www.cadanda.com
Computer-Aided Design AND APPLICATIONS

Research on Multi-person Pose Estimation Technology in Sports Competition Video Based on Big Data Analysis

Shengyu Zhang[1]

[1]Military Physical Education Teaching Department, JiangXi University of Engineering, Xinyu, JiangXi, 338000, China

Corresponding author: Shengyu Zhang, zhangshengyu2017@163.com

Abstract: The research of human pose estimation has changed from the early estimation methods of artificially selected features such as graph models and graph structure models to intelligent image recognition methods based on machine learning. In order to improve the effect of multi-person pose estimation in sports competition video, this paper combines big data technology and image recognition technology to study the multi-person pose estimation technology in sports games. Moreover, starting from the actual situation, this paper enhances the effect of human pose feature recognition in sports competitions, and constructs multi-person pose estimation system in sports competition video based on big data analysis. In addition, this paper combines big data technology to analyze sports competition video data, and uses image processing methods to extract features. The experimental research results show that the multi-person pose estimation system in sports competition video based on big data analysis can realize the feature recognition of multi-person sports.

Keywords: Big data; sports competition; multi-person pose; pose estimation
DOI: https://doi.org/10.14733/cadaps.2023.S12.188-201

1 INTRODUCTION

The rapid improvement of computer computing power and the rapid growth of available training samples provide the possibility for the application of deep learning, and also bring new ideas and new methods for the research of human body pose estimation. The specific problem of the human body pose estimation task is the insufficient training data set. Like many image processing tasks based on deep learning methods, the human body pose estimation method based on deep learning is also supervised learning and requires a large number of data samples for supervised training to obtain a neural network that can handle the corresponding tasks well. However, there are few public body posture data sets, especially human body posture data sets in complex outdoor scenes. The reason is that the calibration of the human body posture data sets is relatively difficult. When there is a person in a given image, the data set usually needs to provide the coordinates of the human

body's head, neck, left and right shoulders, left and right elbows, left and right wrists, left and right hips, left and right knees, and left and right ankles, a total of 14 human body joint points.

In order to easily test the different performance of the human body posture estimation algorithm, the human body posture data set is also required to provide the marked rotation angle, the size of the human body or the category information of the human body posture. When the human body image has posture distortion and occlusion, especially in complex scenes, such as low illumination and complex background objects, it is difficult for the human eye to give a clear joint point position annotation. These reasons will increase the difficulty of annotating human poses, which greatly increases the cost of human pose annotations. It is these reasons that have led to no such large-scale human pose datasets as ImageNet. This brings huge challenges to the development of human body pose estimation methods based on deep learning. Complex background interference issues. The background of part of the image data is relatively simple or relatively single. When performing human pose estimation, the local weight sharing and pooling structure of the convolutional neural network can be used to directly extract the local features of the invariance of the local pose and illumination, thereby regressing Better coordinate map of human body joints. For images with complex backgrounds, there are often image areas in the background that have similar image features with human body parts, which are extremely similar to human body parts. Only local feature information cannot be used to infer whether the area belongs to human parts, and it is easy to cause estimation. Ambiguity.

This paper combines big data technology and image recognition technology to study the multi-person pose estimation technology in sports competitions. Moreover, starting from the actual situation, this paper improves the human pose feature recognition effect in sports competitions, and on this basis, further improves the fairness and reliability of sports competitions

2 RELATED WORK

Estimating human poses with three-dimensional spatial position information from multi-view sequences is widely used in the field of computer vision, such as pose recognition [4], behavior recognition [11], motion capture [9], and human-computer interaction [20]. Although this research has achieved major breakthroughs, there are still many challenging and urgent problems to be solved. First, there is a semantic gap between complex three-dimensional human motion and two-dimensional images, and the lack of depth information leads to ambiguity in the estimated human pose. Secondly, the appearance and contour of the human body are quite different from frame to frame, which brings difficulties to the positioning of the limbs. Finally, the occlusion of limbs, the high dimensionality of pose data, and the change of scene light all make human body pose estimation a difficult task. At present, the common methods are divided into 3 categories. One is the model-based method, which relies on a human body model based on prior knowledge to estimate the human body pose by optimizing the objective function, but the amount of calculation is relatively large [18]. The second is a learning-based method, which directly learns the mapping from the feature space to the pose space [7], but its judgment of the pose is based on huge training data. The third is a sample-based method, which retrieves the most similar data to the input in the training database, and uses the data to interpolate to obtain the result. This method also relies on huge training data, and requires the training data to cover as much freedom as possible in the pose space [14]. Recently, an algorithm called sparse coding [19] has emerged in the field of machine learning and pattern recognition, such as face recognition [17], target classification, and human body pose estimation.

