RESEARCH   Open Access

Basketball shooting technology based on acceleration sensor fusion motion capture technology

Binbin Zhao[1] and Shihong Liu[2*]

* Correspondence: lsh@cuit.edu.cn
[2]Department of Physical Education,
Chengdu University of Information
Technology, Chengdu 610225,
Sichuan, China
Full list of author information is
available at the end of the article

Abstract

Computer vision recognition refers to the use of cameras and computers to replace the human eyes with computer vision, such as target recognition, tracking, measurement, and in-depth graphics processing, to process images to make them more suitable for human vision. Aiming at the problem of combining basketball shooting technology with visual recognition motion capture technology, this article mainly introduces the research of basketball shooting technology based on computer vision recognition fusion motion capture technology. This paper proposes that this technology first performs preprocessing operations such as background removal and filtering denoising on the acquired shooting video images to obtain the action characteristics of the characters in the video sequence and then uses the support vector machine (SVM) and the Gaussian mixture model to obtain the characteristics of the objects. Part of the data samples are extracted from the sample set for the learning and training of the model. After the training is completed, the other parts are classified and recognized. The simulation test results of the action database and the real shot video show that the support vector machine (SVM) can more quickly and effectively identify the actions that appear in the shot video, and the average recognition accuracy rate reaches 95.9%, which verifies the application and feasibility of this technology in the recognition of shooting actions is conducive to follow up and improve shooting techniques.

Keywords: Computer vision, Visual identity, Motion capture, Support vector machine, Shooting skills

1 Introduction

In recent years, with the development of the era of big data, motion capture technology developed by computer vision recognition has become a hot topic in research. In the course of basketball training and competition, coaches should develop corresponding training programs for different athletes to improve their basketball skills. Traditional training methods are coaches based on their theoretical training and experience, combined with the technical level of basketball players to develop training programs. This training mode takes a long time and may cause a waste of coaching resources. Modern sports should be precise and efficient. Computer vision recognition motion capture

technology can more accurately identify athletes' shooting skills, which can effectively improve coaches' teaching efficiency and athlete's shooting skills.

Leo M believes that in the past few decades, there has been a tremendous increase in the demand for computer-assisted technologies, which can help overcome individual functional limitations and improve people's quality of life. Therefore, different research papers on the development of computer-aided technology have appeared in the article, which promotes the necessity of organizing and categorizing them while considering the purpose of visual recognition aids. However, this user needs-oriented classification method treats each technology as a whole and then provides in-depth and critical explanations of the technical knowledge used to construct operational tasks and the discussion of its cross-context applicability. Therefore, existing investigations are unlikely to inspire technical improvements and explore new technological frontiers. In order to overcome this key flaw, the article introduces a new task-oriented method to classify the state of motion capture: it relies on decomposing the final visual recognition aid target into tasks, and then these tasks are used as pointers to motion capture, and every detail in the action is used as a component. The article pays special attention to a set of cross-application computer vision recognition tasks. These tasks are set as pivots to establish a classification that has been used to assist users in visual recognition. For each task, Leo M analyzed the computer vision algorithms involved in the current development of computer vision recognition, trying to see some possible ways in the short and medium term, so that the visual recognition aid results can be truly improved, and from athletes and shots from a technical perspective, the possible impact of visual recognition motion capture evaluation was discussed. However, this research technology setting is more ideal, more difficult, and more difficult to apply in practice [1]. Antonio AD believes that it is difficult for coaches, team analysts, and players to use classic analysis methods to understand the amount of data available in the sports field. There must be new methods to help users decompose relevant action information and analyze them at a deeper level. Therefore, Antonio AD's visual analysis system tool, BKViz, is proposed, which combines a variety of interactive visualization methods and allows users to give feedback in time, allowing analysts, coaches, and athletes to analyze a single basketball game and help the basketball team improve its technical and combat performance. This tool has a high application cost and is not suitable for popularization in China [2]. Sutcliffe M described the development of an automatic defect recognition system suitable for full matrix capture (FMC) imaging data. They used the principle of computer vision to extract the features of the reconstructed FMC image, combined with the multi-layer perceptron artificial neural network for classification. The system uses a variety of single V-shaped weld training samples to train the artificial neural network and test its accuracy. The automatic classification of actual single-v weld defects has achieved a high success rate. It uses computer vision technology to automatically extract the position, size, and location information of the defect through training artificial neural networks, which proves that there is little or no user intervention. Next, automatic action defect classification is possible. The ability to automatically determine the characteristics of movement defects provides a significant advantage for shooting technique detection. The calculation and classification method of this system is more complicated, which may cause errors in the results and affect the training analysis [3].

