Research on the Application of Behavioral Image Feature Capture in Basketball Game Video

Yan Zhang[1], *Wei Wei*[2*]
[1] *Department of Sports, Zhongnan University of Economics and Law, China*
[2] *School of Computer Science and Engineering, Xi'an University of Technology, China*
Z0001697@zuel.edu.cn, weiwei@xaut.edu.cn

Abstract

In order to realize intelligent image recognition of foul behavior in basketball games, this paper designs a feature capture method of video foul behavior based on improved bilateral filtering algorithm. Adaptive bilateral filtering is used to denoise the video image, and the optical flow feature and HOG feature of the behavior in the denoised image are obtained by image combination feature extraction method, which is fused to a combined eigenvector. The combined features were taken as the target recognition samples, and the multi-back propagation neural network was used to identify the foul behavior features. The particle filter was used to capture the video features and identify the location of the video behavior features. The experimental results show that this method can accurately capture the changes of video behavior characteristics, and has applicable performance in the identification of foul behavior in basketball matches.

Keywords: Bilateral filtering algorithm, Behavioral characteristics, Multiple BP neural networks, Particle filter, Optical flow

1 Introduction

Human Motion behavior Capture [1] or Human Motion Capture (HMC) refers to the technology of recording and processing Human Motion. The combination of human motion capture and artificial vision technology can achieve specific motion matching in the field of film and game production. When it captures subtle movements like faces or fingers, it's often called performance capture. In the field of computer vision, human motion feature analysis is one of the most popular research, and an important part of it is the image region capture of human behavior features. Human behavior identifies vulnerable to light, noise, perspective, multi-scale, and keep out the influence of such factors as [2]. Due to the large amount of noise interference [3] and the great difficulty in real-time and robust recognition of multiple behaviors in complex scenes, the recognition and capture of specific behaviors in videos is still one of the challenging research topics at the present stage.Mahdi et al. [4] proposed a human abnormal behavior detection method in surveillance video scenes. In this literature, human activity in public and sensitive areas is monitored through visual surveillance. Deep learning model is used to extract low-level features of human body from video images and select the best features. By applying deep learning technology to image classification, the detection ability of deep learning technology is optimized. This technology omits the manual feature extraction step and reduces the monitoring time of human behavior. However, this method can only be applied to surveillance videos, but its practicability is restricted due to the limited application environment. Ghadi et al. [5] proposed a human behavior prediction and recognition method based on maximum entropy Markov model and graph feature mining. The group data of human behavior is used as input to complete the noise reduction pretreatment. The human body contour is extracted, and the group analysis method and group clustering method are used to predict human behavior more accurately. Then the force interaction matrix and force flow characteristics are extracted by feature extraction method. Combined with graph mining technology, the maximum entropy Markov model is constructed and applied to the classification and prediction of human behavior. However, the existing methods are generally limited to simple motion capture of single person or a small number of people in simple scenes. Its robustness, accuracy and processing speed are far from the requirements of practical applications, so there are still many problems to be solved.

Bilateral filtering is a nonlinear filtering method, which combines the spatial proximity and pixel value similarity of images, and considers spatial information and gray similarity at the same time to achieve the purpose of edge-preserving denoising. Bilateral filtering is simple, non-iterative and local. In order to solve the above problems, in this paper, the video images of basketball games are taken as the research object, and the problem of capturing the features of foul behavior images is studied. The video feature capturing method based on bilateral filtering algorithm is designed, and the experimental process is completed.

The main innovations of this paper are:

(1) Adaptive bilateral filtering is adopted to denoise the video image, and the features in the denoised image are obtained, so as to improve the accuracy of extracting the feature information of basketball foul behavior.

(2) The multi-back propagation neural network is used to identify the characteristics of basketball foul behavior, which greatly reduces the number of parameters in the network, avoids the loss of characteristic information and reduces the error of foul behavior identification.

*Corresponding Author: Wei Wei; E-mail: weiwei@xaut.edu.cn
DOI: 10.53106/160792642024012501013

(3) Capture the video features of basketball game through particle filter, determine the position of foul behavior, and complete the capture of basketball foul behavior features, which effectively improves the accuracy of automatic detection and identification of players' foul actions in basketball matches.

This method can effectively extract the characteristic information of foul behavior in basketball game video, and group the data segments to judge whether the behavior is foul or not. Therefore, this method can convey the sports attributes of basketball video group behavior in the form of continuous basketball video frames, which lays a theoretical foundation for intelligent management of basketball video data.

2 Design of Video Feature Capture

In this paper, a video feature denoising method based on improved bilateral filtering algorithm is proposed to capture video feature. Due to the dynamic characteristics of human behavior in video images, and the influence of lens scanning equipment, the video inevitably has noise information, which has a bad influence on image quality. Therefore, in order to ensure the capture accuracy, the video image is denoised first to ensure the image quality, and then the behavior pixel features in the image are extracted. Finally, the foul behavior features are identified and their position changes are captured.

2.1 Image Denoising

Natural images change slowly in space, so adjacent pixels will be closer, which will lead to blurred image edges. Therefore, KPCA and deep learning methods will be used to extract features in image area capture, and pixels will be used to supplement the edges on both sides of pixel values. This image capture method has a wide range of applications, which can be used in the feature classification of lithium-ion batteries [6-9], the feature recognition of mechanical defects, and the feature capture of motion behavior images. Bilateral filtering algorithm needs to consider both spatial information and pixel value information. According to the pixel value classification filtering neighborhood, the category weight of pixel points is obtained, and the final pixel value result is obtained by weighted summation neighborhood.

