Research Article

Football Players' Shooting Posture Norm Based on Deep Learning in Sports Event Video

Guangliang Huang,¹ Zhuangxu Lan,¹ and Guo Huang²

¹Guangxi College of Physical Education, Nanning, 530012, Guangxi, China

²Zhejiang Wanli University, Ningbo 315100, Zhejiang, China

Correspondence should be addressed to Guo Huang; 2002900057@zww.edu.cn

Received 2 September 2021; Revised 10 October 2021; Accepted 11 October 2021; Published 25 October 2021

Academic Editor: Le Sun

Copyright © 2021 Guangliang Huang et al. 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.

Football is one of the favorite sports of people nowadays. Shooting is the ultimate goal of all offensive tactics in football matches. This is the most basic way to score a goal and the only way to score a goal. The choice and use of shooting technical indicators can have a great impact on the final result of the game. Therefore, how to improve the shooting technique of football players and how to adjust the shooting posture of football players are important issues faced by coaches and athletes. In recent years, deep learning has been widely used in various fields such as image classification and recognition and language processing. How to apply deep learning optimization to shooting gesture recognition is a very promising research direction. This article aims to study the football player's shooting posture specification based on deep learning in sports event videos. Based on the analysis of target motion detection algorithm, target motion tracking algorithm, target motion recognition algorithm, and football shooting posture classification, KTH and Weizmann data sets are used. As the experimental verification data set of this article, the shooting posture of football players in the sports event video is recognized, and the accuracy of the action recognition is finally calculated to standardize the football shooting posture. The experimental results show that the Weizmann data set has a higher accuracy rate than the KTH data set and is more suitable for shooting attitude specifications.

1. Introduction

The recognition of human behavior is a hot topic of new research in the field of machine vision and artificial intelligence. Computer algorithms are used to automatically recognize human behavior from collected videos, that is, to classify and label video clips of human motion behavior [1, 2]. Compared with the research of still image recognition, behavior recognition pays more attention to how it perceives the temporal and spatial dynamic changes of the human body in the image sequence [3, 4]. Deep learning is a kind of machine learning, and machine learning is the only way to realize artificial intelligence. The concept of deep learning comes from the research of artificial neural network. Multilayer perceptron with multiple hidden layers is a deep learning structure. Deep learning combines low-level features to form more abstract high-level representation

attribute categories or features so as to find the distributed feature representation of data. The motivation of studying deep learning is to establish a neural network simulating the human brain for analytical learning, which simulates the mechanism of the human brain to interpret data.

In recent years, many scholars have conducted related research on action recognition and have achieved good results. Some scholars believe that the convolutional neural network is a deep model that can be directly used as the original input. However, this type of model is currently limited to a type specifically used to process 2D action input, so we have developed a special 3D-CNN model to process action recognition. This type of feature representation model is mainly used to perform 3D convolution from space. The corresponding features are extracted from the dimensions of time and time so as to capture and encode the motion data and information in multiple adjacent frames. The developed model generates

multiple channel feature information from multiple input frames at the same time, and the final feature expresses that the model combines all channel information [5]. In addition, some scholars have given a set of training videos. First, by extracting local motion and its appearance characteristics, they are quantified into a visual vocabulary, and then the vocabulary related to nearby points of interest and its direction is formed. The candidate neighborhood composed of central interest points shows how to learn in this way to form the most useful configuration class-specific distance function [6]. Some researchers have a method of recognizing human movements based on gesture primitives. In the learning mode, the parameters representing gestures and activities are estimated from the video. In the running mode, the method can be used in video or still images. In order to recognize the gesture primitives, the researcher extended the method based on the orientation gradient histogram. The descriptor proposed in [7] can better deal with joint poses and cluttered backgrounds. Because the current deep learning network has broken through the bottleneck of many traditional methods in many application fields and has achieved great success, it has received more and more attention from various fields of research and wide application and the international market, and they have built many efficient deep learning network frameworks themselves. The convolution neural network is constructed by imitating the visual perception mechanism of biology, which can carry out supervised learning and unsupervised learning. The parameter sharing of convolution kernel in its hidden layer and the sparsity of interlayer connection enable the convolution neural network to learn lattice features, such as pixels and audio, with less computation, stable effect, and no additional feature engineering requirements for data.