The innovations of this paper are (1) proposing to denoise the background of images created by computer vision recognition, (2) establishing a support vector machine model for motion capture recognition of basketball shooting techniques, and (3) establishing a Gaussian mixture model to perform motion capture processing of basketball shooting techniques.

2 Method of basketball shooting technique based on computer vision recognition fusion motion capture technology

2.1 Method of recognition of shooting action

Basketball action recognition is a kind of human posture recognition. Human motion gesture recognition has always been a research hotspot in various fields. Domestic research has also been conducted on basketball motion recognition, which will increase the analysis of complex motions, such as turning to catch the ball, running, and doing a layup, and expand the recognition of basketball shooting motions. At this stage, there are two main human body gesture recognition methods: inertial sensor gesture recognition and image acquisition gesture recognition [4]. Visual recognition based on image acquisition can be divided into single video recognition and multiple video recognitions [5]. The general idea of image acquisition and visual recognition is to use sensors, such as cameras to collect images or videos of athletes, and then perform hidden functions hidden in the images and videos. Finally, design a classifier to recognize the athlete's motion posture [6]. The basic idea of sensor inertia is to install a data acquisition sensor externally from the athlete's body, send the collected data to the terminal in real time, and identify the position of the athlete based on various data [7].

2.1.1 Background difference

The background difference method is suitable for use when the camera is in a stationary installation, and has the characteristics of accurate detection, simple algorithm, and easy implementation [8]. In this way, through further processing, you can fully extract important data quickly and accurately extract the motion characteristics of the moving target [9]. However, in actual application scenarios, the background reference model is very sensitive to changes in external scenes, such as weather changes, lighting, and emergency situations. In the process of using the background difference method to derive the motion area, a good background reference model should be created [10].

2.1.2 Optical flow method

The advantage of the optical flow method is that it can derive the position information of the moving target in the video sequence relatively completely and it can also support the motion state of the camera [11]. Therefore, the optical flow method can detect moving targets from the camera. This method is more suitable for accurate analysis and processing and solves the problem of overlapping objects and obstacles in traditional moving target detection [12, 13].

2.1.3 Frame difference

The frame difference method can be used to detect moving targets in dynamic scenes and permanently install cameras [14]. This method may lead to the inability to fully

derive all the features and detection results related to the moving target, and the results obtained may be slightly wrong. Generally speaking, in this case, the next processing step is required, which is not conducive to the further analysis and processing of the image [15]. In addition, if the moving target does not move at a constant speed during the movement, it may move at a variable speed. Therefore, using the frame difference method may result in the detection of moving targets or only relatively small and shallow boundaries [16]. However, although the karate diff method cannot accurately derive the moving target, this method is usually used as the original algorithm to quickly determine whether the target enters the scene [17].

2.2 Image processing method of shooting action

2.2.1 Graphic gathering

The process of classifying and digitizing continuous image signals and then sending the generated digital signals to frame memory or computer memory is called image acquisition [18]. Generally speaking, image acquisition can be divided into two categories: one is to capture static images, to obtain images at a given time [19]; the other is to capture static images, which are dynamic images, to obtain a specific time period [20]. Still image acquisition is mainly taken by a camera, and the captured image is stored in the camera as a digital signal or directly transmitted to a computer for subsequent processing [21]. The collection of dynamic images is mainly by digitally storing the images taken by the camera on the hard disk of the camera or directly transmitting to the computer for processing [22].

2.2.2 Image denoising

This intermediate process of removing and suppressing noise in the image is called image denoising, and image denoising generally exists in the image preprocessing process [23]. With the rapid development of digital image processing technology, image denoising methods can generally be divided into two categories: mean filtering and median filtering [24]. The mean image filtering method directly operates on the original image to be processed. According to this operation method, mean filtering denoising can be divided into direct operation on each pixel in the image and direct operation on the adjacent area of the pixel to be processed [25].

3 Mean filter

In image processing, the average neighborhood method is the most intuitive, simple, and easy to apply denoising method, and it is widely used in image noise processing [26]. The average filtering method replaces the gray value of pixels in the area with the average value of several pixels in the standard, eliminates the pixels that cannot represent the environmental pixel value, and makes the image smoother [27]. Assuming that the image to be processed is $m(a, b)$, T represents the kernel, the total number of pixels in the kernel is represented by S, and the average filtered image is $n(a, b)$, which can be expressed as:

$$n(a, b) = \frac{1}{s} \sum_{meT} m(a, b) \tag{1}$$

1. Median filter

Median filtering sorts each pixel in a certain neighborhood of the image and selects an intermediate value to replace all pixels around the neighborhood, instead of simply replacing the average value of these pixels [28]. Assuming the mathematical formula is used and assuming it is the median value of all pixels in the neighborhood of x, then:

$$x = Med\{x_1, x_2, x_3 \Lambda x_n\} = \begin{cases} \dfrac{x_{k+1}}{2} \\ \left(\dfrac{X_k}{2} + \dfrac{x_{k+1}}{2}\right)/2 \end{cases} \tag{2}$$

When k is an odd number, x is equal to the above formula; when k is an even number, the value of x is equal to 1/2 of them.

