

Research Article

Background Modeling Based on Statistical Clustering Partitioning

Biao Li ^{1,2,3,4} Xu Zhiyong ^{1,3} Jianlin Zhang ^{1,3} Xiangru Wang ²
and Xiangsuo Fan ⁵

¹Institute of Optics and Electronics, Chinese Academy of Sciences, Guangdian Avenue, Chengdu 610209, China

²School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, North Jianshe Road, Chengdu 610054, China

³University of Chinese Academy of Sciences, Yuquan Road, Beijing 100049, China

⁴Key Laboratory of Optical Engineering, Chinese Academy of Sciences, Guangdian Avenue, Chengdu 610209, China

⁵School of Electrical and Information Engineering, Guangxi University of Science and Technology, Donghuan Avenue, Liuzhou 545006, China

Correspondence should be addressed to Jianlin Zhang; jlin_zh@163.com

Received 16 June 2020; Revised 15 November 2020; Accepted 10 December 2020; Published 16 January 2021

Academic Editor: Jerzy Baranowski

Copyright © 2021 Biao Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to effectively detect dim-small targets in complex scenes, background suppression is applied to highlight the targets. This paper presents a statistical clustering partitioning low-rank background modeling algorithm (SCPLBMA), which clusters the image into several patches based on image statistics. The image matrix of each patch is decomposed into low-rank matrix and sparse matrix in the SCPLBMA. The background of the original video frames is reconstructed from the low-rank matrices, and the targets can be obtained by subtracting the background. Experiments on different scenes show that the SCPLBMA can effectively suppress the background and textures and equalize the residual noise with gray levels significantly lower than that of the targets. Thus, the difference images obtain good stationary characteristics, and the contrast between the targets and the residual backgrounds is significantly improved. Compared with six other algorithms, the SCPLBMA significantly improved the target detection rates of single-frame threshold segmentation.

1. Introduction

Usually in a remote detection system, the image of the target is relatively weak and occupies a small scale; i.e., it is a dim-small target. Here, the terms “dim” and “small” are defined in terms of the signal strength and scale of the target. There is no uniform standard for the definition of the dim-small target, and there are some differences of the definition in different historical periods and application backgrounds. The International Society of Optical Engineering defines the dim-small target as follows: in terms of signal energy, the local signal-to-noise ratio of the target is no greater than 5 dB; in the scale, the target occupies no more than 9×9 pixels, or no more than 0.12% of a 256×256 pixel image. Whether the dim-small target can be detected and tracked effectively is the key performance index of the remote imaging detection system. Therefore, detecting dim-small target stably and effectively to realize tracking and early

warning is a challenging research topic in the field of remote target detection. Classical infrared dim-small target detection algorithms can be classified into two categories: detect-before-track (DBT) [1, 2] and track-before-detect (TBD) [3, 4]. DBT is composed of single-frame target detection and multiframe trajectory correlation. Firstly, the candidate targets are detected through the single frame. Then, the multiframe trajectory correlation is carried out according to the feature of the continuity and consistency of the target motion, so as to eliminate the false target and confirm the real target. For TBD, it first searches the possible motion trajectories of the dim-small target. Then, it accumulates the signal energy of each trajectory, so as to calculate the probability of each trajectory to confirm the real target trajectory to detect the real target.

The influence of clouds, atmosphere, climate, trees, light, and other natural environments makes the background complex, and the target is almost submerged by background

clutter, which directly impairs the system's stable and effective detection of the target in postprocessing. Therefore, image preprocessing, namely, background modeling, is often needed before target detection. Background modeling methods, such as top-hat algorithm [5], two-dimensional least mean square error algorithm (TDLMS) [6], bilateral filtering algorithm [7], anisotropy algorithm [8], Gaussian Mixed Model (GMM) [9], ViBE [10, 11], and Pixel-Based Adaptive Segmenter (PBAS) [12], obtain their background model by using the neighborhood information of the central pixel, which is an approximate estimation method. Therefore, the generality of these methods is poor, and each algorithm can only achieve good results for specific background. In recent years, neural network algorithms have attracted much attention, and some algorithms have been proposed, for example, the convolutional neural network (CNN) algorithms [13, 14] and the deep convolutional neural network (DCNN) algorithms [15]. However, they need a large number of datasets for training and are usually time consuming. In recent years, the principal component analysis (PCA) method has been paid more attention again. Some new algorithms, such as robust PCA (RPCA) [16, 17], robust subspace learning (RSL) [18], and robust subspace tracking (RST) [19], have been proposed. The common feature of these algorithms is that the columns of the data matrix are composed of the whole image data, which makes good use of the structural information of the whole image, instead of only considering the local information. Therefore, the generality of these algorithms is better than the traditional methods being mentioned above. However, with the increase of image size and data matrix size, the amount of memory and iterative computation increase sharply [20].

In order to make good use of the structural information of the image and reduce the size of data matrix, a statistical clustering partitioning low-rank background modeling algorithm (SCPLBMA) is proposed in this paper. SCPLBMA only processes single-frame image and uses statistical clustering region to form data matrix, so the size of the data matrix is reduced and the structure information of the image can still be utilized. Compared with DCNN algorithms, SCPLBMA is a kind of algorithm that does not need supervised training, and is of low computational complexity and good real-time performance. SCPLBMA can effectively suppress the background and enhance the contrast between the targets and the residual background.

The remainder of the paper is arranged as follows. Section 2 mainly discusses the traditional typical background modeling algorithms and the methods based on PCA theory. Section 3 gives the basis and the model of the SCPLBMA. Experiments of the SCPLBMA are made with three typical scenes to test and evaluate its performance in Section 4. Finally, the conclusion is presented in Section 5.

2. Related Works

Background clutter is the main factor interfering with the dim-small target detection. To detect the dim-small target effectively, background suppression is needed before target

detection. Many scholars have done a lot of works on how to reduce the background clutter interference.

From the perspective of the image structure, the dim-small target image can be assumed to be composed of three components: background, target, and noise; i.e., it can be described as

$$f(x, y) = f_b(x, y) + f_t(x, y) + f_n(x, y), \quad (1)$$

where $f_b(x, y)$ is the background; $f_t(x, y)$ refers to the object; and $f_n(x, y)$ denotes the noise due to the imaging system itself, the imaging environment, and other sources. $f_n(x, y)$ can be considered to be additional interference superimposed on the background, with the characteristics of randomness, occupied few pixels, and isolation relative to the background. A small target $f_t(x, y)$ is also an isolated point relative to the background, but it is different from the noise in two characteristics: the target usually contains more pixels, and the target has a continuous trajectory in time domain. Based on the model of (1), many background modeling algorithms have been proposed [11, 20, 21], for example, the median filtering algorithm, TDLMS, bilateral filtering algorithm, anisotropy algorithm, top-hat filtering algorithm, GMM, ViBE, PBAS, and DCNN. The median filtering algorithm and the top-hat filtering algorithm are greatly influenced by the structure element that needs to be set according to the characteristics of images. The TDLMS algorithm estimates the background according to the error value between the predicted pixels and the real pixels. Therefore, the prediction of the TDLMS algorithm is better for the regions with high degree of correlation, but worse for the regions with large fluctuation. The bilateral filtering algorithm estimates the background based on the spatial distance and gray value of the pixel. It obtains good effect in the background that has large area with good correlation, but it is poor in the fluctuating edge area. The anisotropic filtering algorithm estimates the background based on the difference of the local gradient of the background and the target, so it is poor at the edge of the background and the region with rich textures. These methods are mainly based on the difference between background, noise, and target and employ a filtering method to estimate the central pixel from neighboring pixels so as to highlight the targets by suppressing the background and noise. Therefore, it is essentially a kind of spatial smoothing process, which may cause damage to the target information or even smooth out the target. At the same time, it leaves many residual points in the difference images after subtracting the background, which interferes with target detection in postprocessing. The GMM algorithm uses several Gaussian distributions with different parameters to fit the background points. As the number of Gaussian distributions involved in fitting increases, more parameters need to be updated. This makes the computation of parameter update increase sharply and the real-time performance worsen. At the same time, the statistical distribution of background does not always strictly conform to the Gaussian distribution, so the actual effect is imperfect. ViBE and PBAS are background modeling algorithms based on pixel level, and background model initialization can be

carried out with single image. Therefore, these algorithms are easy to implement and have good real-time performance. However, due to the defects of background model updating method, ViBE and PBAS algorithms have obvious disadvantages: One is ghostly phenomena. The other is that stationary or slowly moving targets may be absorbed or incomplete. The algorithms based on DCNN need a large number of datasets for training. Since the infrared dim-small targets have no obvious geometric features, gray value features, and texture features and the targets are almost immersed in the background clutter, it is difficult for GMM, ViBE, PBAS, and DCNN algorithms to effectively separate the background and dim-small targets.

From the perspective of matrices, an image matrix \mathbf{I} can be divided into two components: the low-rank matrix part and the sparse matrix part [16, 18, 20, 22], just as Figure 1; i.e.,

$$\mathbf{I} = \mathbf{L} + \mathbf{S}, \quad (2)$$

where \mathbf{L} denotes the large continuous area in the background, which is regarded as a low-rank matrix because of its high correlation. \mathbf{S} refers to the target and noise, regarded as a sparse matrix. Usually, the target and noise are assumed to be independent of the background and are superposed on the original low-rank backgrounds. Background modeling based on low-rank matrix reconstruction theory assumes the sparse components \mathbf{S} as disturbance or outliers superposed on the background. The low-rank part and the sparse part can be separated from the original video image by a low-rank reconstruction algorithm. In 1901, although Karl Pearson [16, 23] proposed principal component analysis (PCA), it was rarely applied to high-dimensional data due to its sensitivity to outliers, high memory requirements, and time consumption for computation. From 1980s on, some algorithms have been proposed to improve the robustness of PCA. Representative works, such as Campell [24], used a robust estimator instead of the standard estimation of the covariance matrix, and Croux and Ruiz-Gazen [25] used projection pursuit techniques. With the development of the PCA theory, in recent years, background modeling based on low-rank matrix restoration theory has attracted increasing attention. To improve the robustness of PCA against a grossly corrupted observation matrix \mathbf{I} , new algorithms have been proposed, such as the works of De la Torre and Black [26] and Qifa and Kanade [27]. However, none of these works gives a polynomial-time algorithm with strong and robust performance under a wide range of conditions. Inspired by applications in system identification and graphical models, Chandrasekaran et al. [28] proposed a new algorithm based on sparsity of matrix rank. However, this algorithm neglected the missing entries in the observation matrices. Since there has been no precise mathematical definition of the term “outlier” for a long time, the robust PCA problem was not clearly defined, which affects the development of the robust PCA theory [18]. It was not until 2010 that Wright and Ma [29] put forward the idea that an outlier was regarded as an additive sparse corruption. Based on the definition of Wright and Ma, Candes et al. [17] proposed robust PCA (RPCA), which is more robust than the previous algorithms because it can recover a low-rank matrix \mathbf{L}

even when there are highly erroneous or some missing entries in the measurement matrices. Gao et al. [30] proposed a low-rank matrix restoration method based on an infrared patch-image model. In this method, a single-frame image was processed. However, the method using a fixed regular shape sliding window to get the infrared image patches cannot ensure that each patch is statistically stationary for complex background; i.e., in this way two absolutely different statistic regions may be in one patch. Low-rank background reconstruction methods can be classified into two kinds. In the first one, each frame of a sequence of video images is written into a new matrix as a column vector to form a data matrix, and then low-rank background modeling and matrix decomposition are carried out. This method is widely used in facial recognition, video surveillance, and so on. The advantage of this kind of algorithms is that the whole image is directly taken as the data to be processed, which preserves the data structure. However, the amount of data is large, and therefore a large amount of computation and memory resources is required. The second method processes single image which slides a fixed-scale window to obtain different areas of an image, and then stacks these different background areas into a new data matrix. The data matrix is then decomposed into a low-rank part and sparse part. Finally, the background reconstruction is performed. However, these methods regularly cut the image into many strips of a predetermined size. Therefore, they may break the natural structure of the image, and the reconstructed background may be locally nonstationary.

Actually, the core of the above two kinds of background modeling methods based on low-rank matrix reconstruction is to equalize the background with strips of different images or strips of different areas of one image, so that the background is stationary and of low rank. Based on the insight of these methods, in this study, instead of equalizing the background area, this paper presents a new method, the statistical clustering partitioning low-rank background modeling algorithm (SCPLBMA), which separates the image into several areas according to their statistics so that each area is consistent statistical distribution. Evidently, different from the first kind of methods, SCPLBMA only processes one image. Not just as the second kind of methods that arbitrarily cut an image into strips, SCPLBMA can well preserve the natural structure of the image by taking into account the statistics of each area.

3. Statistical Region Low-Rank Background Modeling

This section introduces the SCPLBMA. First, the stationary characteristics of the image are discussed, and a conclusion is drawn that the statistical clustering region is stationary. Second, low-rank background modeling is carried out for statistical clustering region: after the statistical clustering, the clustering region is extracted; then, the low-rank and sparse characteristics of clustering regions are analyzed, and a low-rank modeling model suitable for PCA method is given. Finally, the detailed solution of background estimation is given: according to PCA theory, the model is converted into PCP equation; then the appropriate solving method of the



FIGURE 1: Example of low-rank matrix reconstruction.

model is discussed in detail; soft-threshold operator [31] and singular-value-threshold operator [32] are determined to seek the optimal solution.

3.1. Stationarity of the Image. Mathematically, a stochastic process $\xi(t)$ is stationary in the time domain if its arbitrary finite-dimensional distribution function $f_n(\cdot)$ is independent of the time starting point (the subscript n denotes the dimension); i.e., for any positive integer n and any real Δ , the distribution function $f_n(\cdot)$ must meet the following equation:

$$\begin{aligned} f_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) \\ = f_n(x_1, x_2, \dots, x_n; t_1 + \Delta, t_2 + \Delta, \dots, t_n + \Delta). \end{aligned} \quad (3)$$

Strictly, for a second order stationary stochastic process, its mathematical expectation is not related to time t and its autocorrelation function is only related to time interval τ [33, 34]. They would be

$$\begin{cases} E[\xi(t)] = \int_{-\infty}^{\infty} x f_1(x, t) dx = a, \\ R(t_1, t_2) = E[\xi(t_1)\xi(t_1 + \tau)] \\ = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 f_2(x_1, x_2; \tau) dx_1 dx_2 = R(\tau), \end{cases} \quad (4)$$

where $E[\xi(\cdot)]$ is the mathematical expectation function, a is a constant, and $R(\cdot)$ is the autocorrelation function. Through extending the definition to the image, by denoting the random variable position of the pixel as s and the pixel value as v , the image can be regarded as the spatial domain stochastic process of position variable s . Then, its definition and numerical characteristics are

$$\begin{aligned} f_n(v_1, v_2, \dots, v_n; s_1, s_2, \dots, s_n) \\ = f_n(v_1, v_2, \dots, v_n; s_1 + \Delta, s_2 + \Delta, \dots, s_n + \Delta), \end{aligned} \quad (5)$$

$$\begin{cases} E[\xi(s)] = \int_0^{\infty} v f_1(v, s) dv = a, \\ R(s_1, s_2) = E[\xi(s_1)\xi(s_1 + \tau)] \\ = \int_0^{\infty} \int_0^{\infty} v_1 v_2 f_2(v_1, v_2; \tau) dv_1 dv_2 = R(\tau). \end{cases} \quad (6)$$

It can be seen from (6) that the autocorrelation coefficient of the stationary image is only related to the inspection interval τ ; that is, the shorter the distance from the current point is, the greater the impact on the current point is, and vice versa. Reflected in the autocorrelation curve, it means that the more stationary the image is, the smoother the

autocorrelation curve is and the faster it decays to zero, while the more nonstationary the image is, the rougher the autocorrelation curve is and the slower it decays [33–35].

