Feature Extraction and Target Recognition of Moving Image Sequences

PENGWEI SONG1,2,3, HONGYU SI1, HUA ZHOU4, RUI YUAN1,5, ENQING CHEN6, AND ZHENDONG ZHANG1

1School of Physical Education (Main Campus), Zhengzhou University, Zhengzhou 450001, China
2Division of Physical Education, Keimyung University, Daegu 42601, South Korea
3Division of Physical Education, Guangxi Science and Technology Normal University, Laibin 546199, China
4Department of Physical Education, Guangxi University of Finance and Economics, Nanning 530007, China
5Department of Marine Sports, Pukyong National University, Busan 48513, South Korea
6School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China

Corresponding author: Hua Zhou (zhouhua2222@126.com)

This work was supported by the Research and Practice Project of Higher Education Teaching Reform in Henan Province, Construction of Comprehensive Reform Mode and Practice Research of College Physical Education Teaching Based on Intelligent Management, under Grant 2019S1GLX005.

ABSTRACT The detection and recognition of moving objects in image sequence images involve many aspects, such as pattern recognition, image processing, and computer vision. The main difficulties of target detection and recognition are complex background interference, local occlusion, real-time recognition, illumination changes, target size type changes, etc. However, it is very difficult to solve these problems in practical applications. This article introduces image pre-processing for the pre-processing of image sequences. Selectively we highlight the visually obvious features that are helpful for target detection in the image, weaken the image background and features that are not related to the target, and improve the quality of the image sequence. A multi-information integrated probability density estimation kernel integrating gray scale, spatial relationship and local standard deviation information is designed, and the multi-information integrated kernel is used to extract the feature of the moving target. In terms of moving target recognition, Naive Bayes is used as a weak learner. In order to avoid the over-fitting of the classifier caused by high-noise moving image sequence features, the regularized Adaboost recognition model is introduced as a moving target recognition classifier. In order to completely separate the target and the background, we propose a moving target extraction method based on multi-information kernel density estimation, and input relevant target feature description vectors into the regularized Adaboost-based moving target recognition framework. Robust target recognition performance is obtained, and the reliability of target recognition under high noise data is improved.

INDEX TERMS Moving image sequence, multi-information integration, feature extraction, regularized Adaboost, target recognition.

I. INTRODUCTION
The actual scene where the moving target is located is generally a more complicated scene [1]–[3]. The image sequence contains a lot of interference noise, such as shadows cast by the target, noise from the sensor, obstruction by obstacles, and ripples on the water surface [4], [5]. The swaying of the branches in the scene where the moving target is located, the reflection interference of the water surface, and the light changes in the cloudy and sunny days will cause unpredictable changes in the pixels of the image [6]. It is precisely because of these changes that the recognition and tracking of moving targets exist. However, it is the existence of these complex scenes and unknowable changes that make target recognition and tracking more widely applicable [7], [8].

According to the different physical characteristics described, the characteristics of image targets mainly include spectral characteristics, texture characteristics and shape characteristics [9]. Using spectral features to identify targets, the main tools are gray-scale histograms of single-band images and color histograms of multi-spectral images. Compared with other features, spectral features have a simple...
and intuitive description, and are not sensitive to size or direction [10]. In some cases, it is quite robust, but it is difficult to completely and accurately describe a specific object using only spectral characteristics, because many different targets may exhibit the same spectral characteristics, which makes its application greatly restricted. According to the different regions of feature description, image features can be divided into two categories: global features and local features [11], [12]. In recent years, the research on local features has been very active, and new methods have emerged [13]. Local feature extraction generally includes two parts: feature region detection and feature region description. Most local features require certain invariance to brightness, scale, translation, and rotation [14]. The dense selection method is widely used in the sliding window model. Its advantage is that there is basically no loss of image details, and very rich local features can be obtained. However, a large part of the feature area has too little information, which has no effect or even interferes with the later recognition, which increases the burden of feature optimization in the next step. The number of feature areas detected by the sparse selection method is generally around 200 to 3,000, and its main advantages are simplicity and compactness [15]. The key points of the image are far less than the pixels of the image, so that the subsequent recognition process can be greatly accelerated. However, many feature area detection algorithms are often related to the characteristics of the image [16]. When applied to general target recognition, there may be certain limitations.