3.1 Classification method of shooting

3.1.1 Template method

The core idea of the template shooting action classification method is to convert the action sequence into a static pattern or a set of static patterns and match it with a known template. Through similarity calculation, the most matching template category is used as the classification result. According to whether the matching object is static mode or static mode, the time series is further divided into template matching and dynamic time warping [29]. Template matching directly compares static templates with existing examples. The features available in the process include spatial features such as contours, gradients, and optical flow, as well as temporal features containing timing information, such as trajectories.

3.1.2 Statistical modeling

Statistical models can generally be divided into two categories: production models and discriminative models. In the model training stage, the production model is trained, the model parameters of different action categories are extracted from the training sample set, and then the observation features to be classified in each model obtained from the previous training are input, and the degree of correspondence with the model is calculated, which is the potential for creation. The final classification result is the behavior category model with the highest matching degree: the discrete model directly models the operator category for the given conditional probability. The most commonly used discrete models are support vector machines and random fields [30].

4 Experiment of basketball shooting technique based on computer vision recognition and motion capture technology

4.1 Constructing a basketball shot recognition model

There are many complicated human structure movements involved in basketball shooting. Before designing a basketball shooting technique recognition model, it is necessary to classify the basic basketball shooting postures. In order to effectively determine the basketball movement according to the physical condition of the basketball player, it is first divided into sports state and static state. The state of the game corresponds to the state of the players when they complete various basketball actions. At present, the limbs

of athletes are in motion; statistics refer to the situation where the limbs of athletes are completely still. The focus of basketball gesture recognition is to recognize various sports gestures. In order to effectively recognize different sports postures in basketball, the sports postures are gradually divided into two stages. First, according to whether the motion state is periodic, the posture of the human body is divided into two categories: continuous action and instantaneous action. The second step is to divide the body's posture into seven postures of walking, running, dribbling, jumping, shooting, passing, and catching the upper or lower limbs according to whether the state of the action is exercise. The basketball position recognition model automatically recognizes the seven sports positions of basketball players.

4.2 Sensor signal collection

Many sensor devices including an accelerometer, a gyroscope, an angular velocity meter, a pressure sensor or the like, in the data collection phase collect body posture information and perform different actions. The basic method of human body movement posture recognition is to install sensors on the key parts of the human body to detect the limb movement information of the human body. In basketball shooting action recognition, the movement information of the legs and arms of the human body is mainly collected. The sensor node formed by the combination of multiple sensor devices can convert the action information during the completion of the action into electrical signals for uploading and fulfill the requirements of subsequent logic operations, data storage, and communication. According to actual application requirements, it is difficult for a single sensor module to meet the work requirements. The information required in human posture recognition is complex and diverse, including physical and physiological information such as acceleration, angular velocity, or heart rate. The internal analysis and processing of the node needs to be completed, so the design of the node needs to include multiple sensor modules, which can be used in conjunction to complete the work requirements of the system. Generally, a sensor node includes four modules, which are mainly composed of four parts: processor module, power module, sensor module, and communication module. The processor module controls the normal operation of each functional module of the sensor node and performs the related processing of each signal; the sensor module realizes the function of detecting the movement information of the object, and realizes the transformation of the movement information to the electrical signal; the communication module is responsible for signal transmission, n nodes transmit wireless data to other devices; the power supply provides the energy for the normal operation of the entire sensor. At present, mobile devices such as mobile phones have also begun to integrate various sensor modules, which have the function of wireless communication. They will replace sensor nodes worn on key parts of the human body for signal collection. Compared with sensor nodes, mobile devices are worn at different locations. Fixed, this will have an impact on the recognition result of the system. When the sensor detects motion information, the device can be placed in a fixed position to avoid this impact.