To intuitively illustrate the statistical characteristics of an image, several different areas were cropped for analysis from the image in Figure 2(a), which is composed of three different regions: the dark sky region **A**, the houses and trees on the left constituting area region **B**, and the relatively large house and its shadow on the right area region **C**. Eight local blocks (**a–h**) were selected. Local blocks **a** and **b** belong to region **A**, **c** and **d** belong to region **B**, **e** and **f** belong to region **C**, **g** is at the boundary between **A** and **B**, and **h** is at the boundary between **B** and **C**. Their expectations are shown in Table 1.

It can be seen from Table 1 and Figure 2(b) that the statistical characteristics of **A**, **B**, and **C** are different: First, the mean values of two different local blocks in the same region are roughly the same, such as blocks **a** and **b** in region **A**, and **c** and **d** in region **B**, while the mean value of each local block in two different regions is quite different. Second, the autocorrelation function curves of local blocks in the three different regions **A**, **B**, and **C** are evidently different. The autocorrelation function curves of different local blocks in the same region are similar, and all of them decay quickly to a near-zero value, e.g., the pairs **a** and **b**, **c** and **d**, **e** and **f**. With the increase of the interval, their autocorrelations all tend to be below 0.2 or decay to zero quickly. However, the autocorrelation function curves of the local blocks at the junction of two regions decay much more slowly, for example, the blocks **g** and **h**. This indicates that the stationarity in the same region can be maintained, while the stationarity at the boundary of different regions may be broken; i.e., the area can be considered to be approximately stationary in the same region, while it is not stationary at the junction area of different regions.

According to the aforementioned analysis, it can be found that the statistical characteristics of different regions are different and those of the same region are similar and stationary. Therefore, segmenting images into several clusters according to statistical characteristics will effectively reduce the nonstationarity of the background and reconstruct a low-rank background.

3.2. Low-Rank Background Modeling Based on Statistical Region. It can be seen from Table 1 and Figure 2 that the mathematical expectation and autocorrelation of one statistical region satisfies (6), so the region is stationary. Therefore, to make good use of the stationarity, effectively

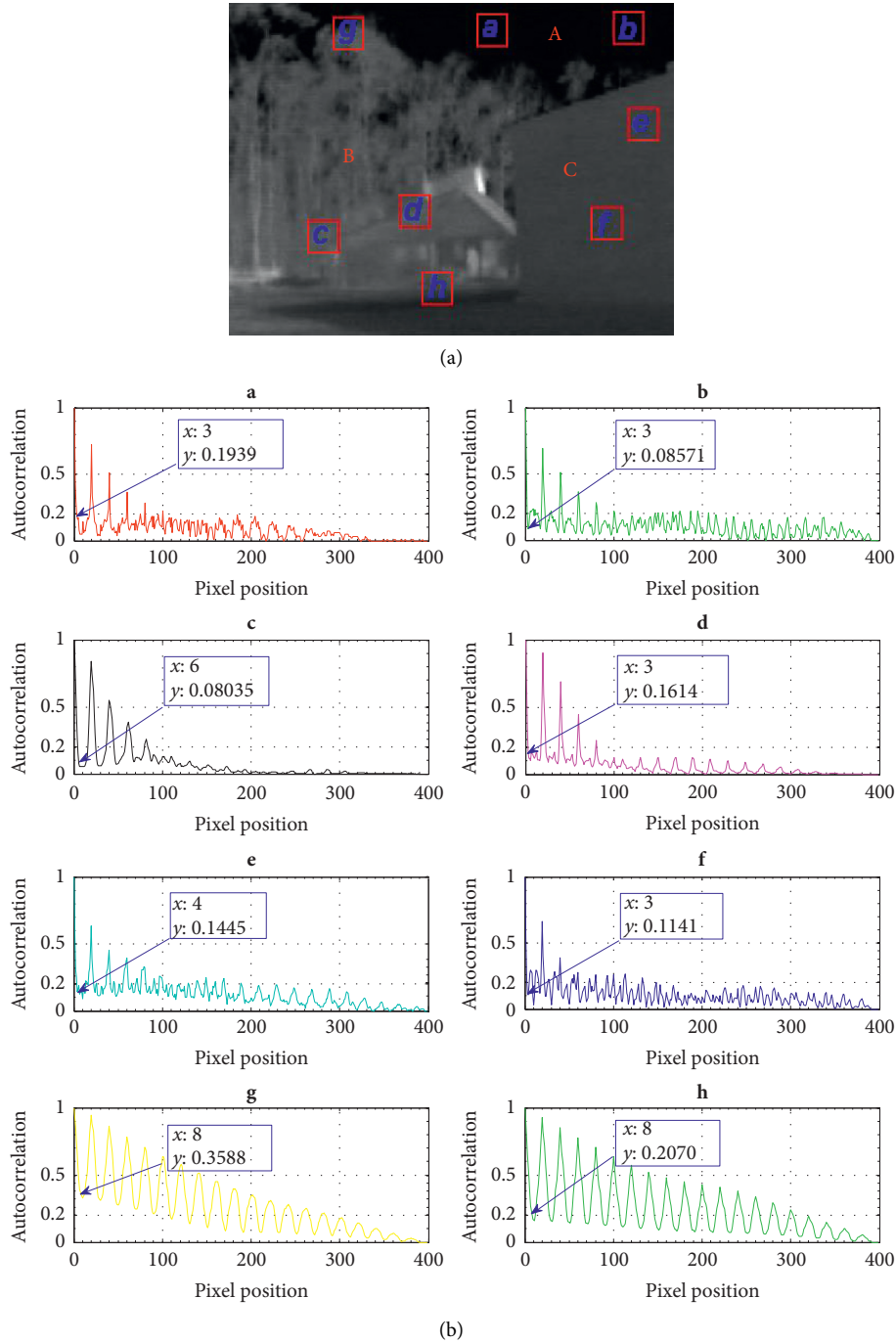


FIGURE 2: (a) Local blocks **a** to **h**; (b) autocorrelation function curves of local blocks **a** to **h**.

TABLE 1: Expectations of local blocks **a** to **h**.

	a	b	c	d	e	f	g	h
Mathematical expectation	3.55	3.75	94.71	91.76	68.55	65.51	70.62	68.26

reduce nonstationarity, and obtain low-rank component, this paper proposes a statistical region low-rank background modeling method by clustering the image regions according to their statistics. This method does not use sliding windows

to capture image blocks and only processes single-frame images. It not only effectively reduces nonstationarity, but also effectively avoids the problem of aliasing caused by image occlusion in front of and behind sliding windows in

the second type of low-rank background modeling algorithms mentioned above. In addition, it can be applied to situations where the background structure is more complex.

3.2.1. Statistical Clustering. To cluster the image into several regions so that each of them is statistically consistent, a k -means statistical clustering algorithm based on the statistics of an image region was adopted.

The contours of different regions in the background image are random, the gray levels of different regions are different, and the gray levels within the same region are continuously distributed. Therefore, compared with spatial attributes, the gray level is more effective for describing the statistical characteristics of background pixels. Thus, in the k -means clustering algorithm, the gray levels of all the pixels in the image are taken as the dataset $\mathbf{Q} = \{q_1, q_2, \dots, q_n\}$. These gray levels are divided into k clustering subsets $\mathbf{S} = \{S_1, S_2, \dots, S_k\}$. k, n are positive integers, and $k \leq n$. Then, the similarity of the gray levels is measured using the gray level distance, as shown in (7). The sum of squared error is used as a criterion to evaluate the clustering performance; that is, when (8) is satisfied, the optimal clustering result is considered to be obtained:

$$d(q_i, q_j) = \sqrt{(q_i - q_j)^2}, \quad (7)$$

$$\arg \min_{\mathbf{S}} \sum_{i=1}^k \sum_{q_e \in S_i} \|q_e - \mu_i\|^2, \quad (8)$$

$$\mu_i = \frac{1}{M_i} \sum_{e=1}^{M_i} q_e, \quad (9)$$

where $d(q_i, q_j)$ is the gray level distance, q_i, q_j are any two gray values of \mathbf{Q} . μ_i is the mean of the clustering subset S_i . M_i is the total number of gray-scale levels of S_i , and $q_e \in S_i$. When (8) is satisfied, the pixels with gray levels belonging to S_i in the image constitute the i th clustering region K_i , $1 \leq i \leq k$.

With image region clustering, the image matrices of the statistical region can be built as in Figure 3. First, clusters in an image are extracted and denoted as K_i , namely, K_1, K_2, \dots, K_k , where $i = 1, 2, \dots, k$. Second, cluster regions are converted into column vectors separately as \mathbf{C}_i , namely, $\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_k$. Third, each column vector is divided into m equal segments as \mathbf{C}_{ig} , namely, $\mathbf{C}_{i1}, \mathbf{C}_{i2}, \dots, \mathbf{C}_{ig}, \dots, \mathbf{C}_{im}$, where $g = 1, 2, \dots, m-1, m$, and then each segment is written into a column in a new matrix. Note that, in this step, the number m of each column vector \mathbf{C}_i could be different depending on the actual situation. Through these steps, an original video frame image can be converted into k statistical region image matrices according to its statistical characteristics, denoted by \mathbf{F} , namely, $\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_k$.

Figure 4 shows the analysis of the logarithmic value of singular-value gradient of the background matrices \mathbf{B} of the statistical region image matrices \mathbf{F} . The third row in Figure 4 shows the logarithmic value curves of singular-value

gradient of different \mathbf{B} , which rapidly decrease, and the gap between the second and the first gradient value is the most significant. This means that the second singular value and those after it are much smaller than that of the first one. Thus, \mathbf{B} can be considered to be a low-rank matrix, which satisfies the following equation:

$$\text{Rank}(\mathbf{B}) \leq r, \quad (10)$$

where $\text{Rank}(\cdot)$ represents the rank of a matrix and r is a positive integer.

As the scale of the small targets is within the range of 9×9 pixels and the noise is distributed randomly and is of a very small scale, the two together make up a very small proportion of \mathbf{F} . Therefore, the noise and the targets together constitute the sparse matrices \mathbf{P} , which satisfy the following equation:

$$\|\mathbf{P}\|_0 \leq z, \quad (11)$$

where $\|\mathbf{P}\|_0$ represents the l_0 norm, which is the number of nonzero elements. z is a positive integer.

Thus, the statistical region image matrix can be decomposed into the low-rank part and sparse part as

$$\begin{aligned} & \underset{\mathbf{B}, \mathbf{P}}{\text{minimize}} \text{Rank}(\mathbf{F}) \\ & \text{subject to} \quad \mathbf{F} = \mathbf{B} + \mathbf{P}, \\ & \quad \text{Rank}(\mathbf{B}) \leq r, \|\mathbf{P}\|_0 \leq z, \end{aligned} \quad (12)$$

where \mathbf{F} represents the statistical region image matrix (SRIM), \mathbf{B} is the low-rank matrix of SRIM, and \mathbf{P} is the sparse matrix of SRIM.

3.2.2. Background Estimation. According to the property analysis in Section 2.2.1, (12), the statistical region image matrix model is a typical PCA model, so the existing robust PCA (RPCA) [16–18, 22, 28] can be used to obtain its optimal solution.

According to [17] and [28], the low-rank and sparse decomposition model of \mathbf{F} can be transformed into the principal component tracking algorithm (PCP) for the optimal solution:

$$\begin{aligned} & \text{minimize} \|\mathbf{B}\|_* + \lambda \|\mathbf{P}\|_1 \\ & \text{subject to} \quad \mathbf{F} = \mathbf{B} + \mathbf{P}, \end{aligned} \quad (13)$$

where $0 < \lambda < 1$ is a weighting parameter to balance the two terms. When the sparsity of \mathbf{P} increases, we will get more suitable results with the increase of λ . Both mathematical theoretical analysis and practical experiments show that appropriate results can be obtained when $\lambda = 1/\sqrt{\max(o, q)}$, where o and q are the scales of the matrices $\mathbf{F} \in \mathbf{R}^{o \times q}$ [17]. $\|\cdot\|_*$ denotes the kernel norm of the matrix, which is the sum of singular values of the matrix; i.e., $\|\mathbf{M}\|_* := \sum_i \sigma_i(\|\mathbf{M}\|)$. $\|\cdot\|_1$ represents the l_1 norm of a matrix; i.e., $\|\mathbf{M}\|_1 = \sum_{ij} |\mathbf{M}_{ij}|$.

