IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 18, 2020, accepted December 7, 2020, date of publication December 14, 2020,
date of current version December 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3044371

Adaptive Recognition of Motion Decomposition Image Based on Adaptive Principal Component Extraction Algorithm

WEI YANG
Sports University, Pingdingshan University, Pingdingshan 467000, China
e-mail: yangwei_li@tom.com

ABSTRACT Moving target detection and recognition methods are the foundation and key of modern intelligent video recognition systems. It combines advanced technologies in many fields such as image processing, pattern recognition, and artificial intelligence, and is a research hotspot in computer vision technology. Therefore, it is of great significance to study moving target detection and recognition algorithms. The non-local image extraction algorithm proposed in this paper uses an adaptive clustering method to perform fine cluster analysis on non-local image blocks with different feature types. Through the step-by-step principal component approximation method, we carefully find the features in each class. This progressive principal component approximation implements singular value hard threshold processing based on the Marchenko-Pastur (MP) theorem to select the main part of the feature, and uses special soft thresholds in the principal component transform domain to further improve the extraction performance. The Lower Bound-Based Within-Class Maximum Division (LBWCMD) is proposed, and this method is used as a preprocessing step of robust principal component analysis in moving target detection. This article applies LBWCMD to the video frame set based on the position information of the moving target, and the obtained frame subset meets the signal requirements of Robust Principle Component Analysis (RPCA) to the greatest extent. On this basis, we add frames with smaller motion amplitudes to each frame subset to increase the proportion of background pixels in each subset. Frame set division and low-rank decomposition realize the detection of moving targets under a unified framework. The detection rate of the proposed method is higher than that of the current popular methods in sports video data sets, and the detection accuracy is improved compared with the original RPCA method.

INDEX TERMS Action recognition, adaptive principal component, moving target detection, random matrix theory.

I. INTRODUCTION

As the field of computer vision continues to mature and develop, human society is gradually becoming intelligent and advanced. Computer vision is widely used in engineering and life, including image processing, pattern recognition, artificial intelligence, computer graphics, and artificial neural networks [1]–[3]. The main purpose of computer vision research is to realize the use of computers to replace the brain and human eyes to capture object information in the environment, and ultimately solve high-level vision problems, and truly realize the description, storage, recognition and understanding of the content in the image. The visual processing system is the main tool used by humans to observe and perceive the outside world [4], [5]. In today's society, with the continuous improvement of computer processing capabilities, engineers hope that computers can replace human eyes and brains to recognize, observe, and interact with external things and the objective world like humans. This requires computers to have human visual processing systems. Due to the continuous improvement of computer hardware processing capabilities and the rapid development of computer vision technology, this expectation is closer to becoming a reality. The main content of computer vision technology research is how to use computer vision technology to solve related human-centered problems, including

The associate editor coordinating the review of this manuscript and approving it for publication was Jenny Mahoney.

human body detection and recognition, face recognition, and human motion analysis [6], [7].

Nowadays, it is assumed that each action input by the system is independent of each other, does not interfere with each other, or has been pre-segmented. The idea is applied in many vision-based action recognition systems. Generally speaking, this assumption is that the action to be tested is in a resting state except for the exercise period, that is, before and after the end, no action is taken [8], [9]. Based on this assumption, it will limit the naturalness of human-computer interaction to a certain extent, but it greatly reduces the difficulty of existing behavior recognition tasks. In the real environment, actions usually do not exist independently, but are embedded in the continuous action process, which means that if you want to recognize human behavior, segmentation of continuous actions is one of the necessary preparatory work [10]–[12]. Relevant scholars use monocular cameras to collect video information, and use the method of convolution surface to calculate each connected bone to construct the required human bone model [13], [14]. The model can be deformed by changing the radius, polynomial and other parameters. Finally, the mapping relationship between the two-dimensional framework and the three-dimensional posture is obtained by analyzing the corresponding relationship between the curve in the image and the surface generated by the convolution calculation [15]. This relationship can be used to express the motion behavior of the human body in the three-dimensional space. Researchers use a large number of experimental samples to construct a three-dimensional human body model based on the representation of the head, body trunk and limbs under the framework of Naive Bayes, and then use this model to segment the human body region in the image [16], [17]. Related scholars pre-made basic image templates. The template materials are derived from trained silhouette images, and the human silhouette is segmented by iterative matching of silhouettes in the target image [18]. Relevant scholars use Grab Cut segmentation algorithm to achieve human body segmentation based on the human body database established by themselves [19]. Human motion behavior is a non-rigid motion. In order to enable the computer to better imitate and continuously approach or even surpass all the functions of the human visual processing system, by processing the video to recognize and analyze the action, it needs to go through motion detection [20]. Actions are composed of a series of static postures, which are generally continuous, and there is no obvious boundary between different actions [21]. Therefore, it is difficult to segment continuous actions to regulate the changes within and between types of actions [22]. Nowadays, it is assumed that each action input by the system is independent of each other, does not interfere with each other, or has been pre-segmented. The idea is applied in many vision-based action recognition systems [23]. The rest state before and after the end, no action is taken. Based on this assumption, it will limit the naturalness of human-computer interaction to a certain extent, but it greatly reduces the difficulty of existing behavior recognition tasks [24]. Due to the diversity of human action categories, with the continuous refinement of the action representation, the similar distance between different actions becomes closer and closer, which greatly increases the difficulty of action recognition [25]–[30]. Taking the action of the characters in the video as the research object, we can see by comparing running and jogging that the postures of these two actions are very similar in most cases [31]. This actual situation forces us to reduce the dispersion between different action types. The degree of recognition is more accurate; at the same time, it can be seen from the video that due to differences in the human body, the same action in different people is very different [32]–[34]. This actual situation forces us to increase the divergence between the same action types. Therefore, how to identify the intra-class changes between the same actions and the changes between different action types is a key problem that the action recognition algorithm needs to solve. The motion scene is complex and changeable, and the target that needs to be tracked and recognized is always in motion. Therefore, when partial occlusion occurs, the tracking algorithm will drift due to the lack of a discrimination mechanism for incomplete targets, which makes tracking or recognition errors [35]. There are also cases where the target will be completely occluded. When the target reappears in the lens range, it may be judged as a new target, which affects subsequent recognition [36]. The same target will be recognized repeatedly, which affects the efficiency of recognition. Therefore, it is necessary to adopt reasonable features and models to increase the local feature description of the target object. Target objects in various states can be tracked and identified correctly. An effective target loss discrimination mechanism can ensure the smooth progress of recognition in video surveillance.

In order to make up for the shortcomings of the existing subspace methods, this paper makes full use of the position information of the moving target to detect the moving target in the video. According to the location information, we designed the Lower Bound-Based Within-Class Maximum Division (LBWCMD) method that divides the video frame into different subsets. This can enhance the algorithm's detection effect of moving targets in high-density crowds. Then we constructed an augmented set. The augmented set can expand each of the previously obtained subsets to generate a group set. And the frames with less motion amplitude contain more real background pixels, which is more conducive to the recovery of low-rank parts. After the above optimization processing, the location information corresponding to each group set will satisfy the uniform distribution and wide coverage in time and space. In short, we combine motion position information and low-rank decomposition to improve the existing Robust Principle Component Analysis (RPCA) method. Specifically, the technical contributions of this article can be summarized as follows:

First: In order to obtain the image extraction effect of detail preservation, this paper applies the random matrix theory to the estimation of feature level and the research of image

extraction algorithm, and obtains a robust and more accurate feature level estimation effect based on adaptive clustering and progressive PCA transform domain approximation texture detail preservation algorithm.

Second: We solve the eigenvalues of the covariance matrix of the matrix by dividing the image into partially overlapping small blocks and stacking the vector composed of pixels in the small blocks into a matrix. The Marchenko-Pastur (MP) rule is used to obtain the feature level to be estimated through heuristic greedy search calculation. A clustering algorithm based on ''over-clustering-iterative merging'' is proposed to adaptively divide image blocks into different classes with significant feature differences relative to feature interference.

Third: Experiments in various video environments show that the proposed method has a good detection effect. The number of elements in the subsets divided by the LBWCMD method is not much different. The proposed Adaptive Principal Component Extraction Algorithm (APCEA) can obtain good results under strong light and crowded crowd environments. The proposed method is an improvement of the RPCA method, which aims to improve the effect of moving target detection. Although the overall computing time is only slightly lower than the RPCA method, experiments on multiple videos show the effectiveness of our method.

The rest of this article is organized as follows. Section 2 discusses the related theories of moving target detection and recognition. Section 3 proposes an extraction algorithm based on adaptive clustering and progressive principal component approximation. In Section 4, simulation experiments and result analysis are carried out. Section 5 summarizes the full text.