Literature [8] proposed a sparse coding algorithm, which overcomes the overfitting caused by the small sample problem by learning two over-complete dictionaries. Literature [1] proposed a sparse coding algorithm to solve the problem of data location and similarity information loss, but the original space features are often noisy, and the regular term constructed in this space may not

accurately reflect the inner manifold of the data. Literature [13] proposed a kernel sparse coding algorithm that can capture the nonlinear similarity of features, breaking the mode of coding only in the original space. However, it is unreasonable to use a single kernel function to process different types of inputs, and it is faced with the problem of kernel function and parameter selection. The samples in human pose estimation are high-dimensional and non-linear. Although the nearest neighbors of the samples can be found at linear distances, the nearest neighbor graph constructed by them cannot accurately reflect the inner manifold of the data, which is used in many applications. It's very important. The kernel technique implicitly maps the original data to the Hilbert space to overcome this problem, but it also faces the problem of choosing kernel functions and parameters. Although the cross-validation method can solve this problem, the amount of calculation is too large. Literature [10] proposed a multi-core sparse coding algorithm. By introducing multi-core learning, it not only solves the "curse of dimensionality" of pose data, but also copes with the nonlinearity of samples. Among them, the optimal kernel is derived from the linear combination of the kernel functions in the kernel function set, so there is no problem of choosing kernel functions and parameters.

Deep convolutional neural networks have improved the performance of many computer vision tasks to a new level. The overall trend is to build deeper and more complex networks to achieve higher accuracy [6]. However, this will greatly increase the amount of parameters and calculations of the network, making the deep convolutional neural network completely unable to meet the requirements of edge devices or mobile devices in terms of scale and speed. In particular, for some ultra-deep networks with hundreds of layers, the amount of parameters and calculations are huge. In order to embed the neural network structure of different tasks into public security cameras or some mobile devices, the image processing tasks are directly performed in the edge devices, and the structure information of the video image is formed, and then the image and video structure information is transmitted. This deployment can reduce the cost of network deployment and save a lot of transmission bandwidth [15].

3 MULTI-PERSON POSE ESTIMATION ALGORITHM IN SPORTS COMPETITION VIDEO

We are given a target sample f, which is composed of N-dimensional feature vectors, that is, $f = \{f_d\}_{d=1:N_d}$. For a two-dimensional image, N_d is the dimension of the extracted visual feature. $f(n) \in R^d$, and n is each point in the bounding box of the target object. Our goal is to learn the correlation filter $h = \{h_d\}_{d=1:N_d}$ for each feature dimension to make the final response result close to the predefined ideal response g. f_d, h_g and g have the same dimension size. In the two-dimensional case, the heights of these three are all d_n and the widths are all d_n. We use the L2 norm to calculate the error to optimize the following objective function [16]:

$$\varepsilon = \left\| \sum_{d=1}^{N_d} f_d \cdot h_d - g \right\|^2 + \lambda \sum_{d=1}^{N_d} \left[h_d \right]^2 \tag{3.1}$$

Among them, · represents the circular convolution operation, λ is the regularization term coefficient, The ideal response g is generally a Gaussian distribution in N-dimensional space, the expectation is 0, and the standard deviation is σ, as shown in Figure 1. The circular convolution operation essentially treats all circular displacement results of the target object as negative samples.

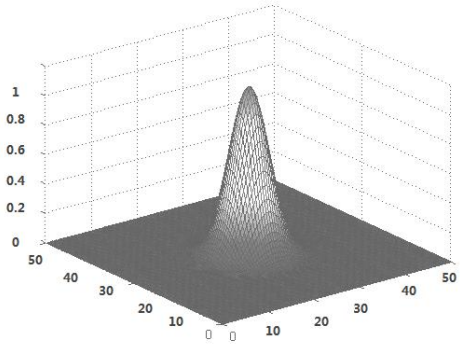

Figure 1: Two-dimensional Gaussian ideal response.

$$\varepsilon = \left\| \sum_{d=1}^{N_d} diag\left(f_d \right) \overline{h_d} - \hat{g} \right\|^2 + \lambda \sum_{d=1}^{N_d} \left[h_d \right]^2 \tag{3.2}$$

Among them, $a = vec\left(\mathcal{F}\left(a\right)\right)$ means to perform Fourier transform on a first and then expand it into a one-dimensional column vector, diag(a) means transforming a into a diagonal matrix, and $\overline{a}$ means conjugate operation. The minimization function (3.2) can solve the corresponding correlation filter for each characteristic channel. We can find the mathematical closed solution of the least squares problem by letting its first derivative be zero:

$$h_d = \frac{\overline{g_d}\, f_d}{\sum_{d=1}^{N_d} f_d\, \overline{f_d} + \lambda} \tag{3.3}$$

Among them, the circumferential displacement of the target object image machine is shown in Figure 2.