To solve the convex optimization problem of (13), the augmented Lagrange multiplier (ALM) algorithm is applied [31, 36], which needs fewer iterations to obtain better results

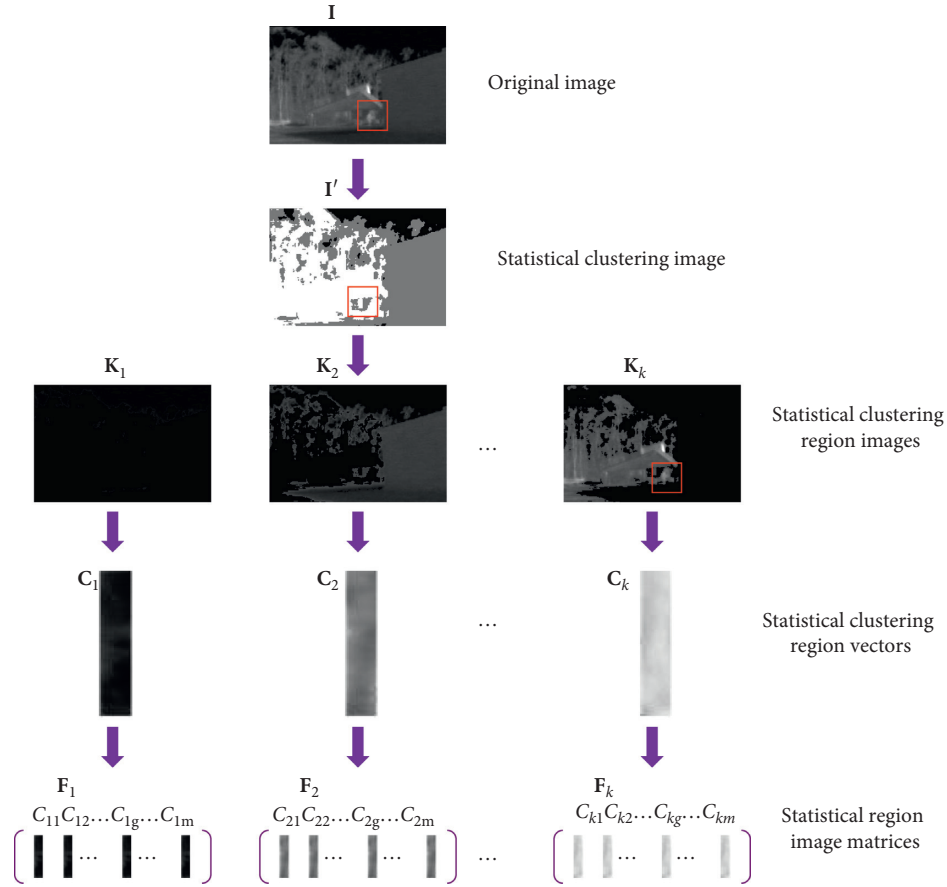


FIGURE 3: Schematic diagram of acquisition of statistical region image matrices.

than other algorithms. Using the ALM, (13) can be rewritten as

$$l(B, P, Y) = \|B\|_* + \lambda \|P\|_1 + \langle Y, F - B - P \rangle + \frac{\mu}{2} \|F - B - P\|_F^2, \quad (14)$$

where $\mu > 0$ is a penalty parameter to penalize the violation of the linear constraint, and it also affects the convergence rate of iterative operations. Theoretically, the larger the μ is, the faster the algorithm converges. However, in practice, this kind of situation should be avoided as extremely large μ leads to numerical difficulty empirically. Both mathematical theoretical analysis and practical experiments show that appropriate results can be obtained when $\mu = o * q/4 \|F\|_1$ [17, 36]. To solve (14), by iteratively minimizing the ALM expression and updating the Lagrange multiplier Y , the optimal solution can be achieved as follows:

$$\begin{aligned} (B_k, P_k) &= \arg \min_{B, P} L(B, P, Y_k), \\ Y_{k+1} &= Y_k + \mu (F - B_k - P_k). \end{aligned} \quad (15)$$

To prevent repeatedly carrying out iterative convex optimization operations, the soft-threshold operator $S_\eta[x]$ [31] and the singular-value-threshold operator $H_\eta(X)$ [32] are applied. They are defined as follows:

$$\begin{aligned} S_\eta[x] &= \operatorname{sgn}(x) \max(|x| - \eta, 0), \\ H_\eta(X) &= US_\eta(\Sigma)V^*. \end{aligned} \quad (16)$$

In (16), $x \in \mathbf{R}$, $\eta > 0$, and $X = U\Sigma V^*$ represents any singular-value decomposition of the matrix X . For two matrices of the same scale $A = (a_{ij})$ and $B = (b_{ij})$, $S_A[B]$ is the new matrix W that is obtained by iterating throughout the corresponding elements; that is, $W = S_A[B] = (S_{a_{ij}}[b_{ij}])$. Thus, the optimal solutions for B and P are obtained concisely and efficiently.

$$\begin{aligned} \arg \min_B l(B, P, Y) &= H_{1/\mu}(F - P + \mu^{-1}Y), \\ \arg \min_P l(B, P, Y) &= S_{\lambda/\mu}(F - B + \mu^{-1}Y). \end{aligned} \quad (17)$$

In the actual solution process, a concise and efficient alternative optimization operation technique [17] is adopted, which first minimizes B in the case of fixed P , then optimizes P in the case of fixed B , and then updates Y .

In the SCPLBMA, the most critical step is obtaining the optimal solution B . Therefore, as presented in Figure 5, first, the original video frame images I are statistically clustered by carrying out (7)–(9), and then each statistical cluster is extracted to obtain the statistical region image matrix F . Second, by solving (14) to obtain the optimal solution B , B_1, B_2, \dots, B_k are obtained in this step. Thirdly,

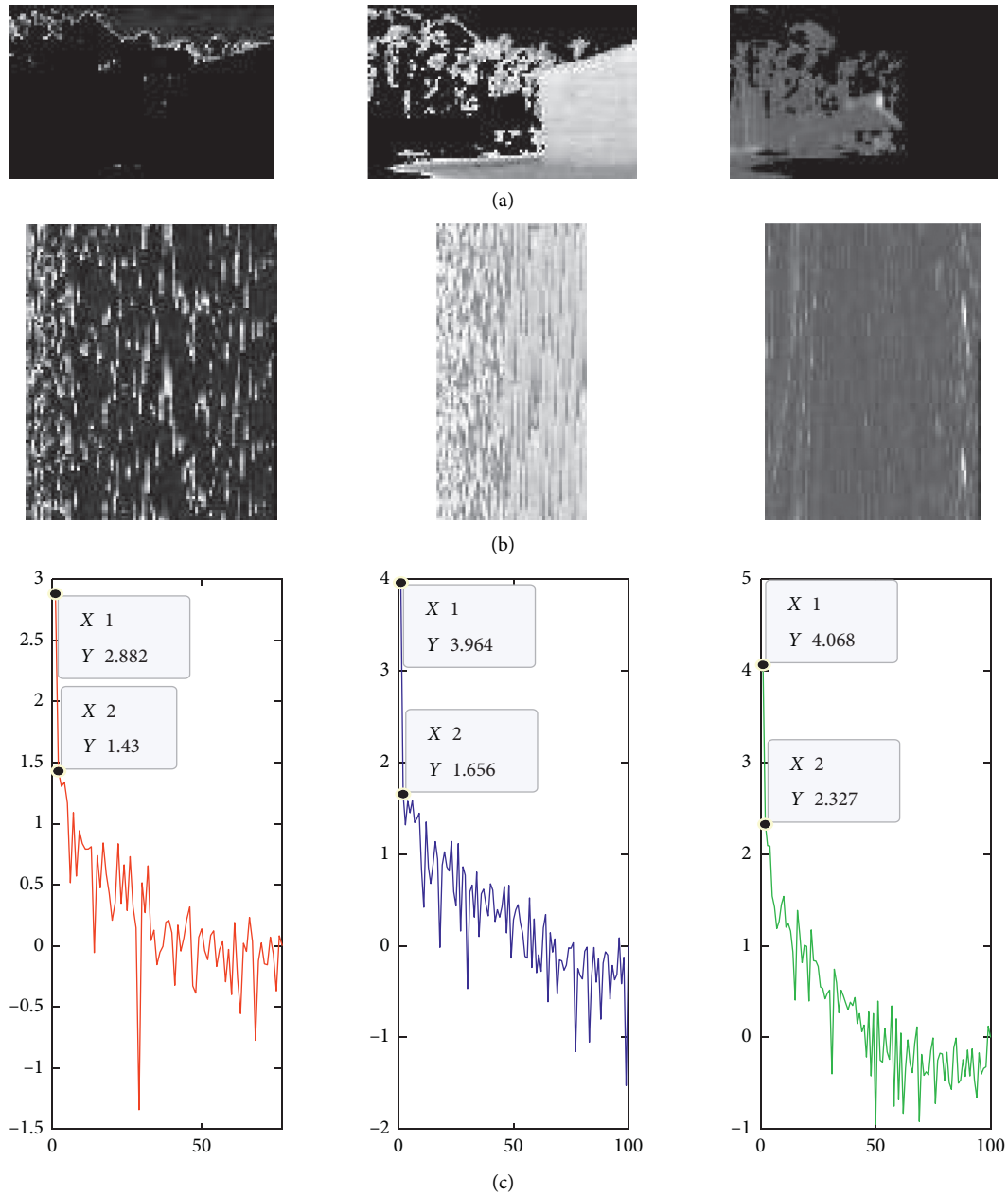


FIGURE 4: Logarithmic value of singular-value gradient of the background matrices \mathbf{B} : (a) the statistical clustering region images; (b) the background matrices of the corresponding statistical region image matrices \mathbf{F} ; (c) the logarithmic value curves of singular-value gradient of the corresponding background matrices \mathbf{B} .

$\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_k$ are used to reconstruct the low-rank background \mathbf{L} of the original video image. Sparse image, that is, difference images \mathbf{S} , can be obtained by subtracting \mathbf{L} from \mathbf{I} .

For scenes A, B, and C used in the next section, the processing time of SCPLBMA and the other 6 algorithms is shown in Table 2. The experimental environment was the same as that in the next section; that is, it was conducted with OCTAVE on a computer running a 32 bit Windows 7 operating system, with a Core-i5 CPU and 3 GB of RAM. Here is a brief analysis of the computational complexity of the SCPLBMA. From the aforementioned analysis of

Algorithm 1, it can be seen that the computation time is mainly composed of k -means statistical clustering operation in the first step, calculating the optimal solutions for \mathbf{B} and \mathbf{P} in the second step, and background reconstruction in the third step. The computational complexity of k -means statistical clustering operation is mainly determined by the total amount of sample data N , the clustering number k , and the number of iterations i ; therefore the computational complexity is around $O(Nki)$ [37] time. Since $N \gg k, N \gg i$, it is approximately $O(N)$. The computational complexity of calculating the optimal solutions for \mathbf{B} and \mathbf{P} is mainly determined by SVD operation and soft-threshold operation.

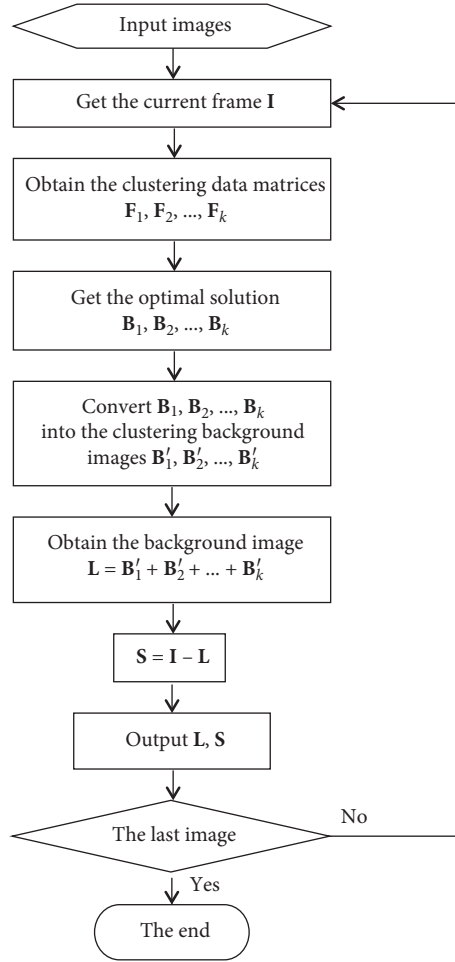


FIGURE 5: The flow diagram of the SCPLBMA.

TABLE 2: The time consumption of SCPLBMA and the other 6 typical algorithms for scenes A, B, and C (the time unit is second).

Scenes	Total frames	Image size	Algorithms	Total time	Average time per frame
A	121	250 * 180	SCPLBMA	51.8181	0.4282
			RPCA	76.3954	0.6314
			GRASTA	2.4909	0.0206
			OSTD	7.7134	0.0637
			THAT	3.1254	0.0258
			Med_FA	2.3198	0.0192
			SGA	2.7475	0.0227
B	361	250 * 180	SCPLBMA	180.8536	0.5010
			RPCA	196.6818	0.5448
			GRASTA	4.7943	0.0133
			OSTD	11.7128	0.0324
			THAT	10.0063	0.0277
			Med_FA	7.4285	0.0206
			SGA	7.9015	0.0219
C	191	250 * 180	SCPLBMA	55.5143	0.2907
			RPCA	95.0086	0.4974
			GRASTA	3.3085	0.0173
			OSTD	6.9870	0.0366
			THAT	7.2066	0.0377
			Med_FA	3.5940	0.0188
			SGA	4.2237	0.0221

```

(1) Input: Original video frame images  $I \in \mathbf{R}^{m \times n}$ . The criterion of convergence was  $\|\mathbf{F} - \mathbf{B} - \mathbf{P}\|_F \leq \varphi \|\mathbf{F}\|_F$ , where  $\varphi = 10^{-7}$  [17].
(2)  $I$  are clustered and segmented to get the statistical region image matrices  $\mathbf{F}$ .
    while not converged do
        compute  $d(q_i, q_j) = \sqrt{(q_i - q_j)^2}$ 
         $\arg \min_S \sum_{i=1}^k \sum_{q_e \in S_i} \|q_e - \mu_i\|^2$ 
         $\mu_i = (1/M_i) \sum_{e=1}^{M_i} q_e$ 
    end while
(3) //Minimize the Lagrange function in equation (14)
    while not converged do
        compute  $\mathbf{B}_{k+1} = H_{1/\mu}(\mathbf{F} - \mathbf{P}_k + \mu^{-1}\mathbf{Y}_k)$ ;
        compute  $\mathbf{P}_{k+1} = S_{\lambda/\mu}(\mathbf{F} - \mathbf{B}_{k+1} + \mu^{-1}\mathbf{Y}_k)$ ;
        compute  $\mathbf{Y}_{k+1} = \mathbf{Y}_k + \mu(\mathbf{F} - \mathbf{B}_{k+1} - \mathbf{P}_{k+1})$ ;
    end while
(4)  $\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_k$  are superimposed to reconstruct  $\mathbf{L}$ .
(5) Let  $\mathbf{S} = \mathbf{I} - \mathbf{L}$ , the sparse images  $\mathbf{S}$  are obtained.
(6) Output  $\mathbf{S}, \mathbf{L}$ 

```

ALGORITHM 1: Statistical clustering partitioning low-rank background modeling algorithm.

Different SVD method has different computational complexity. The use of fast SVD technique can make the calculation be implemented in $O(\text{rop})$ time [38, 39]. Here, r is the number of nonzero singular values, and o, p are the size of the matrix to be processed. The computational complexity of soft-threshold operation is around $O(op \log(op))$ [31]. Then, the computational complexity of calculating the optimal solutions for \mathbf{B} and \mathbf{P} is around $O(i_2 \text{rop} \log(op))$, where i_2 is the iteration number. Image reconstruction can be realized by directly writing the data into a matrix with the same size as the original video frame image; thus, the computational complexity is approximately $O(mn)$. Therefore, the entire computational complexity of SCPLBMA is approximately $O(N + i_2 \text{rop} \log(op) + mn)$.

4. Experiments

In this section, three representative scenes of infrared small target detection, i.e., ground scene A, deep space scene B, and forest scene C, were used to verify the performance of the SCPLBMA. First, the background reconstruction experiment was carried out. Seven background modeling algorithms, that is, the SCPLBMA, Principal Component Pursuit (RPCA) [17], Grassmannian Robust Adaptive Subspace Tracking Algorithm (GRASTA) [40], Online Stochastic Tensor Decomposition (OSTD) [41], top-hat transformation algorithm (THTA), median filtering algorithm (Med_FA), and single Gaussian algorithm (SGA), were compared and evaluated. Then, to further verify the performance of the SCPLBMA, targets were extracted by threshold segmentation method from the single difference image, and then the detection performance of the seven algorithms was given.