As the name implies, in the single classifier method, all target categories share a classifier, and the multi-classifier method sets a classifier for each category [17]–[19]. But the multi-classifier method will bring a very serious “rejection” problem. If the similarity between a target and all target categories is less than the corresponding threshold, the target cannot be identified. In this case, you have to call the single classifier method and set its category to the category with the highest similarity. Therefore, the method of setting a classifier for each category is not widely used [20]–[22]. According to the degree of manual participation in the classifier training process, it can be generally divided into supervised and unsupervised recognition [23], [24]. The essential difference between them is whether the training data has known class labels. Unsupervised recognition is mainly used to determine the “similarity” between two feature vectors and a suitable measure, and to select an algorithm scheme to cluster (group) the vectors based on the selected similarity measure [25]. Generally, different algorithmic solutions may lead to different results, which must be explained by experts [26]. Supervised recognition can design classifiers by learning labeled data and mining known information, which can obtain a higher accuracy model with a smaller training set [27]. In recent years, the recognition method of hybrid generation-discrimination learning has received extensive attention [28]. This method combines the advantages of generative models and discriminant models. Generally, a generative model (probability density model or structural model) is established for each type of model, and then the parameters of the generative model are optimized using discriminant learning criteria. The learning criterion can be a weighted combination of generative model learning criterion (such as maximum likelihood criterion) and discriminant learning criterion (such as conditional likelihood) [29]. Reinforcement learning has been deeply researched and widely used in the field of pattern recognition in recent years [30], [31]. It is essentially an online learning. The most obvious difference from supervised learning is that there is no need to specify the label of the target category, only the outside world needs to give “right” or “wrong” feedback on the completion of this classification task.

In this paper, a multi-information integrated probability density estimation kernel that integrates grayscale, spatial relationship and local standard deviation information is designed. The multi-information integration kernel is used to extract the moving target; Naive Bayes is used as the weak learner, and the regularized Adaboost recognition model is introduced as the moving target. The recognition classifier improves the reliability of target recognition under high-noise data. Specifically, the technical contributions of this article can be summarized as follows:

First: A complete framework for moving target recognition is proposed. Starting from discussing and studying the kernel density estimation of the moving target area and the background sampling area, a kernel density estimation method integrating multiple information is proposed. The designed kernel density estimation method based on multi-information integrated kernel can distinguish the target from the background sampling area obviously, even if the image target is not located in the center of the processed image, the target can be effectively extracted.

Second: For the extracted target, we use the area descriptor of the invariant moment base and the edge descriptor of the Fourier descriptor base to describe, which can effectively integrate the gray distribution and edge shape information of the extracted target.

Third: By regularizing the Adaboost classifier, the target classification is achieved. Regularization ensures that even if the extracted target contains more noise, the classifier will not overfit.

The rest of this article is organized as follows. Section 2 discusses related theories of moving target detection and recognition. Section 3 describes the features of the extracted target and builds a learning model. Section 4 carried out experimental simulation and result analysis. Section 5 summarizes the full text and points out future research directions.

II. RELATED THEORIES OF MOVING TARGET DETECTION AND RECOGNITION

A. IMAGE PREPROCESSING

The image sequence contains a lot of information, and the image sequence per second includes 25-30 frames.
In addition, because the environment in which the image sequence is collected may be subject to conditional restrictions and random interference, resulting in the degradation of the image sequence, it is necessary to perform grayscale correction and noise filtering on the image in the image sequence, so it is not practical to directly detect the image sequence. Image preprocessing refers to selectively highlighting visually obvious features that are helpful for target detection in the image, and weakening features that are not related to the target in the image background. In addition, the image caused by noise interference is subjected to image enhancement, image smoothing and other operations.

1) GRAYSCALE COLOR IMAGE
Color is a very important feature in an image. Color is a visual response of human eyes and brain to light. There are thousands of colors in the image, which can be used to detect and segment the target directly. But for an image sequence that generates a large number of images per second, direct processing of color images requires a lot of calculations. Gray-scale image refers to an image that contains brightness information and discards other color information. Because gray-scale images are simple to calculate and sensitive to changes in illumination.

2) IMAGE FILTERING
It is possible to highlight the inherent characteristics of a certain part of the image. For example, if you want to eliminate the noise of the image and obtain a smooth image, you can construct a low-pass filter to remove the high-frequency components in the image, while retaining the low-frequency components in the image. The schematic diagram of image filtering is shown in Figure 1.