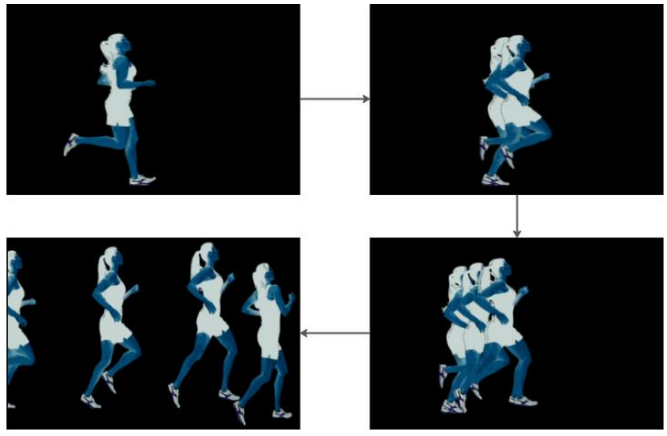

Figure 2: The image of the target object and its circumferential displacement.

The formula (3.3) is based on the closed solution of the correlation filter obtained from a single time point sample, that is, learning and training are performed based on a positive sample and multiple negative samples. In order to obtain a more robust correlation filter, the samples at multiple previous time points can be considered, and the updated spatial objective function is as follows [12]:

$$\varepsilon = \sum_{k=1}^{N_k} \left\| \sum_{d=1}^{N_d} f_d^k * h_d - \hat{g} \right\|^2 + \lambda \sum_{d=1}^{N_d} \left\| h_d \right\|^2 \tag{3.4}$$

Among them, Nk is the number of sample time points that need to be considered, αk is the weight of samples at different time points, and f knot is the d-th dimension feature of the h-th time sample. Kiani derives the mathematical closed solution of this problem in his article, and finally needs to solve N linear systems with the size of $N_d \times N_d$, where N is the number of elements in h_d. For a two-dimensional image, N is the number of pixels in h. However, this kind of thinking will bring a lot of computational overhead. Therefore, it is often used in offline object tracking algorithms, such as CCOT and other algorithms.

In order to ensure the running speed and robustness of the algorithm at the same time, this paper refers to the idea of dynamic update in MOSSE algorithm. For equation (3.3), the d-th dimension feature at time r is f_d^t. If the numerator and denominator of h_d are A_d^t and B_d^t, the dynamic update strategy is as follows [5]:

$$A_d^t = (1-\eta) A_d^{-1} + \eta \overline{g} \cdot f_d^t \tag{3.5}$$

$$B^t = (1-\eta) B_d^{t-1} + \eta \sum_{d=1} f_d^t \overline{f_d^t} \tag{3.6}$$

Among them, η is the learning rate. This dynamic update strategy puts more weight on the filter model obtained by learning and solving at the latest time point, and the filter model at the previous time point decays exponentially with time. In this way, calculation speed and robustness can be taken into account at the same time.

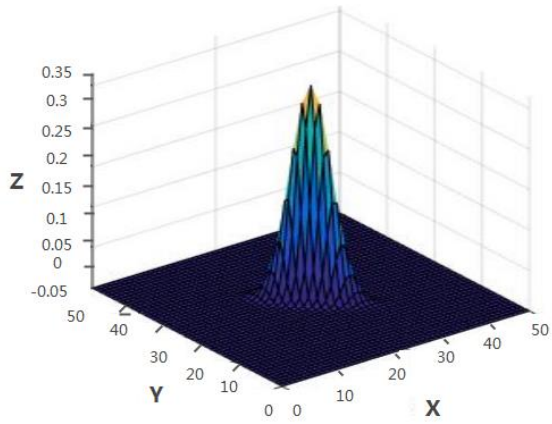

Figure 3: Correlation filter response during object tracking in two-dimensional image.

In order to apply the correlation filter for calculation, at the previous point in time, the neighborhood of the state of the target object is extracted as a new sample $u = \left\{u_d^t\right\}_{d=1:N_d}$. For two-dimensional images, that is, taking a position on the target object as the center, extracting partial images in a certain area of space and calculating visual features. After that, it interacts with the correlation filter to generate a response, and the calculation method is as follows[2]:

$$y^t = \frac{\sum_{d=1}^{N_d} A_d^{t-1} \cdot u_d^t}{B^{t-1} + \lambda} \tag{3.7}$$

After that, we use the inverse Fourier transform to convert the current frequency domain response y^t to the spatial domain y^t, and the maximum point of the spatial response is the current state of the target object. As shown in Figure 3, the bee point is the current position of the target. In essence, the algorithm assumes that the state of the target object changes little in adjacent time, so it can be detected and searched in the vicinity of the target object state at the previous point in time. It further finds the new state of the object and makes full use of the continuity of spatial information changing over time, which is more efficient than training a global detector. For the problem of tracking the position of a target object in a two-dimensional image, the algorithm execution flow is as follows.