4.3 Shot recognition

The essence of the basketball gesture recognition stage is to construct a classification model process that meets the basketball action data division. For each specific

basketball action, after data collection, data preprocessing, data division, and feature extraction, a description of the specified basketball action can be obtained. The attribute set is the feature vector set. These feature vector sets are abstract data sets of basketball actions, and their corresponding classifications can be obtained through calculations in the classifier model. The attributes contained in the feature vector are complex. In order to eliminate irrelevant and redundant attribute values in the feature vector, it is necessary to perform feature selection on the feature vector. In the attribute selection, the first priority search algorithm and principal component analysis method are used. The feature selection realizes the dimensionality reduction of the feature vector, reduces the complexity of the classification calculation process, and improves the work efficiency of the system. In this experiment, sensor nodes are respectively fixed on the lower leg and forearm of the subject to detect the movement behavior information of different limbs. According to the different placement positions of the nodes, the data set of each movement is divided into upper limb movement data set and lower limb movement data set. In the action data set, classifiers are constructed for different sample sets to realize the specific division of the actions of the upper and lower limbs. The combination of the results of the upper and lower limbs can obtain the basketball movement posture of the current subject.

In this paper, support vector machine model (SVM) is used to identify basketball shooting techniques. When (SVM) solves two types of classification problems, it will look for an h-dimensional hyperplane in the $h-1$-dimensional sample feature space as the segmentation plane for the two types of samples. Usually, this plane is called a linear classifier. When the samples can be distinguished correctly, they are said to be linearly separable. When it is necessary to deal with the case of linear inseparability, SVM will map sample points to higher-dimensional or even infinite-dimensional space. At this time, this mapping is nonlinear, so sample points will become linearly separable in high-dimensional space. In this case, using the $k(x, y)$ function that satisfies the Mercer condition as the inner product operation of the two sample features is equivalent to mapping the sample from the original feature space to a new feature space. Suppose the sample feature is x_i, the sample category label is y_i, and the Lagrangian coefficient is a_i, b can be obtained by any support vector, then the corresponding optimal classification function is defined as:

$$f^*(x) = \text{sgn}\left(\sum_{i=1}^{N} a_i y_i \bullet k(x_i, x) + b\right) \qquad (3)$$

This article also tries to use the Gaussian mixture model to eliminate interference from the background image of basketball shooting action recognition. In this model, assume that the pixel value of the recognized video at a certain moment t is Y_t, and k is the Gaussian distribution number (generally 3 5), ϖ_i is the i-th Gaussian distribution weight, $\mu_{i,\ t}$ and $\sigma_{i,\ t}$ represent the mean and variance, respectively, g is the Gaussian distribution function, then the random probability corresponding to Y_t is:

$$P\left(Y_t = \sum_{i=1}^{k} \varpi_{i,t} g\left(Y_t, \mu_{i,t}, \sigma_{i,t}\right)\right) \qquad (4)$$

5 Results and discussion

5.1 Results

Use the basketball shooting recognition model to capture the shooting situation, select 20 basketball players for analysis, use sensors to collect signals, and generate related action images from the model, including walking, running, dribbling, jumping, shooting, passing, and receiving. There are a total of seven postures. In this experiment, a total of 140 sets of data are collected, and the shooting situation is drawn into a table, as shown in Table 1:

Table 1 shows the goals of each of the 20 basketball players. The actual goals and judgment goals of each player are different. The basketball shooting recognition model is used to capture the players' shooting conditions. The shooting accuracy range is 40–95%.

In order to display the experimental results more intuitively, the data in the table is drawn into a graph, as shown in Fig. 1:

Divide the 20 athletes into four groups, each with five people, and plot their basketball goal percentages as a line chart, as shown in Fig. 2:

It can be seen from the chart that according to the data collected by the basketball shot recognition model established in this article, among the twenty basketball players, only one has a shooting accuracy higher than 90%, and the shooting accuracy is between 80 and 90%. There are four athletes with shooting accuracy between 70 and 80%, six athletes with accuracy between 60 and 70%, and seven athletes with accuracy below 60%.

Table 1 Basketball shots

Number	Layups	Actual goals	Judging goals	Goal rate
1	15	11	5	73.3%
2	15	8	3	53.3%
3	15	9	4	60%
4	15	6	2	40%
5	15	12	6	80%
6	15	9	4	60%
7	15	7	3	46.7%
8	15	10	4	66.7%
9	15	13	6	86.7%
10	15	12	6	80%
11	15	11	5	73.3%
12	15	10	4	66.7%
13	15	9	4	60%
14	15	8	3	53.3%
15	15	13	6	86.7%
16	15	7	3	46.7%
17	15	7	3	46.7%
18	15	14	7	93.3%
19	15	9	4	60%
20	15	8	3	53.3%