B. MOVING TARGET DETECTION METHOD
1) BACKGROUND SUBTRACTION
The background subtraction method is to use a frame image in the image sequence sequence as the background, and then subtract the pixel value of each pixel of the subsequent frame and the corresponding pixel of the background to obtain the background subtraction image. Because the background subtraction method is simple to implement, fast, and easy to implement on hardware, it is widely used in applications such as image sequence monitoring and moving target detection with a fixed camera.

2) INTER-FRAME PHASE DIFFERENCE METHOD
The calculation formula of the phase difference method is as follows:

\[
D_i(x, y) = \begin{cases} 
0 & T \geq |I_i(x, y) - I_{i-1}(x, y)| \\
1 & T < |I_i(x, y) - I_{i-1}(x, y)| 
\end{cases}
\]  

(1)

where \(T\) represents the judgment threshold, \(I_i(x, y)\) represents the image of frame \(i\), and \(D_i(x, y)\) represents the moving target area of the image of frame \(i\).

3) OPTICAL FLOW CALCULATION METHOD
Frequency-domain-based methods use frequency filters to obtain the phase information of moving targets in sequence images to obtain high-precision initial optical flow estimates. However, frequency-domain calculations are required, so the amount of calculation is often large and it is difficult to perform reliability evaluation.

4) FEATURE MATCHING METHOD
The features commonly used for target detection are point feature, outline feature and area feature. The feature matching method has good adaptability to occasions such as local occlusion and light changes, but feature matching requires a lot of calculations, so the feature matching method often needs optimization before it can be applied to real-time applications. Multi-feature fusion image target recognition system architecture is shown in Figure 2.

The regional feature matching method first extracts the matched target regional features, and then matches the regional features of the control image.

C. TARGET RECOGNITION METHOD
According to the number of classifiers, it is divided into a single classifier method and a multi-classifier method. A single classifier means that the entire target recognition system
has only one classifier. The advantage of this method is that only one classifier’s classification parameters need to be stored. There are too many classification parameters of the classifier, which is difficult to realize; the multi-classifier refers to designing a classifier for each category, and using the classifier to determine whether the target object belongs to the class, but the multi-classifier has a “refusal to identify” situation. When the similarity of all classifiers is less than the set threshold, the target cannot be identified. In view of the characteristics of single classifier and multiple classifiers, the two classifiers are usually used in combination. Boosting and Bagging are methods that use multiple weak classifiers to form a strong classifier. They both train multiple classifiers by resampling or weighting the training sample set. At the same time, they have a classification error rate upper bound that is stable as training increases. Features such as descent and no overfitting are suitable for various classification scenarios. Bagging is a serial method, and Boosting is a parallel method.

III. FEATURE DESCRIPTION AND LEARNING MODEL OF THE EXTRACTED TARGET

A. DETERMINATION OF TARGET CONSTRAINT AREA

Suppose that the temperature of the moving target we want to extract is higher than the background temperature, that is, it is a bright target imaged on the moving image sequence, and the average grayscale of the imaging target is higher than the average grayscale of the background. This assumption is true in most cases, because the targets to be extracted are usually people, croquet, rolling balls, missiles and other targets. We can find the approximate position of the target in the moving image sequence according to this characteristic, and get the target constraint area. The target constraint area here refers to a small image area located in the original image that completely or partially contains the target, but also contains part of the background.

For a frame of N×M-sized moving image sequence, a number of n×m-sized image blocks are divided in order in a non-overlapping manner (usually n = m = 8). After dividing
the image blocks, then we calculate the mean and variance of each block. We can quickly locate the approximate location of the target area, and a priori knowledge of the target size obtains the target constraint area.

After the target constraint area is determined, due to the existence of background high-gray area, we do not use the area growth method to extract the target, but propose a novel target extraction method based on multi-information integrated kernel density estimation.

**B. KERNEL DENSITY ESTIMATION METHOD BASED ON MULTI-INFORMATION INTEGRATED KERNEL**

Because the texture information of the moving target is relatively poor, other characteristic parameters should be considered to describe the target. The concept of local standard deviation is introduced here. For a moving image sequence, the local standard deviation value \( S(x_i) \) at pixel point \( x_i \) is defined as follows:

\[
S(x_i) = \sqrt{\frac{1}{M-1} \sum_{X \in M} [I(x_i) - I(X)]^2}
\]  

(2)

Among them, \( I(x_i) \) and \( I(X) \) represent the gray value at pixel point \( x_i \) and point \( X \) (pixel point \( X \) refers to the neighborhood pixel point centered on pixel point \( x_i \) in a predefined window), \( M \) is the number of pixels in the neighborhood.