This paper proposes the SRDCF algorithm to perform spatial constraints, and the modified optimization function is as follows [3]:

$$\varepsilon = \left\| \sum_{d=1}^{N_d} f_d * h_d - \hat{g} \right\|^2 + \sum_{d=1}^{N_d} \left\| w \cdot h_d \right\|^2 \tag{3.8}$$

Among them, w is the spatial regularization constraint weight. From the target object area to the background area, the corresponding value of w gradually increases. The correlation filter is constrained by a larger weight value in the area outside the target object, and the spatial weight value restricted in the target object area is smaller. The correlation filter calculated in this way has a better ability to discriminate the target object. The distribution of w is shown in Figure 4

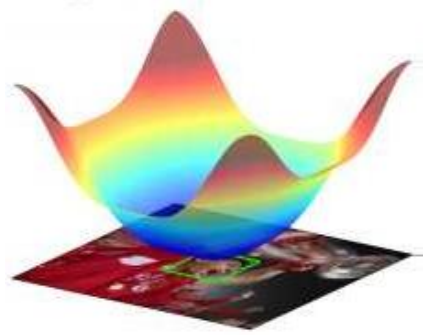

Figure 4: SRDCF algorithm spatial weight constraint**.**

However, this kind of airspace constraint method only gradually increases from the center to the edge according to the pre-defined rules. It has a better constraint effect on square or round target objects, but it is not suitable for irregular target objects, especially easily. Deformable objects. This

paper refers to the work of Lukezic et al. to construct the spatial mask, and uses the ADMM algorithm to optimize the correlation filter model.

The airspace mask m has the same dimensional size as f_d, h_d, and g, the height is d_h, and the width is d_h. Each pixel of the spatial mask is m(n)∈{0,1}. 1 represents that the pixel position belongs to the target object, and 0 represents that the pixel position belongs to the background. With the constraints of the spatial mask, the search area of the correlation filter can be made larger than the bounding box of the target object, similar to that shown in Figure 4. The area in the bounding box of the target object is the foreground, and the surrounding area is the background. Construct a histogram for the foreground area and the background area as the appearance model y of the current target object. If the appearance model y of the target object and the pixel point x are given, the probability that x belongs to the target object can be expressed as follows:

$$p\left(m\left(x\right)=1\left|\, y_a, x\right.\right) \propto p\left(y_a \left| m\left(x\right)=1, x\right.\right) p\left(x \left| m\left(x\right)=1\right.\right) p\left(m\left(x\right)=1\right) \tag{3.9}$$

Among them, the first appearance similarity $p\left(y_a \left| m\left(x\right)=1, x\right.\right)$ can be solved by Bayesian total probability model using histogram reverberation projection algorithm. The third mask prior probability $p\left(m\left(x\right)=1\right)$ is expressed by the ratio of the foreground area of the target object to the background area. The second term $p\left(x \left| m\left(x\right)=1\right.\right)$ is the prior probability of the spatial distribution of the target object.

Using a bounding box aligned with the coordinate axis to indicate the position of the target object, it can be considered that the center area of the bounding box must belong to the target object itself, and is not affected by factors such as rotation or deformation. At the same time, the farther the position from the center point of the bounding box, the greater the probability that it belongs to the background area. The calculation method of $p\left(x \left| m\left(x\right)=1\right.\right)$ is:

$$p\left(x \left| m\left(x\right)=1\right.\right) = k\left(x;\sigma\right) = 1-\left(r/\sigma\right)^2 \tag{3.10}$$

Among them, $k\left(x;\sigma\right)$ is the Epanechnikov kernel function, and r is the distance between the pixel point x and the center point of the bounding box. In order to constrain the influence of the prior probability of the spatial distribution, $p\left(x \left| m\left(x\right)=1\right.\right)$ can be restricted to the interval of $\left[p_{low}, p_{up}\right]$. Among them, $p_{low} \in \left[0,1\right], p_{up} \in \left[0,1\right]$, which is different from the scheme in the literature, and the interval setting can be done flexibly. In this way, the prior probability of the spatial distribution of the center of the bounding box is p_{up} · gradually attenuates to p_{low} around the periphery, which can change this interception interval to adjust the influence of the prior probability of the spatial distribution.

According to equation (3.9), the spatial mask calculated by multiplying the three terms is noisier. In order to better segment the foreground objects, based on the work of Kristan et al., Markov random field is used for smoothing, and the final spatial mask is solved by maximizing the posterior probability. If only a few pixels belong to the target object in the end, the spatial mask is set to be uniformly distributed in the target object's storage box, so as to avoid the influence of too

small foreground area. In addition, an expansion operation is performed on the resulting spatial mask to make the edges smoother. In each image frame, the histogram of the front background of the target object is dynamically updated according to the predefined learning rate by linear interpolation method, and a new spatial mask is solved.