Bilateral filtering algorithm shows that there are edge image weights and the final filtering results, and we can see the boundary of weights at the boundary. Weighting the pixels on one side of the edge to complete the calculation of image edge features. Bilateral filtering algorithm combines spatial neighborhood information and gray similarity value [10-11] to perform filtering on video images, which retains image edge features during filtering.

$$\hat{g}(i,j)=\sum_{x,y\in\beta}\varpi_S(x,y)\varpi_r(x,y)J(i,j), \tag{1}$$

where, $\hat{g}(i, j)$ is the video image after denoising, $\varpi_S(x, y)$, $\varpi_r(x, y)$ are the spatial domain weight of bilateral filter, gray domain weight. β is the neighborhood range at the pixel (i, j).

Then, we have:

$$\varpi_S(x,y)=\exp\left[-\frac{|x-i|+|t-j|}{2\alpha_S}\right], \tag{2}$$

$$\varpi_r(x,y)=\exp\left[-\frac{\left|J(i,j)\right|}{2\alpha_r}\right], \tag{3}$$

$$J(i,j)=g(x,y)+m, \tag{4}$$

where, $g(x, y)$ represents the region where there is no noise in the image. This paper assumes that m is gaussian white noise and $J(i, j)$ is the noise region, i, j represents pixel. In Equations (2) and (3), the spatial domain weight $\varpi_S(x, y)$ and gray domain weight $\varpi_r(x, y)$ of bilateral filters adopt gaussian function. Combining equation (1), it can be seen that the weight coefficient of bilateral filtering is jointly determined by two parameters: spatial variance α_S and gray variance α_r, x is an arbitrary starting scale, t is the image discrete time.

Because of the dynamic feature of video image space, the adaptive setting of spatial variance is extremely important. The inverse distance weighting method is selected to create the source data set, and the spatial variance setting is completed by adjusting the influence degree of surrounding pixels on the current pixel, so as to improve the influence degree of distant pixels on the central pixel and make the filtering result smoother. The global polynomial interpolation method is used to input the data fitting value, and the internal control grey number is obtained through cumulative calculation, and the development coefficient and grey action amount are established to complete the grey variance setting, and the calculation stage ratio and smoothness ratio are optimized. While preserving the basic geometric features of the image, the image is denoised, which effectively improves the information content and visual quality of the denoised image [12]. Therefore, bilateral filtering is improved in terms of local adaptive setting of spatial variance.

In bilateral filtering [13], α_S the value of determines the expansion degree of Gaussian curve in the video image filtering window. The larger the value α_S is, the slower the Gaussian curve drops, the more obvious the Gaussian smoothing effect is, and the video image becomes fuzzy. Therefore, a reasonable value of α_S can improve the ability of bilateral filter to retain image edge. In this paper, a local adaptive setting method of α_S is proposed. By calculating the target scale of each pixel in the image, the smooth region range around the pixel is obtained to control the spatial variance α_S of the point. In the video image edge and texture region, the smaller the target scale, the smaller the value of α_S. In the smooth region, the larger the target scale, the larger the value of α_S.

(1) Target dimension

The target scale in the video image is defined as the radius of the hypersphere with a certain pixel as the center, so all pixels in the hypersphere belong to the same target body [14]. The target scale can reflect the size in the morphological sense of local structure, which is smaller in the area or near the boundary with rich details and larger in the homogeneous

smooth area. While retaining the detailed features of the image, the image is segmented, which effectively improves the information content and visual quality of the image [15]. Therefore, image segmentation [16-17] is used to limit the size of spatial variance α_S. For (i, j) is pixel point in the video image, the segmentation neighborhood of the point is defined as $M_{i,j}(S)$:

$$M_{i,j}(S)=\{(i,j)|i-x|\le S,|j-y|\le S\}, \tag{5}$$

where, S is the size of segmented neighborhood.

In the video image, the known segmentation neighborhood of pixel point (i, j) is $M_{i,j}(S)$, then its boundary region $A_{i,j}(S)$ is:

$$A_{i,j}(S)=\{(i,j)\in M_{i,j}(S)-M_{i,j}(S-1)\}, \tag{6}$$

where $S-1$ is greater than 0.

For pixel point (i, j) in the video image, the similarity degree $V_{i,j}(S)$ between it and neighborhood boundary region $A_{i,j}(S)$ is defined as:

$$V_{i,j}(S)=\frac{\sum_{(i,j)\in A_{i,j}(S)}\exp\left[-\frac{|J(i,j)|}{2\alpha_r}\right]}{|A_{i,j}(S)|}, \tag{7}$$

where, $|A_{i,j}(S)|$ is the cardinal number of $A_{i,j}(S)$, that is, the number of pixels in $A_{i,j}(S)$; α_v is the statistical parameter representing the gradient distribution of video images.

For pixel point (i, j) in the video image, its target scale $S_{i,j}$ is:

$$S_{i,j}=\arg\max_{s\ge 1}\{V_{i,j}(S)\}, \tag{8}$$

where, argmax{·} represents the largest function independent variable that satisfies the conditions in parentheses. H_S is the threshold function. In bilateral filtering calculation, gray scale encryption factor is limited [18-19], and the target scale needs to be appropriately enlarged.