4.1. Background Reconstruction Experiment. The experiment was conducted with OCTAVE on a computer running a 32 bit Windows 7 operating system, with a Core-i5 CPU and 3 GB of RAM. The main parameters of GRASTA were the

estimated rank = 1, ADMM constant step = 1.8 [40, 42, 43]. For OSTD, the tradeoff parameters were $\lambda_1 = 1/\sqrt{\text{max size (tensor data)}}$ and $\lambda_2 = 10^* \lambda_1$ [41–43]. For THAT and Med_FA, the structure element scales were $7 \times 7 \sim 9 \times 9$. The variance value of SGA was determined by the variance of background region; here, it was $7 \times 7 \sim 9 \times 9$.

The SCPLBMA was compared with the RPCA, GRASTA, OSTD, THTA, Med_FA, and SGA. The autocorrelation coefficient, contrast, and contrast mean were used as evaluation indexes. The definition for the autocorrelation coefficient is shown in (6) in Section 3.1. The contrast and contrast mean indexes are defined as the following equations:

$$\text{Cont} = \frac{\text{Target}_{\text{mean}}}{\text{Image}_{\text{mean}}}, \quad (18)$$

$$\text{Cont}_M = \frac{\sum_{i=1}^{\text{Tol}} \text{Cont}}{\text{Tol}}. \quad (19)$$

In (18), Cont is the contrast between the target and \mathbf{S} (difference image) of the i th frame, $\text{Target}_{\text{mean}}$ is the gray-scale mean of the target, and $\text{Image}_{\text{mean}}$ is the gray-scale mean of the difference image. In (19), Cont_M is the contrast mean for all frames of a scene, and Tol is the total frames of a scene. There were 121, 361, and 191 frames in scenes A, B, and C, respectively. Two frames of each scene were randomly selected to illustrate the comparative analysis. Experimental data are shown in Tables 3–7 and Figures 6–11. Tables 3–5 present the contrast values of the 2 randomly selected frames of each scene. Table 6 shows the contrast mean data of each scene. Table 7 shows the gray-scale mean of the difference image, i.e., the noise floor's means of scenes A, B, and C. In each figure of Figures 6–11, the first row has only one image, which is the original video frame image; the second to eighth rows correspond to the SCPLBMA, RPCA, GRASTA, OSTD, THTA, Med_FA, and SGA algorithms, respectively; in the second to eighth rows, the first column contains the difference images, the second

TABLE 3: Contrasts between targets and residual backgrounds in frames 30 and 93 of scene A.

Algorithms	Cont for frame 30	Cont for frame 93
SCPLBMA	7.7505	9.6276
RPCA	3.5215	2.0216
GRASTA	—	1.0022
OSTD	0.9692	1.1965
THAT	3.4633	4.0770
Med_FA	3.7481	2.5104
SGA	3.1043	3.1712

TABLE 4: Contrasts between targets and residual backgrounds in frames 33 and 260 of scene B.

Algorithms	Cont for frame 33	Cont for frame 260
SCPLBMA	6.9340	4.0207
RPCA	1.8826	1.5602
GRASTA	—	—
OSTD	1.4033	2.1592
THAT	3.1623	2.6149
Med_FA	4.3640	2.2522
SGA	4.4648	2.7689

TABLE 5: Contrasts between targets and residual backgrounds in frames 10 and 129 of scene C.

Algorithms	Cont for frame 10	Cont for frame 129
SCPLBMA	7.6233	7.4859
RPCA	1.6091	1.6677
GRASTA	—	—
OSTD	—	1.5018
THAT	0.8779	0.8439
Med_FA	0.8338	0.9146
SGA	0.7225	0.8041

TABLE 6: Contrast means between targets and residual backgrounds of scenes A, B, and C.

Algorithms	Cont_M of scene A	Cont_M of scene B	Cont_M of scene C
SCPLBMA	10.4446	4.7972	6.4491
RPCA	2.9313	1.6472	1.5504
GRASTA	1.0231	1.1088	1.0286
OSTD	1.1071	2.6224	1.2581
THAT	4.0374	2.6125	0.8058
Med_FA	3.1405	2.9945	0.8140
SGA	3.5058	3.0653	0.7367

TABLE 7: The gray-scale means of the difference images ($\text{Image}_{\text{mean}}$), i.e., the noise floor's means of scenes A, B, and C.

Algorithms	$\text{Image}_{\text{mean}}$ of scene A	$\text{Image}_{\text{mean}}$ of scene B	$\text{Image}_{\text{mean}}$ of scene C
SCPLBMA	8.9887	14.2220	11.8174
RPCA	46.7715	75.5719	84.6576
GRASTA	125.0220	118.4166	122.3634
OSTD	131.2614	38.4685	100.4039
THAT	9.1105	4.6739	28.7464
Med_FA	4.7316	3.0439	15.2762
SGA	6.9862	3.6145	21.6499

column contains the autocorrelation curves of the difference images, and the third column contains the three-dimensional energy distribution of the difference images.

Experimental results show that these seven algorithms could effectively suppress the background, but the effect of each algorithm is different. In the following, four aspects will

be discussed: residual texture, residual noise, integrity of target signal, and contrast between target and residual background. For scenes A, B, and C, except for SCPLBMA, RPCA, GRASTA, and OSTD algorithms, the difference images that were obtained from the THAT, Med_FA, and SGA algorithms all leave a large number of textures and fixed

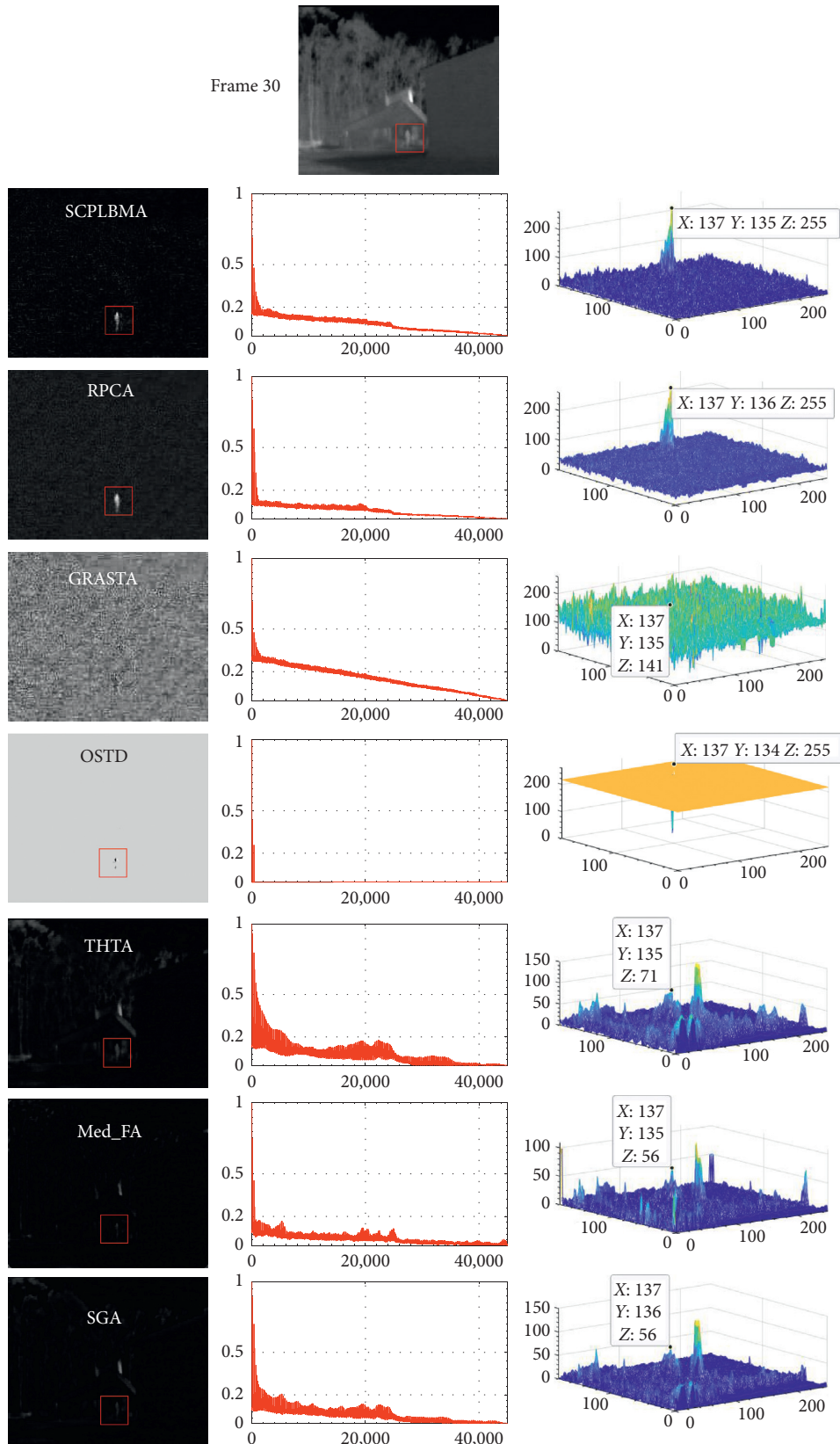


FIGURE 6: Background suppression evaluation for frame 30 of scene A.

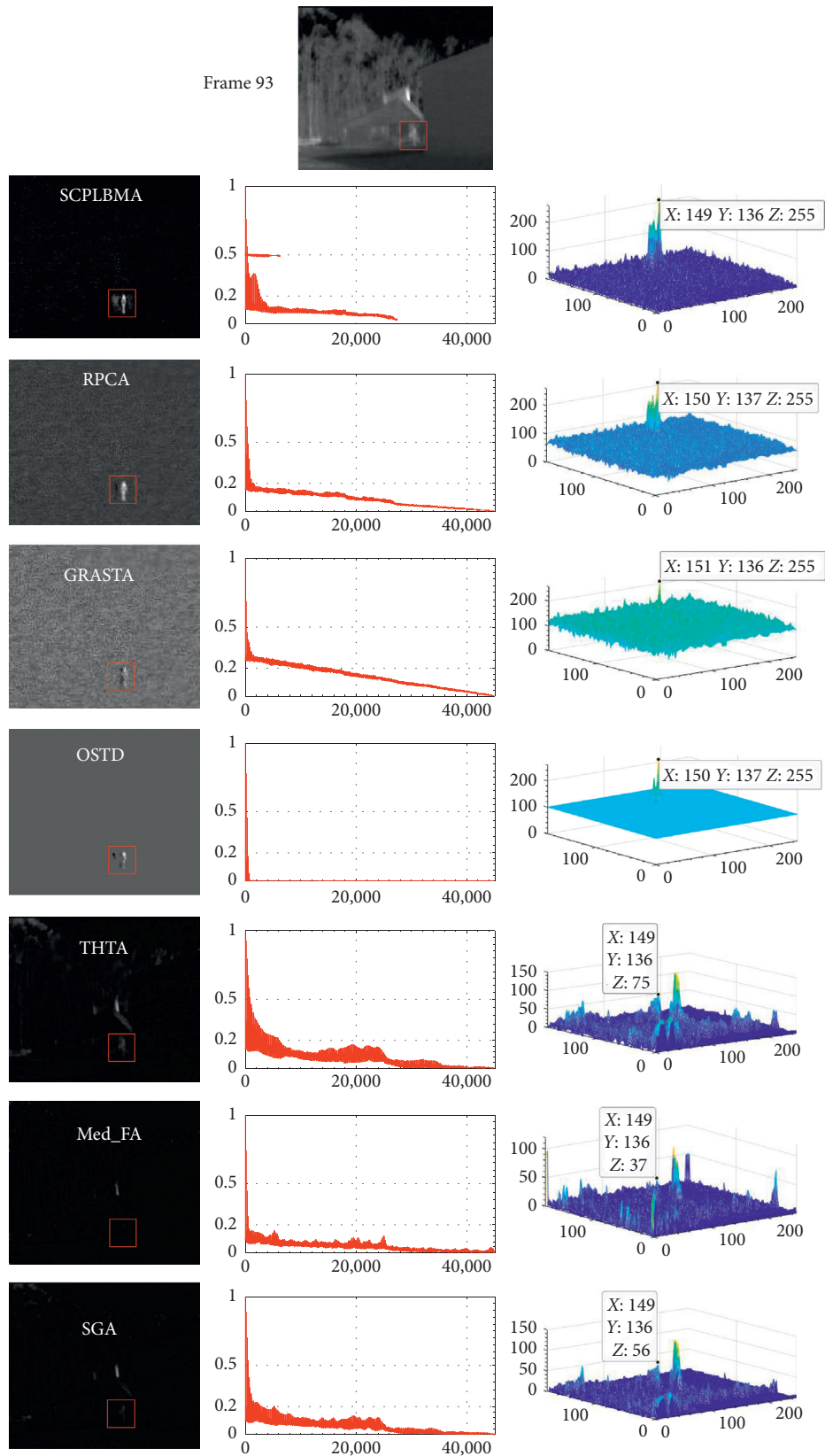


FIGURE 7: Background suppression evaluation for frame 93 of scene A.

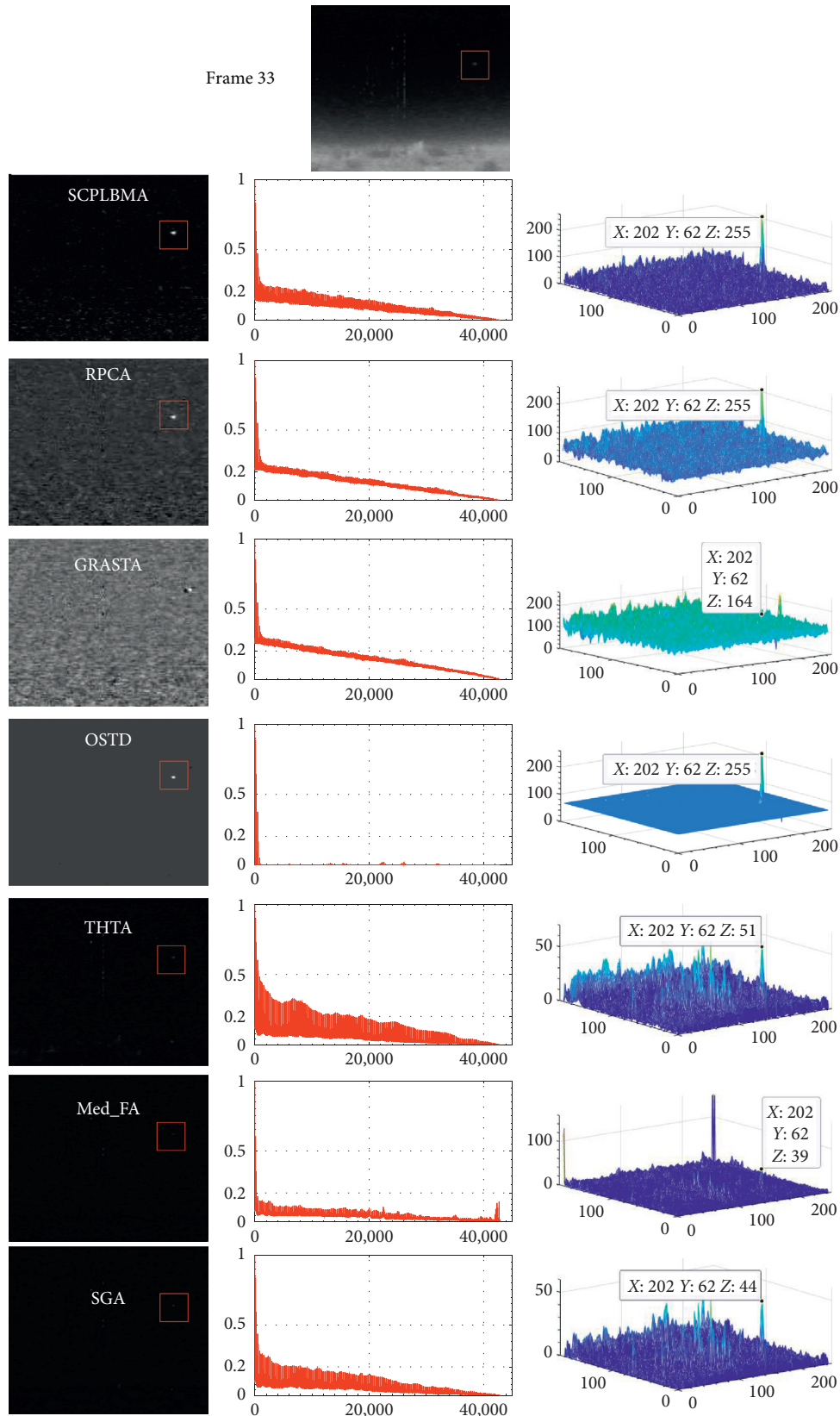


FIGURE 8: Background suppression evaluation for frame 33 of scene B.