The solution process of the spatial mask is shown in detail in Figure 5, and here we set $p_{low} = 0.5$. $p_{up} = 0.9$.

Note that the algorithm takes the center point of the target object as the center and extracts the sample image with a certain size. For pixels beyond the boundaries of the original image, repeated processing operations are used to fill the boundary pixels.

After the spatial mask of the target object is obtained, the correlation filter model can be constrained. The constraint equation is:

$$h_d = m \cdot h_d \tag{3.11}$$

That is, the correlation filter of each feature channel only works in the area where the spatial mask m is 1. In the area where the spatial mask m is 0, the value of h_d is zero. On the one hand, the correlation filter is no longer affected by the background area in the bounding box of the target object, on the other hand, the background area outside the bounding box of the target object will not affect the solution of the correlation filter, and the search area can be arbitrarily large. This constraint method is more accurate than (3.8), and can effectively solve many problems caused by edge effects. However, this constraint makes the formula (3.1) no longer have a mathematically closed solution such as formula (3.3), and an iterative method must be used to optimize the solution.

In order to simplify the problem, we first perform channel separation on equation (3.1) and transform it into the following objective function:

$$\varepsilon = \sum_{d=1}^{N_d} \left\| f_d * h_d - \hat{g} \right\|^2 + \lambda \sum_{d=1}^{N_d} \left\| h_d \right\|^2 \tag{3.12}$$

That is, it is considered that the characteristics of each channel are independent of each other, and each correlation filter model can be solved separately. Then, the objective function of each independent correlation filter is converted to the Fourier frequency domain according to Paseval's theorem. Under the constraint of formula (3.11), the augmented Lagrangian function is as follows:

$$L\left(h_c, h, \hat{I} \mid m\right) = \left\| diag\left(\hat{f}\right)\overline{h_c} - g \right\|^2 + \frac{\lambda}{2}\left\|h_m\right\|^2 + \left[\hat{I}^H\left(h_c - h_m\right) + \overline{\hat{I}^H\left(h_c - h_m\right)}\right] + \mu\left\|h_c - h_m\right\|^2 \tag{3.13}$$

Among them, due to the independence of each channel, we omit the subscript d. h_c is a dual variable that satisfies the constraint $h_c - m \cdot h = 0$, $h_m = m \cdot h$, and I are Lagrangian multipliers, and λ and μ are regularization constraints.

Equation 3.13) can be solved quickly iteratively by using the ADMM (alternating direction method of multipliers) algorithm. The ADMM algorithm is widely used to solve the problem of two optimization variables and an equation constraint term. The specific selection process of (13) is as follows:

$$h_c^{i+1} = \arg\min_{h_c} L\left(h_c, h_i, I^i \mid m\right) \tag{3.14}$$

$$h^{i+1} = \arg\min_{h} L\left(h_c^{i+1}, h, I^i \mid m\right) \tag{3.15}$$

$$I^{i+1} = I^i + \mu\left(h_c^{i+1} - h^{i+1}\right) \tag{3.16}$$

$$\mu^{i+1} = \beta\mu^i \tag{3.17}$$

That is, in each iteration, first fix h to solve h_c that minimizes the objective function, then fix h to solve h that minimizes the objective function, and finally update the Lagrangian multiplier I and the regularization constraint term μ. The ADMM algorithm alternates iteratively in the directions of the two variables, and the iteration speed is faster.

Equations (3.14) and (3.15) are two optimization sub-problems, and the mathematical closed solution can be obtained by taking the first derivative of the objective function and making it zero;

$$h_c^{i+1} = \frac{\hat{f} \cdot \overline{g} + \mu^i h_m^i - I^i}{f - \overline{f} + \mu^i} \tag{3.18}$$

$$h^{i+1} = \frac{,\cdot\mathcal{F}^{-1}\left(\mu^i h_m^i - I^i\right)}{\mu^i + \lambda / 2D} \tag{3.19}$$

Among them, $D = d_w \times d_h$ is the number of elements in the correlation filter

4 MULTI-PERSON POSTURE ESTIMATION SYSTEM IN SPORTS COMPETITION VIDEO

This article will summarize and design all the components of the lightweight and refined deformation frame. They are: preprocessing module, target detection module, human body pose estimation module, post-processing module. The overall process of multi-person pose estimation in this paper is shown in Figure 5.