This paper uses deep learning algorithms to extract the features of football players' shooting posture in the KTH and Weizmann data sets, analyze and calculate the accuracy of their action recognition, and then compare them to select a data set that is more suitable for standardizing football players' shooting posture. In addition, deep learning emphasizes the depth of model structure. The importance of feature learning is clarified. In other words, the feature representation of the sample in the original space is transformed into a new feature space through layer by layer feature transformation so as to make classification or prediction easier. Compared with the method of constructing features by artificial rules, using big data to learn features can better describe the rich internal information of data.

The remainder of this paper is organized as follows. Section 2 introduces the different detection strategies in detail. Experimental results are given and deeply analyzed in Section 3. Finally, we summarize our contributions in Section 4.

2. Research on Football Players' Shooting Posture Norm Based on Deep Learning in Sports Event Video

2.1. Moving Target Detection

2.1.1. Light Field Flow Method. Optical flow is based on the instantaneous speed of the object at the imaging level,

reflecting the speed of the actual object's displacement in space. This is a way to find the correlation of object motion by using the time change of pixels in the video sequence and the correlation information between adjacent images and calculate the motion information of the object [8].

The field of view of moving objects tends to be continuous and smooth. Based on the continuous smoothness of the field of view, the overall smoothness of the field of view is considered a limiting condition and becomes a variable problem. The following is the global energy formula (1) based on the two-dimensional image:

$$E = \int \int [(I_x u + I_y v + I_t)^2 + \alpha^2 (\|\nabla u\|^2 + \|\nabla v\|^2)] dx dy. \quad (1)$$

Among them, for the image at one point, I_x , I_y , and I_t are used to represent the difference in the corresponding direction.

Besides, the advantage of optical flow method is that it can accurately detect and identify the position of moving heliostat without knowing the scene information, and it is still applicable when the camera is moving. The disadvantages of optical flow method are mainly reflected in the large amount of calculation and long time-consuming, which is not applicable in the case of strict real-time requirements; because the changing light will be incorrectly recognized as optical flow, this method is sensitive to light, which will affect the recognition effect.

2.1.2. Difference between Frames. The frame difference method is currently the most widely used method for moving target detection and extraction. The principle is to place two or three image frames next to the video frame sequence, distinguish the pixels corresponding to the two or three images, and extract the trajectory or contour of the moving target. First, it removes the pixels at the corresponding positions in the adjacent frame images to obtain the difference between the images and converts the difference image into a grayscale image. Then, if there is a small change in the environment, appropriate limits are set. If the value of a pixel is less than the threshold, it is considered as the background [9, 10]. The process of interframe differential transmission is basically illustrated in Figure 1.

The basic process of the frame difference method is as follows:

- (1) The difference image is obtained as follows:

$$D_k(x, y) = |f_k(x, y) - f_{k-1}(x, y)|. \quad (2)$$

- (2) Binarization processing of difference image is as follows:

$$R_k(x, y) = \begin{cases} 0, & D_k(x, y) < T, \\ 1, & D_k(x, y) \geq T. \end{cases} \quad (3)$$

In addition, the advantages of the interframe difference method are simple algorithm implementation and low programming complexity. It is not

FIGURE 1: Flow chart of the interframe difference method.

sensitive to scene changes such as light and can adapt to various dynamic environments with good stability. Its disadvantages are as follows: it can not extract the complete area of the object but only the boundary. It also depends on the selected interframe time interval. For fast-moving objects, it is necessary to select a smaller time interval. If the selection is inappropriate, when the objects do not overlap in the front and back frames, they will be detected as two separate objects; for slow-moving objects, a larger time difference should be selected. If the time selection is inappropriate, when the objects almost completely overlap in the front and back frames, the objects will not be detected.