For a two-dimensional discrete local standard deviation image of the target constrained region, its zeroth moment can be defined as:

\[
M_{00} = \sum_{i=0}^{\text{rows}} \sum_{j=0}^{\text{cols}} S(i, j)
\]  

(3)

Here, \( \text{cols} \) and \( \text{rows} \) are the number of horizontal and vertical pixels of the target constraint area analyzed, and \( S(i, j) \) is the local standard deviation value at the pixel point \((i, j)\). Similarly, the two first-order moments in the horizontal and vertical directions of the local standard deviation image in the target constrained area are defined as:

\[
M_{10} = \sum_{i=0}^{\text{rows}} \sum_{j=0}^{\text{cols}} i \cdot S(i, j)
\]  

(4)

\[
M_{01} = \sum_{i=0}^{\text{rows}} \sum_{j=0}^{\text{cols}} j \cdot S(i, j)
\]  

(5)

Here we use the zero-order moment information of the local standard deviation image to set the kernel bandwidth size of the kernel density estimation. Assuming that the maximum local standard deviation of the target constraint region is \( \delta_{\text{max}} \), the core bandwidth can be set as:

\[
h_x = \min 0.5 \cdot (\text{rows}, \alpha \cdot \sqrt{\frac{M_{00}}{\delta_{\text{max}}}})
\]  

(6)

\[
h_y = \min 0.5 \cdot (\text{cols}, \alpha \cdot \sqrt{\frac{M_{00}}{\delta_{\text{max}}}})
\]  

(7)

Among them, \( \alpha \) is a weighting factor, which depends on our prior knowledge of the target.

The above formula uses \( \delta_{\text{max}} \) to remove the zero-order moment of the local standard deviation of the pixel. In fact, it is equivalent to normalizing the local standard deviation of the pixel, and then finding the zero-order moment is equivalent. The normalized zero-order moment information is used to set the scale of the tracking window, that is, the local standard deviation value of the background is relatively low, the normalized value is very small, and the normalized value of the local standard deviation of the bright target is relatively large, and the zero-order moment is a simple sum of these normalized values. If the bright target area is larger, the value of the moment is smaller. Therefore, the zero-order moment information can fully reflect the general shape of the bright target area. In addition, the above formula calculates the square root value to ensure that the result can reflect the horizontal or vertical information of the target distribution. Therefore, the kernel function histogram of the target constrained area image at the feature \( u \) can be expressed as:

\[
q_N = C \cdot \sum_{i=0}^{N} k_s \cdot k_r \cdot (\delta - v)
\]  

(8)

Among them, \( k_s \) and \( k_r \) are the spatial relations and feature relation kernel function profiles, respectively, and \( v \) is the mean value of the gray feature quantification in the target constraint area.

The background sampling area refers to the image area that surrounds the target but does not contain the target. Usually, the target constraining area expands outward to form the background sampling area. This area plays a key role in removing the background part contained in the target constraining area.

**C. FEATURE DESCRIPTION OF THE EXTRACTED TARGET**

For the extracted target, we always need to describe it with corresponding features, and then we can complete the recognition. The features that can be used in the recognition method based on the extracted target features can be divided into two categories: the first category is based on the shape of the area covered by the object, mainly including area, roundness and moment features, etc., and its description method mainly includes travel coding, quadtree, moment descriptors, etc.; the second type is based on the shape of the target edge, mainly including perimeter, angle, width, height, diameter, etc., and its description methods mainly include boundary chain code, autoregressive model and Fourier descriptors, etc. Here, we use two kinds of feature descriptors, namely the descriptor of the invariant moment basis and the Fourier shape descriptor of the edge basis. In this way, the extracted target features have both the first type of features and the second type of features, that is, the descriptor of the invariant moment base is used to describe the shape of the extracted target.
area, and the edge-based Fourier shape descriptor is used to describe the target edge shape. The block diagram of moving image edge detection is shown in Figure 3.