II. RELATED THEORIES OF MOVING TARGET DETECTION AND RECOGNITION

A. BACKGROUND DIFFERENCE METHOD

The basic idea of the background difference method is to select one or more frames for modeling, establish a background model, and use the model to compare subsequent frames. The part of the image that is similar to the background is regarded as the background, and the different area is regarded as the foreground, and finally the background model is judged and updated based on the image information. This method is only suitable when the monitoring equipment is stationary.

1) STATISTICAL AVERAGE METHOD

The statistical average method is to observe the gray values of consecutive multiple frames of images in the video sequence. The average method is to obtain the average of the gray values of the same pixel in multiple frames of images as the background estimate. The gray value of a pixel is sorted, and the median value is calculated as the estimated value of the background of that point. The calculation process is similar in general. The mean value method is described in detail below. The background extraction by means method can be expressed by the following formula:

$$B(x, y) = \frac{1}{L}\sum_{i=1}^{L} I_i(x, y) \tag{1}$$

Among them, B (x, y) represents the estimated value of the background image at (x, y), I_i (x, y) represents the pixel value of the i-th frame image at point (x, y), and L represents the selected number of frames to extract the background.

The number of frames selected to extract the background L is too large, the preprocessing time will be too long, and the selection of L is small, because the surveillance video is all-weather, the background image cannot always be the same, so the calculated background is not always reliable. The background image needs to be updated gradually. A background update method is proposed, using subsequent frames as increments to add the mean value formula:

$$B(x, y, k) = \frac{1}{k}B(x, y, k) + \frac{k+1}{k}B(x, y, k+1) \tag{2}$$

This method is simple to calculate and very easy to implement, but it is not suitable for environments where the illumination changes significantly.

2) GAUSSIAN DISTRIBUTION BACKGROUND MODEL

The single-mode Gaussian distribution background model uses a single Gaussian distribution to represent the color value of each pixel. You set the model parameters as:

$$u_x = \frac{1}{N}\sum_{i=0}^{N-1} u_{i,t} \tag{3}$$

$$\delta_x^2 = \frac{1}{N}\sum_{i=0}^{N-1} (u_x - u_{i,t})^2 \tag{4}$$

Among them, $u_{i,t}$ represents the pixel value of the i-th pixel at time t, and the discrimination relationship is $|u_x - I(i)| \leq 2.4\delta$.

Among them, I(i) is the gray value of the i-th point pixel. The points that meet this condition are considered as background points, and all such points constitute the background image. As time goes by, the background is not static, so it needs to be updated in real time, as follows:

$$u_{x,t+1} = (1-\alpha)u_{x,t+1} + \alpha I_{x,t} \tag{5}$$

$$\delta_{x,t+1} = \alpha(I_{x,t} - u_{x,t}) + (1-\alpha)\delta_{x,t+1} \tag{6}$$

Among them, α is the update parameter. Depending on how fast the background environment changes, it can be 0.3–0.5. This model is only suitable for small and slowly changing scenes. When the Gaussian model presents abrupt or multi-peak distribution, the single-mode Gaussian background model cannot accurately describe.

The mixed Gaussian model uses multiple Gaussian distribution models. Each Gaussian model represents different characteristics. Generally, 3–5 are selected. The more selected, the stronger the processing power for fluctuations and the longer the corresponding processing time. The mixed Gaussian distribution model can well fit the problem of changing the pixel value of the background image under the

change of natural illumination. P represents the distribution of a pixel value, K represents the number of independent Gaussian distributions selected, and the pixel at time t can be obtained, the probability that the point value is x is:

$$P(x_i) = \sum_{i=0}^{K-1} \frac{w_{t,i}}{2\pi} e^{-n/2(x_i - u_{t,i})} \tag{7}$$

Among them, $w_{t,i}$ represents the weight parameter of the i-th Gaussian distribution at time t, n represents the dimension of the image, and the color RGB image has a dimension of 3. Generally, a single-channel grayscale image is used for modeling.

B. FRAME DIFFERENCE METHOD

1) TWO-FRAME DIFFERENCE METHOD

The two-frame difference method is to subtract the gray image of the current frame image from the gray image of the previous frame, set the current frame image (x, y) point gray value to I (x, y, t). The gray value of the image (x, y) point is I (x, y, t − 1), which can be expressed as:

$$\Delta I_t(x, y) = |I(x, y, t + 1) - I(x, y, t)| \tag{8}$$

Due to the displacement of the moving object in two adjacent frames of images, the gray difference of the area where the moving object is located is generally larger.

2) THREE FRAME DIFFERENCE METHOD

Suppose I(x, y, t − 1), I(x, y, t), I(x, y, t + 1) are three consecutive gray-scale images. The second frame and the third frame of gray-scale image are respectively difference, two difference images are obtained, the two difference images are superimposed, and the binarization process is performed to obtain the binary image $D_t(x, y)$. You obtain the binary image $W_t(x, y)$ of the moving object, as follows:

$$D_t(x, y) = |I(x, y, t) - I(x, y, t + 1)| \otimes |I(x, y, t) - I(x, y, t - 1)| \tag{9}$$

$$W_t(x, y) = \begin{cases} 0 & T \geq D_t(x, y) \\ 1 & T < D_t(x, y) \end{cases} \tag{10}$$

Among them, T represents the threshold. The subsequent calculation is the same as the two-frame difference method. The three-frame difference detection target method is shown in Figure 1. Because there is a comparison between the previous frame and the next frame and the current frame, only the overlap between the current frame and the previous two frames is retained. The stretching of the object's movement direction is removed, making the position detection more accurate. However, due to the superposition of the holes due to the AND operation, the hole area will be enlarged compared with the two-frame difference method, and the approximate color of the object will appear to be truncated, and the front part and the back part of the movement direction are checked as two objects.

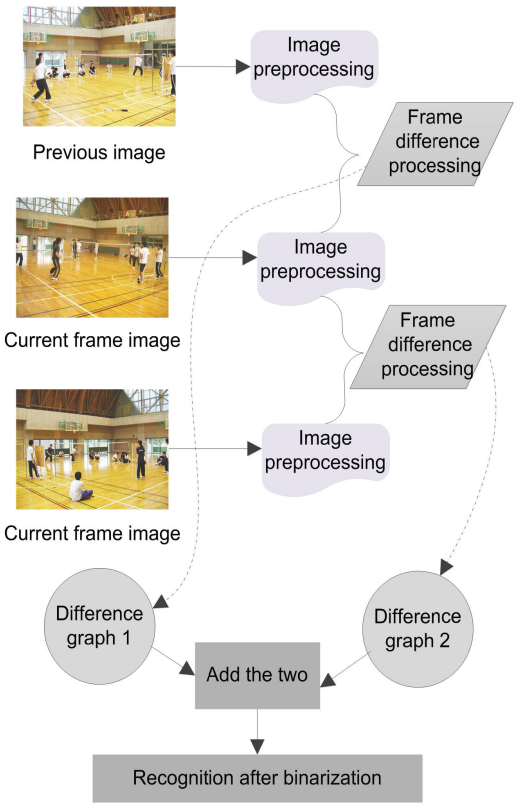


FIGURE 1. Flow chart of three-frame difference method.

C. OPTICAL FLOW METHOD

Optical flow is the instantaneous velocity of the object in space on the observation plane. The motion of an object in three dimensions is called a motion field, and the projection of the motion field in a two-dimensional image is a pixel, and the field of instantaneous velocity of these pixels is called an optical flow field. The optical flow method is to analyze a series of optical flow fields in the image. The optical flow carries the movement information of the object and reflects the change of the movement state of the object at different times. The movement of the target is determined by calculation and analysis.

The optical flow method based on feature matching is divided into local feature matching and global feature matching according to the different feature regions selected. Based on the local feature matching method, the main features of the target are extracted, such as corner points, texture features, and SIFT feature points for detection, which can quickly locate the target location. Based on the global matching algorithm, it is necessary to locate the position of the target corresponding area in the two frames of images in advance, and obtain the required optical flow information by comparing the displacement distance of the target area of the two frames of image. The flow of optical flow method dual-channel network for moving image feature extraction is shown in Figure 2.