- (3) Mathematical morphological processing is performed on binary images.

2.1.3. Background Difference Method. The general concept of the background difference method is to remove the background and leave moving objects, which requires a predefined background pattern. The training image is used to obtain the model parameters. Once the background model is defined, the relative pixel values of the existing image and the background image are subtracted to identify the desired animation target. Once you have the original background, you need to update the background, update the method background model of the method, then calculate the difference between the two images, delete the two images, and use the image threshold to reach the foreground target [11, 12]. Algorithm expression is as follows:

$$f_b(x, y, t) = \begin{cases} 1, & \text{if } |f(x, y, t) - f(x, y, 0)| > T, \\ 0, & \text{others.} \end{cases} \quad (4)$$

Among them, $f_b(x, y, t)$ is the image at time t and $f(x, y, 0)$ is the background image at time t .

It should be noted that in the moving target detection based on the background difference method, the accuracy of

background image modeling and simulation directly affects the detection effect. No matter any moving target detection algorithm, it should meet the processing requirements of any image scene as much as possible. However, due to the complexity and unpredictability of the scene, as well as the existence of various environmental interference and noise, such as the sudden change of illumination, the fluctuation of some objects in the actual background image, the jitter of the camera, and the impact of moving objects entering and leaving the scene on the original scene, it makes the modeling and simulation of background more difficult.

2.2. Moving Target Tracking. The moving target tracking process is shown in Figure 2.

2.2.1. Feature-Based Tracking Method. Feature-based tracking methods mainly rely on the ability of the target to track the target, such as feature extraction and feature matching. First, the target attribute is derived, and then the structure of the video image sequence is used to extract the target attribute, then the structure mapping or search mapping method of the video image sequence is used to find the corresponding attribute point, and finally the location and trajectory of the target information of the traffic matching result are calculated according to the following formula. Because the overall characteristics of the target are not considered, only the local characteristics of the target are identified, and the anti-interference effect is good, and the monitoring target is not affected by scale changes or distortion.

2.2.2. Tracking Algorithm Based on Kernel Density Estimation. The estimation method based on kernel density is currently widely used in the fields of pattern recognition and computer vision. It can create a target model based on the feature probability density distribution of the target image. Since the kernel function has progressive fairness, continuity, and uniform continuity, the gradient estimation of the density function determined by the kernel function can be applied to the entire image space. When estimating the boundary region, there will be boundary effect in kernel density estimation. On the basis of univariate kernel density estimation, the prediction model of value at risk can be established. By weighting the coefficient of variation of kernel density estimation, different var prediction models can be established. Parameter estimation can be divided into parametric regression analysis and parametric discriminant analysis.

2.2.3. Region-Based Matching Algorithm. The algorithm of interest mapping based on grayscale images is also called the algorithm of interest mapping to specific areas. All the gray information based on the image in the area is used to calculate the similarity between gray information in a certain way. The region-based mapping method not only needs to directly extract the features of the region but can directly use all available grayscale image information, which can improve the

FIGURE 2: Moving target tracking process.

accuracy and robustness to a certain extent. One of its main shortcomings is that the amount of calculation errors is large, and it is difficult to meet real-time requirements. Choosing areas with insufficient gray information as feature areas will increase the error rate of matching.

The basic idea of gray matching is as follows: from the statistical point of view, the image is regarded as a two-dimensional signal, and the statistical correlation method is used to find the correlation matching between signals. The correlation functions of the two signals are used to evaluate their similarity to determine the homonymous points. Gray matching determines the corresponding relationship between the two images by using some similarity measures, such as correlation function, covariance function, sum of squares of difference, and sum of absolute values of difference.

