recognition technology of visual sensing, the Kinect bone tracking technology without identification points has been widely developed and applied because of its low cost, portability, easy implementation, and no identification points [4]. In computational mechanics, the introduction of the concept and technology of the Digital Twins (DTs) system can achieve real-time safety monitoring and early warning of solid structures. The DTs system can not only be applied to civil structures. Similarly, it can also be applied to the field of

sports science. The system can not only track and monitor the gait in real time but also optimize and improve the computer simulation recognition technology. It is one of the methods used. DTs are of great help to the modification and improvement of structural system redesign and provide a more reliable basis. Source identification of digital images is an important part of image extraction. Source identification of digital images (i.e., imaging equipment identification) refers to a given graph, which determines imaging and identifies types through scientific and technological means and methods. The camera source is identified by the identification architecture of the DTs network. Firstly, the fingerprint of the camera is extracted from the twin network. Secondly, the residual network of attention mechanism is added to extract features, and the camera source recognition can be realized. Finally, the expected effect of the sample number is achieved.

The application of visual sensing technology mainly focuses on three fields: visual monitoring, interface perception, and content retrieval. In the field of elaboration and analysis, all aspects of the human motion process are discussed and analysed by analogy. The identification and understanding of motion behaviours are analysed, but some studies on motion representation are relatively simple. In the research based on the characteristics of the human model, the angle between the various parts of the human body and the trajectory of bone joints are collected and calculated to achieve a more accurate description of the movement and avoid the error caused by the change of the scene. This method mainly describes the state of each part of the human body in a three-dimensional space [5].

In view of the current development and in-depth development of human visual sensing motion recognition technology, it is increasingly necessary to conduct detailed analysis in related fields. Firstly, after the sample data, algorithm model, and motion behaviour characteristics are analysed, the vision sensor training system is designed. Secondly, the collected experimental data are tested in the training system. Finally, through the DTs network, the camera source identification method is proposed. The research aims to provide a certain experience for the further development of the perspective sensor training system in physical education practice.

2. Materials and Methods

2.1. Construction of Kinect V2 System Using Physical Education Practice

2.1.1. Introduction to Kinect V2. Kinect V2 is a new Kinect for Windows product released by Microsoft in 2014. Kinect is a 3D somatosensory camera with a microphone array, infrared transmitter and receiver, and RGB (RGB colour mode) camera. The image of the human part separated from the background is input into the human part recognition model trained by the cluster system, and the 25-node human model is output, and the bone data are output at the speed of 30 f/s. Its official system includes driver, original sensing data flow development interface, user interface, and file data, which can be developed twice. The device uses Time of Flight technology to obtain

depth image information by calculating the projected infrared and the return time of reflection. Then, through the corresponding algorithm, the coordinate information of human joint points segmented from the background image is estimated. Among them, the bone data output by RGB intelligent camera provides the basis for human motion recognition and solves the problem of data extraction in computer vision.

Kinect V2 components and functions are shown in Table 1.

2.1.2. Kinect V2 Hardware Requirements. System CPU (Central Processing Unit) uses an Intel 7-generation processor, dx11 (DirectX11), with a 64-bit operating system. The hardware composition is shown in Figure 1.

In Figure 1, the CPU is the core of the computer, and the computer acts as an “intermediary” in the Kinect V2 hardware. Kinect V2 stores the bone data related to the human body and the user’s personal information on the card reader and sends it to the computer. The computer sends this information to the cloud space. When users visit the website, they can see their own information on the computer monitor.

2.1.3. Kinect V2 Application Principle. Kinect V2 equipment first uses the infrared receiver to receive the infrared light emitted by the infrared transmitter. By collecting the encoded infrared spectrum, the depth image is processed frame by frame. Then, the device separates the characters and background to obtain the colour image of human motion. Finally, the data are transmitted through Kinect V2. The data flow process is shown in Figure 2.

(1) Depth Imaging. Among the three cameras of Kinect V2, RGB camera is used to collect RGB images, also known as colour camera. Two other cameras (infrared transmitters and infrared cameras) are used to form a depth sensing device.

The depth measurement technology in Kinect V2 is a 3D detection technology based on PrimeSense, which obtains depth information through the optical principle. The infrared ray emitted by the infrared transmitter goes to the different reference planes of the specified scene, and the diffraction grating divides it into multiple beams. The infrared spectrum forms some speckle patterns in these planes, and different speckle patterns show different depth values to determine the depth information.

Assuming that it points to the infrared emitter in the x -axis positive direction and to the object in the z -axis positive direction, the coordinate system conforms to the right-handed spiral rule. Measuring the parallax of t' in the image, there are

$$\frac{C}{c} = \frac{Z_t - Z_o}{Z_t}, \quad (1)$$

$$\frac{C}{e} = \frac{C}{Z_o}. \quad (2)$$

TABLE 1: Kinect V2 components and functions.

Components	Functions
Infrared projector	Infrared emission
Infrared photosensitive camera	Infrared spectral data analysis
Microphone	Get voice data
RGB camera	Collect colour images and video data
Control the camera motor	Get the best angle by controlling the camera
Serial cable	Signal reception and transmission

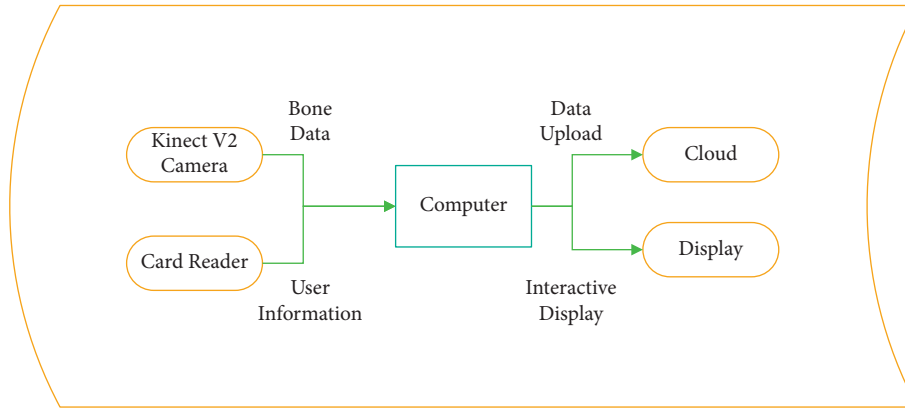


FIGURE 1: Kinect V2 hardware composition diagram.