The (p + q) order moment of the continuous image function $f(x, y)$ is defined as:

$$m_{pq} = \int \int x^p \cdot y^q \cdot f(x, y) \, dx \, dy$$

(9)

Then the central moment of the image $f(x, y)$ is defined as:

$$u_{pq} = \int \int f(x, y) \cdot (x - \bar{x})^p \cdot (y - \bar{y})^q \, dx \, dy$$

(10)

For the shape characteristics of the edge of the target, we use the discrete Fourier transform method to the boundary pixels of the target, which is the Fourier descriptor feature. For the kth boundary point of a target with a total number of boundary points K, if the x and y coordinates are $x(k)$ and $y(k)$, respectively, then each pair of coordinates corresponds to a complex number $s(k)$, namely:

$$s(k) = jy(k) + x(k)$$

(11)

The Fourier transform for discrete $s(k)$ is:

$$a(u) = \frac{1}{K} \sum_{k=1}^{K} s(k) e^{-2\pi j u k}$$

(12)

The complex coefficient $a(u)$ is called the Fourier descriptor of the boundary. The Fourier descriptor has translation, rotation and scale invariance to the target, and is insensitive to the starting point of the boundary. The high-frequency elements of the Fourier descriptor can explain the details well, and the low-frequency elements determine the overall outline shape. Therefore, the low-frequency element of the Fourier descriptor can be used to describe the approximate shape of the target. Here we take the first 9 elements of the Fourier descriptor and find the modulus, namely:

$$r(u) = |a(u)| \quad u = 1, 2, \ldots, 9$$

(13)

According to this, the normalized feature quantity can be obtained. Therefore, combining with the descriptor of the invariant moment base obtained above, we can get the final description feature of the target. This feature component contains not only the regional gray distribution information of the extracted target, but also the edge information of the target.

D. LEARNING MODEL

Adaboost learning/classification algorithm is a very commonly used learning/classification algorithm belonging to
the Boosting class. Its essence is to combine a series of weak learning/classifiers in a weighted sum to form a final learning/classification performance better than any combined weak learning/classifier learning/classifier. This section will start with the analysis of the Adaboost algorithm and gradually introduce its regularization (this article only uses one of the regularization methods in the Adaboost algorithm, namely AdaboostKL, so the regularized Adaboost mentioned below refers to the AdaboostKL algorithm). The regularized Adaboost algorithm proposes a learning/classification framework for moving target recognition.

1) ADABOOST ALGORITHM
The main idea of the Boosting method is to form a final classifier with high classification performance by combining multiple simple learners (called weak learners) with low accuracy. If the calculated combination coefficient is $\alpha_t$, the combined function is:

$$F(I) = \sum_t h_t \alpha_t(I)$$  \hspace{1cm} (14)

This is the final classifier, and its classification performance is better than any weak learners combined. In this way, as long as a weak classification algorithm that is slightly better than random guessing can be obtained, it can be combined into a strong classifier through the Boosting algorithm. Because the original Boosting classification algorithm needs to know the lower limit of the learning accuracy of the weak learner in advance when solving practical problems, this greatly limits the application of the Boosting algorithm. The target recognition process of Adaboost algorithm is shown in Figure 4.

2) REGULARIZATION OF ADABOOST ALGORITHM
Adaboost algorithm is one of the more effective classification methods. When the recognition features used are less affected by noise, it is difficult to overfit. However, when the noise received by the recognition feature is obvious, overfitting will occur. Due to the particularity of moving image sequence imaging, although we adopt a robust target extraction method, it is inevitable that the final feature description contains more noise. Therefore, in order to avoid the occurrence of overfitting, we use the entropy regularization method when obtaining the weight distribution of Adaboost (this method is the more commonly used regularization method when obtaining the probability distribution). If there are $N$ probability density distributions to be sought $w_i$, the regularization condition can be expressed as:

$$g(W) = \sum_{i=0}^{N-1} w_i (\ln w_i - \ln u_i) \quad w_i \geq 0$$  \hspace{1cm} (15)

Among them, $\mu_i$ is a priori distribution, $A > 0$ is a normalized constant.

3) CLASSIFIER BASED ON REGULARIZED ADABOOST ALGORITHM
According to the AdaboostKL algorithm, we propose a moving target recognition framework based on the AdaboostKL algorithm. When the weighted error rate of the weak learner is greater than 0.6, we exit the entire iteration process. In addition, since Naive Bayes is a very effective and simple classifier, it is used as a weak learner in our recognition framework. The final classifier will determine the class number of the sample output by the sign of the linear combination of a single weak learner and the corresponding combination coefficient.