2.3. Moving Target Recognition. The target recognition movement can be simply considered as a combination of feature derivation and classifier recognition. The recognition process is generally shown in Figure 3.

The related recognition algorithm is as follows:

- (1) *Convolutional Neural Network.* For 2D-CNN networks, the convergence function can only receive 2D feature maps from the spatial dimensions of the input information. For video information, it is necessary to record motion information in a series of continuous video frames. In response to this situation, 3D-CNN uses 3D convergence and aggregation functions to simultaneously download 3D feature maps in both time and space dimensions. The 3D concatenation function can be considered as a 2D convergence layer stack, so the generated 3D feature map also connects partially continuous input frames. For the 2D rotation function, the value (x, y) at any position on the j -th feature map of the i -th plane is calculated as follows:

$$v_{ij}^{xy} = \tanh \left(b_{ij} + \sum_m \sum_{p=0}^{P_{i-1}} \sum_{q=0}^{Q_{i-1}} w_{ijm}^{pq} v_{(i-1)m}^{(x+p)(y+q)} \right), \quad (5)$$

FIGURE 3: Moving target recognition process diagram.

where m is the feature map of the $(i-1)$ layer connected with the current feature map.

Convolution layer parameters include convolution kernel size, step size, and filling, which together determine the size of convolution layer output characteristic graph, which is a super parameter of convolution neural network. The size of the convolution kernel can be specified as any value smaller than the size of the input image. The larger the convolution kernel is, the more complex the input features can be extracted.

- (2) *Self-Phase Matrix.* The self-similar table is a graph that can reflect the periodic characteristics of the system. In many dynamic systems, state cycle is a relatively important phenomenon. In order to graphically represent the periodic characteristics of this dynamic system, a recursive diagram can be used to show the retrospective nature of the state in the phase space. The recursive diagram is defined as follows:

$$R(i, j) = H(\varepsilon - \|\vec{x}(i) - \vec{y}(j)\|_2), \quad \vec{x}(i) \in \mathfrak{R}^m, i, j = 1, \dots, N, \quad (6)$$

where N refers to the number of states $\vec{x}(i)$ under consideration and H is a step function.

- (3) *LSTM Neural Network*. The LSTM neural network is designed to solve the problem of long-distance dependence, and it can store long-term information without particularly complex over-tuning parameters. The internal structure of LSTM is more complex than that of RNN. Through this complex calculation process, LSTM has long-term memory, excellent data processing capabilities with time series capabilities, and improved network performance.

The operation process of LSTM is as follows:

$$\begin{aligned} i_t &= \sigma_i(w_i \cdot [h_{t-1}, x_t] + b_i), \\ D_t &= \sigma_c(w_c \cdot [h_{t-1}, x_t] + b_c). \end{aligned} \quad (7)$$

Equation (8) calculates f_t of the value of forget gate at time t .

$$f_t = \sigma_f(w_f \cdot [h_{t-1}, x_t] + b_f). \quad (8)$$

Formula (9) calculates C_t of the new state value of the state information at time t .

$$C_t = f_t * C_{t-1} + i_t * D_t. \quad (9)$$

Among them, C_t represents the state of information transfer, h is the value of the hidden layer, and b represents the bias top.

In addition, LSTM usually performs better than time recurrent neural network and hidden Markov model, such as in piecewise continuous handwriting recognition. LSTM is a kind of neural network containing LSTM blocks or others. In order to minimize the training error, the gradient descent method, such as the sequential back-propagation algorithm, can be used to modify the weight each time according to the error.

2.4. Football Shooting Posture Classification

2.4.1. Instep Shooting. Instep shots have strong firepower and are mainly used for firing when turning around. When the ball falls in front of or slightly away from the body, you can shoot with the inner instep. It can temporarily change the angle of a shot. For example, when the diagonal line is inserted directly, the goalkeeper will have the opportunity to inevitably move to the nearest position to close the nearest corner. At this time, a half-turn shot can be made. Shoot it into the furthest corner.