The frequency-domain-based method obtains the position information of the moving target from the image through

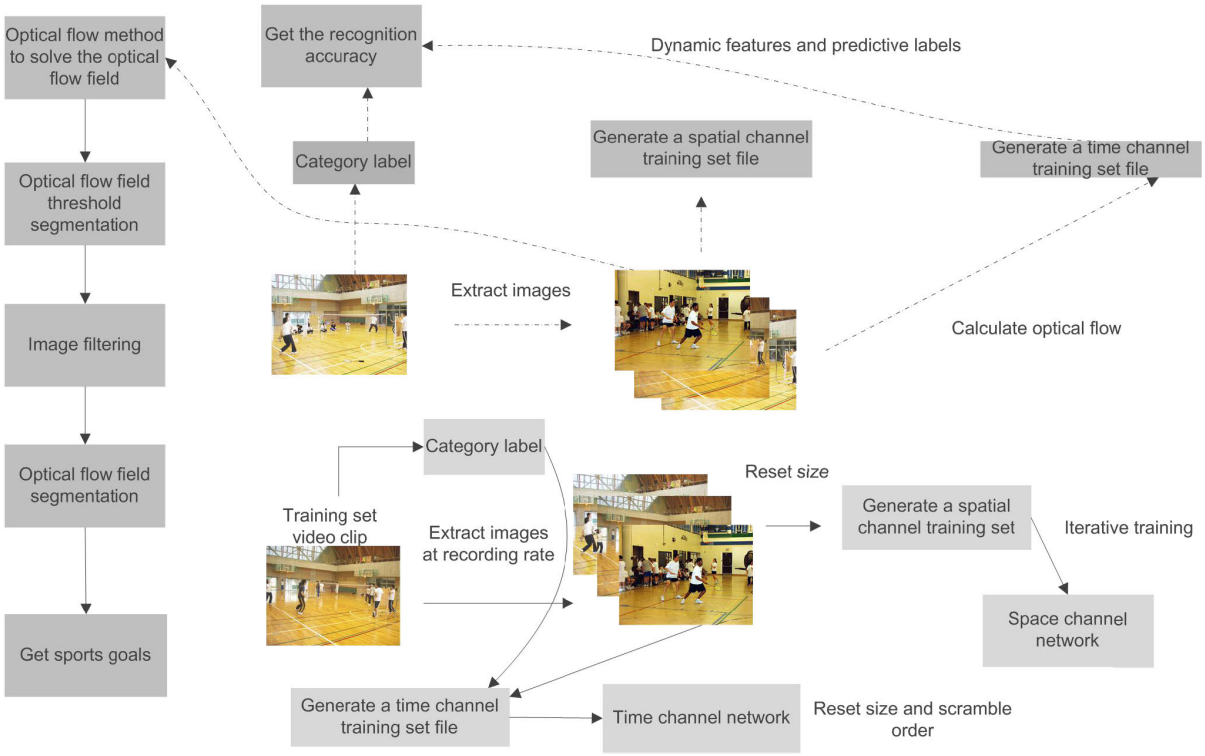


FIGURE 2. The flow of moving image feature extraction with optical flow method dual-channel network.

frequency-domain filtering, and performs high-precision initial optical flow estimation. Because of the calculation in the frequency domain, the amount of calculation is often relatively large.

The gradient-based method is to perform differential calculation on the image in space and time to obtain the optical flow information of the moving target. The calculation of this method is relatively simple, and a better initial estimation of optical flow can be obtained. Its disadvantage is that the parameters need to be adjusted and it is difficult to obtain the evaluation factor for reliability.

Estimating the feasibility hypothesis, we can get:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \tag{11}$$

Among them, I(x, y, t) represents the gray value of the pixel (x, y, t), I(x +Δx, y +Δy, t +Δt) represents the gray value of the pixel after displacement.

D. NO PARAMETER ESTIMATION THEORY

Non-parametric probability density estimation, also known as non-probability density estimation, is developed from the theoretical quotient of non-parametric statistics.

The histogram method is the first proposed method for parameter-free probability density estimation to describe the distribution of image gray values. The histogram is divided into one-dimensional and two-dimensional histograms. The one-dimensional histogram method is more commonly used. Recently, the domain method has poor resistance to features and is susceptible to interference from random features. There are many calculation methods, the simplest of which is the sliding window three-point average method, which selects the values of three pixels and assigns them to the middle point. There are also methods such as mean filtering. These methods filter out while denoising lots of detailed information. In order to ensure that useful information is not filtered out while denoising, a kernel density estimation method is proposed, which is currently the most effective parameter-free estimation method.

The kernel density estimation method is to divide a group of adopted data into the same level, each is called a "bin", which is similar to the idea of histogram, and adds one to the theory of histogram. The kernel function is used to smooth the data.

Kernel density estimation is developed on the basis of the kernel function. Each pixel of the image is regarded as a piece of data, then the image is a set of data points in a one-dimensional space. Then the probability is unknown. Let the probability density be f(x) and the kernel function is $K_h(x)$, you can get the density at point x:

$$f(x) = \frac{1}{n} \sum_{i=0}^{n-1} K_h \left| x - x_i \right| \tag{12}$$

Among them, x is the center point of the kernel function.

E. TARGET RECOGNITION BASED ON IMAGE MATCHING

Image matching is based on the comparison of the texture, gray and feature information of the two images, and the

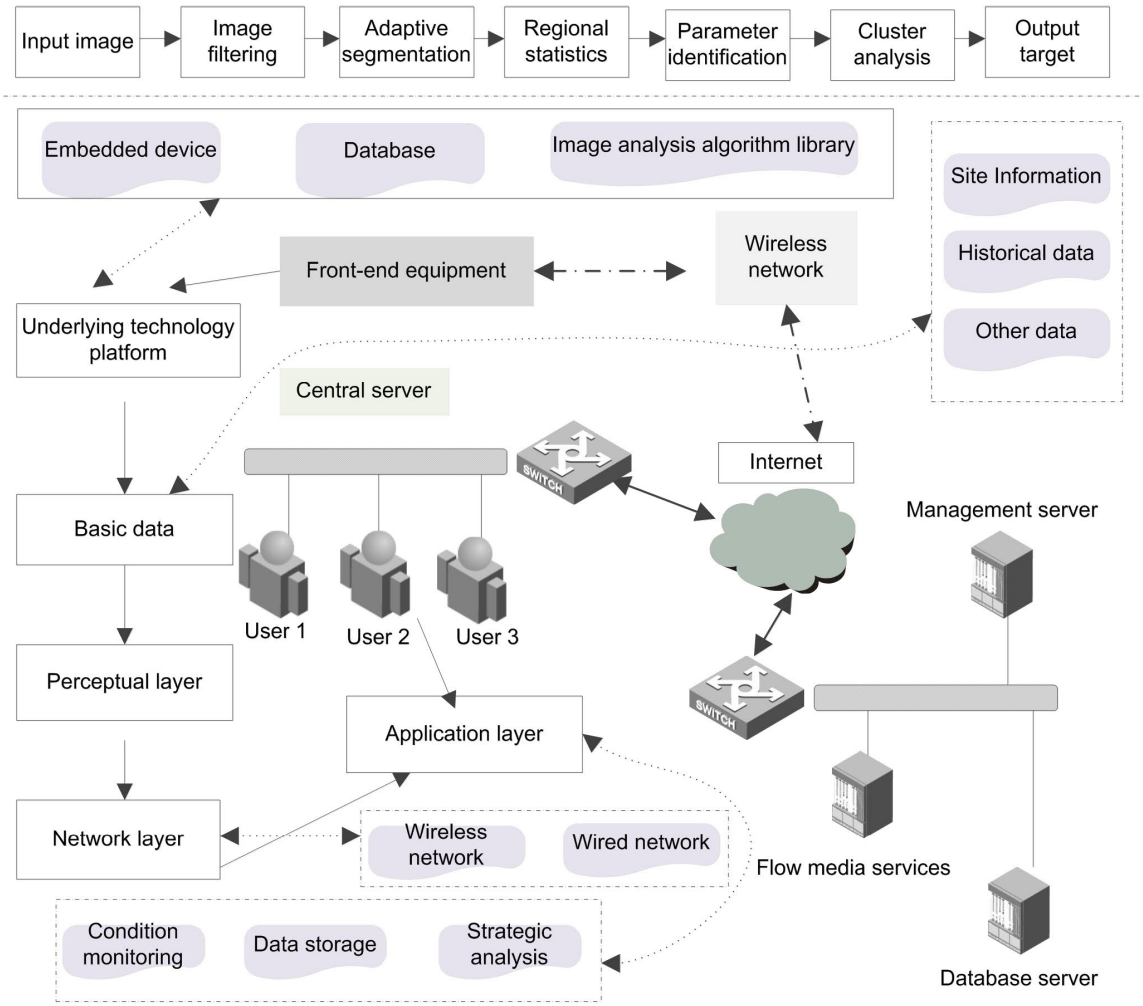


FIGURE 3. Image platform system architecture based on moving image recognition.

similarity between the target window and the window of the same size in the search area is calculated by a matching algorithm, and the matching recognition result is the highest similarity. Image matching can be divided into two types according to the transform domain of the feature. Based on the spatial domain matching, the feature and other information are selected in the original image. The algorithm is relatively simple and intuitive, suitable for recognition in simple backgrounds, and does not have a complex background effect. The frequency domain-based matching is to convert the original two-dimensional image to the frequency domain, and then convert it back to the two-dimensional space according to the frequency domain feature comparison to obtain the desired target area image. This algorithm is more complicated than the spatial domain matching calculation. But it still has a good matching effect for complex backgrounds. The image platform system architecture based on moving image recognition is shown in Figure 3.