Therefore, in the 3*3 neighborhood of pixel (i, j) in this paper, the boundary allows at most two pixel points (i, j) out of the eight field pixels not to belong to the same target region (possibly due to noise). In this case, it is still considered that the whole 3*3 neighborhood and pixel point (i, j) belong to the same target region, thus $H_S = 6/8 = 0.75$.

(2) The calculation of the spatial variance α_S

After the target scale is obtained, the spatial variance α_S is set according to the target scale $S_{i,j}$. The target scale represents the size of the target structure in the morphological sense, that is, the size of the smooth region around the pixel point. $S_{i,j}$ represents the maximum neighborhood radius of the circle within the same target region. Meanwhile, according to the properties of Gaussian function, the Gaussian curve in region $[-2\alpha_S, 2\alpha_S]$ contains more than 95% components, so let $2\alpha_S = S_{i,j}$, namely:

$$\alpha_S=\frac{S_{i,j}}{n}, \tag{9}$$

After the target scale of each pixel is obtained, the spatial variance α_S of each pixel can be calculated according to Formula (9). n is the number of pixels.

To sum up, formula (1) is used to achieve video image denoising. During denoising, the spatial variance A of the bilateral filter will be adjusted adaptively to ensure that the edge features of the video image are not damaged after denoising.

2.2 Video Image Combination Feature Extraction

In section 2.1, the image denoised from the video image is $\hat{g}(i, j)$. In order to capture the characteristics of foul behavior, this paper uses the video image combination feature extraction method to extract the optical flow feature and HOG feature [20] of the behavioral pixel in $\hat{g}(i, j)$, and then describes the behavioral characteristics of basketball players after combination.

2.2.1 Optical Flow Feature Extraction

Assuming that the behavior of $\hat{g}(i, j)$ and the change of pixel point $g(x, y)$ can be represented by vector $E = (\phi, \varphi)$, where ϕ and φ represent the motion components of pixel point in the x direction and y direction respectively, then the optical flow characteristic at moment F is:

$$g(i+dx,j+dy,t+dt)=g(x,y,t)+\frac{\partial E}{\partial x}dx+\frac{\partial E}{\partial y}dy+\frac{\partial E}{\partial t}dt, \tag{10}$$

where, d and ∂ stand for derivative.

2.2.2 HOG Feature Extraction

HOG is called histogram of directional gradient [21], which is often used as a feature descriptor in object detection and is widely used in pedestrian detection. In order to better describe the characteristics of different motion behaviors in video images, this paper proposes an improved HOG algorithm, and the specific calculation steps are as follows:

Step 1. Enter $\hat{g}(i, j)$;

Step 2. One-dimensional gradient template [-1,0,1] and $[-1,0,1]^T$ are used to calculate the gradients F_x and F_y in the X and Y directions of the behavioral pixel of the video image, namely:

$$\begin{cases}F_x=K(i+1,j)-K(i-1,j)\\F_y=K(i,j+1)-K(i,j-1)\end{cases}, \tag{11}$$

where, $K(i, j)$ is the pixel value at pixel point (i, j), and its gradient amplitude $F(i, j)$ and gradient direction $\varepsilon(i, j)$ are respectively:

$$F(i,j)=\sqrt{F_x(i,j)^2+F_y(i,j)^2}. \tag{12}$$

$$\varepsilon(i,j)=\tan^{-1}\frac{F_x(i,j)}{F_y(i,j)}. \tag{13}$$

Step 3. Image $\hat{g}(i, j)$ was divided into several 8*8 cells, and the gradient histogram information in the 9 cells was collected. The gradient direction of each cell was divided into 9 direction blocks, and A cell was represented by A 9-dimensional vector.

Step 4. 2*2 cells are combined to form a large pixel, and all cell vectors are connected in series to obtain HOG feature of the block.

Step 5. Gaussian function is used to weight the gradient features of $\hat{g}(i, j)$, and HOG features of each block are obtained.

Since HOG feature represents the statistical histogram of local small area of A, it can effectively eliminate the influence of shadow and light.

2.2.3 Optical Flow Feature Extraction

Different behaviors can be extracted with features of different dimensions. In this paper, in order to ensure that the dimensions of each behavior feature vector are consistent, the optical flow feature and hog feature are normalized, and the same weight of the feature is set to ensure that each feature is paid equal attention to, so as to achieve the optimal parameter as far as possible, that is, the weight of each feature is 1. the specific calculation process is as follows:

Step 1. The normalized size of image $\hat{g}(i, j)$ is 96*96.

Step 2. Cell size is set to 8*8, and 2*2 cells are selected to form blocks, that is, the size of each block is 16*16, so 4356-dimensional HOG vector can be obtained.

Step 3. Using the Settings in Step 2, 4356-dimensional optical flow feature vectors can also be extracted.

Step 4. The combined feature vector is obtained by connecting HOG feature and optical flow feature.

2.3 Video Feature Recognition Model Based on Multiple Back Propagation Neural Network

2.3.1 Model Building

The feature classification method in this paper is to comprehensively judge the behavior foul situation of basketball game video by all kinds of characteristic values. Each characteristic value will affect the final recognition result, which belongs to the research scope of pattern recognition in the application of neural network. Among commonly used neural network models, convolutional neural network [22] is mainly used for speech recognition and image recognition, and K-nearest Neighbor (KNN) algorithm is mainly used for data classification in data mining. Multi-backpropagation neural network is the pattern recognition neural network model that best meets the classification and recognition requirements of foul behavior characteristics in this paper. Therefore, multi-backpropagation neural network is selected in this paper [23].