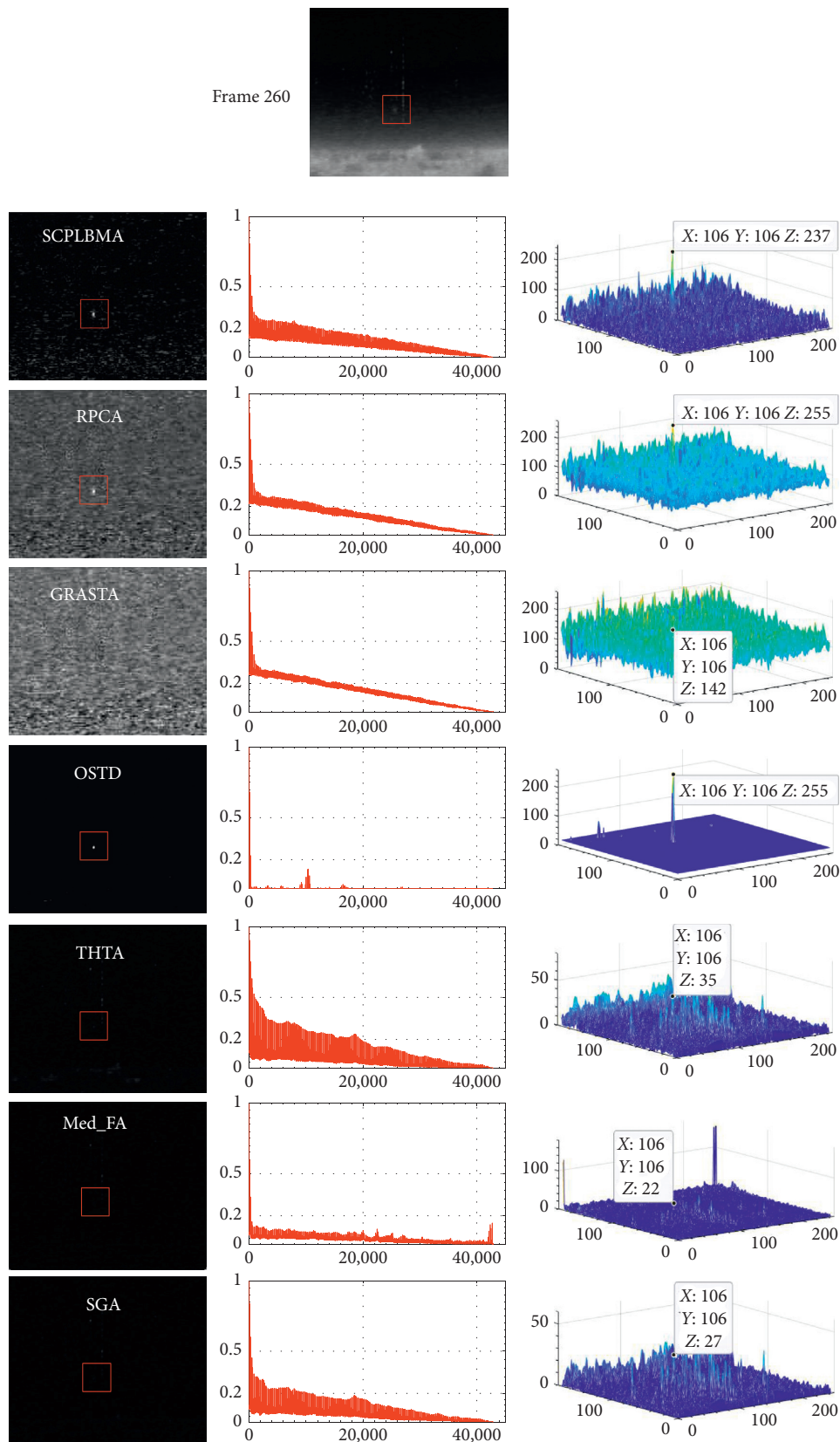


FIGURE 9: Background suppression evaluation for frame 260 of scene B.

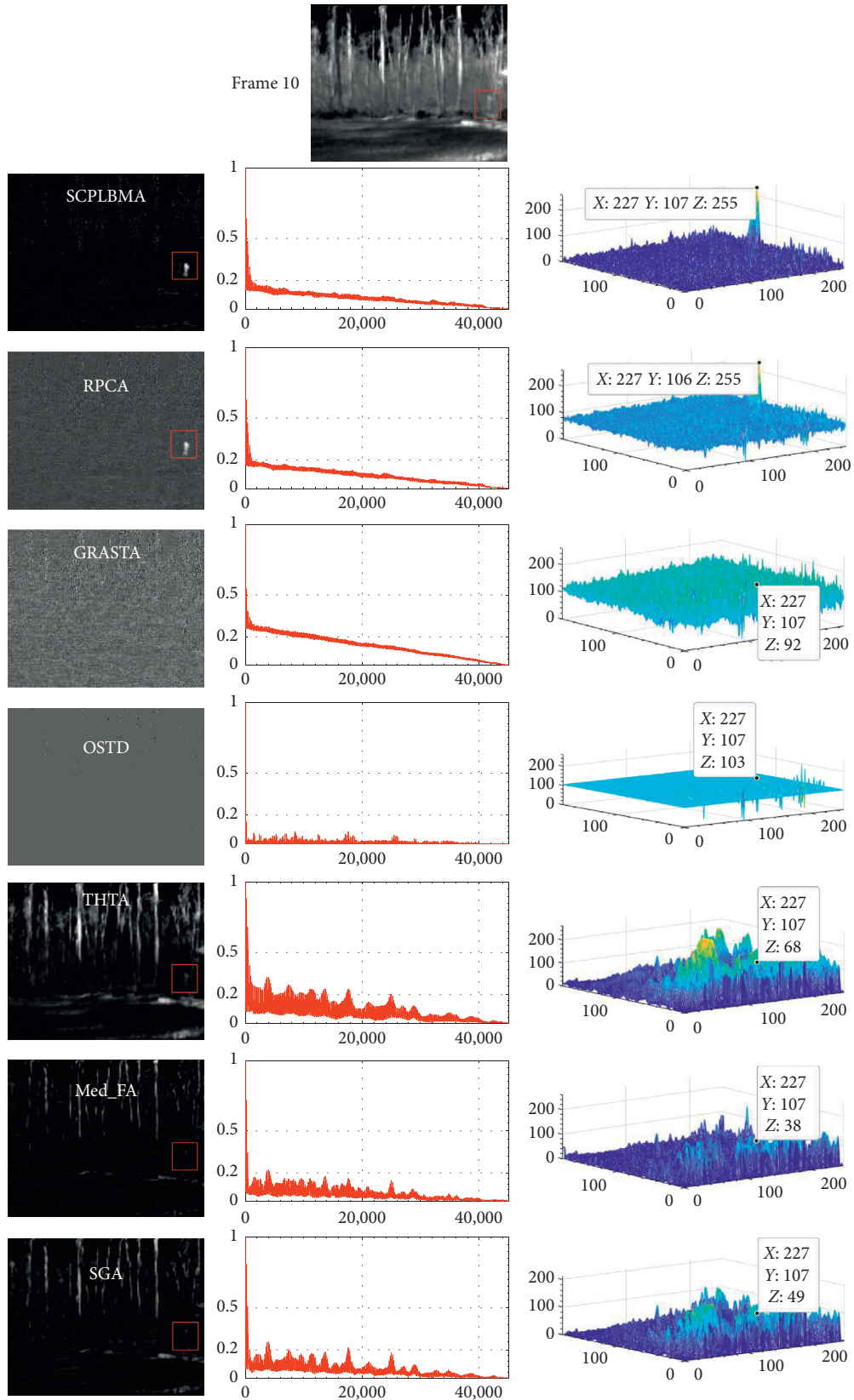


FIGURE 10: Background suppression evaluation for frame 10 of scene C.

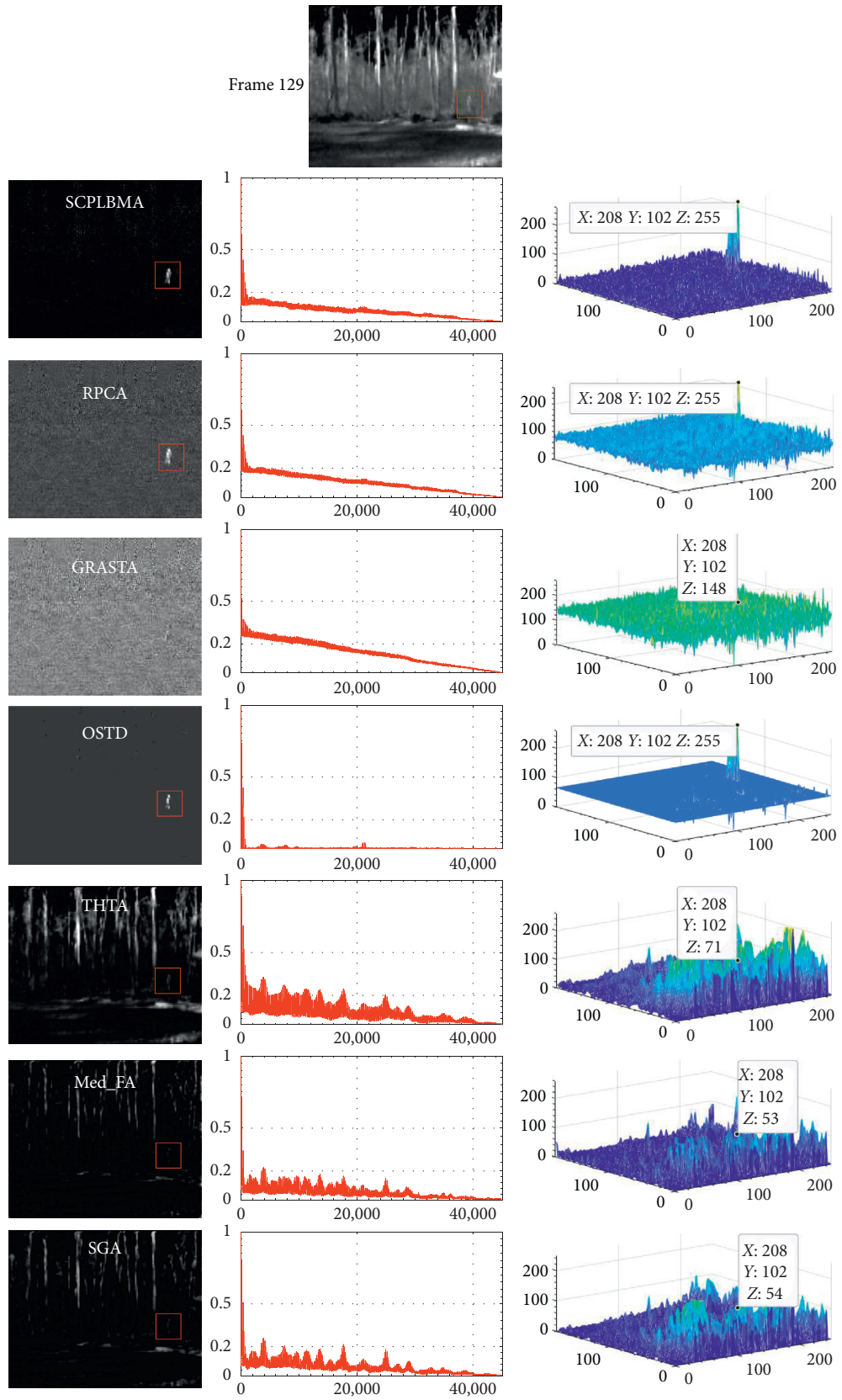


FIGURE 11: Background suppression evaluation for frame 129 of scene C.

noise. Therefore, the stationarity of the difference images obtained by SCPLBMA, RPCA, GRASTA, and OSTD algorithms is better than these obtained by other three algorithms. This can also be seen from the difference images and autocorrelation coefficient curves in Figures 6–11. In these figures, with the less residual textures and the more even distribution of the noises, the corresponding autocorrelation coefficient curves can quickly smooth down to less than 0.2, indicating that the difference images are more stationary. Therefore, it can be seen that the difference image obtained by OSTD algorithm has the best stationarity, those of SCPLBMA and RPCA algorithms have similar stationarity, and those of the other four algorithms have relatively poor stationarity. The experiments show that the target signal integrity of the SCPLBMA and RPCA algorithms' difference images is better than that of the other five algorithms, which can also be intuitively seen from the difference images in Figures 6–11. GRASTA's difference images have the most residual noise, and most of the time the targets are seriously interfered with by the noises or even are submerged in the noises. The OSTD algorithm's difference images have the least residual noises, but the instability in early period leads to poor target signal integrity. For OSTD algorithm, the target signal integrity gradually gets better at the later stage. THTA, Med_FA, and SGA algorithms' difference images have poor target signal integrity due to the operation of a structure element on the background area. Regarding the aspect of contrast, experimental data in Tables 3–5 show that the SCPLBMA has the highest contrast between the target and the background. This is because the gray-scale value of target is very high and noise floor is very low. It can be seen from Table 7 and the three-dimensional energy distribution diagrams of Figures 6–11 that the noise floor of RPCA, GRASTA, and OSTD algorithms' difference images is much higher than that of SCPLBMA, THTA, Med_FA, and SGA algorithms. Therefore, for scenes A, B, and C, the contrast values that were obtained by RPCA, GRASTA, and OSTD algorithms are significantly lower than that of SCPLBMA. From the above analysis, the residual random noises in the difference images that were obtained using the SCPLBMA are evenly distributed, their energy is significantly lower than that of the targets, and there are no residual background textures in the difference images. This gives the difference images of the SCPLBMA the following characteristics: First, their stationarity is good. The autocorrelation attenuates smoothly and decreases quickly to less than 0.2. Second, the targets energy is significantly higher than that of the noises, which can be visually seen from the difference images and their three-dimensional energy diagram in Figures 6–11. Third, the contrast between the targets and the residual backgrounds is high, and targets are significantly enhanced, which can be seen in Tables 3–6. Fourth, there is no evident loss of target information, which can be visually seen in Figures 6–11.