The performance evaluation of the image matching algorithm has three main elements: the first is feature selection, the feature must be selected from all the features that need to be identified, and the number must be large enough to ensure the accuracy of the matching; the second is similarity measurement, that is, the degree of similarity between two images. It is necessary to select an appropriate measurement standard. Some distance or cost functions are commonly used as evaluation standards, such as Hausdorff distance, Bhattacharyya coefficient, etc.; the third is search strategy. There are many search strategies commonly used in engineering, such as hierarchical search method, relaxation iteration method, etc. With the development of artificial intelligence technology, some related strategies have been proposed, such as neural networks and genetic algorithms.

1) AREA MATCHING METHOD

The target area template is obtained through target detection, the center of the target template area is used as the center of the search window, the search window is shifted to scan the image to be processed in turn, and the similarity measurement function is used to calculate the similarity between the search window and each pixel in the area to be matched. The pixel with the largest similarity value is the target center point of

the search result, the updated center point is the center of the target area template, and the template area is corrected according to the similarity measurement function. The similarity measurement function is:

$$c(x, y) = \frac{1}{A} \sum_{i=0}^{n-1} \frac{l_i^2}{r_i} \tag{13}$$

Among them, A represents the template area of the target area, n represents the number of scans of the search window, r_i represents the area of the i-th scan area, and l_i is the overlap area of the i-th scan area and the template area.

In practical applications, the algorithm has very low resolution of similar gray information scenes. Under complex backgrounds, it is difficult to distinguish the targets that need to be identified from the background. Under simple backgrounds, it has good accuracy for rigid objects. And the detection rate for non-rigid objects is very low.

2) FEATURE MATCHING METHOD

Feature matching is to perform feature detection on the detected target area. The selected feature descriptor needs to be able to record the area information around the feature point, and the descriptor must be invariant. The feature matching algorithm is briefly introduced using Harris corner points as an example.

Harris corner points are the key points of some geometric structures in the image, most of which are the intersections between lines. The first-order derivative at the corner of the image is the largest, and the second-order derivative is zero. According to these characteristics, a window can be used to move in the image.

III. EXTRACTION ALGORITHM BASED ON ADAPTIVE CLUSTERING AND PROGRESSIVE PRINCIPAL COMPONENT APPROXIMATION

A. INTERPRETATION OF THE PRINCIPAL COMPONENT EIGENVALUES OF THE CHARACTERISTIC MATRIX N BASED ON THE RANDOM MATRIX THEORY

For a large-size eigencovariance matrix, the asymptotic properties of the principal component eigenvalues can be described by Marchenko-Pastur (MP) law. In the subsection $R_{M\times L}$, each element of the random matrix N is independent of each other and obeys the Gaussian distribution with an average of 0 and a variance of σ^2. Its covariance matrix ØN = 1/LNNT is called the Wishart matrix. If given M, L $\to \infty$, γ = M/L, and variance $\sigma 2 < \infty$, the range of the eigenvalue λn of Wishart matrix ØN is (λn−, λn+), where $\lambda n\pm = \sigma^2(1 \pm \gamma)^2$. Within this range, the eigenvalue λn of ØN has the following probability distribution:

For $\gamma \leq 1$,

$$p_{\sigma,\gamma}(\lambda_n) = \frac{\sqrt{(\lambda_n - \lambda_{n-})}\sqrt{(\lambda_{n+} - \lambda_n)}}{2\pi\lambda_n\gamma} \tag{14}$$

Here 1(·) is an indicator function. For $\gamma > 1$,

$$p_{\sigma,\gamma}(\lambda_n) = \frac{\sqrt{(\lambda_n - \lambda_{n-})}\sqrt{(\lambda_{n+} - \lambda_n)}}{2\pi\lambda_n\gamma} + \frac{1}{\gamma}\delta(\lambda_n) \tag{15}$$

According to the probability distribution described in the equation by MP theorem, we can know that the eigenvalues of the principal components of N satisfy:

$$E(\lambda_n) = \int_{-\infty}^{\infty} \lambda_n p_{\sigma,\gamma}(\lambda_n) d\lambda_n \tag{16}$$

Since MP theorem requires M, L $\to \infty$, the actual size of matrix N is limited. Therefore, the MP theorem cannot accurately describe the characteristics of the eigenvalues of the Wishart matrix. In this article, we only use the MP theorem to approximate the eigenvalues of a Wishart matrix ØN with a finite size. By estimating the range of the principal component eigenvalues of N $\lambda_{n,i} \in$ (λn−, λn+), we can further estimate the numerical rank r of the data matrix X.

We can also get the relationship between the principal component eigenvalue $\lambda_{x,i}$ of X and the principal component eigenvalue $\lambda_{n,i}$(r + 1 ⩽ i ⩽ M) of N:

$$\lambda_{n,i} = \lambda_{x,i}(\sqrt{\lambda_{n+}} > \sqrt{\lambda_{x,i}}) \tag{17}$$

B. GLOBAL FEATURE LEVEL ESTIMATION

The feature level is a key parameter in the entire image denoising process. The selection of low-rank regions causes a heavy computational burden and is unstable at high feature levels. The fast feature level estimation method is proposed based on the phenomenon that image blocks obtained from featureless images are usually located in low-dimensional subspaces. The low-dimensional subspace can be learned by the low-rank approximation of principal component analysis, and the feature level can be estimated by the feature value of the covariance matrix of the feature block. Inspired by this work, and based on the MP theorem, we propose an effective and stable feature level estimation method.

Using the relationship between the feature level σ and the principal component feature value λ of the random matrix N derived based on the MP theorem and the low-rank assumption, we effectively estimate the Gaussian feature level of the entire image. The MP theorem has been used to locally estimate the MRI feature level for a small 3-D block. Our method is based on the low-rank hypothesis and MP theorem and performs global estimation of feature levels. The rationality of the global method lies in the fact that MP theorem describes the properties of large-size matrices more accurately than small-size matrices. If you use the subscript ϕ to represent the entire image. In order to achieve highly accurate estimation, we divide the image of size a × b into overlapping dϕ × dϕ image blocks, and stack all overlapping image blocks together to construct a Mϕ = d2ϕ and Lϕ = (A − dϕ + 1). This large noisy matrix Xϕ can be decomposed into the sum of the low-rank matrix $X_{0,\phi}$ and the characteristic matrix Nϕ:

$$X_\Phi = N_\Phi + X_{0,\Phi} \tag{18}$$

Among them, each column of Nϕ is a vector NMϕ$(0, \sigma^2 I)$ that follows a multivariate Gaussian distribution, with a mean value of 0 and a variance of σ^2I.

The estimation of the characteristic level σ is directly related to the principal component characteristic value of Nϕ. However, Nϕ is unknown, only Xϕ is known. The low rank of $X_{0\phi}$ allows us to study the relationship between the principal component eigenvalues of Nϕ and the principal component eigenvalues of Xϕ.

C. ADAPTIVE CLUSTERING EXTRACTION

We developed an adaptive clustering method through ''overclustering-iterative merging''. The proposed adaptive clustering includes two steps: the over-clustering step ensures that image blocks of different types of image features are completely separated, and the iterative merging step ensures that the same type of image features are in the same class. The adaptive clustering proposed by the ''over-clustering-merge'' method is different in two aspects: 1) the proposed method only takes the feature level σ as a parameter; 2) our iterative merging is based on the distance between classes, not on the data. The distance between the point and the class center is non-iterative merge.

Next, let us introduce the adaptive clustering method in detail. In the clustering process, we need to obtain a large number of classes. In order to accelerate and improve the clustering effect, we use K-means-based ''divide and conquer'' technology.

After clustering, iterative merging is performed to prevent clusters from being too small and scattered. We set a reference value T as the merge threshold. If the distance between any two classes is $//x - y// < T$, we merge these two classes, where $// \cdot //^2$ represents the l^2 norm, and the vectors x and y represent the mean vectors of the two image patch groups. When the number of image block groups no longer changes, the iteration ends. The image extraction based on adaptive image block clustering and progressive PCA domain threshold processing is shown in Figure 4.

Since both classes are obtained by clustering based on K-means, it is almost impossible to have two classes with characteristics of the same type, one of which has a very large size $La = L$ and the other A class only has $Lb = 1$. So the possibility of merging these two image block groups should be $P(DB, A < T) = \epsilon$, where ϵ is a very small number.

The threshold T obtained in the above extreme cases is also applicable to general cases. Since the influence of the feature on the center decreases as the size of the class increases, the T derived above can be regarded as the acceptable upper limit of the influence of the feature on the center of the class. If the distance between two classes is less than T, then it is assumed that there is acceptable similarity between the two classes, so that the two classes can be grouped into one class.

When the size of the two image block groups are large, the center of the two image block clusters suffers very little interference from the feature. As a result, the originally defined threshold (acceptable similarity) will be set too large. In order to correct the threshold in this case, this article assumes that in this case, the distance between the centers of the two image groups is enlarged by a coefficient (equivalent to a decrease in the threshold).