IV. EXPERIMENTAL SIMULATION AND RESULT ANALYSIS
A. EVALUATION OF MULTI-INFORMATION INTEGRATED CORE
In order to verify the effectiveness of the proposed multi-information integration kernel, we conducted experiments with moving targets in different backgrounds. The experiment was implemented on Matlab software platform. The kernel density estimation method that can reasonably separate the target and the background is the most effective method. Therefore, the merits of the designed multi-information integrated kernel density estimation can be evaluated by calculating the ability of the target to distinguish the background. Here, we take the kernel density estimates of the target and background as two different probability densities, and introduce the expressions of relative entropy and discriminant entropy to judge the degree of separation of their probability densities. Suppose that $p$ and $b$ are the kernel density
estimates of the target constrained area and the background sampling area, respectively.

The summation is performed on all possible values of the quantization feature \( u \). To prevent the argument of the log function from being 0, we set it to a very small value when it is 0, such as 0.0001. Relative entropy can be used as a measure of the degree to which a probability distribution density deviates from a given standard distribution. It is a value less than or equal to zero. The smaller the relative entropy, the greater the difference between the two types of probability distributions being compared. When the two types of probability distributions are identical, the relative entropy reaches the maximum value (equal to 0). In order to balance the influence of the kernel density estimation of the target and background, the discriminant entropy \( W(p,b) \) can be used to characterize the degree of separation between the target class and the sampled background class.

Discriminant entropy is the sum of relative entropies between two different classes, so it is also a value less than or equal to 0. The smaller the discriminant entropy, the greater the difference between the two types of probability distributions compared, and the maximum discriminant entropy is also obtained when the two types of probability distributions are identical. We select 8 typical moving image sequences as shown in Figure 5, and calculate the relative entropy value to verify the degree of separation of the estimated kernel density from the target and background. In the experiment, the target constrained area is expanded by 10 pixels along the surroundings to form a background sampling area. The yellow rectangle in Figure 5 represents the target constraint area. A1, B1, C1, D1 target positioning is more accurate, A2, B2, C2, D2 moving image sequence target area positioning is relatively poor.

![Figure 5. Eight typical moving image sequences and marked target constraint regions.](image)

![Figure 6. The degree of separation of target class and sampling background class with different kernel density description methods.](image)
B. MOVING TARGET EXTRACTION BASED ON MULTI-INFORMATION INTEGRATION KERNEL

The moving image sequence has its research difficulties, that is, the background is complex and chaotic, the moving target lacks texture information, and the signal-to-noise ratio is usually very low. Therefore, for a common moving target extraction method, accurate extraction of moving targets is very difficult. Extracting targets based on less accurate methods has a greater impact on the performance of the subsequent recognition process. Therefore, if a robust method of moving target extraction can be designed, it is very beneficial for subsequent recognition to extract the target completely and without errors.

The background weighting method uses the background feature component with a larger feature component to generate a smaller weight to weight the value of the same component of the target constraint region, so that the kernel density component of the target constraint region that is the same as the background gray feature component becomes as small as possible. The kernel density estimation method is used to extract the moving target for the following reasons. First, the target and the background can be separated, that is, the gray values of most pixels of the target and the background are in different intervals. So most of the gray values that appear more frequently in the target must appear less in the background; secondly, the reasonable design of the kernel density can effectively describe the estimated area and can accurately embody the number and distribution of gray values; finally, the difference in the estimation of kernel density is finally reflected in the difference of each component. According to the value of different components and the kernel density component of each pixel, the effective extraction of the target can be achieved. In addition, this method is relatively simple and takes little time. For a sample image and its extracted target, the gray space is regarded as the feature space and quantized to 32 levels (the resolution of target extraction can be determined by the level of quantization). At the same time, in order to facilitate the understanding of the intermediate process of target extraction, Figure 7 also shows the curve of the kernel density estimated by the kernels integrating different information as a function of each component. Figure 7(a) is just the change curve of the two kernel densities estimated by the kernel integrating spatial relationship information, and the results are consistent with the histogram statistical results, that is, there are usually many peaks in the kernel density of the target constraint area and the background sampling area. And there may be overlap of peaks, at this time the background and target are not easy to separate. The kernel used in the kernel density estimation of the curve shown in Figure 7(b) integrates the grayscale, spatial relationship and local standard deviation information, which can highlight the kernel density peaks of the target constrained area and the background sampling area separately. The effective separation of regions is finally reflected in the result of moving target extraction. From the final target extraction results, our target extraction scheme is effective.