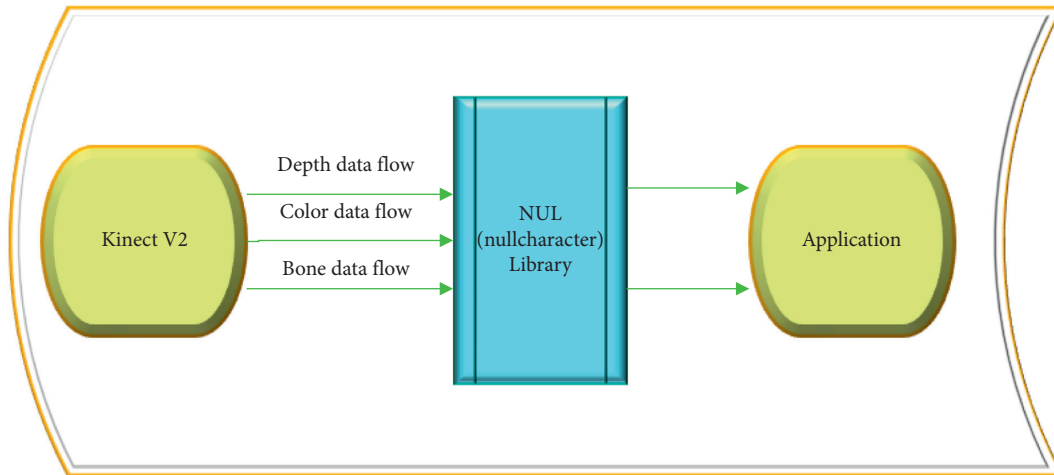


FIGURE 2: The data stream of Kinect V2.

In equations (1) and (2), Z_o is the depth distance from the real sensor to the object, A is the baseline length, e is the focal length of the sensor, C is the displacement distance of point o , and c is the parallax of the infrared image [6]. Combining equations (1) and (2), there is

$$Z_o = \frac{Z_t}{1 + (Z_t/f.b)d} \quad (3)$$

The depth distance Z_o is obtained from the above equation. Using the same principle, there are

$$X_o = -\frac{Z_o}{f} (X_o - X_t + \delta_x), \quad (4)$$

$$Y_o = -\frac{Z_o}{f} (Y_o - Y_t + \delta_y). \quad (5)$$

In equations (4) and (5), the abscissa of o point is X_o , the ordinate of o point is Y_o , the abscissa and ordinate of original coordinate are X_t and Y_t , respectively, and the correction term is δ_y . The three-dimensional coordinates of o point are calculated, Z_o represents the depth distance from the real sensor to the object, and f represents the depth of the reference plane.

(2) *Bone Recognition*. Skeleton recognition is a process of extracting the required information from noise, that is, the process of removing interference information except the human body in the picture. The identification process is shown in Figure 3.

Kinect V2 uses the Poisson equation to filter the noise to determine the existence of feature points on the body surface [7]. By grasping the angle and direction of the surface around the feature point, it determines the spatial position of its existence and forms a distance field near the point to obtain a relatively smooth shape.

2.2. Overall Network Architecture for Source Identification of DTs Camera. DTs play an important role in the camera source identification of the Kinect V2 system. Twins parasitize in digital time and space in computers, such as entity space-time and twin brothers in the corresponding digital space-time. DTs have two evolution paths. One is Product Lifecycle Management (PLM) to DTs. The other is from Physical Twins (PTs) to DTs. The establishment of DTs corresponding to entities marks the improvement of computer numerical simulation and simulation concept and technology and is a breakthrough in computational mechanics.

Generally, DTs have five life stages:

- (1) Design phase: in this stage, DTs can be optimized by finite element decomposition, which determines the design scheme of the solid structure. In DTs, the layout optimization can determine the design scheme of mechanical sensors in the structure.
- (2) Manufacturing phase: in this phase, less computation will focus on manufacturing changes to the design, and DTs will be checked as the entity structure changes to ensure synchronization with the entity.
- (3) Operational phase: in this stage, the calculation of the design is relatively large and the most complex. This stage mainly focuses on Tokyo load identification to carry out calculation related to safety monitoring. Because of high difficulty, large quantity, and high complexity, the real-time repeated calculation is needed.
- (4) Maintenance phase: in this phase, the amount of computation is concentrated mainly before maintenance and is closely related to the previous phase. Real-time safety assessment does not hinder the proposal and proposal of maintenance.
- (5) Retirement phase: in this stage, there are calculation of retirement decision and calculation of rehabilitation. The calculation of decision-making is like the calculation of maintenance suggestions in the previous stage, and the calculation of rehabilitation is like the calculation in the design stage. After the comprehensive analysis of a declaration cycle and the calculation of the optimal design, suggestions for improvement of the redesign are proposed.

In order to realize the identification of the camera source, this paper designs an identification architecture based on a twin network. The overall structure is shown in

Figure 4. The virtual frame is two branch networks of twin network sharing weights. In the designed twin network, two samples are needed to enter, the reference image P1 and the test image P2. The input two images are filtered to obtain the camera reference mode noise and the image to be tested. The pattern noise extracted from the reference image P1 is the “fingerprint” of the camera. Using the residual network with an attention mechanism, two noise image features are extracted. By calculating the distance between the feature vectors, the similarity between the noise information of the image to be measured and the “fingerprint” information of the camera C is measured to determine whether the image P2 can be seen as shot by the camera C and realize the camera source discrimination.

For the accuracy of camera source identification, seven different camera models are selected. The test images with different resolutions are unified into 224×224 image blocks to verify the performance of the designed method. The recognition accuracy and receiver operating characteristic curve are used as evaluation indexes to measure the performance of image source recognition. The recognition accuracy is defined as the ratio of the number of correct images to the number of all recognized images.

2.3. Motion Category and Representation. Human movement can be divided into three types based on three different levels of complexity, namely, action, behaviour, and action [8]. Among them, “action” is the most basic movement, the basic elements of movement, and the necessary basis for the formation of other complex movements. The time sequence is short, and the geometric and statistical methods are commonly used for identification. “Behaviour” refers to the sequence of several continuous actions. Its time scale is larger than that of “action,” and it can clearly show the purpose of human motion. It is usually identified by statistics. “Behaviour” may contain multiple “action,” and the relationship between each action and state needs to be considered and linked in the recognition process. Finally, “action” is the most complex, which is realized through motion. However, it is not a simple mechanical combination of individual motion, but a complete and purposeful motion system with different degrees of complexity. Probability statistics and artificial intelligence algorithms are usually used to identify, among which the Bayesian network is the representative method.