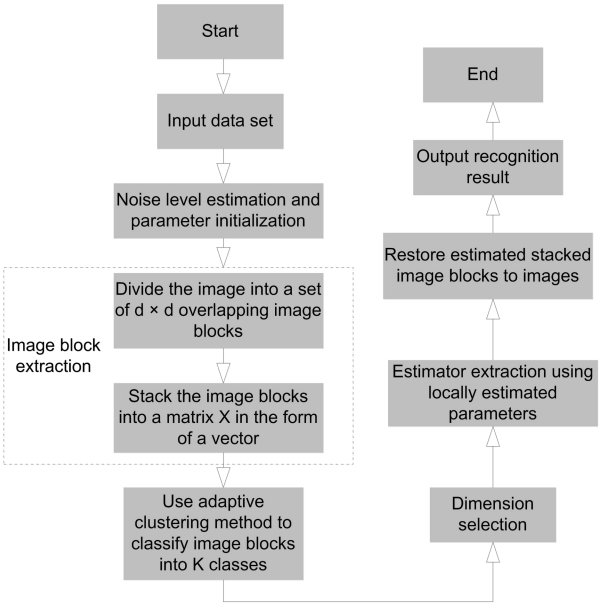


FIGURE 4. Image extraction based on adaptive image block clustering and progressive PCA domain threshold processing.

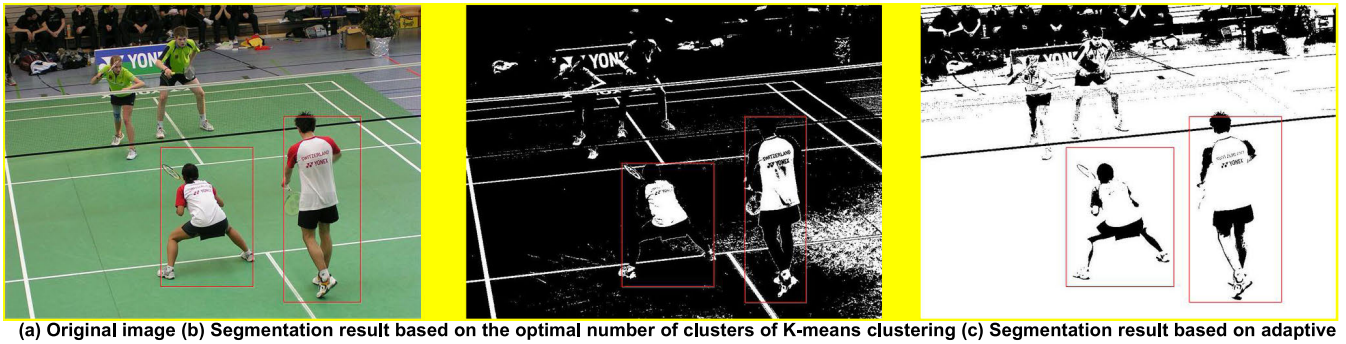
Figure 5 compares the proposed clustering method with traditional K-means clustering initialized with the optimal number of clusters. We see that image segmentation based on the proposed clustering method contains more image details. The segmented images obtained by adaptive clustering show more details like edges and singular points than K-means clustering, and there is no over-segmentation of homogeneous regions.

In the MP-SVD denoising step, we calculate a low-rank approximate matrix based on the MP theorem to select the most feature part of each class matrix. We use $X_{0,j}$ and $X_{j,1} \leqslant j \leqslant K$ to denote the j-th non-characteristic matrix and characteristic matrix in the image. Because we have obtained the characteristic level.

For simplicity, let X denote any characteristic matrix X_j. Then, we approximate the low-rank matrix by hard thresholding the singular values of X.

D. FEATURE EXTRACTION WITH LOCAL PARAMETER ESTIMATION

After searching for most of the features based on the random matrix theory, we get the signal-dominated low-rank image block matrix. Next, we will carefully de-noise the dominant part of the signal. In the principal component transform domain, the signal-dominated part has a total of r_j transform bands. In this algorithm, we process each principal component transformation zone of the dominant part of the signal in a local mode instead of a global mode. This is because we observe that the signal-dominated principal component transformation band has more in-band variation than the feature-dominated transformation band.



(a) Original image (b) Segmentation result based on the optimal number of clusters of K-means clustering (c) Segmentation result based on adaptive clustering

FIGURE 5. Original image and segmentation result.

We use the Linear Minimum Mean Square Error (LMMSE) method to find residual features. We noticed that LPG-pivot also applies this method. However, LPG-pivot applies the LMMSE estimator to the image block group in a global manner. In contrast, our algorithm uses the LMMSE estimator in a local manner, that is, to estimate the parameters of the LMMSE estimator through local averaging. Suppose that sj, (i, k) is an element of the matrix $U^T X_j$ (representing the kth coefficient of the i-th transform band of the jth image block group), and satisfies $1 \leqslant i \leqslant rj$ and $1 \leqslant k \leqslant Lj$.

For the parameters S_j, (i, k), in the LPG-pivot, it is calculated in a ''global'' way. Unlike LPG-pivot, the proposed method uses the ''local'' average of adjacent coefficients to estimate. Since this algorithm is a filtering algorithm combining hard threshold and special soft threshold, we call it an extraction algorithm based on adaptive clustering and progressive principal component thresholding. The local polynomial approximation method combined with the confidence interval intersection rule is a method of point-by-point adaptive estimation of one-dimensional signals. In the standard Local Polynomial Approximation (LPA), there are the following loss functions:

$$\phi(n) = \frac{1}{N} p_h(n)[\, y(n) - m]^2 + p(n) \tag{19}$$

where n is the position of the center of the window and m is the order of the LPA. $\rho(n)$ is a basic window function, $\rho h(n) = \rho(n/h)/h$. Here h gives the window size. In particular, for a square uniform window, if the condition is satisfied $\rho(n) = 1$ when $|n| \leqslant 1$, there is $\rho h(n) = 1$ when $|n| \leqslant h$, otherwise $\rho h(n) = 0$. ICI adaptively determines the maximum window size, in this window, the signal can be well approximated by a local polynomial.

If the original featureless signal is smooth, LPA-ICI can obtain near-optimal signal recovery quality. Since the signal formed by the coefficients in the principal component transformation band need not be smooth, when using LPA to denoise the signal, you may get an over-smooth result. Here, we do not use LPA-ICI to directly denoise the signal, but only use the h calculated by LPA-ICI as the size of the window needed to estimate the parameters of the suboptimal Wiener filter. We set $m = 1$ in order to find the appropriate window size h. In such a window calculated by LPA-ICI, all signals can be approximated by a constant amplitude signal.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. MANUAL POINT SET TEST

In this section, we first test the proposed LBWCMD method on an artificial point set. Then we explain the experimental parameter settings of related methods, and do comparative experiments on sports videos. Finally, the time complexity analysis of APCEA is given. There are two main types of comparison methods: the first type is subspace-based moving target extraction methods, including RPCA, Go Decomposition (Go Dec), PCA, and RPCA-ME; the second type is statistical-based methods, including Mahalanobis Distance-Based Method (MD), Gaussian Mixture Model (GMM), Self-Organizing-Based Method (SOBS) and Kernel Density Estimation (KDE).

For a certain point set, the LBWCMD method can maximize the intra-class distance in each subset while maintaining the spatial distribution of the points in the original set. In this section, we create four sets of artificial points to verify the effectiveness of the algorithm. The four point sets include:

(1) Points that are scattered and aggregated;
(2) Points distributed in a straight line;
(3) Points with random walking distribution;
(4) The points are evenly distributed. We use the objective function to evaluate the performance of LBWCMD.

The lower bound r plays an important role in LBWCMD. It determines the minimum distance between two points in the same subset. If the value of r is too small, then neighboring points cannot be allocated to different subsets with greater probability. On the contrary, if the value of r is too large, then the points in each subset remain almost unchanged, which will cause the scale of each set to not be balanced.

The result of the division is shown in Figure 6. Obviously, while neighboring points are allocated to different subsets, the spatial distribution of points in each subset can still be well maintained. Figure 7 lists the objective function's evaluation of the four types of artificial point set division results. The global similarity index indicates that the convex hull of each subset and the convex hull of the full set overlap in most cases.