Background modeling methods mainly include Gaussian mixture model (MOG), codebook model (CodeBook) and visual background extraction model (ViBe algorithm), etc. The basic principle of the background modeling method is to first model the background of the input image, then compare the pixels of the current frame image and the background model, and determine whether there are moving objects in the image based on the comparison result. The algorithm flow is shown in Figure 6. In this paper, the frequency domain pseudo-color processing enhancement method is used to perform pseudo-color processing on sports video images. This article first transforms the sports video image into the frequency domain by Fourier transform.

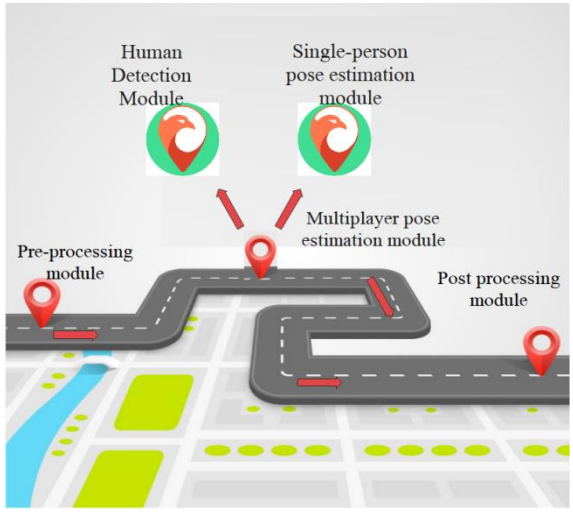

Figure 5: Multi-person posture estimation in sports competitions.

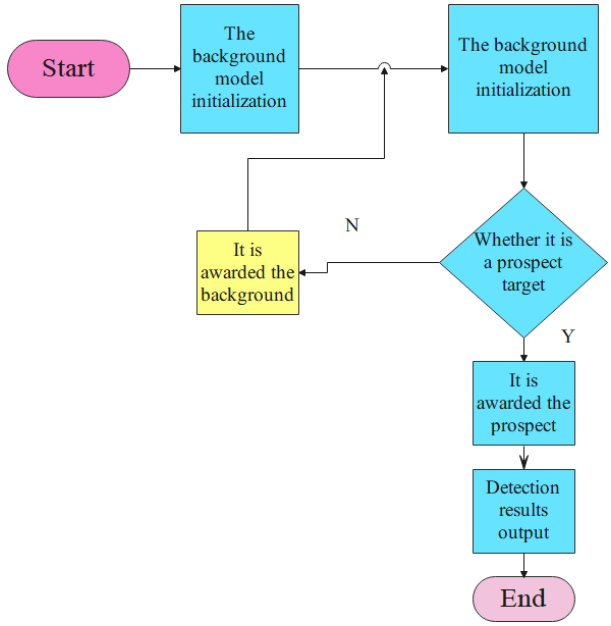

Figure 6: Flow chart of background modelling method.

In the frequency domain, three filters with different transfer characteristics are separated into three independent components, and then the inverse Fourier transform is performed on them to obtain three monochromatic images representing different frequency components. Next, this paper will further process the three images, and finally add them as the three primary color components to the red, green, and blue display channels of the color display, so as to realize the frequency domain segmented false color enhancement. Its block diagram is shown in Figure 7.

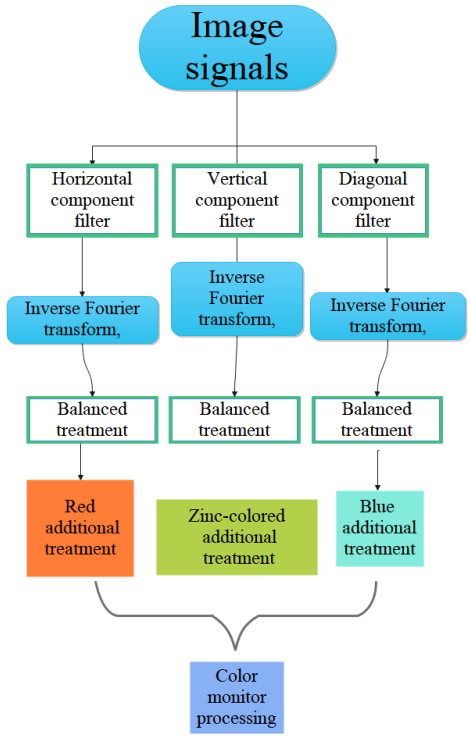

Figure 7: Schematic diagram of frequency domain filtering method to achieve pseudo-color processing.

After constructing the above system model, this paper combines big data technology to analyze sports game video data, and uses image processing methods to extract features. Big data technology can discover patterns from feature extraction. Therefore, this paper verifies the pose recognition and data mining during the experimental research, and obtains the results shown in Table 1 and Figure 8.