The structure of multiple back-propagation neural network is shown in Figure 1. The structure is mainly divided into input layer, hidden layer, output layer, back propagation layer, summary layer.

As shown in Figure 1, the first group of neural networks is composed of h neural networks. The input layer of each neural network is composed of j_h neurons with M eigenvalues extracted ℓ in Section 2.2. The output layer is 1 neuron ρ_h.The hidden layer is composed of n_h neurons, $n_h = \sqrt{j_h + 1} + \mu$, and μ are constants from 1 to 10, which are generally determined according to practical problems. ϖ_h^x and ϖ_h^y represent the weight set of neurons from the input layer to the hidden layer and from the hidden layer to the output layer in the h neural network.γ_h^x and γ_h^y respectively represent the deviation set between neurons from the input layer to the hidden layer and from the hidden layer to the output layer in the h neural network.

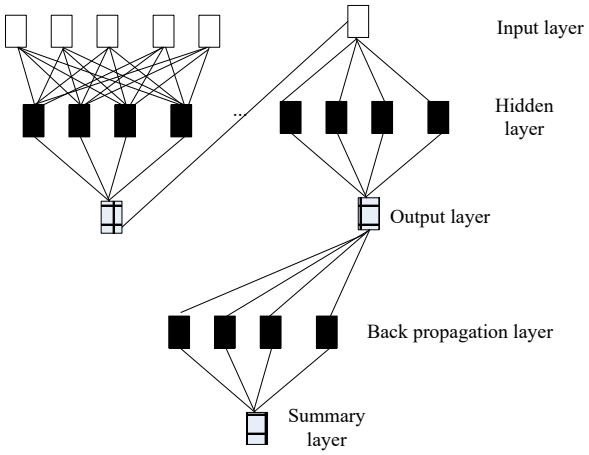


Figure 1. Structure of multi back propagation neural network

The second group is a neural network. its input layer is composed of the output nodes of the first group of neural networks, and the hidden layer is composed of n neurons. The output is 1 neuron Γ. ϖ^x and ϖ^y represent weight sets between neurons from the input layer to the hidden layer and from the hidden layer to the output layer respectively. γ^x and γ^y represent the bias set between neurons from the input layer to the hidden layer and from the hidden layer to the output layer in the h neural network respectively.

2.3.2 Model Training

The training of multi-back-propagation neural network is mainly divided into forward propagation calculation output and back-propagation correction parameters. Before the training, the initial weights and biases are randomly generated.

(1) Forward propagation process

In the process of forward propagation, the output of each neuron in the hidden layer and output layer of each neural network is calculated by the following formula.

$$\Gamma = g\left(\sum_{j=1}^{M} \varpi \ell + \gamma\right), \tag{14}$$

where, $g(\cdot)$ is the activation function in the multi-back-propagation neural network model. M is the total sample of behavior characteristics. ϖ and γ are the weight and bias of each neuron connection in turn.

In this method, it is known from previous experience that Sigmoid function is used as the activation function in the first set of hidden layers. The convergence speed is slightly faster, and the final training and test results are satisfactory.

$$\Gamma\left(\ell\right)=\frac{1}{1+e^{-\Gamma}}, \tag{15}$$

where, e is the error coefficient.

Since the purpose is to judge whether there is foul or not, there are only two cases of foul and standard. If a certain eigenvalue is considered as a possible foul feature, its characteristics need to be amplified through the model. If an eigenvalue is considered standard, its excitation effect on the results of multiple back-propagation neural networks will be gradually reduced. At the same time, in order to improve the convergence speed of the multi-back-propagation neural network, ReLU (Satisfaction Liner Unit) was used as the activation function in the output layer of the first group and the hidden layer of the second group.

In the output layer of the whole model, since it is necessary to output whether the whole measurement result is standard, namely foul probability, the result must be less than 1 and greater than 0, so SoftMax which can map the output of multiple neurons to the interval of (0,1) is used as the activation function.

(2) Back propagation layer

After obtaining the output behavior feature recognition result, training algorithm is used to modify the weight and bias of the multiple backpropagation neural network to make it approximate to the expected value. The selection of training algorithm will affect the final effect and training speed. In this paper, the small-batch stochastic gradient descent algorithm is adopted as an improved algorithm, which reduces the number of iterations of network training, improves the convergence speed of network, and finally trains the model with higher accuracy.

Firstly, the loss function $Cost\left(\partial\right)$ is defined. For the model with M behavior characteristic samples, the formula is as follows:

$$Cost\left(\varpi,\gamma\right)=\frac{1}{M}\sum_{j=1}^{M}\left(\varpi\ell_{j}+\gamma-\Gamma_{j}\right)^{2}, \tag{16}$$

where, Γ_j is the recognition output of the behavior characteristic sample j.

The small batch gradient descent method is used as the training optimization algorithm, which selects n behavior feature samples for updating each time. The number of iterations required for convergence can be greatly reduced, and the convergence result can be closer to the effect of gradient descent. Finally, the weights and biases of each neuron were modified according to the forward propagation and back propagation processes, and a multi-back-propagation neural network model was obtained after the training. The model test set is used to verify the model. Unless the verification accuracy meets the requirements, the batch sample number is updated and a new round of training is conducted until the training process is finally completed. After the training, the results of behavioral feature recognition are summarized through the summary layer, and the behavioral feature score of the video image is obtained. If the output result is 1, it indicates that a certain behavioral feature belongs to the foul behavior.