**Fig. 1** Basketball shots
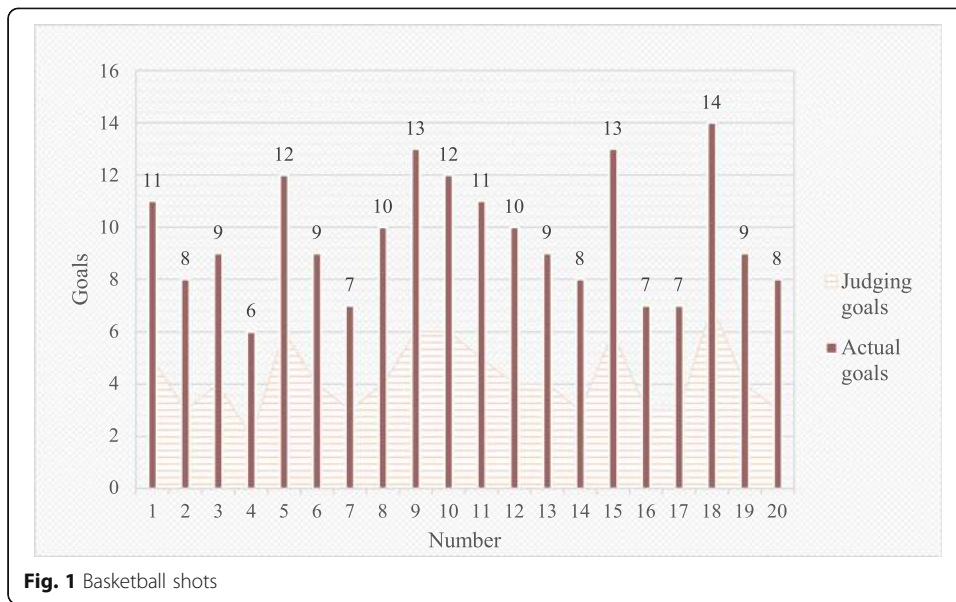
5.2 Discussion

In the process of collecting sports posture data, testers should complete the required basketball actions according to the default posture of the human body and normal exercises. Every basketball stop action includes upper limb movement and lower limb movement. When tracking sports basketball, it is necessary to analyze the upper and lower limb movements of athletes separately. For this reason, according to the characteristics of the athlete's upper and lower limbs, a classifier is constructed to recognize the posture of basketball players. In this paper, two models of support vector machine and Gaussian mixture model are used to recognize basketball players' shooting actions. Two parameter values are set: the first parameter α is the distance critical value, which determines the number of typical postures (representatives); the second important

Fig. 2 Basketball players' shooting accuracy
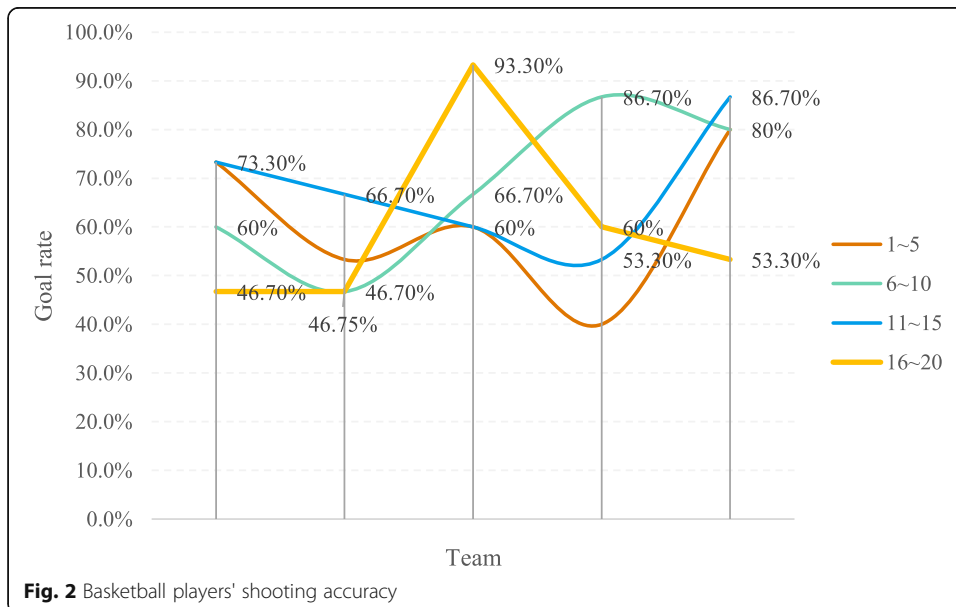
parameter is the assumed number of each action. The distance threshold is used to determine whether two histograms are different, which affects the hypothetical number of an action. Use the "shooting" experiment to study the influence of this parameter. The "shooting" is chosen because it is relatively short and changeable and has a high probability of hypothetical segmentation. It turns out that when the distance threshold is lower than 0.4, the assumed number is still high and unchanged. On the contrary, when the distance threshold exceeds 0.4, the hypothesis number decreases rapidly. Therefore, the distance threshold should be 0.4–0.8, depending on the required level of granularity. For all tests in this article, this parameter is set to 0.4 to maintain most of the intraclass variation. Based on the action recognition of the support vector machine model and the Gaussian mixture model, use the methods mentioned above to test the obtained data, and draw the test results into graphs, as shown in Figs. 3 and 4:

From the above chart and data calculation, it can be seen that in order to obtain the size of the object from the video image and its corresponding position in the image, it must be able to determine the relationship between the corresponding point in the object image and the corresponding point in the image. The commonly used method is image calibration technology. Regardless of whether it is necessary to specify a calibration object, image calibration technology is divided into traditional camera calibration and self-correction methods. Traditional camera calibration methods have certain requirements for camera models. The size and shape of the calibration object should meet certain requirements. Image processing under known conditions, through mathematical transformation and calculation, can obtain the model of the internal and external parameters of the camera. The camera automatic adjustment method does not require a specific calibration object, but is based on the positional relationship between the calibration of the camera's circular image and the corresponding image taken

Fig. 3 Support vector machine accuracy

Fig. 4 Gaussian mixture model accuracy
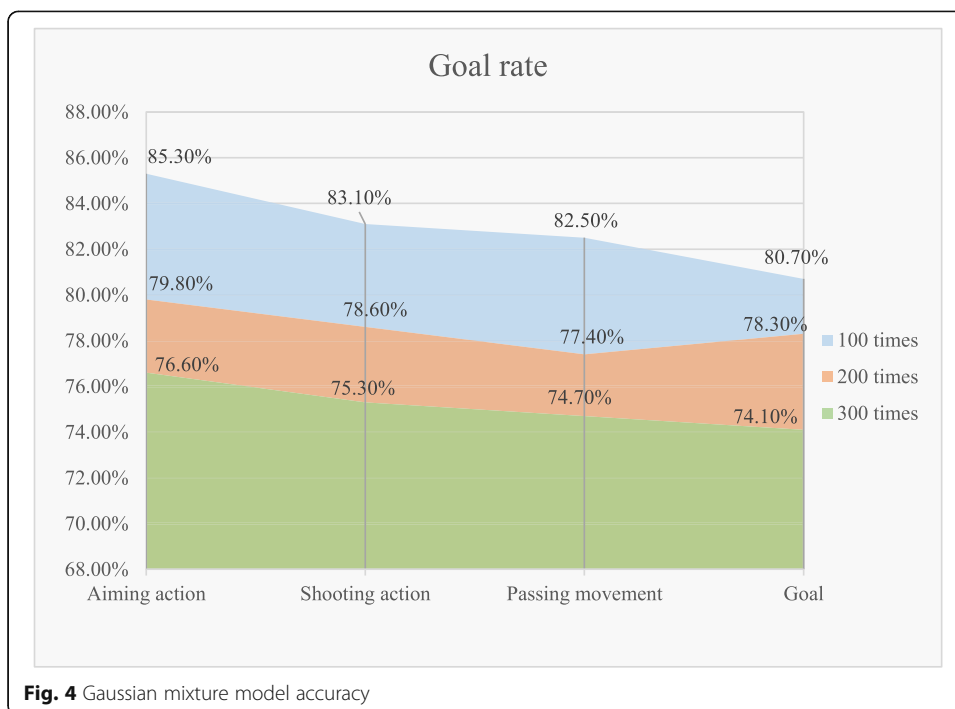
during the movement. Camera self-adjustment methods can be divided into camera automatic calibration technology and basic matrix and automatic matrix calibration technology.

Taking the experimental results of 100 tests as an example, the average accuracy rate of motion capture results for computer vision recognition using the support vector machine model is 95.9%, and the average accuracy rate of motion capture results for computer vision recognition using the Gaussian mixture model is 82.9%. Therefore, the use of support vector machine models for visual recognition and capture of basketball shooting movements has a high accuracy rate. It can be used in the teaching process of basketball coaches and athletes training. It is conducive to more accurately capture shooting-related actions and generate specific images, allowing coaches and athletes observe clearly the defects of the movement and correct them to improve training efficiency.

6 Conclusions

Computer vision recognition motion capture system is a technical device that measures the movement of objects in space. Its principle is based on computer graphics, which uses sensors or trackers to observe and record the trajectory of objects in three-dimensional space. Under the current technical conditions, the fusion motion capture technology of computer vision recognition is used in the research of basketball shooting technology. With the rapid development of computer technology and microelectronics industry, computer vision recognition fusion motion capture technology will be used in sports work. I believe that in the near future, neural network and deep learning technology will be applied to professional sports work. Computer vision recognition system will bring huge changes to traditional basketball teaching and training work. The innovation of this article is to use a variety of methods, such as data analysis

method, background difference method, optical flow method, and frame difference method, and design two classifications of shooting actions: template method and statistical model method, which fully integrate computer vision. The recognized motion capture technology is applied to the teaching of basketball, thereby improving the quality of teaching and promoting the development of basketball.