Number	*Gesture recognition*	*Data mining*	*Number*	*Gesture recognition*	*Data mining*
1	*87.8*	*86.9*	*17*	*81.7*	*91.2*
2	*83.2*	*96.5*	*18*	*92.0*	*86.0*
3	*81.7*	*93.2*	*19*	*88.5*	*92.2*
4	*87.0*	*95.6*	*20*	*93.3*	*89.1*
5	*86.1*	*92.7*	*21*	*91.7*	*83.9*
6	*93.2*	*92.1*	*22*	*90.5*	*86.6*
7	*86.2*	*90.5*	*23*	*86.2*	*84.7*
8	*82.2*	*90.6*	*24*	*83.3*	*93.5*
9	*80.6*	*93.1*	*25*	*79.7*	*86.2*
10	*91.4*	*93.3*	*26*	*93.1*	*83.3*
11	*91.8*	*83.7*	*27*	*87.4*	*89.5*

12	*88.2*	*88.7*	*28*	*87.8*	*85.2*
13	*90.1*	*83.3*	*29*	*83.5*	*90.5*
14	*92.8*	*85.5*	*30*	*93.6*	*92.4*
15	*89.4*	*86.6*	*31*	*92.8*	*95.0*
16	*91.0*	*88.6*	*32*	*88.0*	*88.1*

Table 1: Test statistics results.

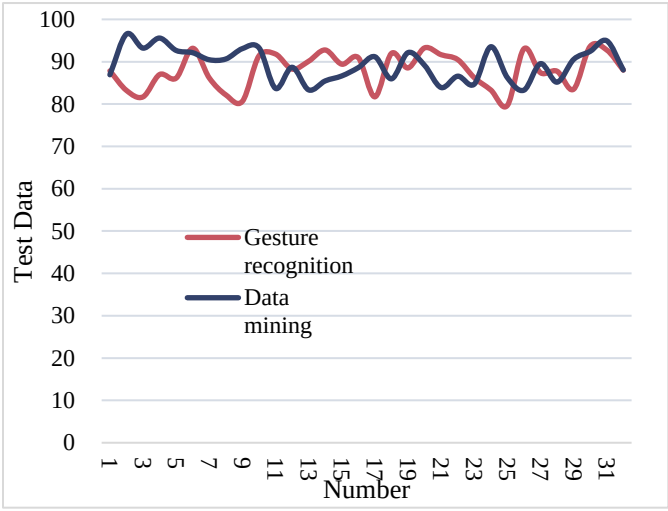

Figure 8: The effect of the multi-person pose estimation system in sports competition video based on big data analysis.

From the above research, we can see that the multi-person pose estimation system in sports competition video based on big data analysis can realize the feature recognition of multi-person sports. Moreover, it can use data mining technology to conduct regular exploration, which has extremely important guiding significance for the control of sports competitions and sports training.

5 CONCLUSION

Human body pose estimation is an important research field of computer vision, which has a wide range of practical application prospects. However, due to the existence of complex scenes and non-rigid changes in human body pose, the estimation accuracy of human body pose estimation is greatly suppressed. Therefore, the task of human body pose estimation at this stage, especially multi-person pose estimation, is still a very challenging subject. Multi-person pose estimation in complex scenarios has important research value. The reason is that in the process of human body pose estimation, there will be many different changes caused by complex scenes, such as human body differences, environmental differences, time differences, and care differences. This paper constructs the multi-person pose estimation system in sports competition video based on big data analysis, combines big data technology to perform sports game video data analysis, and uses image processing methods for feature extraction. The experimental research results show that the multi-person pose estimation system in sports competition video based on big data analysis can realize the feature recognition of multi-person sports.

Shengyu Zhang, https://orcid.org/0009-0004-5043-2845

REFERENCES

[1] Aso, K.; Hwang, D. H.; Koike, H.: Portable 3D Human Pose Estimation for Human-Human Interaction using a Chest-Mounted Fisheye Camera, In Augmented Humans Conference 2021, 2021, 116-120. https://doi.org/10.1145/3458709.3458986

[2] Bakshi, A.; Sheikh, D.; Ansari, Y.; Sharma, C.; Naik, H.: Pose Estimate Based Yoga Instructor, International Journal of Recent Advances in Multidisciplinary Topics, 2(2), 2021, 70-73.