2.4 Feature Capture Method

2.4.1 Parameter Setting

The parameter setting in feature capture method is of great significance, which can reduce the number of features and dimension, make the model more generalized, reduce over-fitting, and enhance the understanding between features and eigenvalues.

The feature weight parameter is set to 8-bit or floating-point single-channel and three-channel images, that is src of InputArray type. The derived variable parameter represents the diameter of each pixel neighborhood in the filtering process and is set to a non-positive number. The image feature parameters of similar regions are set to sigmaColor with double type, that is the sigma value of color space filter. The larger the value of this parameter, the wider the colors in the neighborhood of the pixel will be mixed together, resulting in a larger semi-equal color area. The color vector parameter of the feature pixel position is set to the sigma value of the filter in the sigmaSpace coordinate space with double type, that is the labeling variance of the coordinate space. The larger the variance value, the more distant the image feature pixels in the area will influence each other, so that the larger area is similar enough to obtain the same color. The foul behavior parameter is set to a borderType with int type, which is used to infer a certain boundary pattern of pixels outside the image.

2.4.2 Model Building

Behavioral feature state transition model, also known as target motion model, is one of the basic elements of video feature capture and synthesis. For video feature capture, the range of target behavior feature motion between adjacent frames is small, so this paper uses a second-order autoregressive model to describe its motion law. Assume that the state of the characteristic particle of the foul behavior j at moment $g-1$ (the particle represents the location information of the behavior characteristic) is $\Gamma_{g-1}^{(j)}$, then the state $\Gamma_{g}^{(j)}$ of the characteristic particle of the foul behavior j at moment g is:

$$\Gamma_{g}^{(j)}=d+b_{1}\Gamma_{g-1}^{(j)}+b_{2}\Gamma_{g-2}^{(j)}, \tag{17}$$

where, d is constant, b_1 and b_2 are model parameters.

2.4.3 Characteristic Observation Model

After obtaining the new position of the foul behavior feature through state transfer, the probability of observing the position state of each foul behavior from the real state of the foul behavior location should be calculated, which can be approximated by the Gaussian function of the similarity between the image feature of the region where the particle is located and the image feature of the target behavior. Commonly used feature segmentation information includes color, texture, shape and contour, etc. [24-26]. Since color information does not change much during tracking, and HSV color model can separate brightness information from chroma and saturation value, it is less affected by illumination changes. Therefore, in this paper, the HSV color histogram of foul behavior features of video images is used to establish the observed likelihood function.

The HSV histogram is divided into n small cells, and the quantized value of the color vector with pixel position pl of foul behavior feature is $c_k(pl)$, then the kernel density of the color distribution in the target area is estimated as $H(\Gamma_k)$. Assume that the color reference model is $Ka = Ka^*(m, \mathrm{k})$, and $Ka^*(m, \mathrm{k})$ is the probability that the quantized value of the color vector with pixel position pl of foul behavior feature is $c_k(pl)$. The candidate reference model is Ka^*, which measures the similarity between image behavior features by Means of Barbitan distance. The calculation of The Babbitt distance $bs[Ka^*, Ka]$ is:

$$bs\left[Ka^*, Ka\right] = \left[1 - \sum_{m=1}^{M} \sqrt{Ka}\right]. \tag{18}$$

The observational likelihood function is used to analyze the observational results.

$$q\left(\Gamma_t^{(j)} \middle| \ell\right) \propto e^{-\frac{bs\left[Ka^*, Ka\right]}{2\vartheta^2}}, \tag{19}$$

where, ϑ^2 is the Gaussian variance. The larger $q(\Gamma_t^{(j)} | \ell)$ is, the closer the position represented by the particle is to the color of the foul behavior feature, and the more likely the position represented by the particle is to be the foul behavior feature, so as to complete the capture of the change of the position of the foul behavior feature.

3 Experiment and Result Analysis

This experiment has certain requirements on GPU computing power. The experiment is carried out in Google colab cloud, mounted to a remote server through the network for training, and uses the Computing platform of Compute Unified Device Architecture (CUDA) parallel computing architecture. The specific experimental environment is shown in Information Table 1.

Table 1. Experimental operation environment

Type	Edition
Cloud system	Google colab
Parallel computing architecture and platform	Compute unified device architecture
Development language	C language
Training environment	Linux
Graphics card	Quadroq series

3.1 Video Image Filtering Effect

3.1.1 Subjective Effect

Take Figure 2 as an example. Due to illumination and equipment, the video image of the basketball match in Figure 2 is not clear and has a large number of noise stripes, which affects the image quality. Therefore, the denoising effect of the method in this paper is tested, and the results are shown in Figure 3.

Figure 2 and Figure 3 By contrast, after denoising the video image with noise, the visual effect of the image is significantly optimized. The original noise stripes have been filtered out. From the perspective of vision, the proposed method is proved to be effective in denoising video images with noise.