4.2. Single-Frame Threshold Segmentation Experiment. To further verify the performance of the SCPLBMA, the single-frame threshold segmentation method was adopted for

scenes A, B, and C to extract targets from the difference images that were obtained from the SCPLBMA, RPCA, GRASTA, OSTD, THTA, Med_FA, and SGA, respectively. Values of 0.3 times, 0.4 times, and 0.6 times the maximum gray level of the difference images were used as the segmentation thresholds. One frame of each scene was randomly selected to illustrate the comparative analysis, as shown in Figures 12–14. In each figure of Figures 12–14, the first row only has one image, which is the original video frame image; the second to eighth rows correspond to the SCPLBMA, RPCA, GRASTA, OSTD, THTA, Med_FA, and SGA algorithms, respectively; in the second to eighth rows, the first column contains the difference images, and the second to fourth columns correspond to the segmentation images with segmentation thresholds of 0.3, 0.4, and 0.6 times the maximum gray level, respectively. Equations (20)–(22) are the detection rate, error detection rate, and miss detection rate of the target, respectively, which were used as evaluation indexes. The experimental data are shown in Figures 15–17:

$$\text{Det_rate} = \frac{\text{TP}}{\text{Target_all}}, \quad (20)$$

$$\text{Err_rate} = \frac{\text{FP}}{\text{Target_all}}, \quad (21)$$

$$\text{Mis_rate} = \frac{\text{FN}}{\text{Target_all}}, \quad (22)$$

where Det_rate is the target detection rate; Err_rate is the target detection error rate, that is, in cases where some background or noise is detected as the target; and Mis_rate is the target miss detection rate, that is, in cases where, due to strong clutter disturbance, the targets are declared as background or noise. TP is the total number of frames in which real targets could be declared as targets; FP is the total number of frames in which the background or noise is declared as targets; FN is the total number of frames in which real targets are not clearly distinguished from noises; and Target_all is the total number of original video frames with real targets. In particular, there was a real target in each original video frame of scenes A, B, and C.

As can be seen from Figures 12–17, for scenes A, B, and C, SCPLBMA achieved good results with high detection rate and low missed detection rate, because the gray levels of the targets are significantly higher than that of the residual backgrounds.

For scene A, the gray levels of the roofs in the middle of the difference images that were obtained from THTA, Med_FA, and SGA algorithms are significantly higher than those of the targets. Then, for the three algorithms, only few frames allowed the targets to be segmented out from the difference images under the threshold of 0.3 times the maximum gray level, as can be seen in Figures 12 and 15. Compared with the above three algorithms, the difference images of the four robust PCA algorithms SCPLBMA, RPCA, GRASTA, and OSTD did not leave obvious house and trunk textures. GRASTA's difference images have the

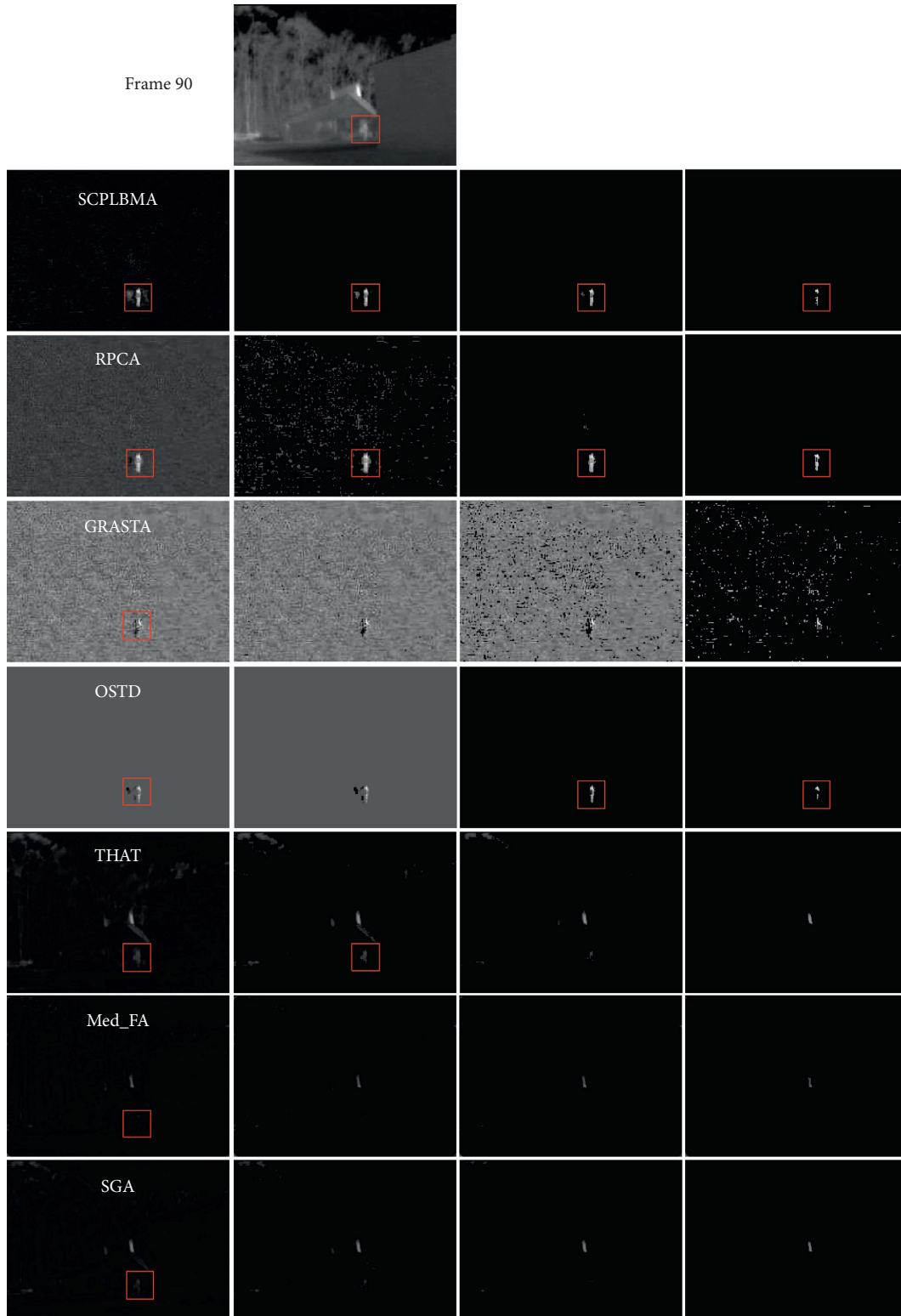


FIGURE 12: Threshold segmentation for frame 90 of scene A.

strongest random noise and the weakest target contrast. During the whole process, the targets appear steadily only after the 95th frame. Therefore, the segmentation operation with threshold values of 0.3, 0.4, and 0.6 times of the maximum gray value could not effectively segment out the

target. Although the OSTD algorithm's difference images have the least textures and noises, targets in the difference images only gradually appear after the 41st frame and stabilize only after the 87th frame. Furthermore, the noise floor of difference images is very high, which can be seen from

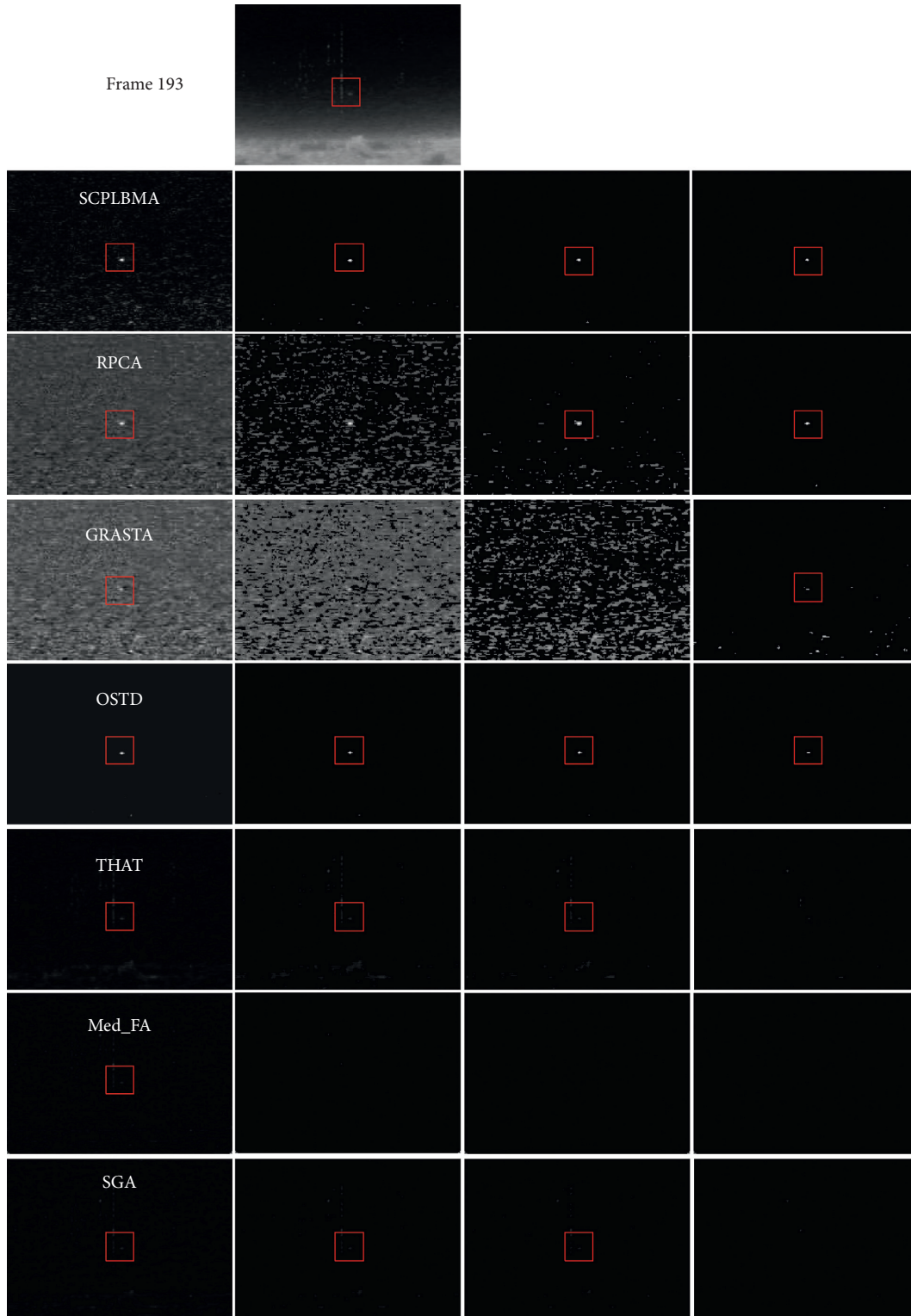


FIGURE 13: Threshold segmentation for frame 193 of scene B.

Table 7 and the three-dimensional energy distribution diagrams of Figures 6 and 7. Therefore, the threshold segmentation effect of OSTD was not good. Both SCPLBMA and RPCA achieved good results, because the residual

random noises in the difference images are significantly weaker than the targets. In Table 7, the noise floor of SCPLBMA and RPCA algorithms' difference images is, respectively, 8.9887 and 46.7715, and Figures 6 and 7 show

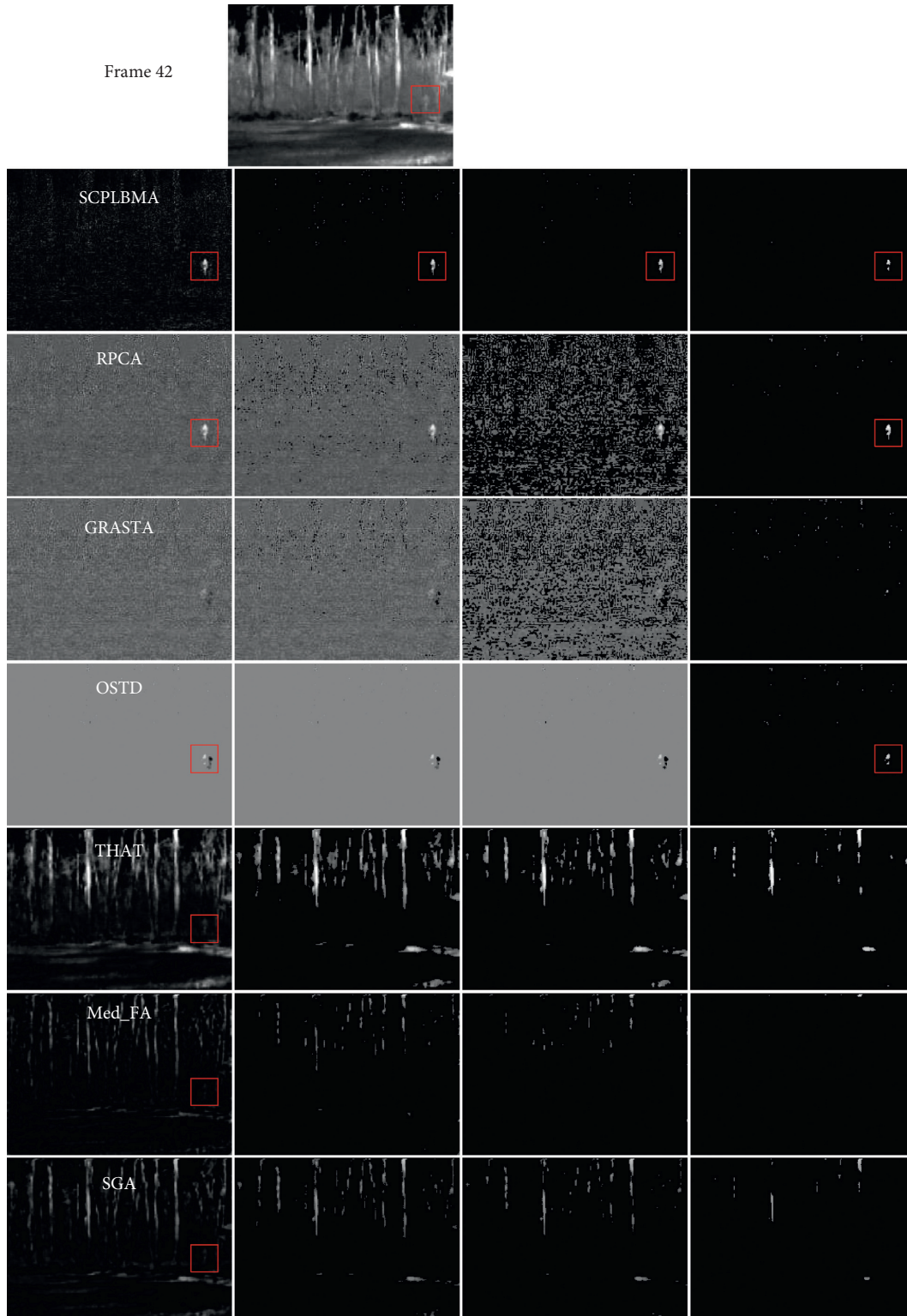
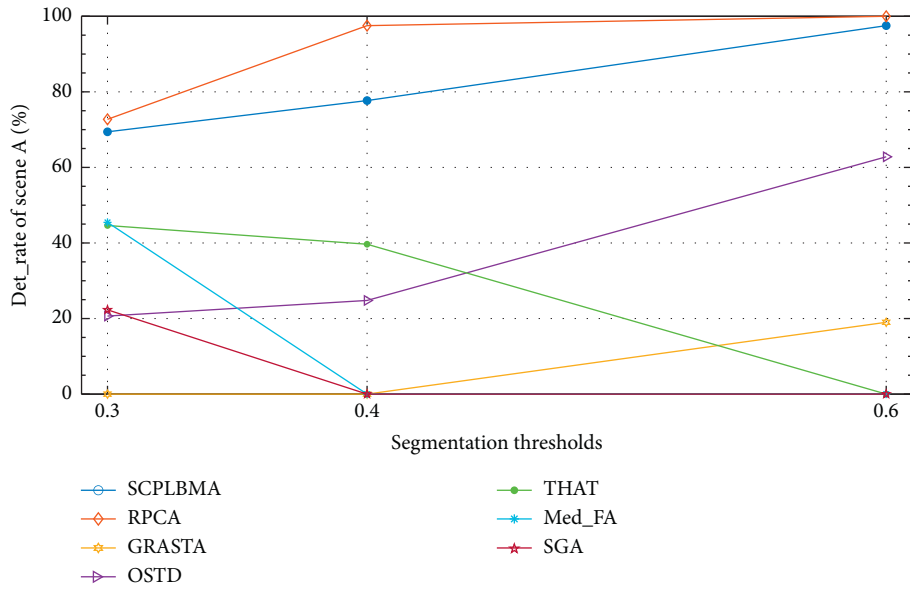