[3] Colyer, S. L.; Evans, M.; Cosker, D. P.; Salo, A. I.: A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system, Sports Medicine-open, 4(1), 2018, 1-15. https://doi.org/10.1186/s40798-018-0139-y

[4] Dang, Q.; Yin, J.; Wang, B.; Zheng, W.: Deep learning based 2d human pose estimation: A survey, Tsinghua Science and Technology, 24(6), 2019, 663-676. https://doi.org/10.26599/TST.2018.9010100

[5] Díaz, R. G.; Laamarti, F.; El Saddik, A.: DTCoach: Your Digital Twin Coach on the Edge During COVID-19 and Beyond, IEEE Instrumentation & Measurement Magazine, 24(6), 2021, 22-28. https://doi.org/10.1109/MIM.2021.9513635

[6] Ershadi-Nasab, S.; Noury, E.; Kasaei, S.; Sanaei, E.: Multiple human 3d pose estimation from multiview images, Multimedia Tools and Applications, 77(12),2018, 15573-15601. https://doi.org/10.1007/s11042-017-5133-8

[7] Gu, R.; Wang, G.; Jiang, Z.; Hwang, J. N.: Multi-person hierarchical 3d pose estimation in natural videos, IEEE Transactions on Circuits and Systems for Video Technology, 30(11), 2019, 4245-4257. https://doi.org/10.1109/TCSVT.2019.2953678

[8] Hua, G.; Li, L.; Liu, S.: Multipath affinage stacked—hourglass networks for human pose estimation, Frontiers of Computer Science, 14(4), 2020, 1-12. https://doi.org/10.1007/s11704-019-8266-2

[9] Li, M.; Zhou, Z.; Liu, X.: Multi-person pose estimation using bounding box constraint and LSTM, IEEE Transactions on Multimedia, 21(10), 2019, 2653-2663. https://doi.org/10.1109/TMM.2019.2903455

[10] Liu, S.; Li, Y.; Hua, G.: Human pose estimation in video via structured space learning and halfway temporal evaluation, IEEE Transactions on Circuits and Systems for Video Technology, 29(7), 2018, 2029-2038. https://doi.org/10.1109/TCSVT.2018.2858828

[11] Martínez-González, A.; Villamizar, M.; Canévet, O.; Odobez, J. M.: Efficient convolutional neural networks for depth-based multi-person pose estimation, IEEE Transactions on Circuits and Systems for Video Technology, 30(11), 2019, 4207-4221. https://doi.org/10.1109/TCSVT.2019.2952779

[12] McNally, W.; Wong, A.; McPhee, J.: Action recognition using deep convolutional neural networks and compressed spatio-temporal pose encodings, Journal of Computational Vision and Imaging Systems, 4(1), 2018,3-3.

[13] Mehta, D.; Sridhar, S.; Sotnychenko, O.; Rhodin, H.; Shafiei, M.; Seidel, H. P.; Theobalt, C.: Vnect: Real-time 3d human pose estimation with a single rgb camera, ACM Transactions on Graphics (TOG), 36(4), 2017, 1-14. https://doi.org/10.1145/3072959.3073596

[14] Nasr, M.; Ayman, H.; Ebrahim, N.; Osama, R.; Mosaad, N.; Mounir, A.: Realtime Multi-Person 2D Pose Estimation, International Journal of Advanced Networking and Applications, 11(6), 2020, 4501-4508. https://doi.org/10.35444/IJANA.2020.11069

[15] Nie, X.; Feng, J.; Xing, J.; Xiao, S.; Yan, S.: Hierarchical contextual refinement networks for human pose estimation, IEEE Transactions on Image Processing, 28(2), 2018, 924-936. https://doi.org/10.1109/TIP.2018.2872628

[16] Nie, Y.; Lee, J.; Yoon, S.; Park, D. S.: A Multi-Stage Convolution Machine with Scaling and Dilation for Human Pose Estimation, KSII Transactions on Internet and Information Systems (TIIS), 13(6), 2019, 3182-3198. https://doi.org/10.3837/tiis.2019.06.023
[17] Petrov, I.; Shakhuro, V.; Konushin, A.: Deep probabilistic human pose estimation. IET Computer Vision, 12(5), 2018, 578-585. https://doi.org/10.1049/iet-cvi.2017.0382
[18] Szűcs, G.; Tamás, B.: Body part extraction and pose estimation method in rowing videos, Journal of Computing and Information Technology, 26(1), 2018, 29-43. https://doi.org/10.20532/cit.2018.1003802
[19] Thành, N. T.; Công, P. T.: An Evaluation of Pose Estimation in Video of Traditional Martial Arts Presentation. Journal of Research and Development on Information and Communication Technology, 2019(2), 2019,114-126.https://doi.org/10.32913/mic-ict-research.v2019.n2.864
[20] Xu, J.; Tasaka, K.; Yamaguchi, M.: Fast and Accurate Whole-Body Pose Estimation in the Wild and Its Applications, ITE Transactions on Media Technology and Applications, 9(1), 2021, 63-70. https://doi.org/10.3169/mta.9.63
[21] Zarkeshev, A.; Csiszár, C.: Rescue Method Based on V2X Communication and Human Pose Estimation, Periodica Polytechnica Civil Engineering, 63(4), 2019, 1139-1146. https://doi.org/10.3311/PPci.13861