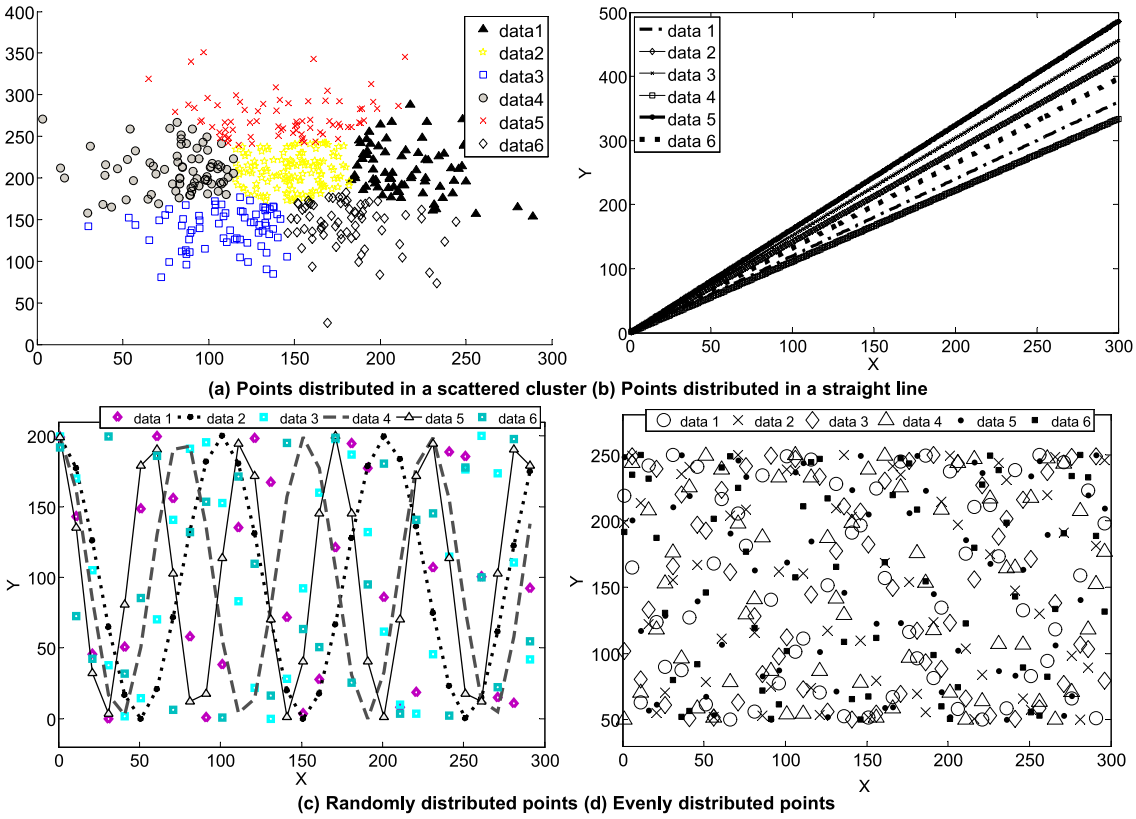


FIGURE 6. LBWCMD division result on point set.

In addition to the global similarity index, the local similarity index shown in Figure 7 shows that the density of the points in each subset tends to be consistent in the local area. The comprehensive similarity index is the comprehensive reflection of global similarity and local similarity. The closer the value of this indicator is to zero, the better the effect of the division. It can be seen from Figure 7 that the APCEA method proposed in this paper can well divide the artificial point set.

B. EXPERIMENTS ON THE VIDEO DATASET

In this section, in order to verify the effectiveness of the proposed algorithm, we conduct comparative experiments on sports videos. In the choice of comparison method, we not only compare the existing RPCA method, but also compare it with related subspace methods. In order to further prove the effectiveness of the algorithm, we also compared statistics-based moving target detection methods.

We define the ideal standard result of moving target detection as the standard value, namely Ground Truth (GT). The resolution of this standard value is the same as the size of the video frame. In GT, the value corresponding to the area where the moving target appears is 255, which is white, and the value corresponding to the background area is 0, which is black. In this experiment, all GTs of the video frame will be used to evaluate the performance of the algorithm. The overlapping part of the target area detected by the algorithm and the foreground area of GT is a true positive, and the non-overlapping part is a false positive. The similarity is sampled in the experiment to evaluate the performance of the algorithm. Definition D is the set of coordinates of all foreground pixels in a frame detected by the algorithm, and G is the set of coordinates of real foreground pixels in the corresponding frame.

If the detected foreground area is almost the same as GT, the similarity will tend to 1. On the contrary, if the detected foreground area does not have any overlap with GT, the similarity will be 0. Therefore, we can easily evaluate the detection effect of APCEA and other methods.

1) ALGORITHM PARAMETER SETTING

In our proposed method, there are three parameters that need to be determined, namely parameters m, r and α. Where m represents the number of divisions of the video frame subset, r represents the distance between two moving targets, and the parameter α reflects the number of frames with smaller motion amplitude. Based on experience, we use fixed parameters $m = 10$ and $r = 7$ for all videos in the experiment. We introduce a weighting factor p into the empirical value to balance the role played by the low-rank part and the sparse part.

For RPCA-ME and principal component methods, set the number of principal components to 10. Since these subspace methods process video frames in a batch manner, we construct a measurement matrix with every 300 frames as a

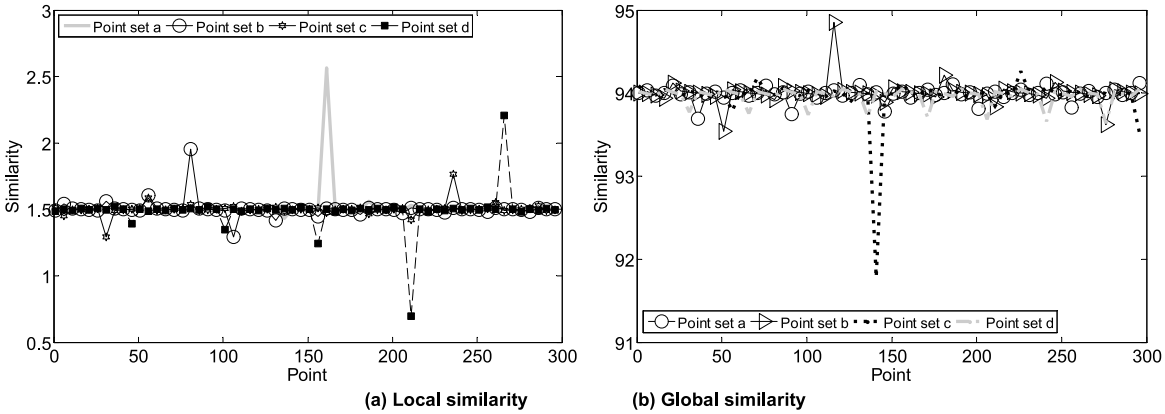


FIGURE 7. The effect of dividing on the point set (%).

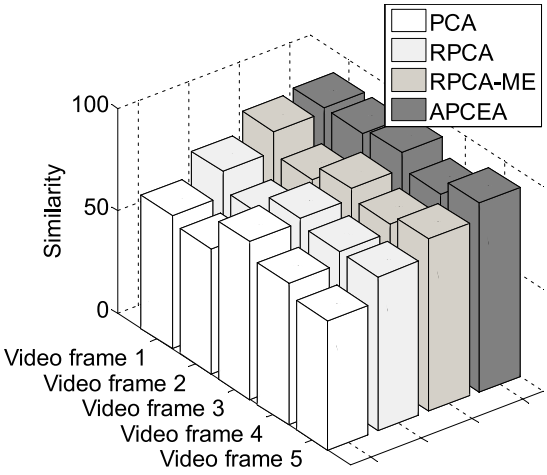


FIGURE 8. The similarity under various videos (%).

group to detect moving targets through component analysis or low-rank decomposition. The generated foreground part is a grayscale image. This experiment converts the grayscale foreground image into a binary foreground image by setting a threshold value of 25.

2) COMPARATIVE EXPERIMENT UNDER CONVENTIONAL VIDEO ENVIRONMENT

In this section, we compare APCEA with subspace-based methods and statistical-based methods in different video environments. The resolution of this video is 352 × 288, and we manually calibrated the GT of the corresponding frame every 10 frames.

We run the comparison algorithm on these 5 video frames. The obtained similarity evaluation results are shown in Figure 8. APCEA works best, that is, compared with all other methods, the results of this method are closest to the standard value. It can be seen from the figure that the APCEA method proposed in this paper performs better than all other methods on 5 video frames.

In order to show the effects of all methods more vividly, we display the binary foreground pictures obtained by each method in Figure 9. Unlike the traditional method, the APCEA method proposed in this paper fully integrates the distribution of sparse features in space and time, as far as possible to make the sparse features of each subset divided evenly distributed, and the coverage is wider. In addition, the algorithm also makes full use of video frames that contain a small amount of motion or no motion occurs. These video frames can provide more real background pixels, which facilitates the separation of foreground and background under the effect of low-rank decomposition. Therefore, the proposed method can obtain better moving target detection effect than some other methods.

3) COMPARATIVE EXPERIMENT UNDER STRONG LIGHT VIDEO ENVIRONMENT

In this section, we will conduct a comparative experiment in a strong light video. The sports video was shot in a strong outdoor environment with a resolution of 360 × 240. The frame number of GT provided ranges from 300 to 304. GT is provided on the 5 key frames of the video as a test. Our test results on the lighting video are shown in Figure 10. Our method achieves better results than all other methods on sports videos.