2.4.2. Shooting outside the Instep. Shooting from outside the instep is very threatening, sudden, and highly concealed. It can shoot the ball from various directions including front, side, and back and can shoot straight and curved balls.

2.4.3. Arch Shooting. Arch shooting is suitable for high precision, low power consumption, various short shots, and free throws.

2.4.4. Shot from the Back of the Foot. Forefoot shooting is more powerful, more accurate, and the most widely used. This is the basic footwork of shooting footwork.

2.4.5. Toe Shot. A tiptoe shot is fast and sudden. If the competition in front of the goal is fierce, you may not have time to move your legs. Shooting with a tiptoe can win surprisingly, but sometimes the accuracy is relatively low.

2.5. The Difference between Existing Works and Our Work. Compared with the traditional machine learning algorithm, the target detection task needs to be divided into multiple subtasks, which is very cumbersome and time-consuming. The deep learning algorithm used in this paper can effectively realize end-to-end training and learning and greatly reduce the working time. In addition, the accuracy of the traditional machine learning algorithm is limited, and the deep learning algorithm can achieve very high accuracy with the help of a large amount of data.

3. Experiment

3.1. Data Collection. With the development of video data acquisition equipment, a large amount of video data is poured into us every day. It provides useful video data for the recognition research of human behavior. Personal collection of video data including actions, consumes a lot of resources, and the collected data are relatively one-sided, so the recognition results may not fully reflect fairness and impartiality. Therefore, this article uses the public data set as the experimental verification set of this article, namely, the KTH and Weizmann data set.

3.2. Feature Extraction. Deep learning is used to calculate the displacement vector field d_t of the pixel (x, y) between two consecutive frames at t and $t+1$ and find the function u and v that minimizes the energy function. The K-proximity calculation has data items and smoothing items to form the optimization of the global energy function. The mathematical form is expressed as follows:

$$E_{\text{Global}} = E_{\text{Data}} + \lambda E_{\text{Smooth}}. \quad (10)$$

Among them, E_{Data} is a data item.

4. Discussion

4.1. KTH Data Set Result Analysis. After the KTH data set is preprocessed, 50 single-frame image data sets of the shooting postures of 5 types of football players are obtained.

TABLE 1: Accuracy of various types of action recognition.

Shot type	Accuracy (%)
Instep shot	91
Outside instep shot	84
Arch shot	90
Forefoot shot	92
Toe shot	94
Average of all test sets	90.2

FIGURE 4: Accuracy of various types of action recognition.

Through many experiments, the final recognition accuracy rate on the KTH data set is 90.2%. It can be seen from Table 1 and Figure 4 that the recognition rate of the four shooting postures exceeds 90%, of which the accuracy rate of instep shooting is 91%, the accuracy rate of arch shooting is 90%, and the accuracy rate of front instep shooting is 92%. The accuracy rate of toe shots is 94%, while the accuracy rate of shots from the outside of the instep does not reach 90%, only 84%. It can be seen from the experiment that the results obtained by the method in this article have a higher accuracy rate, which fully reflects the advantages of deep learning to automatically extract features. The deep learning method is to automatically extract features, which has more generalization ability and reduces the cost of manpower and material resources and greatly saves the cost of human action recognition.

4.2. Analysis of Weizmann Data Set Results. After pre-processing the Weizmann data set, 45 groups of single-frame image data sets of the shooting postures of 5 types of football players are obtained. The form of a chart is used to intuitively reflect the recognition accuracy of each human body action category. The recognition rate of each category is shown in Table 2 and Figure 5.

The average recognition rate of this method on the Weizmann data set is 95.2%, of which the accuracy rate of the inner instep shot is 99%, the accuracy rate of the outer instep shot is 94%, the accuracy rate of the instep shot is 95%, and the accuracy rate of the instep shot is 95%. The accuracy rate is 91%, and the accuracy rate for tiptoe shots is 97%.