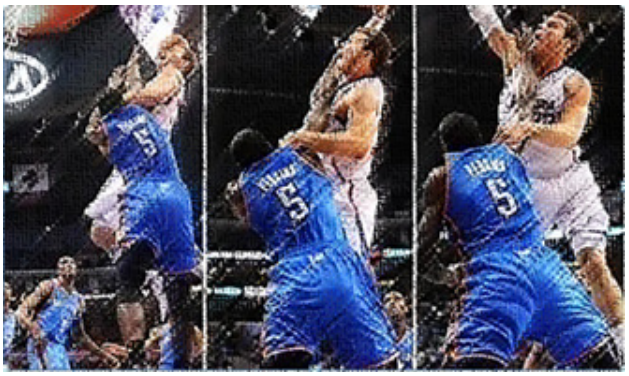


Figure 2. Video image of basketball game with noise

Figure 3. Filtering effect of basketball game video image

3.1.2 Objective Analysis

Figure 2 is transformed into a 256-level grayscale image and denoised by the proposed method using the peak SNR analysis. The larger the peak SNR is, the better the denoising effect of the video image is, and the more complete the image edge and detail features are. The test results are shown in Figure 4.

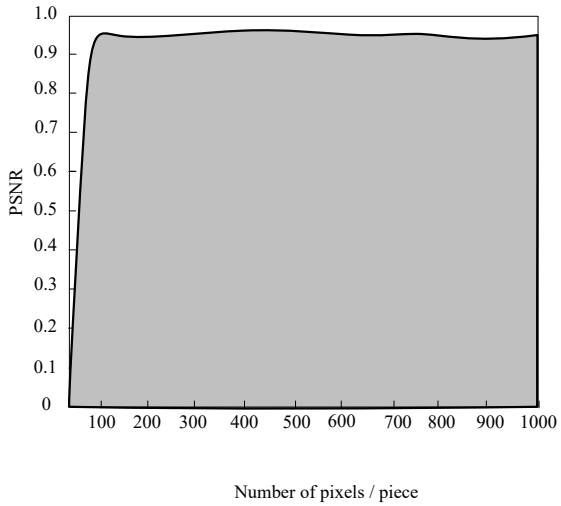


Figure 4. Peak signal-to-noise ratio test results of basketball game video images after de-noising

Analysis result of Figure 4 shows that the method after denoising, video image peak high signal noise ratio (SNR), is greater than 0.9, whose reason is that the method can accord to the video image target dimension values each area and set different spatial variance in the image position to retain more edge and detail characteristics, which explain the method of video image denoising effect is more ideal.

3.2 Capture Effect of Foul Behavior Feature in Video Image

3.2.1 Subjective Effect

The method in this paper is used to capture the foul behavior characteristics of basketball players in Figure 3, and the captured results are marked in the image. The effect diagram is shown in Figure 5.

According to Figure 5, based on subjective visual effect analysis, the proposed method can effectively capture the characteristics of foul behavior in video images, which verifies the ability of the proposed method to capture foul behavior in video images from the perspective of subjective vision.

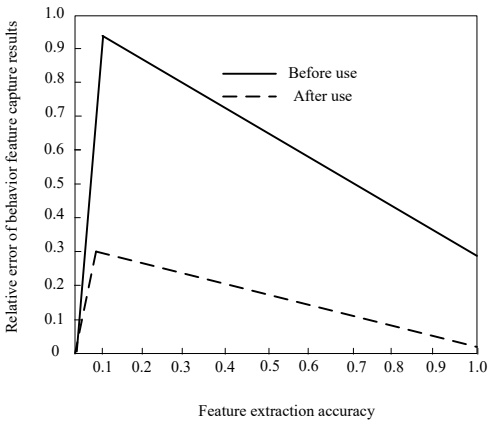
Figure 5. Capture effect of foul behavior characteristics in video images of basketball games

3.2.2 Objective Effect

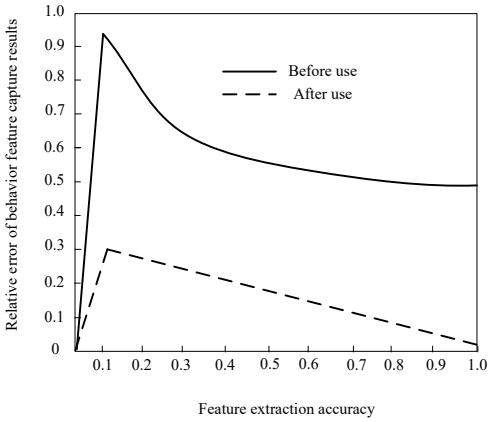
The objective effect is used to test the ability of this method to capture foul behaviors in video images. In Figure 5, there are three kinds of foul behaviors: hand intrusion foul, block foul and pull foul.

The accuracy of feature matching shows that the matched two feature points have the same semantic features in real space. At the same time, the matching results need to ensure as many correct matches as possible to reduce false matches. With the increase of precision, the lower the relative error, which indicates that the closer the model is to the optimal value, the higher the model discrimination, and the better the sample of foul behavior can be distinguished. The evaluation index is the relative error of behavior feature capture results, and the smaller the error result, the better the video processing effect after applying this method.

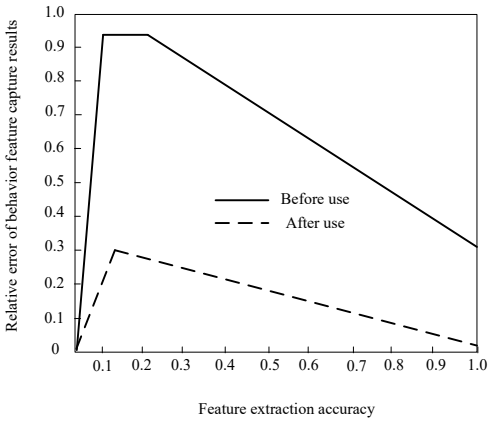
Test the AR curves of the proposed method in capturing three kinds of foul behavior (X axis is foul behavior feature extraction accuracy, the Y axis is foul behavior characteristics capture the results of relative error). The result is shown in Figure 6.