The behaviour process extracted from the video sequence containing human motion that can reasonably and appropriately represent the motion data is called motion representation. In the process of motion recognition, motion representation is an extremely important step. Different occasions and environments need to choose different methods for motion representation. When the motion scene is a large scene, it is necessary to carry out remote camera and monitoring; only the trajectory of the moving target is extracted. When the motion situation is small, such as gesture recognition, it is necessary to model the limb joints of the moving target in two-dimensional (2D) or 3D. In general, four criteria are used to measure the pros and cons of motion representation [9],

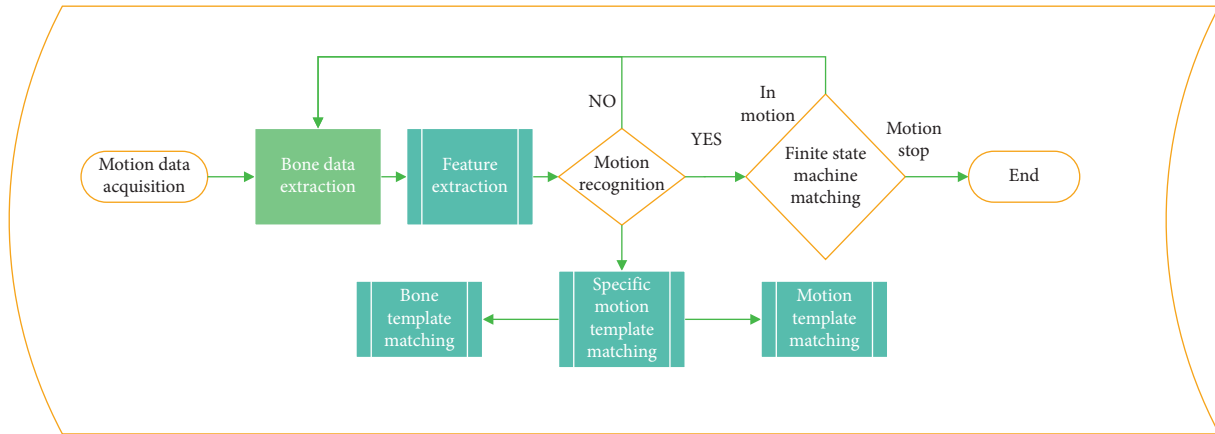


FIGURE 3: Kinect V2 recognition process.

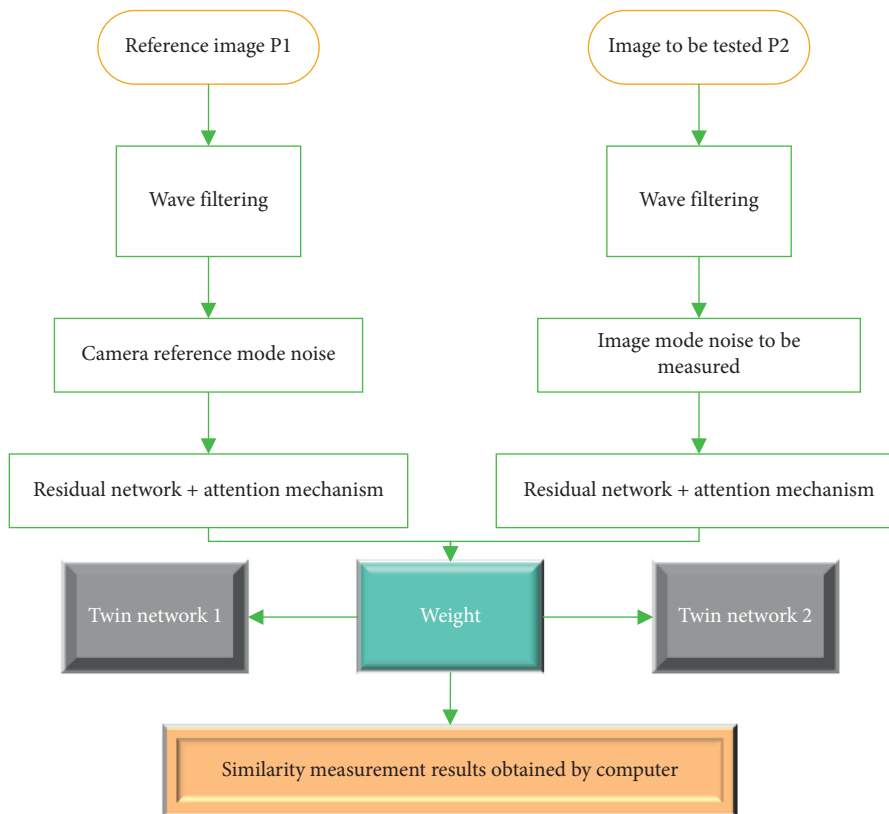


FIGURE 4: DTs network modelling for camera source recognition.

namely, compactness, completeness, continuity, and uniqueness [10]. Most of the current motion representation can only meet part of the above metrics. Generally, there are two representation methods, one based on appearance and the other based on the human model.

The representation method based on representation is to analyse and represent the motion by directly using the colour information or grey information in the image. This method of directly using image information is relatively simple. Yamato et al. (2016) took the two-dimensional network feature as the motion feature [11] and divided the whole image into several networks by binarization of the image extracted from the

moving target. They calculated the proportion of the number of target pixels in each network in the whole grid, thus representing the movement of the target.

Another representation method is based on human contour information or region information. Kale et al. (2017) adopted a gait recognition method based on human contour information [12]. They first extract the human contour, calculate the contour width, and use the width vector as the feature vector to complete the recognition. There is also an apparent representation method based on motion information (optical flow, target trajectory, speed, etc.). However, due to a large amount of calculation and

insufficient robustness of this method, Psarrou et al. (2017) used spatiotemporal trajectory technology to represent behaviour and further used the Markov process to model [13].

2.4. Identification Method Technology

2.4.1. Template-Based. The template-based method is to transform the motion image sequence into a single or a set of static templates and realize recognition by matching the sample template to be identified and the known template. The matching method is to directly match the sample of the template to be identified and the known template and obtain the category of the known template with the smallest distance as the recognition result. Bobick and Davis (2018) transformed the image sequence into MEI (motion energy image) and MHI (motion history image) and used Mahalanobis distance to measure the similarity between templates [14]. MEI represents the coverage and intensity of motion, and MHI represents the time change of motion.

Due to the different duration of motion in the same mode, template time warping becomes very necessary. Arie et al. (2016) linearized the template sequence before matching and then matched it by voting [15]. The duration of action is often random in the process of movement, so the linear time regulation method cannot completely solve the problem.

2.4.2. Based on Probability Network. As the most important motion recognition method at present, the method based on probability network can fully consider the dynamic process existing in the motion process. Different from the template-based recognition method, the method based on probability network can model the subtle changes in time and space scales through the probability method, so it has good robustness. At present, there are two kinds of probabilistic networks: dynamic Bayesian network and hidden Markov model. The hidden Markov model, as a special form of dynamic Bayesian network, has been used as a conditional random field in behaviour recognition [16]. Because it can avoid the independence assumption in the usual probability model, it is currently used most widely.

2.4.3. Grammar-Based Technology. The grammar technique is also called the syntax technique. Because of its advantages of understanding complex structures and utilizing prior information, there are more and more opportunities for motion recognition. Huber et al. (2016) used the deterministic syntax of adding orders to identify discrete events [17]. Cho et al. (2017) used statistical grammatical reasoning to solve automatic identification [18]. Ivanov and Bobick (2016) used the technology of instant grammar to identify the behaviour interaction of multiagents [19] and divided the identification problem into two levels. The bottom layer was the candidate feature detected by the independent probability event detector, which generated random syntactic analysis services for context-free.