FIGURE 14: Threshold segmentation for frame 42 of scene C.

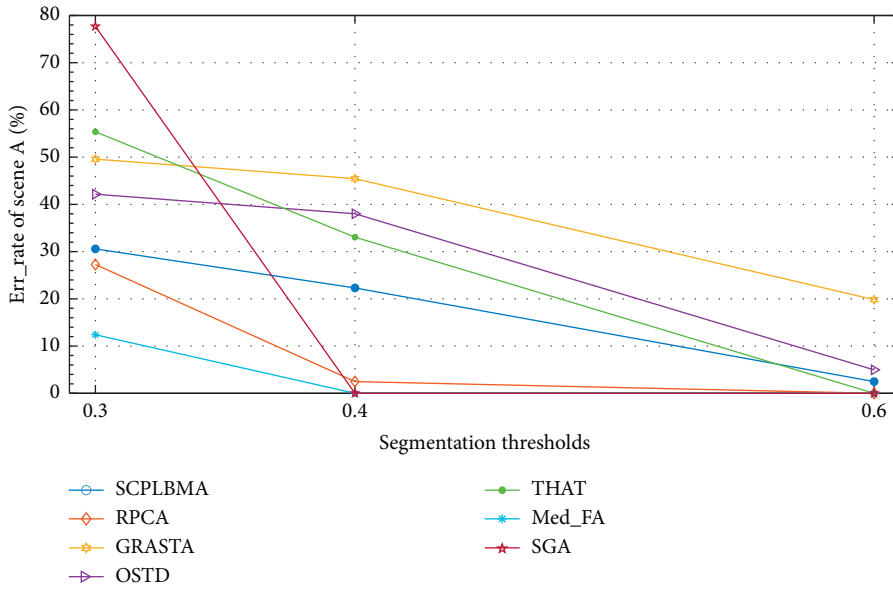
that the highest gray value of the targets is 255, which indicates that the targets are obviously much stronger than the noises. Therefore, these two algorithms had good segmentation results under the three different thresholds. Of course, it can be seen in Figure 12 that targets in difference images

are accompanied by a relatively large noise, so the segmentation result of SCPLBMA is not as good as that of the RPCA algorithm.

For scene B, the gray levels of several vertical lines in the difference images that were obtained from THAT, Med_FA,

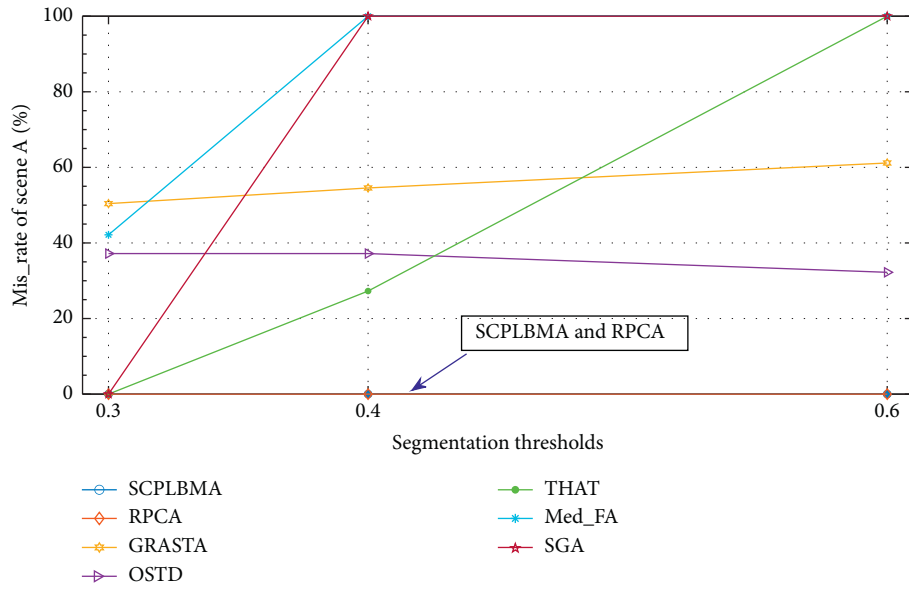


(a)



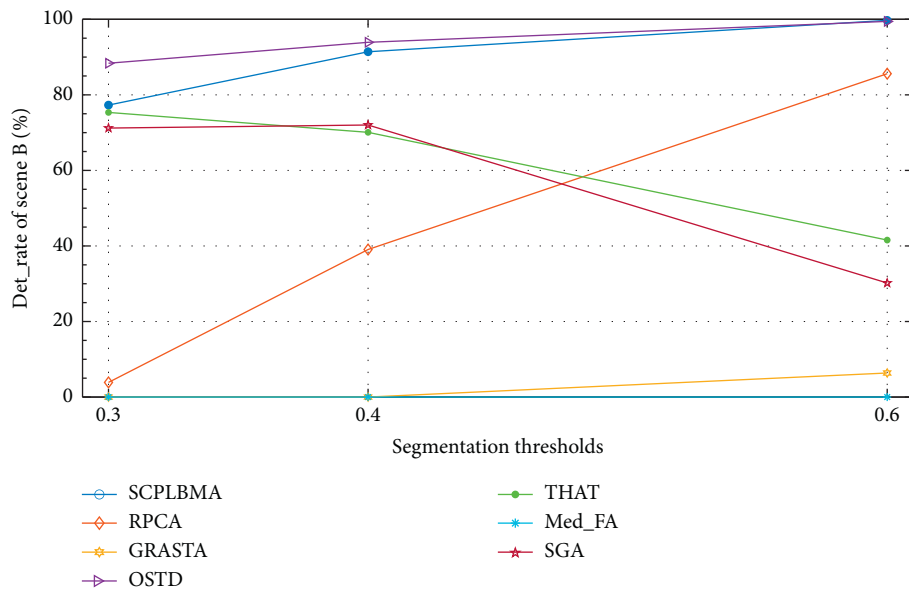
(b)

FIGURE 15: Continued.



(c)

FIGURE 15: Detection rates, error detection rates, and miss detection rates of the targets for scene A.



(a)

FIGURE 16: Continued.

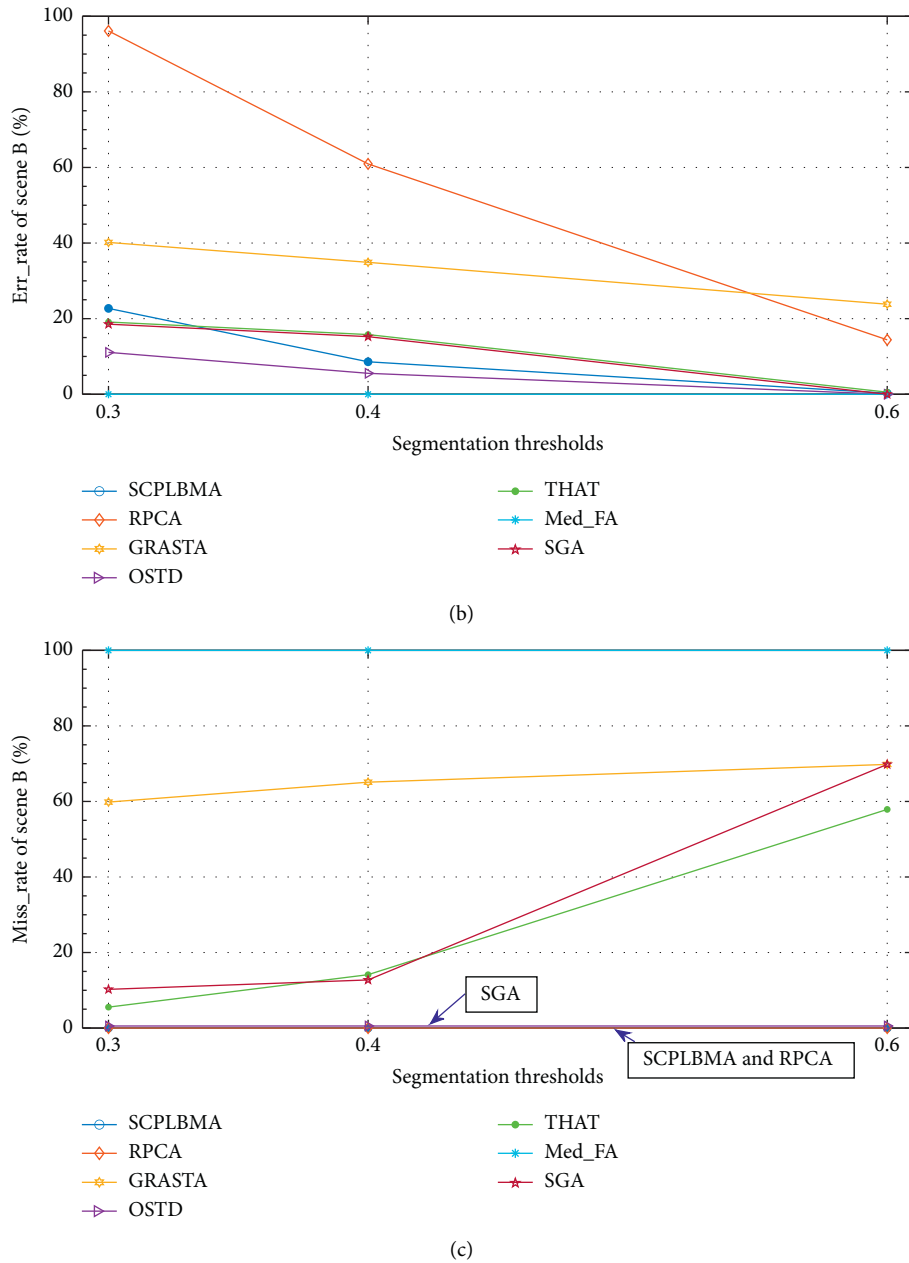
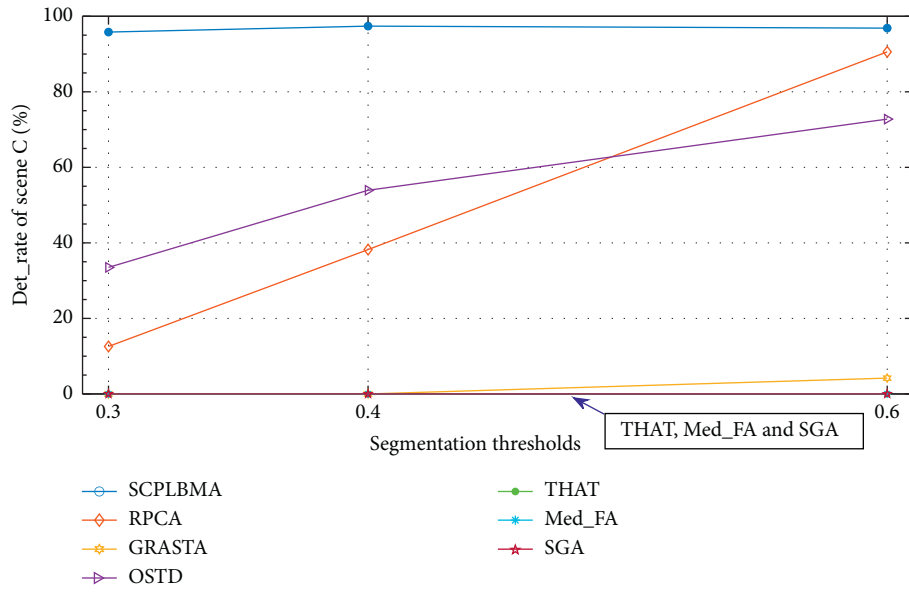


FIGURE 16: Detection rates, error detection rates, and miss detection rates of the targets for scene B.