4) EXPERIMENTS ON IMAGE RECOGNITION OF MOTION DECOMPOSITION IN A HIGH-DENSITY CROWD ENVIRONMENT

In order to further test the performance of all algorithms, we conduct comparative experiments on three challenging videos. For each video, we manually calibrate a frame of GT every 10 frames. The resolution of the video is 360 × 288 and the number of frames is 1000. This video is a complex type of data set. Not only are there more people coming and going, but they are also seriously blocked. Since the video itself does not provide GT as an evaluation criterion, we manually calibrated the GT of the corresponding frame every 10 frames. The test results of the algorithm on all GT corresponding frames on the three videos will be shown in Figure 11(a), 11(b) and 11(c) in the form of similarity curves. From the curves of the three figures, it can be found

FIGURE 9. Comparative experiment results under video frames.

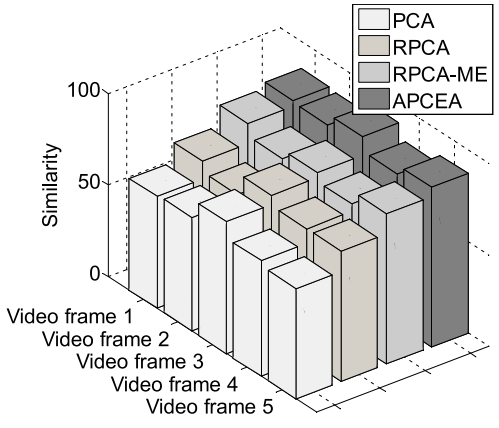
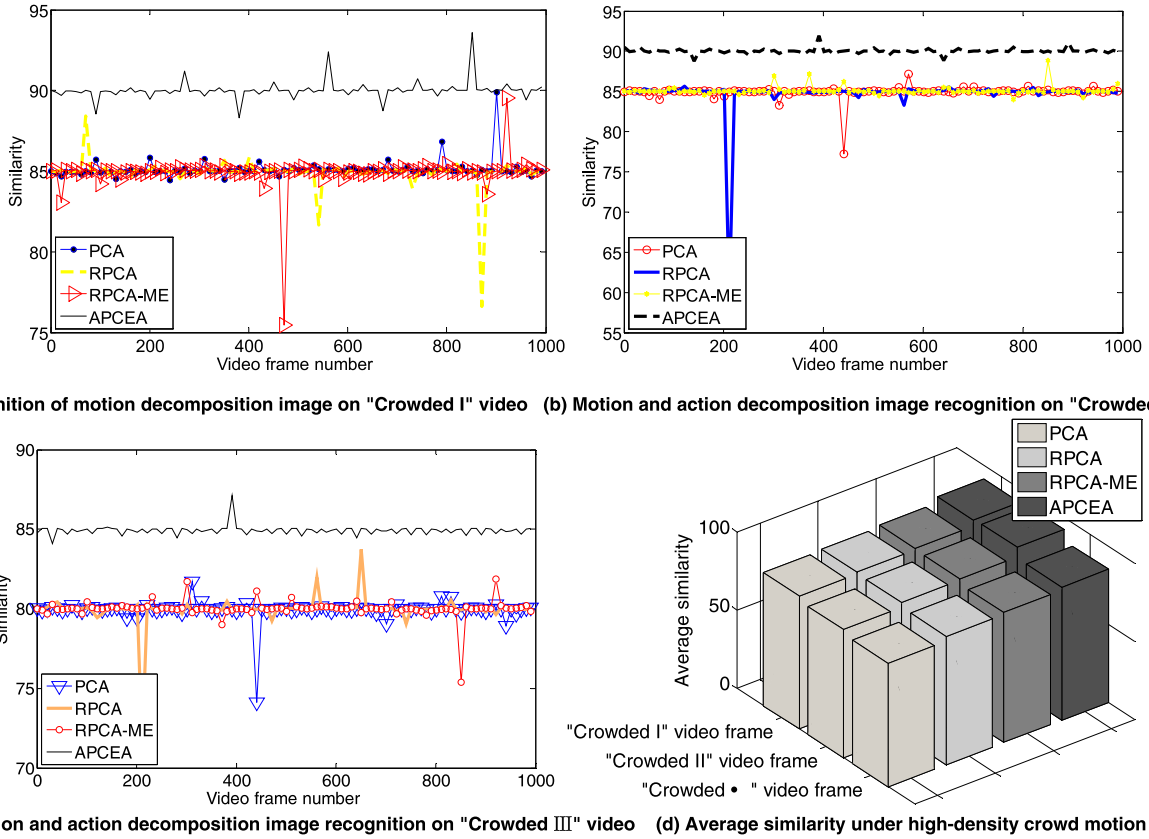


FIGURE 10. Similarity (%) under strong light video.

that the proposed APCEA method achieves almost the best results on all GT corresponding frames. Among the three videos, "Crowded II" is the one with the most crowded and blocked conditions. Even under such circumstances, APCEA can still obtain satisfactory results. We display the average of the results of all GT corresponding frames of each video in Figure 11(d), and we can find that APCEA has the best effect.

Here we further analyze the experimental results. Since APCEA is also a kind of subspace method, let's start with the subspace method. Traditional subspace methods often convert continuous video frames into column vectors of the measurement matrix, and then restore the sparse part of the matrix through a certain principle to achieve the purpose of detecting moving targets. This is a global mode of moving target extraction method. However, in actual situations, especially in crowded video sequences, the frequency of background pixels appearing in the video is not high due to the high crowd density. Therefore, in this case, the moving target detection method based on the global mode will fail. Different from the global mode, the APCEA method proposed



(a) Recognition of motion decomposition image on "Crowded I" video (b) Motion and action decomposition image recognition on "Crowded II" video

(c) Motion and action decomposition image recognition on "Crowded Ⅲ" video (d) Average similarity under high-density crowd motion videos

FIGURE 11. The similarity between experimental results on sports videos and high-density crowd videos (%).

in this paper makes full use of motion prior information, that is, position information, and generates different subsets according to the principle of dividing adjacent frames into different frame subsets. Therefore, the video frames in each subset are discontinuous, that is to say, the measurement matrix of each frame subset is sparse from the perspective of time and space. This processing mode can alleviate the impact of high-density people on moving target detection to a certain extent. In addition, the proposed method also introduces the advantage of augmented set. The video frames in the augmented set provide more real background pixels, which facilitates the separation of the sparse part and the low-rank part. For other statistical-based methods, they often need to go through a training or learning stage to obtain better moving target detection parameters. But for video sequences in crowded situations, it is very difficult to learn a good set of parameters.

C. ALGORITHM COMPLEXITY ANALYSIS

In this section, we present the complexity analysis of the APCEA method. Although the algorithm complexity of the proposed method is at the same level as RPCA, because APCEA needs to perform some preprocessing steps before low-rank decomposition, the total number of operations of the proposed method will be higher than that of RPCA. In fact, after the preprocessing step, the low-rank decomposition on each group set can be performed in parallel. This prompted

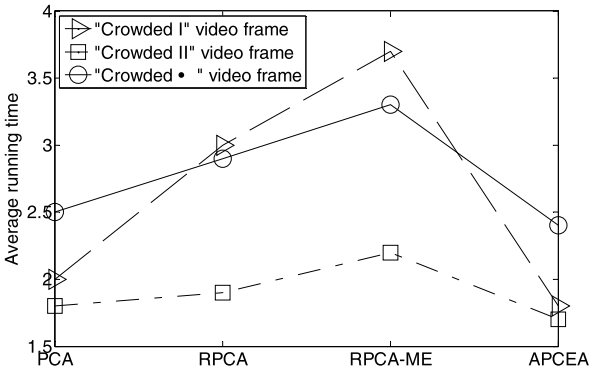


FIGURE 12. The average running time (seconds) of various methods under different videos.

us to apply parallel technology to the APCEA method to speed up its execution. To this end, we have calculated the average execution time of all comparison algorithms under the Windows 10 platform, as shown in Figure 12. Obviously, the algorithm proposed in this article has the fastest execution speed on the three videos. This is because the proposed algorithm performs some optimization steps before low-rank decomposition.

V. CONCLUSION

Based on the low-rank hypothesis and MP theorem analysis, this paper interprets the principal component eigenvalues, and

obtains the eigenvalues of the characteristic matrix and the distribution characteristics of the characteristics in the principal component transform domain. In terms of feature level estimation, in view of the shortcomings of current feature level estimation algorithms that are not stable and accurate under high-level features, we use the MP theorem to propose a new feature level estimation method by analyzing the principal component eigenvalues of all image data matrices. For image feature extraction, in order to overcome the shortcomings of traditional clustering algorithms that it is difficult to adaptively determine the number of classes, we propose an adaptive clustering algorithm suitable for extraction problems to obtain the extraction effect of detail preservation. For each segmented class, different from the traditional simple soft threshold or hard threshold processing method, we propose a progressive principal element domain approximation method that combines hard threshold and special soft threshold of local estimation parameters. We finally achieve the purpose of detail preservation and extraction. According to the principle of maximizing the distance within the class, a frame set division method is proposed, and this method is used as a preprocessing step of low-rank decomposition in moving target detection, so that the average detection accuracy is improved on the sports video data set. This work divides the frame set to make the obtained sparse feature distribution in each subset more conducive to the separation of foreground and background. In order to solve the problem of too small subset size, we use video frames with smaller motion amplitude as a supplement to each subset, thereby increasing the number of real background pixels. A method of combining statistical methods and subspace methods is discussed. While making full use of the advantages of both, the detection algorithm is more robust in complex motion environments such as light changes and severe occlusion.