2.5. Hidden Markov Model. It is the simplest dynamic Bayesian network based on training and evaluation in the probability and statistics model. In this paper, this algorithm is used to study the recognition of motion. In the hidden Markov model, the state is not directly visible but is visible by using the output dependence of a specific observation value. The model includes two random processes, as shown in Figure 5.

The hidden Markov model has two parts. The first is the Markov chain, and the output through this process is the state sequence. The other is the random process of the output corresponding to the observation value sequence and the state sequence. Usually, hidden Markov models are defined by five parameters.

2.5.1. N Hidden States. The hidden Markov model (HMM) is a probability model about time series. It describes the process of randomly generating unobservable state random sequences from a hidden Markov chain and then generating an observation from each state to generate an observation random sequence. Hidden Markov states satisfy Markov properties and are actually hidden states in Markov models. However, these states cannot be obtained by direct observation and can be specifically represented by $\theta_1, \theta_2, \theta_3, \dots, \theta_n$.

2.5.2. M Observable States. M is the number of observed values corresponding to a specific state, which can directly observe the state associated with the hidden state in the model. $V_1, V_2, V_3, \dots, V_m$ are defined as the observed M observations, and Q_t is defined as the observed value of t at any time.

2.5.3. Initial State Probability Matrix Π . The initial state probability matrix represents the probability matrix of the implicit state at the initial time. For instance, when the initial state probability matrix defining $\Pi = (\Pi_1, \Pi_2, \Pi_3, \dots, \Pi_n)$,

$$\Pi_i = P(q_i = \theta_i), 1 \leq i \leq N. \quad (6)$$

2.5.4. Implicit State Transition Probability Matrix A . The implicit state transition probability matrix A describes the transition probability between states in the HMM. If defining $A = (a_{ij})_{NN}$, then there is

$$a_{ij} = P(q_{t+1} = \theta_j | q_t = \theta_i). \quad (7)$$

At time t , when state θ_t is established, Equation (7) indicates the probability of the observation state being at time t and the implicit state of $t + 1$.

2.5.5. Observation State Transition Probability Matrix B . The probability of θ_j is observed when the implicit state is θ_j at t . B represents the observation matrix of the observation value; defining $B = (b_{jk})_{NM}$, there is

$$b_{ij} = P(O_t = V_k | q_t = \theta_j) 1 \leq j \leq N, 1 \leq k \leq M. \quad (8)$$

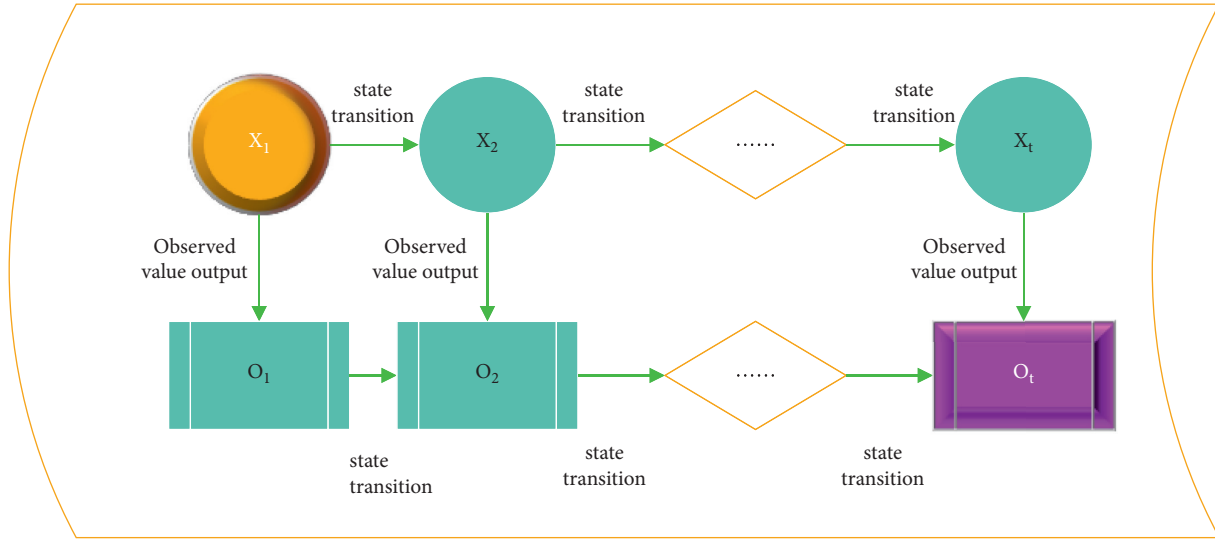


FIGURE 5: The process diagram double random about the hidden Markov model (HMM).

Therefore, in the hidden Markov model, there are

$$\lambda = (N, M, \Pi, A, B). \quad (9)$$

Briefly,

$$\lambda = (\Pi, A, B). \quad (10)$$

2.6. Hidden Markov Algorithm

2.6.1. Viterbi Algorithm. The Viterbi algorithm is an algorithm that uses the dynamic programming idea to calculate the hidden Markov chain model prediction so as to find the optimal path problem. Define the observation sequence $O = O_1 O_2 O_3 \dots O_T$, $\lambda = (A, B, \Pi)$, with $P = (Q, O | \lambda)$ as the maximum premise, and find the specific $Q^* = q_1^*, q_2^*, \dots, q_t^*$ sequence.

$\delta(i)$ is defined as the maximum probability of $O_1, O_2, O_3, \dots, O_t$ obtained at time t at the state q_1, q_2, \dots, q_b when $q_t = \theta_i$.

Initialize:

$$\begin{aligned} \delta_1(i) &= \Pi_i b_i(O_1), \quad 1 \leq i \leq N, \\ \varphi_1(i) &= 0, \quad 1 \leq i \leq N. \end{aligned} \quad (11)$$

Reasoning:

$$\begin{aligned} \delta_t &= \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T, 1 \leq j \leq N, \\ \varphi_t(j) &= \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], \quad 2 \leq t \leq T, 1 \leq j \leq N. \end{aligned} \quad (12)$$

Ending is as follows:

$$\begin{aligned} P^* &= \max_{1 \leq i \leq N} [\delta_T(i)], \\ q_T^* &= \arg \max_{1 \leq i \leq N} [\delta_T(i)]. \end{aligned} \quad (13)$$

Optimal path is as follows:

$$q_{t+1}^* = \varphi_{t+1}(q_t^*), \quad t = T-1, T-2, \dots, 1. \quad (14)$$

2.6.2. Baum-Welch Algorithm. This algorithm can solve the parameter estimation problem in the hidden Markov model [20] and can easily calculate the model-related parameters:

$$P(O | \lambda) = \sum_{i=1}^N \sum_{j=1}^N a_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad 1 \leq t \leq T-2. \quad (15)$$

The idea of the dynamic algorithm can be used to obtain $\lambda = (A, B, \Pi)$ by making the maximum value of $P = (O | \lambda)$ to the local optimal solution.