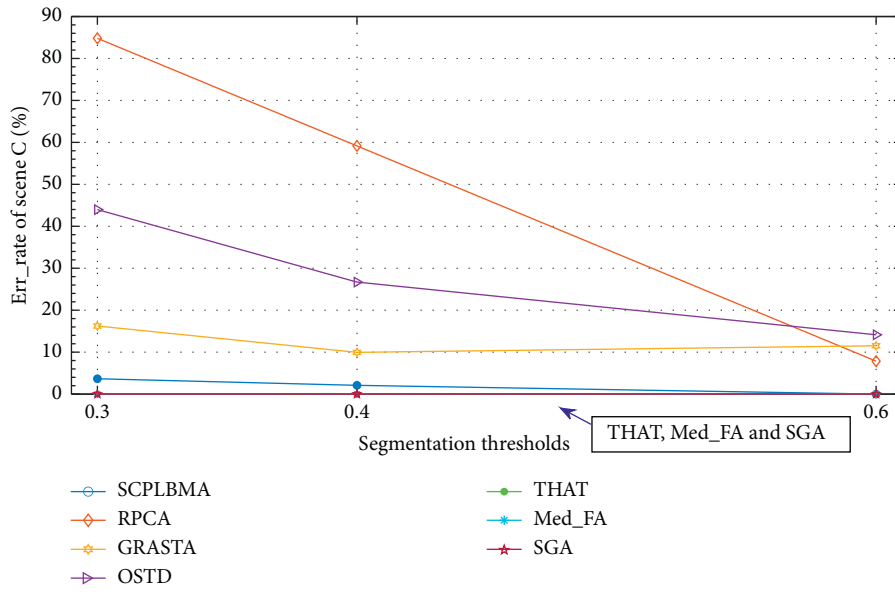
and SGA algorithms are all higher than that of the targets, so when the segmentation thresholds were 0.4 and 0.6 times the maximum gray level, the target detection rates decreased rapidly, which can be visually seen from Figures 13 and 16. In particular, the Med_FA algorithms could not segment out targets from the difference images at all in these two cases. Figures 13 and 16 show that SCPLBMA and OSTD algorithms have the best threshold segmentation results because their difference images have weak random noises and strong targets. Table 7 shows that the noise floor means of their difference images are only 14.2220 and 38.4685, respectively. The highest gray-scale value of their targets is 255, which can be seen from the three-dimensional energy distribution of Figures 8 and 9. There are many strong random noises

remaining in the difference images, and the mean of the noise floor is 75.5719, so the threshold segmentation result of RPCA is not good. The residual noises of GRASTA's difference images are the highest, and the noise floor is 118.4166. From beginning to end, the targets were severely interfered with by the noises. The targets are only faintly displayed in frames 51 to 185, and in other periods the targets are submerged in the noises. Therefore, the GRASTA failed to effectively segment out the targets when the threshold values were 0.3, 0.4, and 0.6 times the maximum gray value.

For scene C, the gray levels of bright tree trunks in the difference images that were obtained from THAT, Med_FA, and SGA algorithms are significantly higher than that of the



(a)



(b)

FIGURE 17: Continued.

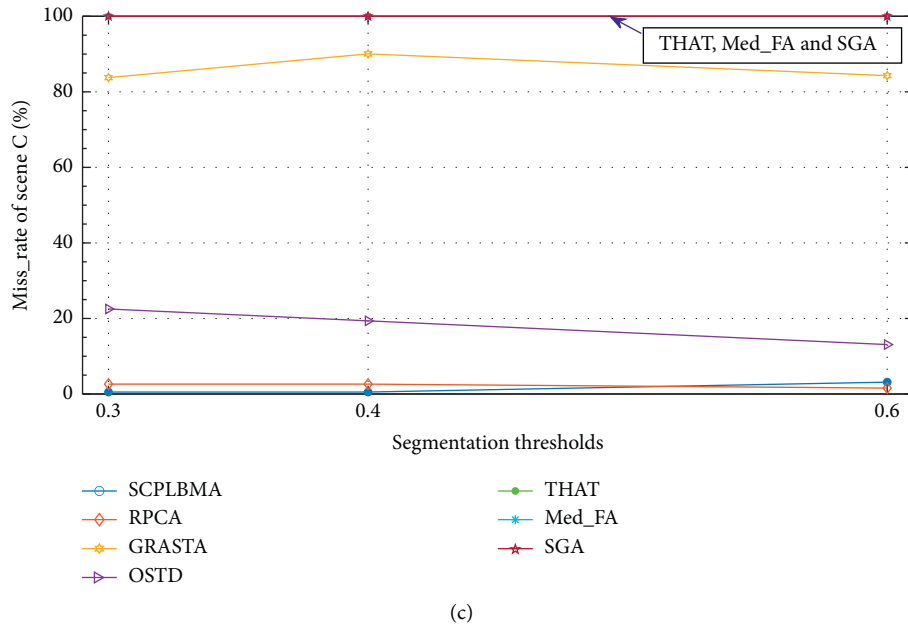


FIGURE 17: Detection rates, error detection rates, and miss detection rates of the targets for scene C.

targets. Then, they failed to segment out the targets when the segmentation threshold was 0.3, 0.4, and 0.6 times the maximum gray level, as can be visually seen from Figures 14 and 17. SCPLBMA, RPCA, GRASTA, and OSTD algorithms did not leave a distinct trunk texture in their difference images. The noise floor of SCPLBMA is the lowest, as shown in Table 7, being only 11.8174. The targets in the difference images are strong, so the threshold segmentation result is the best. Although the targets in the RPCA algorithm's difference images are strong, the residual noises are also very strong, and noise floor mean value is 84.6576. Therefore, only when the threshold value was 0.6 times the maximum gray value could we get good segmentation result. From beginning to end, the random noises in GRASTA's difference images are very strong, and the noise floor mean value is 122.3634. For GRASTA, except for frames 166–185, there were faint targets, and in other periods the targets were submerged by residual noises, so the threshold segmentation operation had no effect. The noise floor mean of OSTD algorithm's difference images is 100.4039, and the targets fully appeared only after frame 96. In other periods, the targets signal was weak and incomplete. Therefore, for OSTD algorithm, threshold segmentation failed to achieve good results.

The experimental results of scenes A, B, and C show that, compared with other algorithms, SCPLBMA stably achieves good performance for all typical scenes. This indicates that the SCPLBMA has good stability and generality.

5. Conclusion

The SCPLBMA for background modeling was proposed in this paper. The difference images that were obtained by the SCPLBMA had no residual textures, and the residual random noise was evenly distributed. Therefore, the difference

images had good stationarity. At the same time, the gray levels of the targets were significantly higher than that of the residual random noise, so the contrast values between the targets and the residual backgrounds were high. Experimental data showed that, for scenes A, B, and C, the SCPLBMA achieves good background suppression: First, the difference images had good stationarity. The autocorrelation coefficient quickly attenuated below 0.2, and the attenuation trend was smooth. Second, the contrast between the targets and the residual backgrounds was high. Experimental data showed that, for scenes A, B, and C, the contrast mean values for the SCPLBMA were 10.4446, 4.7972, and 6.4491, respectively. Third, the target information in the difference images was preserved well. SCPLBMA did not cause a significant loss of target information. Fourth, it was easy to extract out targets using the threshold segmentation. Because the residual random noise was evenly distributed and its gray levels were significantly lower than that of the targets, it was very easy to extract the targets. For scenes A, B, and C, when the segmentation thresholds were 0.3 times the maximum gray level, the target detection rates of the SCPLBMA were 69.42%, 77.29%, and 95.82%, respectively; when the segmentation thresholds were 0.4 times the maximum gray level, the target detection rates of the SCPLBMA were 77.69%, 91.41%, and 97.39%, respectively; when the segmentation thresholds were 0.6 times the maximum gray level, the target detection rates of the SCPLBMA were 97.52%, 99.72%, and 96.86%, respectively. Fifth, SCPLBMA has good stability and universality. Good background suppression effect was obtained for the 3 representative scenes by using SCPLBMA.

The main shortcomings of SCPLBMA are as follows: The method of using gray value distance as the clustering basis cannot well accord with the statistical characteristics of various image pixels. When the target's size is smaller and

the energy is weaker, the adaptability of the SCPLBMA will decrease. Therefore, the next work can be carried out from the perspective of improving the clustering method and clustering characteristics, such as the adoption of mean shift clustering, density-based clustering, GMM, and higher-order cumulant. For smaller and weaker target scenes, the modeling model based on pixel neighborhood would be considered.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest in this work.

Acknowledgments

This work was partly supported by the West Light Foundation of the Chinese Academy of Sciences (ya18k001), the Guangxi Science and Technology Base and Talent Project (Acceptance no. 2019AC20147), and the Doctoral Fund of the Guangxi University of Science and Technology (no. 19Z31).

References

- [1] Z. Wang, C. Mo, and H. Dai, "Target detection method before tracking based on particle filter," *Procedia Computer Science*, vol. 107, no. 1, pp. 141–147, 2017.
- [2] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Small infrared target detection based on weighted local difference measure," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 4204–4214, 2016.
- [3] L. R. Moyer, J. Spak, and P. Lamanna, "A multi-dimensional hough transform-based track-before-detect technique for detecting weak targets in strong clutter backgrounds," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 4, pp. 3062–3068, 2011.
- [4] T. Zhang, M. Li, Z. Zuo, W. Yang, and X. Sun, "Moving dim point target detection with three-dimensional wide-to-exact search directional filtering," *Pattern Recognition Letters*, vol. 28, no. 2, pp. 246–253, 2007.
- [5] X.-Z. Bai and F. Zhou, "Infrared small target enhancement and detection based on modified top-hat transformations," *Computers and Electrical Engineering*, vol. 36, no. 10, pp. 1193–1201, 2010.
- [6] T.-W. Bae, Y.-C. Kim, S.-H. Ahn, and K.-I. Sohng, "An efficient two-dimensional least mean square (TDLMS) based on block statistics for small target detection," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 30, no. 10, pp. 1092–1101, 2009.
- [7] T. W. Bae and K. L. Sohng, "Small target detection using bilateral filter based on edge component," *Journal of Infrared Millimeter and Terahertz Waves*, vol. 31, no. 6, pp. 735–743, 2010.
- [8] Q. Ling, S.-C. Huang, X. Wu, and Y. Zhong, "Infrared small target detection based on kernel anisotropic diffusion," *High Power Laser and Particle Beams*, vol. 27, no. 1, pp. 93–98, 2015.
- [9] Z. Zivkovic, "Improved adaptive Gaussian mixture model for back-ground subtraction," in *Proceedings of the International Conference on Pattern Recognition*, pp. 28–31, IEEE, Cambridge, UK, August 2004.
- [10] O. Barnich and M. Van Droogenbroeck, "ViBE: a powerful random technique to estimate the background in video sequences," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 945–948, Taipei, Taiwan, April 2009.
- [11] O. Barnich and M. Van Droogenbroeck, "ViBe: a universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing*, vol. 20, no. 6, pp. 1709–1724, 2011.
- [12] M. Hofmann, P. Tiefenbacher, and G. Rigoll, "Background segmentation with feedback: the pixel-based adaptive segmenter," in *Proceedings of the Computer Vision & Pattern Recognition Workshops*, IEEE, Providence, RI, USA, June 2012.
- [13] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [14] S. Pierre, E. David, Z. Xiang et al., "Integrated recognition, localization and detection using convolutional networks," in *Proceedings of the International Conference on Learning Representations (ICLR) (Banff)*, Banff, Canada, April 2014, <http://arxiv.org/abs/1312.6229v4>.
- [15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the Computer Science, Computer Vision and Pattern Recognition*, Columbus, OH, USA, October 2014, <http://arxiv.org/abs/1311.2524v5>.
- [16] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo, "On the applications of robust PCA in image and video processing," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1427–1457, 2018.
- [17] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, pp. 1–37, 2011.
- [18] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, "Robust subspace learning: robust PCA, robust subspace tracking, and robust subspace recovery," *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 32–55, 2018.
- [19] H. Guo, C. Qiu, and N. Vaswani, "An online algorithm for separating sparse and low-dimensional signal sequences from their sum," *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4284–4297, 2014.
- [20] B. Garcia-Garcia, T. Bouwmans, and A. J. R. Silva, "Background subtraction in real applications: challenges, current models and future directions," *Computer Science Review*, vol. 35, pp. 1–42, 2020.
- [21] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4–21, 2014.
- [22] T. Bouwmans, N. S. Aybat, and E.-h. Zahzah, *Handbook on Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*, Chapman and Hall/CRC Press, Taylor and Francis Group, Boca Raton, FL, USA, 2016.
- [23] F. R. S. Karl Pearson, "LIII. On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 1901.

- [24] N. A. Campbell, "Robust procedures in multivariate analysis I: robust covariance estimation," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 29, no. 3, pp. 231–237, 1980.
- [25] C. Croux and A. Ruiz-Gazen, "A fast algorithm for robust principal components based on projection pursuit computational statistics," in *COMPSTAT: Proceedings in Computational Statistics*, A. Prat, Ed., Physica-Verlag HD, Heidelberg, Germany, 1996.
- [26] F. De la Torre and M. J. Black, "A framework for robust subspace learning," *International Journal of Computer Vision*, vol. 54, no. 1–3, pp. 117–142, 2003.
- [27] K. Qifa and T. Kanade, "Robust ℓ^1 -norm. factorization in the presence of outliers and missing data by alternative convex programming," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 739–746, San Diego, CA, USA, September 2005.
- [28] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM Journal on Optimization*, vol. 21, no. 2, pp. 572–596, 2011.
- [29] J. Wright and Y. Ma, "Dense error correction via 1-minimization," *IEEE Transactions on Information Theory*, vol. 56, no. 7, pp. 3540–3560, 2010.
- [30] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. G. Hauptmann, "Infrared patch-image model for small target detection in a single image," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4996–5009, 2013.
- [31] Z.-C. Lin, M.-M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2013, <http://arxiv.org/abs/1009.5055>.
- [32] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [33] C.-X. Fan and L. Cao, *Principles of Communications*, National Defense Industry Press, Beijing, China, 2012.
- [34] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*, Xi An Jiao Tong University Press, Xi'an, China, 2018.
- [35] S. Karlin and H. M. Taylor, *A Second Course in Stochastic Processes*, Posts & Telecom Press, Beijing, China, 2009.
- [36] X.-M. Yuan and J.-F. Yang, "Sparse and low-rank matrix decomposition via alternating direction method," *Pacific Journal of Optimization*, vol. 9, no. 1, 2009.
- [37] J. Leskovec, A. Rajaraman, and J. Ullman, *Mining of Massive Datasets*, Stanford University Press, Stanford, CA, USA, 3rd edition, 2014.
- [38] M. P. Holmes, A. Gray, and C. Lee Isbell Jr., "QUIC-SVD: fast SVD using cosine trees," in *Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems*, Advances in Neural Information Processing Systems 21, Vancouver, Canada, December 2008.
- [39] P. Rodríguez and B. Wohlberg, "Fast principal component pursuit via alternating minimization," in *Proceedings of the 2013 IEEE International Conference on Image Processing*, pp. 69–73, Melbourne, Australia, September 2013.
- [40] J. He, L. Balzano, and A. S. Szlam, "Incremental gradient on the grassmannian for online foreground and background separation in subsampled video," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1568–1575, Providence, RI, USA, June 2012.
- [41] A. Sobral, S. Javed, S. K. Jung, T. Bouwmans, and E. Zahzah, "Online stochastic tensor decomposition for background subtraction in multispectral video sequences," in *Proceedings of the 2015 IEEE International Conference on Computer Vision Workshop (ICCVW)*, pp. 946–953, Santiago, Chile, December 2015.
- [42] A. Sobral, T. Bouwmans, and E. Zahzah, "LRSLibrary: low-rank and sparse tools for background modeling and subtraction in videos," *Robust Low-Rank and Sparse Matrix Decomposition: Applications in Image and Video Processing*, CRC Press, Taylor and Francis Group, Boca Raton, FL, USA, 2015, <https://github.com/andrewsobral/lrslibrary>.
- [43] T. Bouwmans, A. Sobral, S. Javed, S. K. Jung, and E. Zahzah, "Decomposition into low-rank plus additive matrices for background/foreground separation: a review for a comparative evaluation with a large-scale dataset," *CoRR*, <http://arxiv.org/abs/1511.01245>, 2015.