In the early stage of the research, this paper puts forward the method of shooting action recognition. Basketball action recognition is a kind of human body gesture recognition, including background difference method, optical flow method, and frame difference method. The background difference method is suitable when the camera is installed in a static state, and has the characteristics of accurate detection, simple algorithm, and easy implementation; the advantage of optical flow method is that through calculation and analysis, the position of the moving target in the video sequence can be more fully extracted Information and support the movement status of the camera. The frame difference method is mainly suitable for the detection of moving targets and cameras in dynamic scenes of fixed devices. The main disadvantage of this method is that it cannot fully output all the features and detection results related to moving targets. This article also proposes a shooting action image processing method, which is divided into image acquisition and image denoising. Image acquisition is for static image acquisition, that is, to take photos, and the purpose is to obtain images at a certain moment, and the other is for dynamic image acquisition. The purpose of video shooting is to obtain continuous images in a certain period of time. Image denoising is an intermediate process of removing and suppressing noise in the image. Two algorithms, mean filtering and median filtering, are proposed in the article. In addition, the article conceives two methods to classify shooting actions, including template method and statistical model method.

In the experimental stage, this paper first builds a basketball shooting recognition model, then uses sensors to collect signals, and finally establishes a support vector machine model (SVM) and a Gaussian mixture model in the field of shooting motion capture recognition to recognize and perform background images for basketball shooting motion recognition interference elimination processing. Based on the analysis of the experimental part, the article concludes that the average accuracy rate of motion capture using the support vector machine model for computer vision recognition is 95.9%, and the accuracy rate is high. It can be used in the teaching process of basketball coaches and athletes, which helps improve teaching. Training efficiency adds boost to the development of basketball career.

Abbreviations
SVM: Support vector machine; FMC: Full matrix capture

Authors' contributions
Binbin Zhao did the writing—editing. Shihong Liu did the data analysis. The authors read and approved the final manuscript.

Authors' information
Binbin Zhao was born in Shenyang, Liaoning, P.R. China, in 1979. She received the Master's degree from Shenyang Sport University, P.R. China. Now, she works in the College of Sport Science, Shenyang Normal University. Her research interests include sports education and training learns, sport management, and sport sociology.

Shihong Liu was born in Chengdu, Sichuan, P.R. China, in 1973, and is an associate professor. He received the Master's degree from Chengdu Sport University, P.R. China. Now, he works in Chengdu University of Information Technology. His research interests include basketball teaching and training.

Availability of data and materials
Data sharing does not apply to this article because no data set was generated or analyzed during the current research period

Declarations

Ethics approval and consent to participate
This article is ethical, and this research has been agreed.

Consent for publication
The picture materials quoted in this article have no copyright requirements, and the source has been indicated.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]College of Sports Science, Shenyang Normal University, Shenyang 110034, Liaoning, China. [2]Department of Physical Education, Chengdu University of Information Technology, Chengdu 610225, Sichuan, China.