REFERENCES

[1] L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2247–2253, Dec. 2007.

[2] J. Xu, M. Jiang, L. Yu, W. Yang, and W. Wang, "Robust motion compensation for event cameras with smooth constraint," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 604–614, Jan. 2020.

[3] Z. Gao, D. Y. Wang, Y. B. Xue, G. P. Xu, H. Zhang, and Y. L. Wang, "3D object recognition based on pairwise multi-view convolutional neural networks," *J. Vis. Commun. Image Represent.*, vol. 56, pp. 305–315, Oct. 2018.

[4] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.

[5] Y. Feng, M. Ji, J. Xiao, X. Yang, J. J. Zhang, Y. Zhuang, and X. Li, "Mining spatial-temporal patterns and structural sparsity for human motion data denoising," *IEEE Trans. Cybern.*, vol. 45, no. 12, pp. 2693–2706, Dec. 2015.

[6] C. Igual, J. Igual, J. M. Hahne, and L. C. Parra, "Adaptive auto-regressive proportional myoelectric control," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 2, pp. 314–322, Feb. 2019.

[7] M. Atzori, A. Gijsberts, I. Kuzborskij, S. Elsig, A.-G. Mittaz Hager, O. Deriaz, C. Castellini, H. Muller, and B. Caputo, "Characterization of a benchmark database for myoelectric movement classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 23, no. 1, pp. 73–83, Jan. 2015.

[8] L. Shao, L. Liu, and M. Yu, "Kernelized multiview projection for robust action recognition," *Int. J. Comput. Vis.*, vol. 118, no. 2, pp. 115–129, Jun. 2016.

[9] B. Zhao, R. Ding, S. Chen, B. Linares-Barranco, and H. Tang, "Feedforward categorization on AER motion events using cortex-like features in a spiking neural network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 1963–1978, Sep. 2015.

[10] X. Lagorce, G. Orchard, F. Galluppi, B. E. Shi, and R. B. Benosman, "HOTS: A hierarchy of event-based time-surfaces for pattern recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 7, pp. 1346–1359, Jul. 2017.

[11] R. Benosman, C. Clercq, X. Lagorce, S.-H. Ieng, and C. Bartolozzi, "Event-based visual flow," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 407–417, Feb. 2014.

[12] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, and D. Tao, "Stacked convolutional denoising auto-encoders for feature representation," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1017–1027, Apr. 2017.

[13] L. Shao, J. Han, D. Xu, and J. Shotton, "Computer vision for RGBD sensors: Kinect and its applications," *IEEE Trans. Cybern.*, vol. 43, no. 5, pp. 1313–1316, Oct. 2013.

[14] Z. Gao, D. Y. Wang, S. H. Wan, H. Zhang, and Y. L. Wang, "Cognitive-inspired class-statistic matching with triple-constrain for camera free 3D object retrieval," *Future Gener. Comput. Syst.*, vol. 94, pp. 641–653, May 2019.

[15] Z. Gao, H. Zhang, G. P. Xu, Y. B. Xue, and A. G. Hauptmann, "Multi-view discriminative and structured dictionary learning with group sparsity for human action recognition," *Signal Process.*, vol. 112, pp. 83–97, Jul. 2015.

[16] G. Gallego, J. E. A. Lund, E. Mueggler, H. Rebecq, T. Delbruck, and D. Scaramuzza, "Event-based, 6-DOF camera tracking from photometric depth maps," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2402–2412, Oct. 2018.

[17] T. Matsubara and J. Morimoto, "Bilinear modeling of EMG signals to extract user-independent features for multiuser myoelectric interface," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 8, pp. 2205–2213, Aug. 2013.

[18] L. Liu, L. Shao, X. Li, and K. Lu, "Learning spatio-temporal representations for action recognition: A genetic programming approach," *IEEE Trans. Cybern.*, vol. 46, no. 1, pp. 158–170, Jan. 2016.

[19] E. Mueggler, H. Rebecq, G. Gallego, T. Delbruck, and D. Scaramuzza, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and SLAM," *Int. J. Robot. Res.*, vol. 36, no. 2, pp. 142–149, Feb. 2017.

[20] P. Zhu, W. Zuo, L. Zhang, S. Chi-Keung Shiu, and D. Zhang, "Image set-based collaborative representation for face recognition," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 7, pp. 1120–1132, Jul. 2014.

[21] A.-A. Liu, N. Xu, W.-Z. Nie, Y.-T. Su, Y. Wong, and M. Kankanhalli, "Benchmarking a multimodal and multiview and interactive dataset for human action recognition," *IEEE Trans. Cybern.*, vol. 47, no. 7, pp. 1781–1794, Jul. 2017.

[22] Y. Dong, S. Gao, K. Tao, J. Liu, and H. Wang, "Performance evaluation of early and late fusion methods for generic semantics indexing," *Pattern Anal. Appl.*, vol. 17, no. 1, pp. 37–50, Feb. 2014.

[23] L. Liu, L. Shao, X. Zhen, and X. Li, "Learning discriminative key poses for action recognition," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1860–1870, Dec. 2013.

[24] G. Orchard, C. Meyer, R. Etienne-Cummings, C. Posch, N. Thakor, and R. Benosman, "HFirst: A temporal approach to object recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 2028–2040, Oct. 2015.

[25] L. Shao, X. Zhen, D. Tao, and X. Li, "Spatio-temporal Laplacian pyramid coding for action recognition," *IEEE Trans. Cybern.*, vol. 44, no. 6, pp. 817–827, Jun. 2014.

[26] M. Yu, L. Liu, and L. Shao, "Structure-preserving binary representations for RGB-D action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 8, pp. 1651–1664, Aug. 2016.

[27] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview spectral embedding," *IEEE Trans. Syst., Man, Cybern. B. Cybern*, vol. 40, no. 6, pp. 1438–1446, Dec. 2010.

[28] S. H. Abdulhussain, S. A. R. Al-Haddad, M. I. Saripan, B. M. Mahmmod, and A. Hussien, "Fast temporal video segmentation based on krawtchouk-tchebichef moments," *IEEE Access*, vol. 8, pp. 72347–72359, 2020.

[29] S. H. Abdulhussain, A. Rahman Ramli, B. M. Mahmmod, M. I. Saripan, S. A. R. Al-Haddad, T. Baker, W. N. Flayyih, and W. A. Jassim, "A fast feature extraction algorithm for image and video processing," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8, doi: 10.1109/IJCNN.2019.8851750.

[30] A. Sasithradevi and S. Mohamed Mansoor Roomi, "A new pyramidal opponent color-shape model based video shot boundary detection," *J. Vis. Commun. Image Represent.*, vol. 67, Feb. 2020, Art. no. 102754.

[31] S. Wan, Y. Zhao, T. Wang, Z. Gu, Q. H. Abbasi, and K.-K.-R. Choo, "Multi-dimensional data indexing and range Query processing via Voronoi diagram for Internet of Things," *Future Gener. Comput. Syst.*, vol. 91, pp. 382–391, Feb. 2019.

[32] G. Orchard and R. Etienne-Cummings, "Bioinspired visual motion estimation," *Proc. IEEE*, vol. 102, no. 10, pp. 1520–1536, Oct. 2014.

[33] F. Lunardini, C. Casellato, A. d'Avella, T. D. Sanger, and A. Pedrocchi, "Robustness and reliability of synergy-based myocontrol of a multiple degree of freedom robotic arm," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 9, pp. 940–950, Sep. 2016.

[34] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.

[35] T. Huynh-The, O. Banos, S. Lee, B. H. Kang, E.-S. Kim, and T. Le-Tien, "NIC: A robust background extraction algorithm for foreground detection in dynamic scenes," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 7, pp. 1478–1490, Jul. 2017.

[36] T. Huynh-The, C.-H. Hua, N. A. Tu, and D.-S. Kim, "Locally statistical dual-mode background subtraction approach," *IEEE Access*, vol. 7, pp. 9769–9782, 2019.

WEI YANG was born in Henan, China, in 1984. He received the bachelor's degree from Wuhan Sports University, in 2006, and the master's degree from Jing De Zhen Ceramic University, in 2018. He is currently working with Pingdingshan University. His research interest includes art and sports.