$\xi(i, j)$ is defined as the probability of θ_i at time t when the training sequence O and the model parameter λ are known, and the Markov chain state θ_i and $t+1$ are defined as follows:

$$\xi(i, j) = P(O, q_t = \theta_i, q_{t+1} = \theta_j | \lambda). \quad (16)$$

Reasoning is as follows:

$$\xi(i, j) = \frac{[a_t(i) a_{ij} b_j(O_{t+1}) \beta_{t+1}(j)]}{P(O | \lambda)}. \quad (17)$$

Thus, it is probable that

$$\xi(i) = P(O, q_t = \theta_i | \lambda) = \frac{\sum_{j=1}^N \xi_t(i, j) = a_t(i) \beta_t(i)}{P(O | \lambda)}. \quad (18)$$

The reevaluation equations are

$$\begin{aligned} \overline{\Pi}_i &= \xi_1(i), \\ \overline{a}_{ij} &= \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \xi_t(i)}, \\ \overline{b}_{ik} &= \frac{\sum_{t=1}^T \xi_t(j) \mathbb{1}_{O_t=V_k}}{\sum_{t=1}^T \xi_t(j)}. \end{aligned} \quad (19)$$

2.7. Motion Recognition under Hidden Markov Model. In the process of restoring motion recognition, all action sequences are divided into N segments on average so that they correspond to the N state in the hidden Markov model one by one and N state transition sequences are generated for the sequences. The transition process is shown in Figure 6.

The number of t frame states q_i in the m sequence and the number of $t + 1$ frame states q_j are represented by equations

$$N_{i,j}^m = N^m(S_{t+1} = q_j | S_t = q_i). \quad (20)$$

The number of frames 1 in the m state sequence and the state q_i is represented by an equation

$$N_i^m = N^m(S_1 = q_i). \quad (21)$$

In the probability of the initial state, there is

$$\Pi_i^0 = P(S_1 = q_i) = \frac{\sum_m N_i^m}{\sum_i \sum_m N_i^m}. \quad (22)$$

According to the above equation, the number of transitions between two adjacent frames in all state sequences is calculated to obtain the transition probability A^0 of the initial state. There is

$$a_{i,j} = P(S_{t+1} = q_j | s_t = q_i) = \frac{\sum_m N_{i,j}^m}{\sum_i \sum_m N_i^m}. \quad (23)$$

The Baum–Welch algorithm is used to learn the parameters, and the initial parameters are input into the hidden Markov model for parameter training. Through this process, the maximum motion action is selected as the recognition result. The training and recognition process is shown in Figure 7.

2.8. Technical Architecture. The system collects the bone motion data of the target in real time through the camera of Kinect V2 [21]. Skeleton data after coordinate mapping to get smooth data, using space vector method to calculate the angle characteristics between bone joints and using HMM model to collect the motion characteristics of data model training [22], select the maximum output probability, so as to realize a reliable real-time training system design and output.

2.8.1. Data Collection. The camera of Kinect V2 is used to obtain the motion data of the target, and the bone data flow is applied to the subsequent skeletal motion recognition.

2.8.2. Bone Algorithm Processing. The data collected from the previous layer are filtered to obtain smoother bone data [23], and the angle characteristics of bone joints by spatial vector algorithm are calculated.

2.8.3. Motion Recognition. The data of predefined motion actions in the previous layer are collected and calculated to obtain the initial parameters of HMM, and the motion action model is trained to obtain the output probability of

many hidden Markov models. The maximum output probability is selected as the recognition result by comparison.

2.8.4. Application. Through the different training movements identified by the previous layer, the data generated in the process (movement time, movement number, performance settlement, etc.) are stored and uploaded to the cloud server [24]. The scientific analysis and management of the system are carried out, and the scientific training plan that meets the target characteristics is finally generated.

The architecture is shown in Figure 8.

2.9. Software Design. The system software architecture is divided into four layers, including data layer, skeleton algorithm layer, motion recognition layer, and motion performance layer. The software client uses TCP/IP (Transmission Control Protocol/Internet Protocol) communication protocol to communicate with the cloud server [25].

The client function design is shown in Table 2.

Figure 9 shows the software design process.

3. Results

3.1. Bone Joint Confusion Experiment. Motion recognition design is based on bone joint data acquisition. In order to ensure that cameras and software can collect and analyse all the joints, five daily behaviours were monitored and analysed, including sitting, standing, pouring water, drinking water, and using the phone. In the experiment, 100 groups of samples were collected for each of the five behaviours, a total of 500 groups. 40 groups of components were extracted from the samples collected by each behaviour, and other samples were tested according to the standard template library. The matrix results obtained according to the test results are shown in Table 3.

According to the results of sample confusion, the actual motion and template have played a good classification and comparison results, and the identification of behaviour representation features is feasible. Due to the similarity between the behavioural characteristics of some actions and the joint angle, there is a low degree of recognition.

3.2. Training System Test. In order to test the feasibility of the system, 20 men and 5 women were invited to participate in the test for 3 days. In the three days, 25 people used this system for training in the morning and afternoon, once a day, and a total of 250 data samples were obtained. Sample analysis results are shown in Table 4.

Through the comparative analysis of the recognition rate and accuracy, it is found that the training system has a high recognition rate of motion and can accurately measure and analyse the joint angle and motion trajectory for motion recognition. Among them, the accuracy of sit-up and squat is the highest, which can meet the requirements of the market. In contrast, the accuracy rate in monitoring and measuring

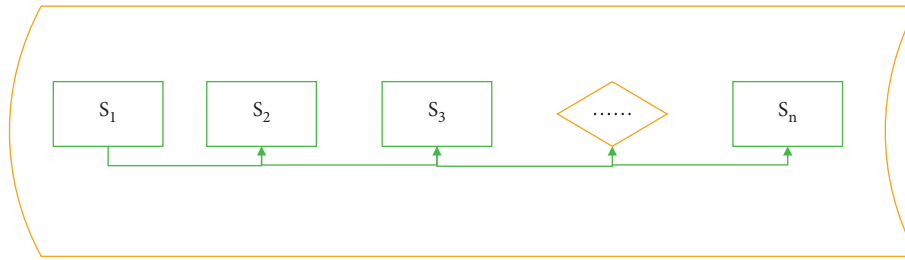


FIGURE 6: The process of sequence conversion (S_1 represents the distance of the first leg, S_2 represents the distance of the second leg, S_3 represents the distance of the third leg, and S_n represents the distance of the n -th leg).