(a) Hand intrusion foul



(b) Block foul



(c) Pull foul

Figure 6. Objective evaluation results of foul behavior characteristics capture in video images of basketball games

By analyzing the test results in Figure 6, it can be seen that the extraction accuracy of the three foul behavior features of hand intrusion foul, blocking foul and pull foul in this method is relatively small, and the maximum relative error of the position captured by the foul behavior feature in the video image is 0.3. When the accuracy of behavior feature extraction increases, the relative error of the position of video image foul behavior feature capture gradually decreases,

whose minimum value is only 0.02. Compared with before using this method, the accuracy of video image foul behavior feature capture is significantly improved. It is verified that the method in this paper can accurately capture the three foul behavior characteristics of hand intrusion foul, blocking foul and pulling foul, and can improve the accuracy of capturing the foul behavior characteristics of basketball game video images.

3.3 Robustness Test Analysis

Randomly, the 15-minute time period in the basketball video are selected, which contain 60 basketball action data groups and 7 fouls. The resolution of the image is reduced to 80 × 100 pixels, and 30dB Gaussian white noise is added, which makes the image processing have great ambiguity in time and space. Under this disturbance condition, the recognition result of basketball violations is obtained. The result is shown in Figure 7.

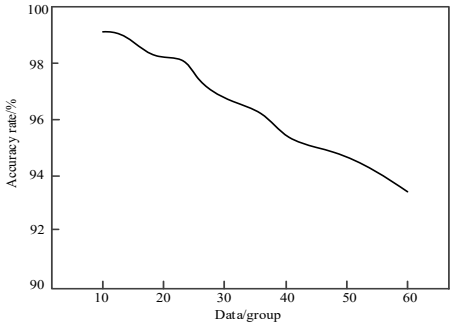


Figure 7. Identification result of foul behavior in disturbed state

According to Figure 7, when the basketball video image is disturbed by noise to a certain extent, the method in this paper can tolerate the disturbance and keep the original output, and can work normally under small and frequent disturbances, with excellent generalization ability. It is proved that this method is robust and can be popularized.

4 Conclusion

It is difficult to detect the foul action of basketball players in the fierce competition by using the technique of detecting the foul action mainly in the basketball match. Therefore, this paper proposes a video feature capture method based on improved bilateral filtering algorithm. Experiments show that the proposed method has application value, which is mainly reflected in the following points:

(1) After denoising the video image of basketball match, the peak SNR of the image is high. The method in this paper can set different spatial variances at different positions of the image according to the target scale value of each region, so as to retain more edge and detail features. The future work is to reduce the computational complexity of the target scale and further study the bilateral filtering kernel function to preserve more edge feature information.

(2) A comparison was made between the methods in this paper and those not in this paper on the capture accuracy of three kinds of foul behavior features, namely, manual trespass, blocking foul and pulling foul. The method in this paper has high extraction accuracy and small relative error in the capture results of foul behavior features, which verifies that the method in this paper can improve the capture accuracy of video features.

Although good research results have been achieved in this paper, the ideal accuracy still cannot be achieved for the occlusion problem in the process of video behavior recognition. In future studies, further discussions are needed on the multi-perspective occlusion problem, behavior representation, feature selection and fusion strategy.

References

[1] M. Miyakoshi, L. Gehrke, K. Gramann, S. Makeig, J. Iversen, The AudioMaze: An EEG and motion capture study of human spatial navigation in sparse augmented reality, *European Journal of Neuroscience*, Vol. 54, No. 12, pp. 8283-8307, December, 2021.

[2] P. Pareek, A. Thakkar, A survey on video-based human action recognition: recent updates, datasets, challenges, and applications, *Artificial Intelligence Review*, Vol. 54, No. 3, pp. 2259-2322, March, 2021.

[3] A. Davy, T. Ehret, J. M. Morel, P. Arias, G. Facciolo, Video denoising by combining patch search and CNNs, *Journal of Mathematical Imaging and Vision*, Vol. 63, No. 1, pp. 73-88, January, 2021.

[4] M. S. Mahdi, A. J. Mohammed, M. M. Jafer, Unusual activity detection in surveillance video scene, *Journal of Al-Qadisiyah for Computer Science and Mathematics*, Vol. 13, No. 3, pp. 92-98, September, 2021.

[5] Y. Y. Ghadi, I. Akhter, H. Aljuaid, M. Gochoo, S. A. Alsuhibany, A. Jalal, J. Park, Extrinsic behavior prediction of pedestrians via maximum entropy Markov model and graph-based features mining, *Applied Sciences*, Vol. 12, No. 12, Article No. 5985, June, 2022.

[6] C. Zhang, S. Zhao, Z. Yang, Y. Chen, A reliable data-driven state-of-health estimation model for lithium-ion batteries in electric vehicles, *Frontiers in Energy Research*, Vol. 10, Article No. 1013800, September, 2022.