Since the images used for recognition training and testing are generally not too large, like this image, after the target constraint area is generally determined, the background sampling area refers to the remaining portion except the target constraint area. Our target extraction scheme is effective, and some details of the target are also extracted. To make it easier to understand the entire target extraction process, Figure 8 shows the change curve of the logarithmic value of the nuclear density ratio between the target and...
FIGURE 8. The logarithmic value of the nuclear density ratio between the target and the background.

the background with the quantized gray value (quantized to 32 gray levels). As can be seen from Figure 8, as long as we set an appropriate threshold, we can distinguish which gray-scale component really belongs to the target area, and which gray-scale component is the background area. In order to make the contour curve of the extracted target change smoothly, proper area expansion and corrosion are also used here to achieve the purpose. Since this part is relatively simple, it will not be detailed here.

C. TARGET RECOGNITION EXPERIMENT RESULTS

In order to verify the effectiveness of our proposed recognition framework based on regularized Adaboost, we tested our algorithm (including moving target extraction algorithm and recognition algorithm) with a series of related experiments. The entire moving target extraction algorithm and recognition framework were tested under the Matlab environment. In the kernel density estimation for target extraction, the gray space is taken as the feature space and quantized to 32 levels. The experiment in this section involves the size of each image in the image database is $68 \times 58$ pixels, and contains three categories of targets, namely badminton, rolling and croquet. In the training and test set, the number of positive samples and the number of negative samples are the same, and the total number of samples is shown in Figure 9.

Figure 10 shows some experimental results obtained with different target extraction methods. Because the size of the training and test images is not large, after the target constraint area is obtained, the remaining part is regarded as the background sampling part. It can be seen from Figure 10 that the regional growth method is more effective for the case where the target gray distribution is relatively uniform. However, when the gray distribution of the target is uneven or even discontinuous, especially when the target contains unconnected areas, the area growth method is not satisfactory for the target extraction, as shown in Figure 10(b), 10(c). In addition, the regional growth algorithm is relatively time-consuming, and the selection of the threshold for regional growth is not easy, and the selection of the threshold has a great influence on the result. For simple background images, the target can be effectively extracted, as shown in Figures 10(a) and 10(b). However, it is not very effective for target extraction in complex scenes. It will divide too much background into targets, and sometimes the background and target cannot be distinguished, as shown in Figure 10(b). However, from the three test images in different scenes in Figure 10, it seems that our proposed method of moving target extraction based on multi-information integration kernel is very effective for the target extraction of the tested images. Even if the target is in a complex environment, our algorithm can still extract the target completely, as shown in Figure 10(c).

For moving targets, if the wavelength of the acquired image is different, the imaging of the same target is also very different. The first two images of the positive sample are very different from the latter images, but in fact they belong to the same category. Obviously, the recognition results obtained by different area description methods are different. In the training and testing of regularized Adaboost, we use different feature descriptors for experiments. The corresponding experimental results are shown in Figure 11. The experimental results in Figure 11 show that the feature vector using the combination of the feature descriptor of the invariant moment base and the Fourier descriptor is better than the performance of using a single feature for recognition. This is obvious, because the recognition performance is naturally better than the classifier trained by a single feature. At the same time, we also found that using the Fourier descriptor as a feature to describe the target, the recognition result is better than the descriptor of the invariant moment base. This is mainly because the descriptor of the invariant moment base contains the statistical information of the gray distribution, which is greatly affected by the noise, and the Fourier descriptor can reflect the contour similarity of the imaging of the moving target at different wavelengths, so that the recognition performance is good.

Figure 12 compares the recognition results of the moving target recognition algorithm with the weak learner...
FIGURE 10. Experimental results (from left to right are the contours of the original moving image sequence, the target extracted by the area growth method, the target extracted by the Otsu algorithm and the target extracted by the proposed algorithm).

FIGURE 11. Results of different feature descriptors. (a) Descriptor of invariant moment base (b) Fourier descriptor (c) Combination of two features.