References
1. M. Leo, G. Medioni, M. Trivedi, et al., Computer vision for assistive technologies. Computer Vision Image Understanding **154**(Jan.), 1–15 (2016)
2. A.D. Antonio, BKViz: a basketball visual analysis tool. Comput Rev **58**(7), 435–436 (2017)
3. M. Sutcliffe, J. Lewis, Automatic defect recognition of single-v welds using full matrix capture data, computer vision and multi-layer perceptron artificial neural networks. Insight - Non-Destructive Testing and Condition Monitoring **58**(9), 487–493 (2016)
4. A. Issac, M.K. Dutta, C.M. Travieso, Automatic computer vision-based detection and quantitative analysis of indicative parameters for grading of diabetic retinopathy. Neural Comput & Applic **32**, 15687–15697 (2020)
5. A.R. Di Rosa, F. Leone, F. Cheli, et al., Fusion of electronic nose, electronic tongue and computer vision for animal source food authentication and quality assessment – a review. J Food Eng **210**(OCT.), 62–75 (2017)
6. Rameshan R, Arora C, Dutta Roy S. [Communications in Computer and Information Science] Computer Vision, Pattern Recognition, Image Processing, and Graphics Volume 841 || Classification of Indian Monuments into Architectural Styles. 2018, 10.1007/978-981-13-0020-2(Chapter 47):540-549.
7. A. Cuzzocre, E. Mumolo, G.M. Grasso, et al., An effective and efficient approximate two-dimensional dynamic programming algorithm for supporting advanced computer vision applications. J Visual Languages Computing **42**(oct.), 13–22 (2017)
8. M.K. Gregersen, T.S. Johansen, Corporate visual identity: exploring the dogma of consistency. Corporate Communications An International Journal **23**(3), 342–356 (2018)
9. W. Bolhuis, M.D.T.D. Jong, A.L.V.D. Bosch, Corporate rebranding: effects of corporate visual identity changes on employees and consumers. J Marketing Communications **24**(1), 3–16 (2018)
10. I. Ramírez, A. Cuesta-Infante, J.J. Pantrigo, et al., Convolutional neural networks for computer vision-based detection and recognition of dumpsters. Neural Comput & Applic **32**, 13203–13211 (2020)
11. C. Gorman, The Role of Trademark Law in the History of US Visual Identity Design, c.1860-1960. Journal of design history **30**(4), 371–388 (2017)
12. X. Li, L. Huang, Z. Wei, et al., Adaptive multi-branch correlation filters for robust visual tracking. Neural Comput & Applic **33**, 2889–2904 (2021)
13. E. Go, S.S. Sundar, Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions. Comput Hum Behav **97**(AUG.), 304–316 (2019)
14. M. Ochkovskaya, V. Gerasimenko, Buildings from the Socialist Past as part of a City's Brand Identity: The case of Warsaw. Bulletin of Geography. Socio-economic Series **39**(39), 113–127 (2018)
15. S.K. Jeong, Y. Xu, Behaviorally relevant abstract object identity representation in the human parietal cortex. Journal of Neuroscience **36**(5), 1607–1619 (2016)
16. Z. Xu, C. Cheng, V. Sugumaran, Big data analytics of crime prevention and control based on image processing upon cloud computing. J Surveill Secur Saf **1**, 16–33 (2020)
17. M.D. Vida, A. Nestor, D.C. Plaut, et al., Spatiotemporal dynamics of similarity-based neural representations of facial identity. Proceedings of the National Academy of Sciences **114**(2), 388–393 (2017)
18. D. Howell, S. Cox, B. Theobald, Visual units and confusion modelling for automatic lip-reading. Image Vision Computing **51**(6), 1–12 (2016)
19. M.L. Smith, B. Volna, L. Ewing, Distinct information critically distinguishes judgments of face familiarity and identity. J Exp Psychol Hum Percept Perform **42**(11), 1770–1779 (2016)

20. D.J. Humphries, F.M. Finch, M.B.V. Bell, et al., Vocal cues to identity: pied babblers produce individually distinct but not stable loud calls. Ethology **122**(7), 609–619 (2016)
21. A.L. Michal, D. Uttal, P. Shah, et al., Visual routines for extracting magnitude relations. Psychonomic Bulletin & Review **23**(6), 1802–1809 (2016)
22. K. Fizza, A. Banerjee, K. Mitra, et al., QoE in IoT: a vision, survey and future directions. Discov Internet Things **1**, 4 (2021)
23. P. Rahimian, J.K. Kearney, Optimal camera placement for motion capture systems. IEEE Transactions Visualization Comput Graphics **23**(3), 1209–1221 (2017)
24. L.D. Van, L.Y. Zhang, C.H. Chang, et al., Things in the air: tagging wearable IoT information on drone videos. Discov Internet Things **1**, 6 (2021)
25. W. Xu, A. Chatterjee, M. Zollhfer, et al., Mo2Cap2: Real-time mobile 3D motion capture with a cap-mounted fisheye camera. IEEE Transact Visualization Comput Graphics **25**(5), 2093–2101 (2019)
26. K.A. Mazurek, D. Richardson, N. Abraham, et al., Utilizing high-density electroencephalography and motion capture technology to characterize sensorimotor integration while performing complex actions. IEEE Transactions on Neural Systems and Rehabilitation Engineering **28**(1), 287–296 (2020)
27. R. Roberts, J.P. Lewis, K. Anjyo, et al., Optimal and interactive keyframe selection for motion capture. Computational Visual Media **5**(002), 171–191 (2019)
28. W. Hu, Z. Wang, S. Liu, et al., Motion capture data completion via truncated nuclear norm regularization. IEEE Signal Processing Letters **25**(2), 258–262 (2018)
29. A. Aissaoui, A. Ouafi, P. Pudlo, et al., Designing a camera placement assistance system for human motion capture based on a guided genetic algorithm. Virtual Reality **22**(1), 13–23 (2018)
30. Huzaifah bin Md Shahrin, M., Wyse, L, Applying visual domain style transfer and texture synthesis techniques to audio: insights and challenges. Neural Comput Applic **32**, 1051–1065 (2020)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.