[7] L. Li, H. Li, G. Kou, D. Yang, W. Hu, J. Peng, S. Li, Dynamic camouflage characteristics of a thermal infrared film inspired by honeycomb structure, *Journal of Bionic Engineering*, Vol. 19, No. 2, pp. 458-470, March, 2022.

[8] S. Zhao, C. Zhang, Y. Wang, Lithium-ion battery capacity and remaining useful life prediction using board learning system and long short-term memory neural network, *Journal of Energy Storage*, Vol. 52(B), Article No. 104901, August, 2022.

[9] C. Zhang, S. Zhao, Y. He, An integrated method of the future capacity and RUL prediction for lithium-ion battery pack, *IEEE Transactions on Vehicular Technology*, Vol. 71, No. 3, pp. 2601-2613, March, 2022.

[10] P. Naveen, P. Sivakumar, Adaptive morphological and bilateral filtering with ensemble convolutional neural network for pose-invariant face recognition, *Journal of*

Ambient Intelligence and Humanized Computing, Vol. 12, No. 11, pp. 10023-10033, November, 2021.
[11] G. U. Bhargava, S. V. Gangadharan, FPGA implementation of modified recursive box filter-based fast bilateral filter for image denoising, *Circuits, Systems, and Signal Processing*, Vol. 40, No. 3, pp. 1438-1457, March, 2021.
[12] B. Goyal, A. Gupta, A. Dogra, D. Koundal, An adaptive bitonic filtering based edge fusion algorithm for Gaussian denoising, *International Journal of Cognitive Computing in Engineering*, Vol. 3, pp. 90-97, June, 2022.
[13] W. Wei, B. Zhou, D. Polap, M. Wozniak, A Regional Adaptive Variational PDE Model for Computed Tomography Image Reconstruction, *Pattern Recognition*, Vol. 92, pp. 64-81, August, 2019.
[14] S. Wang, X. Liu, S. Liu, K. Muhammad, A. A. Heidari, J. D. Ser, V. H. C. Albuquerque, Human Short Long-Term Cognitive Memory Mechanism for Visual Monitoring in IoT-Assisted Smart Cities, *IEEE Internet of Things Journal*, Vol. 9, No. 10, pp. 7128-7139, May, 2022.
[15] S. Liu, J. Ma, Y. Yang, T. Qiu, H. Li, S. Hu, Y. Zhang, A multi-focus color image fusion algorithm based on low vision image reconstruction and focused feature extraction, *Signal Processing: Image Communication*, Vol. 100, Article No. 116533, January, 2022.
[16] A. Al-Ahmad, O. S. Almousa, Q. Abuein, Enhancing steganography by image segmentation and multi-level deep hiding, *International Journal of Communication Networks and Information Security*, Vol. 13, No. 1, pp. 143-150, April, 2022.
[17] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, K. H. Maier-Hein, nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation, *Nature Methods*, Vol. 18, No. 2, pp. 203-211, February, 2021.
[18] A. Girdhar, H. Kapur, V. Kumar, A novel grayscale image encryption approach based on chaotic maps and image blocks, *Applied Physics B*, Vol. 127, No. 3, Article No. 39, March, 2021.
[19] O. P. Singh, A. K. Singh, G. Srivastava, N. Kumar, Image watermarking using soft computing techniques: A comprehensive survey, *Multimedia Tools and Applications*, Vol. 80, No. 20, pp. 30367-30398, August, 2021.
[20] W. Wei, H. Song, W. Li, P. Shen, A. Vasilakos, Gradient-driven parking navigation using a continuous information potential field based on wireless sensor network, *Information Sciences*, Vol. 408, pp. 100-114, October, 2017.
[21] M. N. Khan, A. Das, M. M. Ahmed, S. S. Wulff, Multilevel weather detection based on images: A machine learning approach with histogram of oriented gradient and local binary pattern-based features, *Journal of Intelligent Transportation Systems*, Vol. 25, No. 5, pp. 513-532, 2021.
[22] S. Liu, G. Liu, H. Zhou, A robust parallel object tracking method for illumination variations, *Mobile Networks and Applications*, Vol. 24, No. 1, pp. 5-17, February, 2019.
[23] S. Liu, D. Liu, G. Srivastava, D. Połap, M. Wozniak, Overview and methods of correlation filter algorithms in object tracking, *Complex & Intelligent Systems*, Vol. 7, No. 4, pp. 1895-1917, August, 2021.
[24] S. Liu, Z. Pan, H. Song, Digital image watermarking method based on DCT and fractal encoding, *IET Image Processing*, Vol. 11, No. 10, pp. 815-821, October, 2017.
[25] W. Wei, X. Xia, M. Wozniak, X. Fan, R. Damasevicius, Y. Li, Multi-sink distributed power control algorithm for cyber-physical-systems in coal mine tunnels, *Computer Networks*, Vol. 161, pp. 210-219, October, 2019.
[26] S. Liu, M. Lu, G. Liu, Z. Pan, A novel distance metric: generalized relative entropy, *Entropy*, Vol. 19, No. 6, Article No. 269, June, 2017.

Biographies

Yan Zhang now act as Lecturer in Sports Department, Zhongnan University of Economics and Law. Her main research domain is physical intelligence and deep learning in physical analysis.

Wei Wei is an associate professor of School of Computer Science and Engineering, Xi'an University of Technology, China. He is a senior member of ACM & IEEE. He has published around one hundred research papers. He is an editorial board member of FGCS, IEEE Access, AHSWN, etc.