(Naive Bayes), the standard Adaboost algorithm and the proposed recognition framework based on the regularized Adaboost algorithm (AdaboostKL). The feature vectors used by these classifiers are a combination of the feature descriptors of the invariant moment base and the feature descriptors of the Fourier descriptor base. It can be seen from Figure 12 that the proposed recognition algorithm based on regularized Adaboost (AdaboostKL) has better recognition effect than the other two algorithms compared. The main reason is that the regularization method improves the performance of the classifier, and the test results are better than the non-regularization method. Therefore, for the recognition of moving targets, the regularization process is still very necessary.

We also show the 60 iterations of the three algorithms as classifiers in the badminton target recognition process, as shown in Figure 13. It can be seen from Figure 13 that
V. CONCLUSION

This paper has carried out a series of researches around the problem of detection and recognition of moving targets based on computer vision. Combined with multi-scale space theory, moving target detection theory and target recognition theory for in-depth research, practical solutions for practical applications are given. This article introduces the current commonly used target recognition model. Several main target recognition model methods are described, and production models and discriminant models are introduced in a targeted manner. Based on the analysis of the imaging characteristics of the moving target, the multi-information integration check is used to extract the moving target, that is, the moving target extraction method based on the kernel density estimation on the integrated gray scale, spatial relationship and local standard deviation information is effective for the target and the background. We especially pointed out the guiding role of the local standard deviation information on the adjustment of the nuclear bandwidth. In order to improve the reliability of target recognition under high-noise moving data, the recognition model Adaboost is regularized, and an effective moving target recognition algorithm is proposed based on the regularization. This paper proposes an algorithm for moving target recognition. For the description of the target, we use the invariant moment base descriptor to describe the shape of the extracted target area, and the boundary base Fourier shape descriptor to describe the edge shape of the target. However, such characterization is not complete. We need to collect and analyze the characteristics of the target and the possible movement methods, and extract various useful features, establish a target and environmental characteristic database, and filter the extracted features to facilitate classification. What kind of features to extract and what kind of feature selection method to adopt, and how to build the target and environmental characteristic database are yet to be further studied.

REFERENCES

[1] F. Golabi, M. Shamsi, M. H. Sedaaghi, A. Barzegar, and M. S. Hejazi, “Development of a new oligonucleotide block location-based feature extraction (BLBE) method for the classification of riboswitches,” Mol. Genet. Genomics, vol. 295, no. 2, pp. 525–534, Jan. 2020.
[2] P. Wang, H. Liu, L. Wang, and R. X. Gao, “Deep learning-based human motion recognition for predictive context-aware human-robot collaboration,” CIRP Ann., vol. 67, no. 1, pp. 17–20, May 2018.
[3] P. Helber, B. Bischke, A. Dengel, and D. Borth, “EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 12, no. 7, pp. 2217–2226, Jul. 2019.
[4] W. Wu, H. Li, X. Li, H. Guo, and L. Zhang, “PoSAR image semantic segmentation based on deep transfer Learning—Realizing smooth classification with small training sets,” IEEE Geosci. Remote Sens. Lett., vol. 16, no. 6, pp. 977–981, Jun. 2019.
[5] C. F. Baumgartner, K. Kannm bias, J. Matthew, T. P. Fletcher, S. Smith, L. M. Koch, B. Kainz, and D. Rueckert, “SonoNet: Real-time detection and localisation of fetal standard scan planes in freehand ultrasound,” IEEE Trans. Med. Imag., vol. 36, no. 11, pp. 2204–2215, Nov. 2017.
[6] K. Z. Haider, K. R. Malik, S. Khalid, T. Nawaz, and S. Jabbar, “Deeppen: Real-time gender classification using deep learning for smartphones,” J. Real-Time Image Process., vol. 16, no. 1, pp. 15–29, Feb. 2019.
ENQING CHEN was born in Tianjin, China, in 1977. He received the B.E. and M.S. degrees from Zhengzhou University, in 2000 and 2003, respectively, and the Ph.D. degree in communication and information systems from the Beijing Institute of Technology, China, in 2007. Since 2007, he has been with the School of Information Engineering, Zhengzhou University, where he is currently a Professor. In 2015, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, Ryerson University, Canada. His research interests include machine learning and image processing, including computer vision, human action recognition, and target detection.

ZHENDONG ZHANG was born in Beijing, China, in 1961. He received the bachelor’s degree from Wuhan Sports University, in 1983. He currently works with Zhengzhou University. He has published a total of 12 articles on CSSCI. He hosts a general project of the China National Social Science Foundation. His research interests include sports training and social sports.