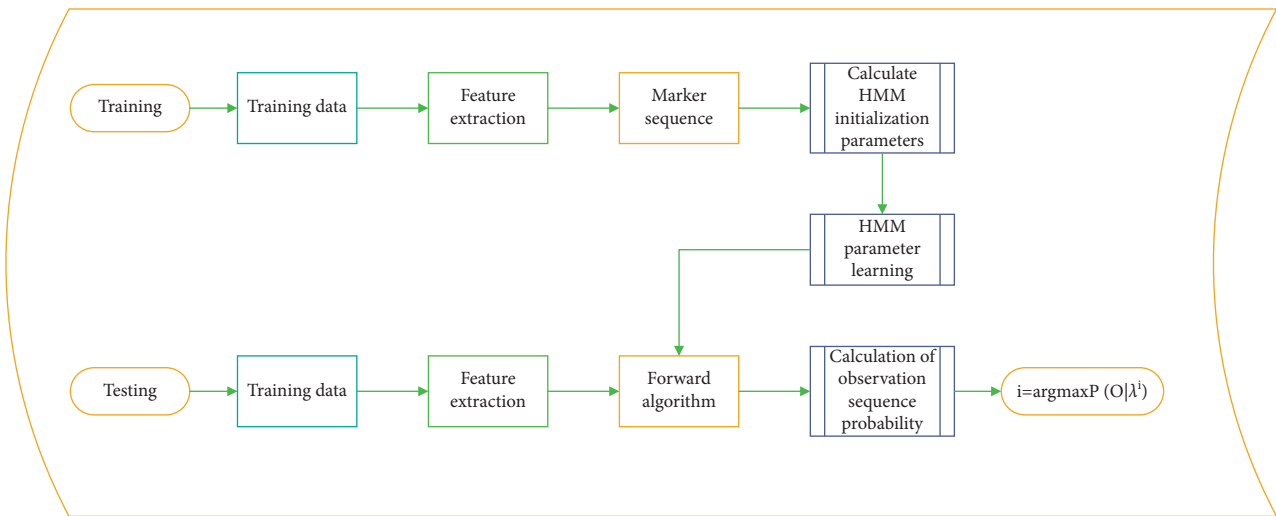


FIGURE 7: The process of training and recognition.

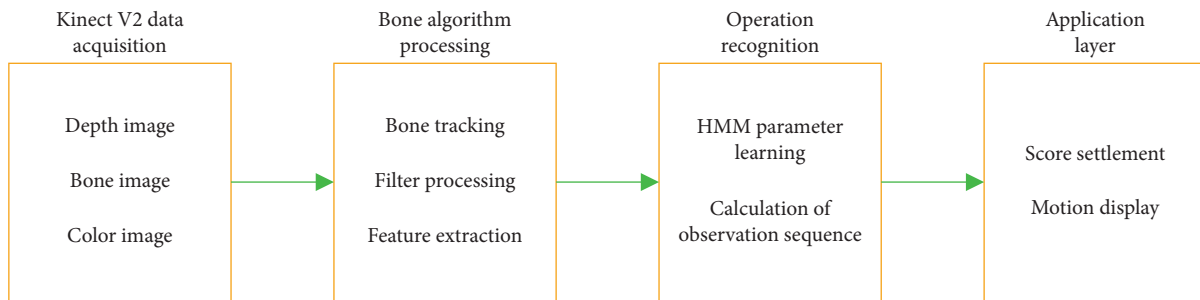


FIGURE 8: The system of architecture.

TABLE 2: Client functionality.

Name	Concrete functions
Information acquisition	Through the information collection of students' identity cards (campus card and meal card), the card number is used as the number, and the USB (Universal Serial Bus) is used to read the information and save the record
Motion recognition	4% use USB connection Kinect V2 for training motion data acquisition, through model design in the client for motion recognition, judge the basic information, and save its records
Feedback and interaction	Give feedback on the basic information of the target and real-time motion training data to the user through the front-end display device and realize the interactive function
Cloud upload	Upload the user's basic information and training data to the cloud and save its records. Concurrent information analysis to customize training plans

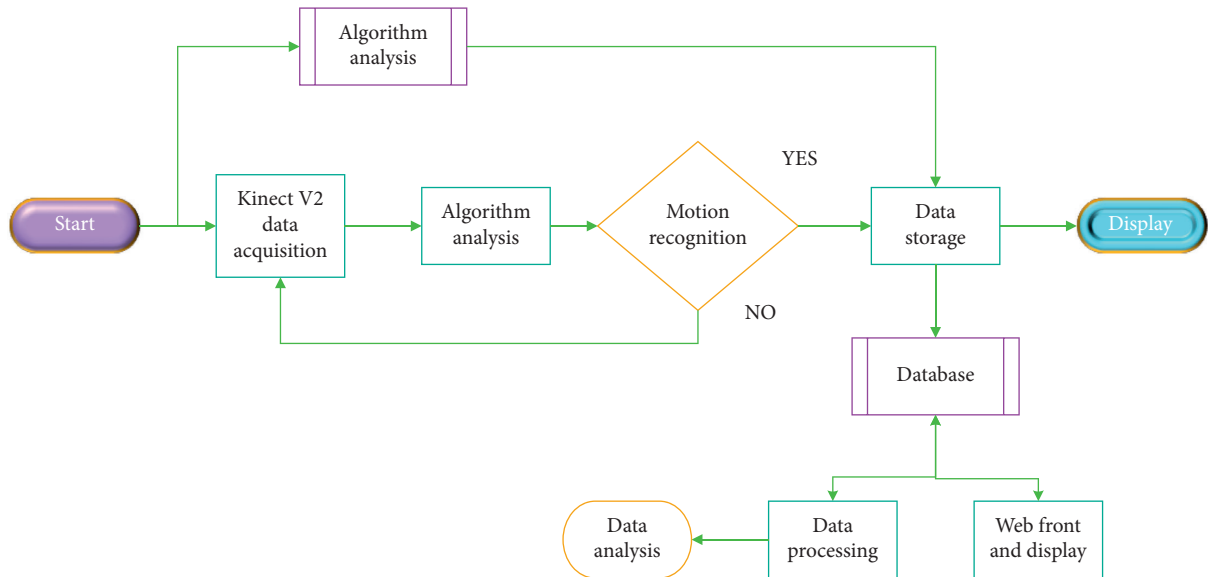


FIGURE 9: The process of software design.

TABLE 3: Matrix results.

Final display Actual action	Sitting down	Standing up	Water pouring	Drinking water	Using mobile phone
Sitting down	96%	3.9%	—	—	—
Standing up	4%	97%	—	—	—
Water pouring	—	—	92.5%	5.3%	2.1%
Drinking water	—	—	7.2%	97%	6.7%
Using mobile phone	—	—	—	—	89.9%

TABLE 4: Test results.