TABLE 2: Accuracy of various types of action recognition.

Shot type	Accuracy (%)
Instep shot	99
Outside instep shot	94
Arch shot	95
Forefoot shot	91
Toe shot	97
Average of all test sets	95.2

FIGURE 5: Accuracy of various types of action recognition.

5. Conclusions

As modern society has put forward higher requirements for human-computer interaction technology and effects, whether it is widely used in people's leisure and entertainment, medical care, intelligent monitoring, and other technologies and applications closely related to people's daily life or some specialized motion correction and other aspects of motion recording technology has greatly promoted the development of human motion recognition research. This paper uses deep learning algorithms to identify and analyze football shooting postures in sports event videos in the KTH and Weizmann data sets. The results show that the Weizmann data set has higher accuracy than the KTH data set and is more suitable for shooting posture specifications. Deep learning is the internal law and representation level of learning sample data. The information obtained in the learning process is very helpful to the interpretation of data such as text, image, and sound. Its ultimate goal is to make machines have the ability of analysis and learning like people.

Data Availability

The experimental 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 to report regarding the present study.

References

- [1] A. Akula, A. K. Shah, and R. Ghosh, "Deep learning approach for human action recognition in infrared images," *Cognitive Systems Research*, vol. 50, pp. 146–154, 2018.
- [2] Y. Tian, Y. Kong, Q. Ruan, G. An, and Y. Fu, "Hierarchical and spatio-temporal sparse representation for human action recognition," *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1748–1762, 2018.
- [3] H.-H. Pham, L. Khoudour, A. Crouzil, P. Zegers, and S. A. Velastin, "Exploiting deep residual networks for human action recognition from skeletal data," *Computer Vision and Image Understanding*, vol. 170, pp. 51–66, 2018.
- [4] J. Basavaiah, C. Patil, and C. Patil, "Robust feature extraction and classification based automated human action recognition system for multiple datasets," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 1, pp. 13–24, 2020.
- [5] S. Kumawat, M. Verma, Y. Nakashima, and S. Raman, "Depthwise spatio-temporal STFT convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1, 2021.
- [6] A. Sarabu and A. K. Santra, "Distinct two-stream convolutional networks for human action recognition in videos using segment-based temporal modeling," *Data*, vol. 5, no. 4, p. 104, 2020.
- [7] B. Gu, W. Xiong, and Z. Bai, "Human action recognition based on supervised class-specific dictionary learning with deep convolutional neural network features," *Computers, Materials & Continua*, vol. 62, no. 3, pp. 243–262, 2020.
- [8] M. M. Zhdanova, O. S. Balabaeva, V. V. Voronin, E. A. Semenishchev, and A. A. Zelensky, "Improvement of the human action recognition algorithm by the pre-processing of input data," *IOP Conference Series: Materials Science and Engineering*, vol. 1029, no. 1, Article ID 012119, 2021.
- [9] A. Sarabu and A. K. Santra, "Human action recognition in videos using convolution long short-term memory network with spatio-temporal networks," *Emerging Science Journal*, vol. 5, no. 1, pp. 25–33, 2021.
- [10] J. Kusch, M. Burgdorff, H. Filaretov, and J. Krüger, "Inertial measurement unit based human action recognition for soft-robotic exoskeleton," *IOP Conference Series: Materials Science and Engineering*, vol. 1140, no. 1, Article ID 012020, 2021.
- [11] R. Liu, Z. Liu, and S. Liu, "Recognition of basketball player's shooting action based on the convolutional neural network," *Scientific Programming*, vol. 2021, no. 3, 8 pages, Article ID 3045418, 2021.
- [12] M. Navarro, J. van der Kamp, P. Schor, and G. J. P. Savelsbergh, "Implicit learning increases shot accuracy of football players when making strategic decisions during penalty kicking," *Human Movement Science*, vol. 61, no. 61, pp. 72–80, 2018.