Projects	Numbers	Success rate (%)	Accuracy rate (%)
Setting	600	95.7	89
Squatting	270	96	94.8
The lead upward	200	94.8	92.5
Sit-up	240	97	96.8

TABLE 5: Accuracy at different take-off points and monitoring points.

Take-off point	Monitoring point	Distance between	Accuracy (%)
1.21	0.532	0.678	92.36
1.963	1.0236	0.9394	91.01
2.635	1.268	1.367	89.26

the standing long jump is low, which may be due to the error in the range of sight. Additionally, the distance between the take-off points and the monitoring point also has a certain impact on the accuracy. The specific results are shown in Table 5.

Table 5 shows that the closer the distance between the take-off point and the monitoring point, the higher the accuracy obtained. The greater the distance between the take-off point and the monitoring point, the lower the accuracy obtained. The comparison of recognition rate and accuracy is shown in Figure 10.

In view of the low error and accuracy in the test of this standing long jump project, the actual situation and test situation of the standing long jump test are compared. The details are shown in Figure 11.

According to the results, the data results can maintain a certain accuracy between 1.25 and 1.75 meters, while the data are not accurate beyond the subrange. The analysis

shows that the scene measurement environment may bring light interference to the Kinect V2 camera based on the TOF (Time of Light) principle, resulting in a lack of accuracy. Meanwhile, there may also be site installation location defects and test groups of individual body differences in force majeure.

The skeletal confusion experiment is to ensure that the camera and simulation software can collect and analyse data. The training system test experiment is to check the accuracy of all the collected data in the test system. After analysis, it is found that the accuracy of the standing long jump is the lowest. And then, the standing long jump is tested again. Figure 12 is the comparison of the proposed method and the baseline method for camera source recognition accuracy.

Figure 12 shows that the average accuracy of camera source identification using the basic sensor mode noise

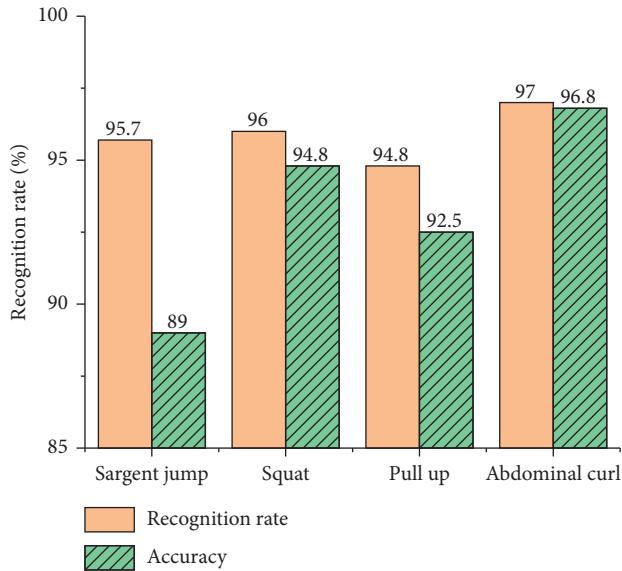


FIGURE 10: Comparison of recognition rate and accuracy rate.

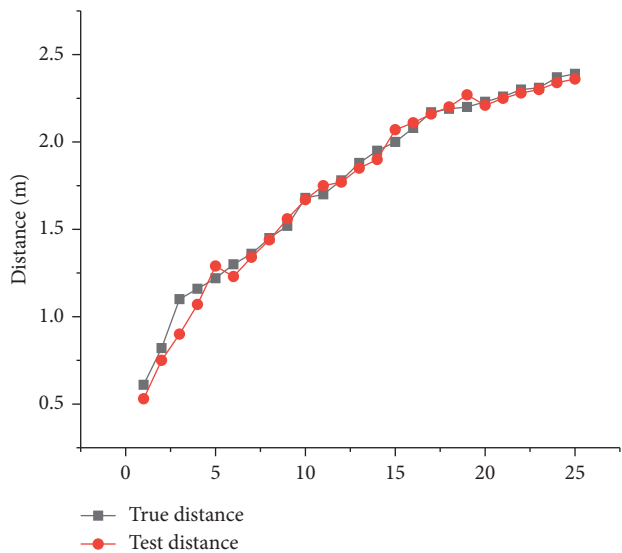


FIGURE 11: Comparison of standing long jump sample results.

identification method is 93%. When Alexnet is used as the branch network of the twin network for camera source recognition, the recognition accuracy is generally low. The method is greatly affected by the number of samples in the dataset, and the network is more inclined to learn the features related to image content. It is proved that the use of the neural network without pretreatment for camera source recognition may not be as good as the effect of camera source recognition by sensor mode noise. The method designed is used for experiments, and the average accuracy reaches 97%, which is 4% higher than that of the sensor-based noise identification method. The recognition method designed not only is more convenient and automatic than traditional methods but also has a certain improvement in recognition rate.

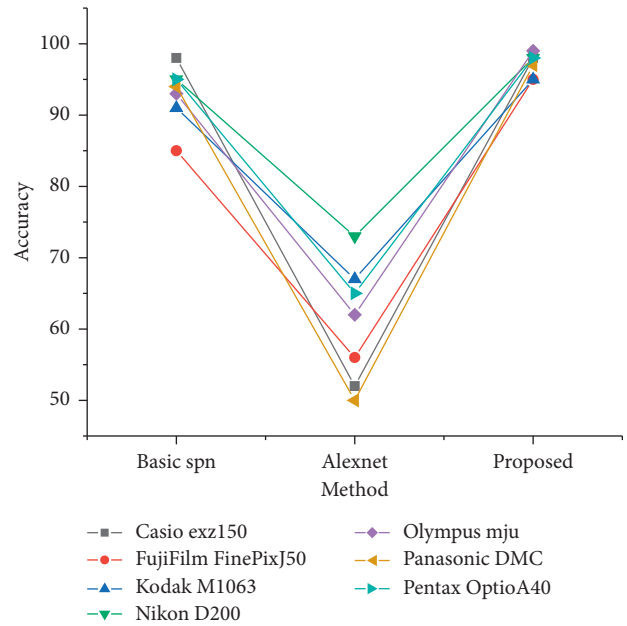


FIGURE 12: Comparison of recognition accuracy between traditional camera source recognition method and visual sensor training system (%).

4. Conclusions

With the continuous development of the global economy and the reduction of trade barriers, it provides a broad platform for the rapid development of international trade. International trade is the participation of countries in the world in the international division of labour and an important means to realize the smooth progress of social reproduction. Meantime, international trade is also an important medium for economic, political, and cultural exchanges between countries, and it plays an important role in production and life. So, it is necessary to add the use of data to educational practice. Besides, with the rapid development of visual sensing technology, it has the research and development conditions to realize nonwearable sensor monitoring and training system. In this paper, a visual sensor training system is designed by using joint data and an algorithm model to collect and analyse the characteristics of motion behaviour. Finally, the visual sensing training system for physical education practice that can complete data collection without wearing is realized, and the collected data are uploaded to the cloud through the Internet. Meanwhile, the proposed method combines the identification method of sensor mode noise with deep learning and proposes a camera source identification method by DTs network. The recognition results are judged by the similarity measure of the twin network and compared with the experimental results, which proves that the method has better recognition rate advantages. The data are analysed using the national health standards, and reasonable and scientific training plans are developed according to different sports data sources. The software designed in this paper also has defects. The software only simulates the research scene, and there are some differences between the simulated scene and the real

conditions. Therefore, when the simulation experiment is carried out, the problems that appeared are not very comprehensive. The corresponding hardware equipment needs to be under suitable temperature and light, not exceeding 30 degrees; otherwise, it will affect the operating speed of the hardware. The hardware equipment mentioned in this paper needs to be in a suitable light environment to ensure complete accuracy. Therefore, the necessary conditions, such as shading of hardware equipment, will be improved in the subsequent research. The management system developed in this paper is intended to solve the problems of old mechanism and malpractice in physical education in colleges and universities and is committed to realizing the scientific, integrated, and standardized management of physical education and sports training for college students.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] J. Ting, H. Ting, and D. Tan, "Kinect-based badminton motion analysis using intelligent adaptive range of movement Index," *IOP Conference Series: Materials Science and Engineering*, vol. 7, pp. 12–17, 2019.
- [2] S. X. Zuo, "Exploration on the reform of college physical education curriculum and the promotion of students' physical health," *Industrial & Science Tribune*, vol. 20, no. 66, pp. 170–171, 2021.
- [3] L. Wang, Y. A. Zhang, and L. Wang, "Foreign research progress on the influence of screen time on Teenagers' physical health in recent 10 years," *Journal of Physical Education*, vol. 23, no. 2, pp. 138–144, 2016.
- [4] F. Yusuke, M. Keisuke, and A. Isao, "Development of virtual reality cycling training and assessment system to investigate the effect of cycling motions in rehabilitative training," *The Proceedings of JSME annual Conference on Robotics and Mechatronics (Robomec)*, vol. 2, pp. 37–40, 2017.
- [5] L. Li, L. Han, and J. G. Xu, "Human model structure segmentation based on semantic and geometric features," *Application Research of Computers*, vol. 33, no. 1, pp. 316–320, 2016.
- [6] T. Kato, H. Hino, and N. Murata, "Multi-frame image super resolution based on sparse coding," *Neural Networks*, vol. 66, pp. 64–78, C, 2015.
- [7] S. D. Du, "Solution of the first boundary value problem of Poisson equation based on finite difference method," *Bulletin of Science and Technology*, vol. 34, no. 4, pp. 21–24, 2018.
- [8] M. U. Islam, H. Mahmud, and F. B. Ashraf, "Yoga posture recognition by detecting human joint points in real time using Microsoft Kinect," in *Proceedings of the IEEE Region 10 Humanitarian Technology Conference*, pp. 668–673, Dhaka, Bangladesh, December 2017.
- [9] V. Parameswaran and R. Chellappa, "Quasivariants for humanaction representation and recognition," in *Proceedings of the 2002 International Conference on Pattern Recognition*, vol. 7, pp. 307–310, IEEE Computer Society Press, Quebec City, QC, Canada, August 2017.
- [10] B. Reily, H. Zhang, and W. Hoff, "Real-time gymnast detection and performance analysis with a portable 3D camera," *Computer Vision and Image Understanding*, vol. 159, pp. 154–163, 2017.
- [11] J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time sequential images using Hidden Markov model," in *Proceedings of the CVPR, IEEE*, vol. 18, pp. 379–385, Champaign, IL, USA, June 1992.
- [12] A. Kale, N. Rajagopalan, N. Cuntoor, and V. Kruger, "Gait-based recognition of humans using continuous HMMs," vol. 21, pp. 321–326, in *Proceedings of the Fifth IEEE International Conference on Automatic Face Gesture Recognition*, vol. 21, pp. 321–326, IEEE Computer Society Press, Washington D.C., USA, 2017.
- [13] A. Psarrou, G. Gong, and M. Walter, "Recognition of human gestures and behaviour based on motion trajectories," *Image and Vision Computing*, vol. 20, no. 5, pp. 349–358, 2017.
- [14] F. Bobick and W. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 3, pp. 257–267, 2018.
- [15] B. Arie, Z. Wang, P. Pandit, and S. Rajaram, "Human activity recognition using multidimensional indexing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1091–1104, 2016.
- [16] J. Lafferty, A. Mccallum, and F. Pereira, "Conditional random fields: probabilistic models for segmenting and labelling sequence data," in *Proceedings of the 18th International Conference on Machine Learning*, vol. 7, pp. 282–289, Williamstown, MA, USA, January 2001.
- [17] M. Huber, A. L. Seitz, M. Leeser, and D. Sternad, "Validity and reliability of Kinect skeleton for measuring shoulder joint angles: a feasibility study," *Physiotherapy*, vol. 101, no. 4, pp. 389–393, 2016.
- [18] K. Cho, H. Cho, and K. Um, "Human action recognition by inference of stochastic regular grammars," *Lecture Notes in Computer Science*, vol. 3138, pp. 388–396, 2017.
- [19] A. Ivanov and F. Bobick, "Recognition of visual activities and interactions by stochastic parsing," *IEEE Trans PAMI*, vol. 22, no. 8, pp. 852–872, 2016.
- [20] G. Nguyen and T. Phung, "Reducing over-smoothness in HMM-based speech synthesis using exemplar-based voice conversion," *EURASIP Journal on Audio Speech and Music Processing*, vol. 1, pp. 76–89, 2017.
- [21] M. Gu and L. Q. Ge, "Natural gamma spectrum processing based on variable parameter double exponential smoothing method," *Computer techniques for geophysical and geochemical exploration*, vol. 6, pp. 506–509, 2018.
- [22] S. Yohei, Y. Takeshi, and T. Hiroshi, "Generation of point light source data for animated hologram with motion capture using KinectV2," *ITW Technical Report*, vol. 40, no. 11, pp. 67–69, 2016.
- [23] R. Noel, A. Salekin, and R. Islam, "A natural user interface classroom based on Kinect," *IEEE Learning Technology*, vol. 13, no. 4, pp. 59–61, 2017.
- [24] G. Rogers and H. Christensen, "A conditional random field model for place and object classification," in *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 300, pp. 766–772, Saint Paul, MN, USA, May 2012.
- [25] J. Yin, G. Tian, and F. Zhou, "Human activity representation and recognition based on temporal order histogram in intelligent space," *Chinese Journal of Computers*, vol. 37, no. 2, pp. 470–479, 2016.